



UNIVERSIDADE FEDERAL DO CEARÁ
DEPARTAMENTO DE CIÊNCIAS SOCIAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM SOCIOLOGIA

OSCAR ARRUDA D'ALVA

**ESTATÍSTICAS OFICIAIS E CAPITALISMO DE PLATAFORMA: A TRANSIÇÃO
PARA UM REGIME DE DATAFICAÇÃO NO BRASIL**

FORTALEZA
2024

OSCAR ARRUDA D'ALVA

**ESTATÍSTICAS OFICIAIS E CAPITALISMO DE PLATAFORMA: A TRANSIÇÃO
PARA UM REGIME DE DATAFICAÇÃO NO BRASIL**

Tese apresentada ao Curso de Doutorado em Sociologia do Programa de Pós Graduação em Sociologia da Universidade Federal do Ceará, como parte dos requisitos para obtenção do título de Doutor em Sociologia.

Orientador: Prof. Edemilson Paraná
Co-Orientador: Prof. Fábio Gentile

FORTALEZA
2024

Dados Internacionais de Catalogação na Publicação Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

A819e Arruda d'Alva, Oscar.

Estatísticas Oficiais e Capitalismo de Plataforma : A transição para um regime de dataficação no Brasil / Oscar Arruda d'Alva. – 2024.

381 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Humanidades, Programa de Pós-Graduação em Sociologia, Fortaleza, 2024.

Orientação: Prof. Dr. Edemilson Paraná.

Coorientação: Prof. Dr. Fabio Gentile.

1. Estatísticas Oficiais. 2. Plataformas. 3. Big Data. 4. Ciência de Dados. 5. Algoritmos. I.
Título.

CDD 301

OSCAR ARRUDA D'ALVA

**ESTATÍSTICAS OFICIAIS E CAPITALISMO DE PLATAFORMA: A TRANSIÇÃO
PARA UM REGIME DE DATAFICAÇÃO NO BRASIL**

Texto apresentada ao Curso de Doutorado em Sociologia, do Programa de Pós Graduação em Sociologia, da Universidade Federal do Ceará como parte dos requisitos para obtenção do título de Doutor em Sociologia.

Orientador: Prof. Edemilson Paraná
Co-Orientador: Prof. Fábio Gentile

Aprovado em 02/07/2024.

BANCA EXAMINADORA:

Prof. Dr. Edemilson Paraná (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Jawdat Abu-El-Haj
Universidade Federal do Ceará (UFC)

Prof. Dr. Alexandre de Paiva Rio Camargo
Instituto Universitário de Pesquisas do Rio de Janeiro
da Universidade Candido Mendes (IUPERJ-UCAM)

Profa. Dra. Cecília Rikap
University College of London (UCL)

Prof. Dr. Rodrigo Santaella Gonçalves
Instituto Federal de Educação, Ciência e Tecnologia do Ceará

Aos trabalhadores e trabalhadoras do
IBGE.

Em memória de meu pai, Francisco d'Alva.

AGRADECIMENTOS

À minha companheira Carolyna Cabral, por tudo que passamos juntos neste período, pela sua luz e seu amor. Sem você esta jornada não teria sido possível.

À minha mãe Maria, minhas irmãs Alina, Cinira e Rita e meu filho Léo. Meu farol e porto seguro ao longo dessa viagem.

Ao meu orientador Prof. Dr. Edemilson Paraná, por sua incrível capacidade intelectual, compromisso e humanidade. Camarada, sua amizade foi o grande acontecimento deste percurso.

Ao meu analista Dr. Antonio Secundo, por nosso trabalho psicanalítico e sua ajuda em trazer à luz a porção inconsciente desta tese.

Ao meu Babalorixá Lynconly Jesus e a todo o povo do terreiro escola Ilê Obá Oladeji, a casa do rei que veio multiplicar o conhecimento, por todo axé e cuidado.

À querida prof. Mônica Martins e querido prof. Daniel Zirker, por regarem as primeiras sementes desta pesquisa no período de estudos na Nova Zelândia.

Ao Prof. Alexandre Camargo, pelo diálogo intelectual fraterno e debates nos GTs e SPGs da ANPOCS ao longo do percurso.

Aos profs. Tomás Rotta, Cecília Rikap e Evelyn Ruppert, pela acolhida no período de estudos em Londres e valorosas contribuições para a pesquisa.

Ao Prof. Fábio Gentile, por quem estendo meus agradecimentos a todos(as) os(as) professores(as) do PPGS da UFC, que acolheram esse *outsider* e abriram importantes portas para o saber e prática sociológica.

Aos colegas doutorandos e mestrandos do PPGS da UFC, pelas boas trocas e camaradagem neste conturbado período pandêmico.

Aos companheiros do NETS – Núcleo de Estudos em Economia, Tecnologia e Sociedade, pelos debates que abriram rumos de investigação nesta pesquisa.

Aos colegas do IBGE, por seu apoio e valiosas contribuições à pesquisa.

Ao IBGE, pela licença para a realização da pesquisa em um período tão desafiador de sua história.

Ao Projeto CAPES-PRINT-UFC, pela bolsa de pesquisa que viabilizou a realização do doutorado sanduíche na Goldsmith University of London.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Sem dúvida, realmente não se pode esperar nenhuma grandeza, quando se trata de pensamento, senão do conhecimento da 'miséria'. E talvez, segundo a mesma dialética, tipicamente pascaliana, da inversão entre o pró e o contra, a sociologia, forma de pensamento detestada pelos "pensadores", pelo fato de dar acesso ao conhecimento das determinações sociais que incidem sobre eles, por conseguinte sobre seu pensamento, está em condições de lhes oferecer — bem mais do que as rupturas de aparência radical que, frequentemente, deixam as coisas intactas — a possibilidade de livrar-se de uma das formas mais comuns da miséria e da fraqueza a que está muitas vezes condenado o pensamento pela ignorância e pela recusa ativa de saber (Bourdieu, 2001, p.160).

RESUMO

Com este trabalho buscamos compreender em que medida os processos contemporâneos de dataficação e a introdução de big data e ciência de dados no Instituto Brasileiro de Geografia e Estatística (IBGE) reconfiguram o campo de produção e distribuição de estatísticas oficiais no Brasil. Utilizando ferramentas do estruturalismo genético de Bourdieu e da sociologia da quantificação conduzimos uma investigação comparativa que aborda a implantação do Hub Regional de Big Data para a América Latina no IBGE, à luz dos processos históricos de estruturação do campo estatístico e de transição de regimes estatísticos, nos contextos do Norte-Global e do Brasil. A pesquisa consistiu na análise de dados secundários, cotejados com dados primários, obtidos por meio de entrevistas e observação não participante. O tema é relevante enquanto, no contexto do Capitalismo de Plataforma, a incidência das Big Tech no campo estatístico transnacional implica em riscos à soberania dos países do Sul-Global. Nossa hipótese é a de que a introdução de novas fontes de dados e métodos nas estatísticas oficiais implica em uma interseção entre o campo estatístico, tradicionalmente vinculado ao Estado, e o campo algorítmico, associado a corporações privadas. Este engajamento conduz a uma transição para um novo regime de dataficação. Nossa tese é a de que esta transição é caracterizada por dois obstáculos. Um obstáculo ontológico, relacionado à economia política dos dados, tensionada entre as concepções de bem público e mercadoria. E um obstáculo epistemológico, relacionado aos métodos, tensionados entre as interpretações frequentista e bayesiana da ciência estatística. Estes obstáculos conduzem a um duplo movimento no campo estatístico. Esforços para criar mercados de dados para estatísticas oficiais e gerar cientistas de dados nos Institutos Nacionais de Estatística se contrapõe a iniciativas que visam a regulamentação, proteção e acesso aos dados como bens públicos, assegurando a sua incorporação e a preservação dos métodos e fontes estatísticas tradicionais, sob a governança dos estatísticos de Estado. No caso do Brasil e países do Sul-Global, destaca-se a relevância dos contramovimentos nacionais, protetivos do caráter público das estatísticas oficiais e da soberania de dados, na determinação do grau de autonomia ou dependência que seus sistemas estatísticos poderão assumir na transição para um novo regime de dataficação.

Palavras Chave: estatísticas oficiais; plataformas; big data; ciência de dados.

ABSTRACT

With this work we seek to understand to what extent contemporary processes of datafication and the introduction of big data and data science at the Brazilian Institute of Geography and Statistics (IBGE) reconfigure the field of production and distribution of official statistics in Brazil. Using tools from Bourdieu's genetic structuralism and from the sociology of quantification, we conducted a comparative investigation that addresses the implementation of the Regional Big Data Hub for Latin America at IBGE, in light of the historical processes of structuring the statistical field and transition of statistical regimes, in the contexts of the Global North and Brazil. The research consisted of the analysis of secondary data, compared with primary data, obtained through interviews and non-participant observation. The topic is relevant to the extent that, in the context of Platform Capitalism, the incidence of Big Tech in the transnational statistical field implies risks to the sovereignty of countries in the Global South. Our hypothesis is that the introduction of new data sources and methods in official statistics implies an intersection between the statistical field, traditionally linked to the State, and the algorithmic field, associated to private corporations. This engagement leads to a transition to a new datafication regime. Our thesis is that this transition is characterized by two obstacles. An ontological obstacle, related to the political economy of data, tensioned between the concepts of public good and commodity. And an epistemological obstacle, related to the methods, tensioned between the frequentist and bayesian interpretations of statistics. These obstacles lead to a double movement in the statistical field. Efforts to create data markets for official statistics and to breed data scientists at National Statistics Offices are opposed to initiatives aimed at regulating, protecting and accessing data as public goods, ensuring their incorporation with the preservation of methods and traditional statistical sources, under the governance of State statisticians. In the case of Brazil and countries in the Global South, the relevance of national counter-movements, protective of the public nature of official statistics and data sovereignty, stands out in determining the degree of autonomy or dependence that their statistical systems may assume in the transition to a new datafication regime.

Keywords: official statistics; platforms; big data; data science.

LISTA DE ILUSTRAÇÕES

Figura 1 - Esquema Analítico da Investigação Sociológica.....	33
Figura 2 - Esquema Analítico da Investigação Histórica	36
Figura 3 - Mapa de Campo	38
Figura 4 - Mapa de Variáveis dos Informantes.....	41
Figura 5 - Modelo de Formação do Estado	50
Figura 6 - A Urna de Bernoulli.....	60
Figura 7 - Teorema de Bayes e Árvore Bayesiana	61
Figura 8 - A Curva Normal	62
Figura 9 - Modelo de Formação do Estado-Nação	67
Figura 10 - Quetelet e a Curva Normal	80
Figura 11 - A Eugenia de Galton e a Curva Normal.....	101
Figura 12 - Indução e Dedução: Abordagens de Modelo e de Desenho.....	208
Gráfico 1 - Análise de Correspondências Múltiplas - Informantes.....	42
Gráfico 2 - O Brasil e as Big Tech em 'Valores de Mercado'.....	167
Gráfico 3 - Estatísticos, Cientistas da Computação e Cientistas de Dados	173

LISTA DE TABELAS

Tabela 1 - Variáveis dos Informantes.....	41
Tabela 2 - Classificações de Fontes de Big Data (UNECE, 2013).....	182
Tabela 3 - Big Data: Oportunidades, Limitações e Riscos	184
Tabela 4 - Modelos de Acesso: Compartilhamento de Dados/Computação	201
Tabela 5 - Corte Epistemológico: Abordagens Frequentista e Bayesiana.	207
Tabela 6 - Quadro Analítico dos Regimes Estatísticos	331

LISTA DE ABREVIATURAS E SIGLAS

ACM – Análise de Correspondências Múltiplas
ADP – Automatic Data Processing
AIS – Automatic Identification System
API – Interface de Programação de Aplicações
ASSIBGE – Sindicato Nacional de Trabalhadores em Fundações Públicas Federais de Geografia e Estatística
AWS – Amazon Web Service
BID – Banco Interamericano de Desenvolvimento
BIRD – Banco Internacional de Reconstrução e Desenvolvimento
CBS - Netherlands Statistics
CDDI - Centro de Documentação e Disseminação de Informações
CDR – Call Detail Records
CEBD - Comitê de experts em big data para estatísticas oficiais das Nações Unidas
CEPAL - Comissão Econômica das Nações Unidas para a América Latina e o Caribe
CES – Comitê de Estatísticas Sociais
CES – Conferência dos Estatísticos Europeus
CIE – Congresso Internacional de Estatística
CLT – Consolidação das Leis Trabalhistas
CNE – Conselho Nacional de Estatística
COMEQ - Coordenação de Metodologia e Qualidade
CONCAR – Comissão Nacional de Cartografia
CONPLAN – Comissão Nacional de Planejamento
COVID – Doença do Coronavírus
CSO – Central Statistics Office do Reino Unido
DANE - Departamento Administrativo Nacional de Estatísticas da Colômbia
DFID - Departamento para o Desenvolvimento Internacional do Reino Unido
DESA – Departamento das Nações Unidas para Assuntos Econômicos e Sociais
DGC – Diretoria de Geociências
DGE – Diretoria Geral de Estatística
DGINS – Diretores Gerais dos Institutos Nacionais de Estatística da Europa
DI – Diretoria de Informática
DPA – Data Pop Alliance

DPE – Diretoria de Pesquisas Estatísticas

ENCE - Escola Nacional de Ciências Estatísticas

ENDEF - Estudo Nacional de Despesa Familiar

ENSAE - École Nationale de la Statistique et de l'Administration Économique

ESCAP - Comissão Económica e Social das Nações Unidas para a Ásia e o Pacífico

ESGAB - Conselho Consultivo de Governança das Estatísticas Europeias

ESS – Sistema Estatístico Europeu

Eurostat - Serviço de Estatística da Comissão Europeia

FOC - Grupo de Amigos da Presidência em Estatísticas Económicas

GFT – Google Flu Trends

GWG - Grupo de trabalho global em big data para estatísticas oficiais das Nações Unidas.

CGBigData - Comitê gestor de Big Data

GPSDD – Parceria Global de Dados para o Desenvolvimento Sustentável

HLG-MOS - Grupo de Alto Nível para a Modernização das Estatísticas Oficiais

HLG-PCCB - Grupo de Alto Nível para Parcerias, Coordenação e Capacitação para estatísticas da Agenda de 2030

HLP - High-Level Panel of Eminent Persons

IA – Inteligência Artificial

IEAG - Independent Expert Advisory Group on a Data Revolution for Sustainable Development

IAOS - International Association for Official Statistics

IBGE - Instituto Brasileiro de Geografia e Estatística

IDH – Índice de Desenvolvimento Humano

INDE - Infraestrutura Nacional de Dados Espaciais

INE - Instituto Nacional de Estatística

INEGI – Instituto Nacional de Estatística e Geografia do México

INSEE - Institut Nationale de la Statistique et des Études Économiques

IPEA - Instituto de Pesquisa Económica Aplicada

ISEB - Instituto Superior de Estudos Brasileiros

ISI – International Statistical Institute

LGPD – Lei Geral de Proteção de Dados

MPD – Mobile Positioning Data

MLOPs – Machine Learning Operations

NPM – New Public Management
NRC – National Research Council
OCDE - Organização de Cooperação e Desenvolvimento Económico
ODM – Objetivos de Desenvolvimento do Milênio
ODS – Objetivos de Desenvolvimento Sustentável
OI – Organização Internacional
ONE – Oficina Nacional de Estatísticas da República Dominicana
ONS – Office for National Statistics do Reino Unido
ONU - Organização das Nações Unidas
OPAL – Open Algorithms
PET – Tecnologia de Melhoria de Privacidade
PIB – Produto Interno Bruto
PME – Pesquisa Mensal de Emprego
PNAD – Pesquisa Nacional por Amostra de Domicílios
PND – Plano Nacional de Desenvolvimento
PPP – Parceria Público-Privada
PT – Partido dos Trabalhadores
RDH – Relatório de Desenvolvimento Humano
TCU – Tribunal de Contas da União
TIC – Tecnologia de Informação e Comunicação
UN – United Nations
UNECA - Comissão Econômica das Nações Unidas para a África
UNGP – Plataforma Global da ONU
UNSD – Divisão Estatística da ONU
UN-StatCom – Comissão de Estatística das Nações Unidas
UFC – Universidade Federal do Ceará
UNECE – Comissão Econômica das Nações Unidas para a Europa
UNGP – Plataforma Global da ONU
UNSD – Divisão Estatística das Nações Unidas
UN-StatCom – Comissão Estatística das Nações Unidas
USAID - Agência dos Estados Unidos para o Desenvolvimento Internacional

SUMÁRIO

INTRODUÇÃO	18
PARTE I	46
1. A ESTATÍSTICA OFICIAL E O CAMPO DE PRÁTICAS ESTATÍSTICAS	47
1.1 O Estado Territorial e o regime estatístico de contabilização.....	49
1.2 A Razão probabilística e os fundamentos cognitivos da estatística	56
1.3 O Estado Nação e a transmutação da estatística em bem público	65
1.4 Governamentalidade liberal e mais-valia: a estatística oficial e o mercado.....	73
1.5 População e normalização: a estatística social e o regime de enumeração.....	78
1.6 Entre a nação e o internacionalismo: a construção política do campo estatístico	85
1.7 A estatística desigual e combinada dos Estados Pós-Coloniais	90
1.8 Crise de governamentalidade liberal e transição de regime estatístico.....	97
1.9 Os contramovimentos nacionais e a consolidação do regime de enumeração no Brasil	107
1.10 O capitalismo administrado e o regime estatístico de precisão.....	112
1.11 Entre a soberania e a dependência: a consolidação do regime de precisão no Brasil.....	117
Conclusão: A constituição da estatística oficial como um bem público e a estruturação do campo estatístico como um espaço de lutas entre a dinâmica internacional e a soberania nacional	127
2. A TRANSIÇÃO PARA UM REGIME DE DATAFICAÇÃO	132
2.1 A crise da governamentalidade keynesiana e a emergência do neoliberalismo	134
2.2 O regime estatístico de comensuração: financeirização e política de indicadores	138
2.3 Virada empresarial dos Institutos Nacionais de Estatística e contramovimentos protetivos.....	145
2.4 Espaços transnacionais de quantificação e governança por indicadores.....	148
2.5 As Tecnologias de Informação e Comunicação e seus efeitos no campo estatístico	153
2.6 Democratização e resiliência do IBGE às reformas neoliberais	156
2.7 Dataficação, governamentalidade algorítmica e capitalismo de plataforma	163
2.8 A dataficação e as novas 'pressões' sobre o campo estatístico.....	173
2.9 O big data e a 'modernização' das estatísticas oficiais.....	181
2.10 O obstáculo ontológico: a nova economia política dos dados e o duplo movimento do campo estatístico.....	185

2.11 O obstáculo epistemológico: não representatividade do big data, modelagem e estatística bayesiana	202
2.12 Habitus algorítmico e mudança cultural: a introdução da ciência de dados nos INEs.....	210
Conclusão: Obstáculos ontológicos (dados) e epistemológicos (métodos) à transição para o regime de dataficação conduzem a um duplo movimento no campo estatístico.....	213
PARTE II	216
3. O PROJETO ONU BIG DATA E A PLATAFORMIZAÇÃO DAS ESTATÍSTICAS OFICIAIS	217
3.1 A abordagem neoliberal na agenda de desenvolvimento global pós-2015.....	218
3.2 O discurso da revolução de dados.....	221
3.3 Incidência de corporações privadas no campo estatístico transnacional	224
3.4 A Plataforma Global da ONU.....	227
3.4.1 O modelo de mercado e extrativismo de dados da Plataforma Global.....	229
3.4.2 Controvérsias na discussão do modelo da Plataforma Global.....	233
3.4.3 The hacker, the hipster and the hustler: a construção de um mercado global de estatísticas oficiais.	237
3.4.4 Entre a Comunidade e as Big Tech: controvérsias na implantação da Plataforma Global... ..	243
3.5 O elefante na sala: controvérsias de alto nível no campo estatístico transnacional.....	248
Conclusão: Plataformização das estatísticas oficiais e disseminação do regime de dataficação alinhado à estratégia econômica e política das Big Tech	258
4. O IBGE E O HUB REGIONAL DE BIG DATA NO BRASIL	262
4.1 Antecedentes: O Brasil e o IBGE no contexto da crise de governamentalidade neoliberal.....	263
4.2 O big data e a ‘modernização’ do IBGE: controvérsias no Governo Bolsonaro.....	266
4.4.1 A controvérsia do Censo Demográfico.....	267
4.4.3 Modernização subalterna, desestabilização institucional e legado disruptivo.....	274
4.5 O Hub Regional de Big Data no Brasil	275
4.5.1 Hub regional e Plataforma Global: posicionamentos e controvérsias.....	279
4.5.2 Percepções sobre a incidência de Big Techs no campo estatístico	283
4.6 A interface público-privada do Hub Regional.....	287
4.6.1 Viés mercadológico e a incidência da iniciativa privada	290
4.7.2 A empresa Positium e os modelos de acesso aos dados guardados por privados	296
4.7.4 Flowminder e Data Pop Alliance: o modelo de parcerias público-privadas e as tecnologias de ampliação de privacidade.....	301
4.8 Os casos de uso de big data no IBGE e o debate do acesso aos dados.	304
4.8.1 O uso de Web scraping para estatísticas de preço e a dificuldade de acesso aos dados fiscais.	305
4.8.2 O uso experimental de dados de telefonia celular e o debate do acesso aos dados guardados por privados.	308
4.8.3 O uso de web crawling para estatísticas de comércio eletrônico e o debate da compra de dados de intermediários	313
4.9 Percepções sobre à introdução de big data e ciência de dados no IBGE.	317
4.9.1 Estatísticos de Estado versus Cientistas de Dados.....	318

4.9.2 O habitus ibgeano e o obstáculo epistemológico.	321
Conclusão: A disputa pelo Sistema Estatístico Nacional na transição para o regime de dataficação	326
CONCLUSÃO	330
REFERÊNCIAS.....	344
APÊNDICE – TESE E META-TESE: TEXTO DE APRESENTAÇÃO DA DEFESA DA TESE DE DOUTORADO EM 02.07.2024.	370

INTRODUÇÃO

A técnica não é igual a essência da técnica. Quando procuramos a essência de uma árvore, temos de nos aperceber de que aquilo que rege toda árvore, como árvore, não é, em si mesmo, uma árvore que se pudesse encontrar entre as árvores. Assim também a essência da técnica não é, de forma alguma, nada de técnico. Por isso nunca faremos a experiência de nosso relacionamento com a essência da técnica enquanto concebermos e lidarmos apenas com o que é técnico, enquanto a ele nos moldarmos ou dele nos afastarmos. Haveremos sempre de ficar presos, sem liberdade, à técnica tanto na sua afirmação como na sua negação apaixonada. A maneira mais teimosa, porém, de nos entregarmos à técnica é considera-la neutra, pois essa concepção, que hoje goza de um favor especial, nos torna inteiramente cegos para a essência da técnica (Heidegger, 2002, p.11).

Neste trabalho, investigamos possíveis implicações dos processos contemporâneos de dataficação, para a produção e distribuição de estatísticas oficiais no Brasil.

Por dataficação, compreendemos o duplo processo de conversão de diversos aspectos da vida humana, do mundo físico, da economia, da sociedade, dos bens e serviços para o formato de dados digitais, e de geração de diferentes formas de valor a partir dos dados (Couldry; Mejias, 2019). Por um lado, um processo sociotécnico, na medida em que novas modalidades de processamento, quantificação e análise automatizada dos dados transmutam a cognição e ação humanas em uma forma analisável (Mayer-Schönberger; Cukier, 2013). Por outro lado, um novo regime político e econômico impulsionado pela lógica da acumulação do capital e caracterizado pela extração, cercamento e conversão dos dados em mercadorias, ativos e capitais (Sadowski, 2019).

A dataficação foi precedida e viabilizada pelos processos de digitização e digitalização, a partir do desenvolvimento e intensificação das Tecnologias de Informação e Comunicação (TICs). O primeiro, refere-se à criação de artefatos digitais por meio de processos técnicos. O segundo, consiste na transformação da economia e sociedade por meio da utilização destes artefatos (Gradillas; Thomas, 2023). Essa transformação caracteriza-se fundamentalmente pela automação de tarefas por meio de processos computacionais; pela ampliação da conectividade entre pessoas, empresas e máquinas; e pela maior convergência entre diferentes setores e áreas de conhecimento e produção.

Em linha com Paraná (2019, 2024), compreendemos criticamente a

digitalização como um processo multifacetado que engloba uma série de transformações estruturais do capitalismo contemporâneo sob o regime neoliberal, sobredeterminadas principalmente pela tecnologia e financeirização. A financeirização diz respeito a uma reorganização da lógica geral de acumulação do capital, em que o processo produtivo é submetido à lógica da valorização financeira (Chesnais, 1996, 2016; Mollo et. al., 2022, Paraná, 2024). Resultado da remoção de restrições e proteções sociais ao capitalismo, a financeirização é acompanhada pelo desenraizamento do mercado de suas bases sociais, e pela extensão da lógica e disciplina competitivas para outras dimensões da vida social. Conjugados, estes processos implicam em tendências globais como “a aceleração social ligada à compressão dos fluxos espaço-temporais [e] a crescente importância estratégica da propriedade sobre a produção [...] facilitando a utilização da descentralização técnica para a concentração econômica e política” (Paraná, 2024, p.20).

Os processos de digitalização e dataficação, por sua vez, são acompanhados pela crescente utilização de algoritmos para a extração de conhecimento dos dados. Algoritmos são procedimentos codificados para transformar dados em informações e resultados desejados, com base em cálculos especificados. Os computadores são essencialmente máquinas algorítmicas. Na medida em que as ferramentas computacionais se tornam o principal meio de expressão e interação social e econômica, o conhecimento e as práticas humanas são cada vez mais submetidos à lógica e às normas algorítmicas (Gillespie, 2013). A crescente automatização de tarefas e processos de tomada de decisão baseados em regras predefinidas por algoritmos, conduz à emergência de novas dinâmicas de dominação, que alguns autores têm chamado de poder algorítmico, governança algorítmica, governamentalidade algorítmica ou, simplesmente, algoritmização (Beer, 2009; Musiani, 2013; Ruvroy; Berns, 2013, Kitchin, 2016). A algoritmização tem produzido efeitos disruptivos sobre o funcionamento de diversos processos e sistemas (Steiner, 2012; Pasquale, 2015; Lehdonvirta, 2016), especialmente a partir do desenvolvimento do modelo de plataforma (Gillespie, 2010; Srnicek, 2017, 2020; Sadowski, 2019, 2020; Duffy et al., 2019, Morozov, 2020).

Em linha com Srnicek (2017, 2021) e Sadowski (2019, 2020), compreendemos a plataforma como um novo regime de acumulação do capital, caracterizado pela centralização de poder e extração de valor e rendas por meio de plataformas digitais, controladas por corporações privadas. O modelo das plataformas se baseia em três

pilares fundamentais: **i)** desintermediação: eliminação de intermediários tradicionais e conexão direta entre consumidores e produtos/serviços; **ii)** capitalização dos dados: extração, cercamento e conversão dos dados em mercadorias, ativos e capitais; e **iii)** governança algorítmica: uso dos algoritmos para influenciar processos de tomada de decisão. Para Srnicek (2021) e Sadowski (2020), o regime de propriedade das plataformas (intelectual e material) lhes possibilita a apropriação de valor produzido em outros espaços da economia global. Neste sentido, o capitalismo de plataforma (Srnicek, 2017) pode ser compreendido como um processo de atualização e expansão do rentismo e da monopolização no capitalismo contemporâneo, por meio das tecnologias digitais. Este processo passa a englobar a extração de rendas de conhecimento (Rotta e Teixeira, 2018; Rikap, 2021) e uma tendência de plataformização das finanças (Hendrikse et al., 2024) com a crescente interligação entre as grandes empresas globais de tecnologia, Big Techs (Google, Apple, Facebook, Amazon, Microsoft), e o setor financeiro digitalizado (Paraná, 2019, 2024).

A convergência entre plataformização e financeirização aponta para uma concentração sem precedentes de capitais pelas Big Tech, processo que ganha ainda maior impulso com o desenvolvimento da Indústria de Inteligência Artificial (Dyer-Witheyford et al., 2019). Na medida em que estas corporações passam a deter cada vez mais influência na determinação da forma como o valor é capturado no sistema capitalista, ampliam o seu poder econômico e político. A expansão da dominação política das Big Tech se reflete em sua penetração em campos tradicionalmente públicos e estatais e em estruturas nacionais, supranacionais e globais de governo (Morozov, 2020). Nesta tese investigamos a sua incidência nas estatísticas oficiais.

Na esteira dos processos de digitalização, dataficação, algoritmização e plataformização, apontamos para a emergência de novas fontes de dados, métodos e agentes, que passam a conformar um novo campo de práticas algorítmicas, envolvido em uma série de atividades de produção e distribuição de informações. A partir da segunda década do Séc. XXI, o campo algorítmico — associado aos novos mercados de dados, às plataformas digitais e às corporações privadas de tecnologia — passou a exercer considerável incidência, influência e pressão sobre o campo estatal de produção e distribuição de estatísticas oficiais, ou, simplesmente, campo estatístico.

O objeto da pesquisa é o Instituto Brasileiro de Geografia e Estatística (IBGE) e a evolução recente do fenômeno de dataficação das estatísticas oficiais no Brasil,

entre 2019 e 2023. Neste período, sob o Governo Bolsonaro, verifica-se a implantação do Hub Regional de Big Data para a América Latina no Brasil, sediado na Escola Nacional de Ciências Estatísticas (ENCE) do IBGE. O Hub é vinculado a um projeto de big data para estatísticas oficiais da Organização das Nações Unidas (ONU), coordenado globalmente pela Divisão Estatística da ONU. O projeto compreende a implantação de uma Plataforma Global — infraestrutura tecnológica, para compartilhamento de dados, métodos e tecnologias, voltada à produção de estatísticas oficiais — e quatro hubs regionais em Institutos Nacionais de Estatística (INEs) do Sul-Global (Ásia, Oriente Médio, África e América Latina). Além disso, a iniciativa compreende uma série de grupos de trabalho em novas fontes de dados e métodos para estatísticas oficiais.

O problema da pesquisa pode ser resumido na seguinte pergunta: ‘Em que medida os processos de dataficação das estatísticas oficiais, com a introdução de big data e ciência de dados no IBGE, reconfiguram o campo de produção e distribuição de estatísticas oficiais no Brasil, e qual é o sentido desta reconfiguração?’.

Para responder a este problema, a pesquisa persegue os seguintes **objetivos**: (i) investigar transformações contemporâneas (2019-2023) nos processos de produção e distribuição de estatísticas oficiais pelo IBGE, relacionadas ao uso de novas fontes de dados (big data) e métodos (ciência de dados); (ii) pesquisar a implantação do Hub Regional de Big Data no Brasil; (iii) analisar, à luz disso, o significado de transformações no campo de produção e distribuição de estatísticas oficiais no Brasil, relacionadas ao processo de dataficação.

O tema é relevante na medida em que as assimetrias econômicas, políticas e tecnológicas, que caracterizam os processos globais de dataficação, têm implicado na emergência de novas formas de poder geopolítico e em um reposicionamento da condição de subdesenvolvimento e dependência de países periféricos frente aos centros mais dinâmicos da economia mundial.

Para autores como Thatcher et. al (2016), Couldry e Mejias (2019), a utilização dos dados por empresas de tecnologia e governos do Norte-Global, para manter poder e influência sobre os países do Sul-Global, tem configurado formas contemporâneas de ‘colonialismo de dados’. Acreditamos que este conceito ganha maior envergadura analítica quando lido pelas lentes da teoria latino-americana da dependência (Bambirra, 2013; Cardoso; Faletto, 1970; Frank, 1969; Furtado, 1961, 1968; Marini,

2008; Prebish, 1949; Rangel, 2005; Santos, 2011, Tavares, 1982, 1992)¹. Neste sentido, compreendemos que sob uma nova economia orientada aos dados verificam-se alterações na divisão internacional do trabalho, que atualizam e reconfiguram as condições de dependência e periferidade dos países do Sul-Global.

No novo jogo da 'dependência digital e algorítmica', os países periféricos que tradicionalmente se posicionaram como consumidores de tecnologias de segundo nível dos centros dinâmicos, e como fornecedores de matérias primas e força de trabalho, são convocados à ocupar uma nova função nas cadeias globais de produção, enquanto fornecedores de dados. Na medida em que consomem novas tecnologias orientadas pelo modelo de plataformização, intensifica-se o cercamento e a extração de dados e rendas das populações destes países, acentuando-se a condição de dependência e as assimetrias de poderes e capitais entre Norte e Sul-Global.

Notadamente, o avanço das Big Tech em setores tradicionalmente estatais dos países do Sul-Global amplifica as formas de dependência, subordinando atividades e serviços públicos aos regimes de propriedade das plataformas tecnológicas (León, 2016; Rikap, 2021; Muller; Kuerbis, 2022). Este processo implica no cercamento e extração de dados em áreas sensíveis como saúde, educação, pesquisa e desenvolvimento, segurança, mobilidade, etc., ampliando o poder destas corporações, especialmente no contexto de desenvolvimento da indústria de Inteligência Artificial (Verdegen, 2022; Pasquinelli; Joler, 2021; Dyer-Witthford et al., 2019). As novas formas de dominação economia e política por meio dos dados são instrumentalizadas de acordo com os referentes técnicos, políticos e normativos das próprias corporações, ampliando sua ingerência sobre governos, mercados e populações.

Conforme indicam as evidências reunidas em nossa investigação, a incidência de Big Techs e outras corporações privadas sobre o setor tradicionalmente estatal das

¹ Em linha com Silva e Paraná (2017), adotamos aqui uma visão que foca nas convergências e continuidades da teoria latino-americana da dependência em suas correntes estrutural-cepalina, dependentista-associada e marxista. A despeito de diferenças conceituais importantes no que diz respeito à compreensão dos fatores causais do processo de industrialização latino-americano e suas consequências políticas, há considerável convergência entre as correntes no que se refere ao método — mediação entre as dimensões concretas e abstratas pela via da análise histórica — e à caracterização e diagnóstico dos fatores de subdesenvolvimento e dependência na América Latina — reprodução do atraso estrutural vinculado ao descompasso entre os setores moderno e arcaico, aos movimentos da luta de classes e à vinculação e influência dos interesses externos na direção da economia política nacional.

estatísticas oficiais, por meio da agenda do big data, implica em novos riscos à soberania nacional, especialmente para os países do Sul-Global. Isto porque as estatísticas oficiais são informações públicas essenciais, tanto para orientar a ação do Estado e dos agentes econômicos no espaço nacional, como para possibilitar o exercício crítico da democracia pela população, com vistas a sua autodeterminação e desenvolvimento. Isto inclui a cognição de aspectos físicos, sociais, econômicos e ambientais, bem como a avaliação dos governos por meio de informações confiáveis e referentes públicos construídos em um espaço político e cognitivo comum, tradicionalmente vinculado ao Estado e à lei (Desrosières, 1998). Conforme veremos, as novas modalidades informacionais algorítmicas, que emergem no processo de dataficação, tendem a desestruturar estes espaços, submetendo-os à lógica dos mercados de dados e das plataformas.

O capital tem tendências gerais, mas é constantemente atualizado no espaço-tempo e se expressa de modos diferentes de acordo com mudanças estruturais que se processam tanto no centro quanto na periferia (Prebish, 1949; Tavares, 1982). A leitura sociológica da dualidade estrutural destas mudanças e das novas formas de dependência que engendram, pode, portanto, contribuir para o exercício de uma perifericidade crítica e ativa (Rangel, 2005). Não se trata, portanto, de abdicar da importação e da apropriação de tecnologias, pelo contrário. A tarefa que se impõe ao Sul-Global é a de melhor compreender as determinações políticas e econômicas das novas tecnologias, de modo a melhor selecioná-las e adaptá-las, negociando termos mais justos e desenvolvendo capacidades próprias, à luz de um projeto popular, soberano, inclusivo e democrático de desenvolvimento.

Ainda que países periféricos como o Brasil não tenham controle sobre as determinações do capital internacional — e, no caso em análise, sobre o internacionalismo estatístico associado às corporações privadas globais — é essencial que entendam o modo como a importação de capitais (econômicos, tecnológicos, cognitivos) dos centros mais dinâmicos, se integram aos esquemas relativamente autônomos do espaço nacional periférico. Aqui nos referimos especificamente ao nosso objeto, o IBGE e o sistema estatístico e geocientífico nacional. Em nossa pesquisa, esta compreensão é construída sobretudo por meio da análise crítica dos processos históricos envolvidos na estruturação dos campos estatísticos global e nacional, e na transição de regimes estatísticos, conforme resumizamos a seguir.

Entre os séculos XIX e XX, a instrumentalização de práticas estatísticas em ferramentas de governo conduziu a estruturação de campos estatísticos na fronteira entre o Estado e a ciência. Os Estados Nacionais mobilizaram consideráveis investimentos econômicos e cognitivos para a consolidação de sistemas estatísticos nacionais, centralizados por Institutos Nacionais de Estatística (INEs), públicos e estatais. Os 'estatísticos de Estado', detentores de conhecimentos científicos especializados e investidos de um papel e *ethos* públicos, personificam as disposições dos agentes do campo. Estas disposições são pautadas pelo compromisso com a qualidade, confidencialidade, publicidade, precisão, confiabilidade e independência profissional na produção e distribuição de informações públicas e oficiais nos mais diversos aspectos: demográficos, geográficos, econômicos, sociais, ambientais e políticos (Desrosières, 1998).

Entre as décadas de 1940 e 1980, as estatísticas oficiais foram instrumentos essenciais para a operacionalização da gestão desenvolvimentista das economias capitalistas ocidentais pelos Estados Nacionais, de acordo com a macroeconomia keynesiana. Neste período, verifica-se o fortalecimento do caráter público e nacional das estatísticas oficiais e a sua confluência com construções cognitivas da estatística inferencial e econometria (Desrosières, 1998). Esta confluência resultou na construção de novas ferramentas de governo (inquéritos amostrais e contas nacionais) essenciais para a intervenção do Estado no contexto do capitalismo administrado e do regime de acumulação keynesiano-fordista (Chesnais, 1996, Coriat, 2000; Harvey, 2008). A crise deste modelo conduz a importantes reconfigurações no campo estatístico (Desrosières, 1998).

Nos anos 1990, o modelo neoliberal encontrou um terreno político fértil para sua expansão global. Sob um novo regime de acumulação flexível e financeirizada, emergem novas formas de gestão da economia e da sociedade inspiradas no dogma do mercado autorregulado (Chesnais, 1996; Harvey, 2008). A submissão das relações sociais e econômicas ao mercado demanda uma ampla quantificação do mundo social. As práticas de valoração financeira baseadas na precificação dos ativos dão margem a uma proliferação de indicadores que se espriam por outros campos de conhecimento e ação (Chiapello; Walter, 2016; Porter, 2015). A cultura da auditoria financeira é internalizada pelas organizações públicas e privadas (Power, 2000), refletindo-se em um modo de governo baseado na avaliação permanente de metas e rankings que visam medir, modelar e disciplinar o comportamento dos agentes por

meio de recompensas e punições (Bruno; Didier, 2013). Os centros de quantificação se multiplicam com o desenvolvimento das tecnologias de informação e comunicação e passam a envolver novos agentes privados e internacionais (Desrosières, 2008). Os processos de comensuração — transformação de qualidades em quantidades por meio de uma métrica comum (Espeland; Stevens, 1998) — generalizam-se no período neoliberal, configurando, um importante elo de transição para transformações subsequentes no regime estatístico, vinculadas aos processos de dataficação.

A crise do regime de acumulação financeirizada do capital, evidenciada a partir de 2007/2008/2011 (Anglietta, 2015; Paraná, 2022, 2023) é acompanhada por uma deterioração da legitimidade das instituições políticas e econômicas. Para Streeck (2013), adentramos um período de crise do capitalismo democrático, cujo sintoma mais visível é a incapacidade de resolução de conflitos sociais e econômicos pela via política, o que se reflete na emergência de governos de extrema direita em todo o mundo. À reboque dos processos de dataficação e plataformização, o poder assimétrico das Big Tech e suas estratégias de governança global passam a constituir um importante elemento da crise. Para alguns autores, verifica-se, neste contexto, a emergência de uma nova governamentalidade (Foucault, 2008), algorítmica (Ruvroy; Berns, 2013) e autoritária (Saidel, 2023).

Do ponto de vista epistêmico, os processos de dataficação conduziram a uma hibridização entre a ciência estatística e a ciência da computação, resultando na emergência da ciência de dados, como um novo campo de conhecimentos voltado à análise dos grandes volumes de dados digitais (NRC, 2013). Mais do que uma nova fronteira técnica, entretanto, o desenvolvimento de novos métodos analíticos e modalidades informacionais passa a envolver questões econômicas e políticas relevantes. Isto por que a ciência de dados emerge estreitamente vinculada aos novos mercados de dados e às corporações privadas de tecnologia.

Apontamos para a estruturação de um novo campo de práticas dedicado à produção e distribuição de informações — congregando Big Techs, empresas de tecnologia e telecomunicações, *think-tanks*, *startups*, organizações da sociedade civil e um conjunto de agentes intermediários (Hayashi et al. 2021, Richter e Slowinski 2019, Liang et al. 2018, Schomm et al. 2013, Stahl et al. 2014) — que denominamos de campo algorítmico. As novas modalidades informacionais deste campo — alegadamente mais potentes, abrangentes, oportunas e tempestivas que as estatísticas oficiais — produzidas e distribuídas fora do âmbito estatal, passaram a

exercer considerável fascínio e pressão sobre o campo estatístico, levando-o a se engajar com o big data e a ciência de dados.

Nossa hipótese é a de que a introdução de novas fontes de dados (big data) e métodos (ciência de dados) no processo de produção das estatísticas oficiais conduz a uma interseção entre os campos de práticas estatísticas e algorítmicas e a uma transição para um novo regime estatístico-algorítmico de dataficação.

Nossa tese é a de que esta transição é caracterizada por dois obstáculos (ontológico e epistemológico) que produzem importantes efeitos na estruturação do campo estatístico e conduzem a um duplo movimento (movimento pró-mercado e contramovimento protetivo).

Vejamos inicialmente os obstáculos à transição do regime estatístico. Em primeiro lugar, identificamos um obstáculo ontológico (Bhaskar, 2008, 2014) relativo à nova economia política dos dados: enquanto para o campo estatístico os dados são tradicionalmente bens públicos, para o campo algorítmico, tratam-se de mercadorias e ativos. Há, portanto, uma evidente contradição entre a lógica pública e a lógica de mercado neste engajamento. Em segundo lugar, verificamos um obstáculo epistemológico (Bachelard, 2005), que decorre da natureza não estruturada e não representativa do big data. Esta característica desafia a prática estatística tradicional — dedutiva, frequentista e baseada na teoria da amostragem — à qual se opõe uma nova abordagem — indutiva, bayesiana e orientada para a modelagem — que emerge no campo algorítmico.

Tratemos agora do duplo movimento (Polanyi, 2001) no campo estatístico. Na medida em que o novo campo algorítmico ameaça a hegemonia informacional dos Estados Nacionais e a relevância das estatísticas oficiais, observam-se iniciativas de engajamento do campo estatístico com o big data. A maioria destas fontes, entretanto, está sob domínio privado. Os cercamentos de dados (Dean, 2016; Verdegem 2022) vedam o acesso dos Estados aos novos capitais informacionais e frustram as iniciativas do campo estatístico. Para o campo algorítmico, os governos e as estatísticas oficiais passam a constituir um potencial nicho de mercado, o que conduz a uma incidência comercial de interesses privados sobre o campo estatístico. Configura-se assim um espaço de lutas em que identificamos a emergência de um duplo movimento.

Por um lado, a articulação do discurso de ‘modernização das estatísticas oficiais’ com big data e ciência de dados configura um movimento pró-mercado. Este

movimento conduz ao estabelecimento de interseções entre os campos estatístico e algorítmico, por meio de uma agenda de parcerias público-privadas (PPPs) e de incentivos à uma ‘mudança cultural’ no campo estatístico. Uma das estratégias para esta mudança consiste na gestação de uma nova geração de cientistas de dados no interior dos INEs, mais receptivos à inovação e à cooperação com as empresas privadas e mais abertos à lógica de mercado. Por outro lado, verificam-se contramovimentos protetivos da autonomia dos INEs e dos métodos estatísticos tradicionais, por meio da atualização da legislação estatística e da regulamentação do acesso às novas fontes de dados, caracterizando-as como bens públicos e visando a sua incorporação controlada sob a governança dos estatísticos de Estado e dos sistemas estatísticos nacionais.

No caso de países periféricos como o Brasil, os movimentos pró-mercado vinculam-se fundamentalmente ao internacionalismo estatístico, a partir da incidência de corporações privadas sobre o campo estatístico transnacional. Neste sentido, na medida em que a União Europeia gradualmente se afasta do modelo de parcerias público-privadas em big data para estatísticas oficiais e busca reforçar seus sistemas estatísticos nacionais pela via da regulação, verifica-se a disseminação da agenda de PPPs no campo estatístico do Sul-Global, por iniciativa da Divisão Estatística da ONU, com financiamento e participação de Big Techs. A articulação de alto nível de agentes transnacionais dos campos estatístico e algorítmico se materializa no projeto da Plataforma Global da ONU e na implantação dos hubs regionais de Big Data em INEs do Sul-Global.

Embora os processos de dataficação afetem os sistemas estatísticos nacionais dos países do Norte e do Sul-Global, o desenvolvimento desigual e combinado dos países periféricos (Frank, 1969; Trotsky, 1988; Santos, 2011) tende a implicar em uma maior dependência tecnológica e em uma inserção subordinada destes países na economia mundial e nos ciclos tecnológicos que a dinamizam (Rangel, 2005). Neste sentido, a relação dialética entre internacionalismo e nacionalismo estatístico ganha especial relevância nos países do Sul-Global. Isto porque, as transformações tecnológicas destes países se dão como uma resposta a movimentos engendrados nos centros mais dinâmicos da economia mundial no Norte-Global. É, portanto, o nacionalismo estatístico que confere uma organicidade própria aos sistemas estatísticos nacionais, e que, em última instância, determina o modo como estes países organizam suas respostas aos impulsos de ‘modernização’ do

internacionalismo estatístico. Esta resposta é uma importante condicionante da dependência e do caráter ativo ou passivo da condição de periferidade (Rangel, 2005).

No caso brasileiro, verificamos que o contramovimento protetivo — à subordinação das estatísticas oficiais aos interesses políticos e econômicos de agentes externos — reside fundamentalmente no conjunto de disposições relativamente conservadoras dos ‘estatísticos de Estado’. Estas disposições se manifestam em um habitus estatístico, historicamente incorporado pelos agentes e sancionado pelo campo estatístico nacional. É justamente a barreira interposta por este habitus — orientado pelos valores da lógica pública das estatísticas oficiais, soberania nacional e confiabilidade dos métodos estatísticos tradicionais — que os movimentos pró-mercado e os investimentos em uma ‘mudança cultural’ no campo estatístico — por meio da criação de espaços público-privados de articulação em torno do big data e da formação de uma nova geração de cientistas de dados — visam superar.

O método utilizado para responder ao problema e objetivos da pesquisa compreendeu, por um lado, uma investigação temporal e espacialmente abrangente de dados secundários e, por outro, o levantamento de dados primários sobre o objeto empírico em sua configuração atual. Em uma palavra, para a compreensão do processo contemporâneo e local, recorreremos à análise histórica e global. Isto foi necessário para que pudéssemos compreender os macroprocessos envolvidos na (re)configuração dos regimes estatísticos e na estruturação dos campos estatísticos nacional e transnacional. A investigação histórica e macrosociológica, portanto, nos forneceu as ferramentas para a análise microsociológica dos dados empíricos. Na medida em que abordamos um processo que se configura globalmente e se manifesta localmente, a investigação empírica se desdobrou igualmente na compreensão do objeto em suas dimensões global (Plataforma Global da ONU) e local (Hub regional de Big Data no Brasil).

A primeira etapa da pesquisa consistiu na construção de um referencial teórico e analítico para a interpretação histórica e sociológica da estruturação dos campos estatísticos nacional e transnacional. Esta investigação histórica foi necessária para a melhor compreensão dos processos contemporâneos de transformação das estatísticas oficiais no contexto de dataficação. Aqui nos valemos principalmente de trabalhos canônicos da história social da estatística, sociologia da quantificação e historiografia estatística brasileira, além de dados secundários que

cotejamos com abordagens sobre os macroprocessos históricos de configuração e transição dos regimes de acumulação do capital e modos de ação do Estado.

Partimos da premissa de que a investigação sociológica de um campo de práticas não pode prescindir de uma compreensão abrangente do seu desenvolvimento histórico. Isso é particularmente relevante para a abordagem de um objeto polissêmico como a estatística, situado na fronteira entre a ciência e o Estado (Camargo, 2016). Neste sentido, as diferentes conotações que a palavra ‘estatística’ assume no presente não são fortuitas, mas decorrem justamente de uma dupla origem histórica, político-administrativa e cognitiva. De acordo com Desrosières (1998), estes dois percursos e sua convergência conferem à estatística um duplo papel, enquanto ferramenta de prova e de governo:

Em sua presente arquitetura, a estatística se apresenta como uma combinação de dois distintos tipos de ferramentas [de governo e de prova], cujo caminho histórico apenas convergiu e levou a uma construção robusta em meados do século XX. A primeira é político-administrativa: desde o século XVIII, sistemas de registros, codificação, tabulação e publicação de estatísticas foram gradualmente instituídos, estatística compreendida em termos de descrição numérica de vários aspectos do mundo social. A segunda é cognitiva e envolve a formalização de esquemas científicos (médias, dispersões, correlações, e amostra probabilística) desenhados para resumir, principalmente por meio de ferramentas matemáticas, uma diversidade presumida ingovernável [...] Esse duplo caminho deve ser traçado se queremos compreender como um espaço cognitivo de equivalência e comparabilidade veio a ser construído para razões práticas (Desrosières, 1998, p. 79, tradução minha).

A investigação que conduzimos ao longo da tese se norteia pela noção do duplo papel da estatística de Desrosières (1998). Nosso foco principal é o de compreender a estatística enquanto ferramenta de governo, implicada na organização social da produção capitalista e no modo de ação do Estado. Entretanto, a compreensão de determinados aspectos epistemológicos da estatística, enquanto ferramenta de prova, também são relevantes para a pesquisa. Isso porquê as configurações contemporâneas da estatística oficial e do campo de práticas estatísticas são um resultado da confluência histórica do duplo caminho, político e cognitivo.

A investigação da estatística como ferramenta de prova se vale das contribuições da história social da estatística, em especial do grupo de Bielefeld², para

² O grupo de Bielefeld, como ficou conhecido a partir da reunião realizada na cidade alemã em 1982, desenvolveu uma série de estudos relevantes sobre a estatística e probabilidade (Porter, 1986, 1995;

apontar como o desenvolvimento cognitivo do pensamento e ciência estatística está historicamente imbricado ao desenvolvimento da estatística oficial e vice-versa. Para que o trabalho científico pudesse traduzir leis probabilísticas em leis estatísticas, foi necessário que uma avalanche de números sociais (Hacking, 2004) fossem codificados, registrados e publicados pelas burocracias estatais. Para que o trabalho burocrático pudesse ter eficácia na concentração e controle do capital informacional viabilizado pela estatística, foi preciso incorporar conhecimento científico à uma prática de Estado. Neste sentido, a estatística oficial desenvolveu-se historicamente como um campo de fronteira e tensões na interface dos campos burocrático e científico.

A análise da estatística como ferramenta de governo utiliza o conceito de governamentalidade de Foucault (2008, p.143-144) entendido como “o conjunto formado pelas instituições, os procedimentos, análises e reflexões, os cálculos e as táticas”, que permitem ao Estado exercer um determinado poder de governar, “que tem por alvo principal a população, por principal forma de saber a economia política, e por instrumento técnico essencial os dispositivos de segurança”. Segundo Foucault (2008), a estatística fornece uma articulação instrumental dos dois conjuntos tecnológicos que constituem os mecanismos de segurança dos Estados Nacionais, a polícia e os aparelhos diplomático-militares. Por um lado a estatística opera como uma ferramenta de racionalização e um dispositivo biopolítico para a avaliação e controle das forças internas do Estado, e, por outro, constitui um instrumento para o cálculo da dinâmica de forças e posições entre Estados inseridos em um espaço de competição global. A este duplo papel da estatística como tecnologia de governo corresponde um elemento de tensões recorrentes no campo estatístico, que decorrem da articulação entre o seu papel interno na produção de estatísticas oficiais, subordinado ao princípio da soberania nacional, e o seu papel externo, orientado pela lógica da competição e articulado a uma complexa rede que compreende o internacionalismo estatístico.

Trata-se, portanto, de acessar os processos históricos e sociológicos de

Hacking, 2004, 2007; Daston, 1988), abordando-os do ponto de vista da filosofia da ciência e de seu desenvolvimento histórico, aplicações e consequências para a ciência, o Estado e o mercado. Os principais conceitos que emergem dos trabalhos do grupo de Bielefeld, relacionam a estatística e o cálculo de probabilidades a três importantes macroprocessos históricos: (i) a erosão do determinismo e domesticação do acaso por meio das leis estatísticas; (ii) a produção de subjetividades por meio de classificações estatísticas, e (iii) o desenvolvimento de uma objetividade mecânica nas ciências, na economia e na sociedade.

estruturação do campo de produção e distribuição de estatísticas oficiais, em seus aspectos ontológicos e epistemológicos, para compreender a sua atual configuração. Para esta tarefa nos valem do estruturalismo genético/gerativo de Bourdieu e de seus conceitos analíticos de campo, capital, habitus e práticas. Para Bourdieu (1986, 1990, 1998, 2008), as estruturas sociais são o resultado de uma série de processos históricos e de interações entre os agentes e as instituições. Na medida em que as estruturas são internalizadas e incorporadas pelos agentes em esquemas de pensamento e ação (habitus) impõem-se como estruturas estruturantes, que moldam os seus comportamentos. Por outro lado, a estrutura social é ao mesmo tempo estruturada pela ação dos agentes em espaços sociais objetivamente determinados (campos). A interação dinâmica entre a estrutura estruturante e a estrutura estruturada, por sua vez, abre um espaço de possibilidades para alterações na estrutura social. As práticas, posicionamentos e relações entre os agentes no espaço social estruturado são objetivamente determinados pela lógica particular de cada campo, os quais se constituem ao mesmo tempo como espaços de disputas e de competição por capitais (econômicos, culturais, informacionais, sociais e simbólicos).

As noções de ontologia e epistemologia que adotamos neste trabalho se referenciam tanto no realismo crítico de Roy Bhaskar (1944-2014) como no racionalismo aplicado de Gaston Bachelard (1884-1962). Embora as abordagens destes filósofos da ciência apresentem divergências significativas, esta aproximação se beneficia da importância conferida por ambos sistemas filosóficos às estruturas e mecanismos gerativos na explicação dos fenômenos sociais. Acreditamos que para nossa pesquisa, esta síntese seja relevante ao nos possibilitar depurar a abordagem epistemológica da ‘falácia epistêmica’, segundo a qual asserções acerca do ser poderiam ser reduzidas a asserções do conhecimento. Neste sentido, assumimos a distinção categórica de Bhaskar (2014) entre as dimensões intransitivas ou ontológicas e transitivas ou epistemológicas do conhecimento. Esta aproximação crítica é especialmente relevante para uma releitura do estruturalismo genético/gerativo de Bourdieu (Vandenbergh, 2010).

Embora Roy Bhaskar não se refira a um ‘obstáculo ontológico’, este termo se alinha aos seus conceitos de ontologia estratificada (Bhaskar, 2014)³ e reais ausentes

³ A ontologia estratificada de Bhaskar (2014) postula que a realidade é composta de múltiplas camadas ou estratos, cada um com suas propriedades emergentes: o real —mecanismos, estruturas e poderes geradores, o atual — eventos e processos concretos que ocorrem no mundo, e o empírico — nossas

(Bhaskar, 2008)⁴. Em sua interpretação da dialética, Bhaskar (2008) introduziu o conceito de ausência para se referir a aspectos da realidade que não são diretamente acessíveis à nossa percepção ou conhecimento (estrato empírico), mas que, no entanto, têm um impacto significativo nos fenômenos (estrato atual) e nos mecanismos causais dos poderes geradores (estrato real). Bhaskar (2008) compreende a dialética como a 'lógica da mudança', mas não trata da mudança apenas como uma simples negação do real, mas como uma copresença do ausente (negativo) e do presente (positivo), que conduz à emergência de novas propriedades. A mudança envolve a negação transformativa, e, portanto, a ausência/superação de obstáculos (sobre desejos, vontades, necessidades e interesses), especialmente àqueles que fluem das relações de poder. Em nosso caso, concebemos os campos estatístico e algorítmico como espaço-temporalidades distintas (Estado/tradição/estatísticas x Mercado/Inovação/algoritmos) que se interceptam no processo contemporâneo de dataficação. Esta relação dialética ganha materialidade na 'atualização' dos dados no 'real', enquanto bens públicos ou enquanto mercadorias. É esta contradição na ontologia dos dados, que apontamos como um obstáculo aos processos geradores de dataficação das estatísticas oficiais.

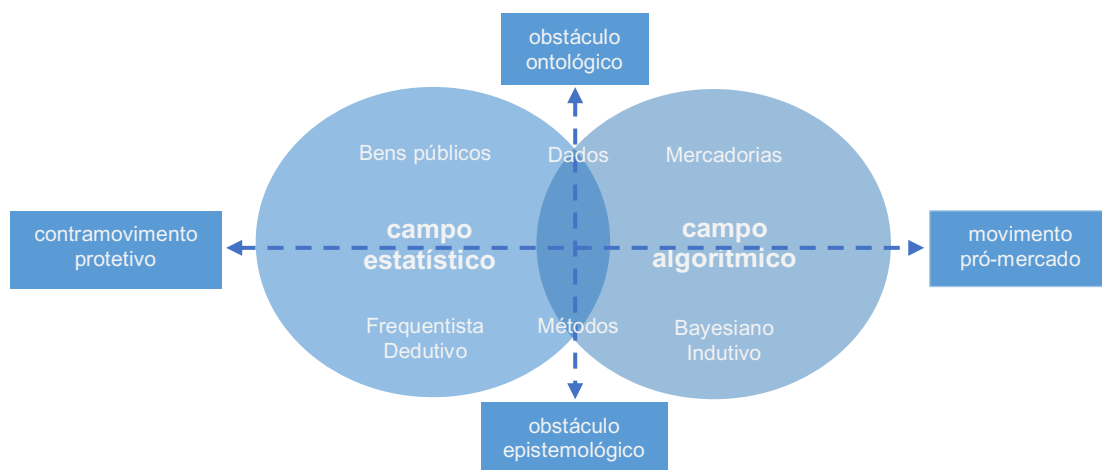
Já o conceito de 'obstáculo epistemológico', que tomamos de Bachelard (2005), refere-se aqui menos à noção de crenças ou hábitos que limitam, impedem ou retardam o conhecimento científico (como o conhecimento primeiro, o senso comum ou a doxa) e mais às crises que implicam em uma reorganização de um determinado sistema de saber ou campo de práticas. Neste sentido, também descartamos a noção de 'progresso científico' que Bachelard (2005) associa à superação dos obstáculos epistemológicos e nos concentramos em seu conceito de 'ruptura'. Compreendemos que a 'ruptura' que se efetua no confronto ou superação de um determinado obstáculo

experiências e percepções do mundo. Cada estrato da realidade emerge daquele que está abaixo, esta emergência, entretanto, não é aleatória, mas guiada pelos mecanismos subjacentes e poderes geradores no domínio do 'real', que comporta 'aquilo que é' e 'aquilo que poderia ser'. Os estratos da realidade não são estáticos, mas dinâmicos e interligados, interagindo e influenciando-se constantemente. Esta abertura permite a mudança, a transformação e o surgimento de novos fenômenos.

⁴ Para Bashkar (2008), os reais ausentes podem ser mecanismos ocultos — estruturas causais subjacentes e poderes geradores que operam abaixo da superfície dos fenômenos observáveis, ou potenciais inexplorados — o potencial para novas formas, estruturas e resultados que existem no mundo, mas que ainda não foram realizados.

epistemológico, apesar de especificamente vinculada ao conhecimento, é concretamente realizada no domínio das práticas, ou seja, na aplicação técnica e na produção de novos objetos que incorporam as abstrações teóricas objetivas. Em nosso caso, o caráter não representativo das fontes de big data impõe um obstáculo epistemológico objetivo às práticas estatísticas tradicionais. Este obstáculo, por sua vez, expõe os limites e tensões entre interpretações distintas da ciência estatística (frequentista e bayesiana) que foram historicamente apropriadas pelos campos estatístico e algorítmico. Neste sentido, a incorporação de novos métodos, que derivam da confluência entre os dois campos, emerge como um fator de ruptura.

Figura 1 - Esquema Analítico da Investigação Sociológica



Do ponto da investigação histórica, nossa análise apoia-se, por um lado, na estratificação das ferramentas estatísticas desenvolvida por Desrosières (2008a) no quadro “O Estado, o mercado e as estatísticas”, e por outro, no conceito de ‘regime estatístico’ de Beaud e Prévost (2011).

Em seu quadro analítico, Desrosières (2008) sintetizou suas investigações sobre as transformações históricas das ferramentas estatísticas, de acordo com mudanças na teoria econômica e nos modos de ação do Estado. Menos que uma tipologia cronológica, o quadro de Desrosières funciona como uma estratificação de modos de pensamento e ação engendrados na relação Estado-mercado, que oferecem uma grade de leitura das acumulações e rupturas históricas que contribuíram para conformar e transformar as ferramentas estatísticas. Seu modelo adota cinco configurações típicas de Estado que correspondem ao desenvolvimento

de diferentes ferramentas estatísticas: Estado Engenheiro (pós Séc. XVII), Estado Liberal (pós Séc. XVIII), Estado do Bem Estar (após o final do Séc. XIX), Estado Keynesiano (após anos 1940) e Estado Neoliberal (após 1990).

O conceito de 'regime estatístico' foi desenvolvido e utilizado por Beaud e Prévost (2011) em sua investigação sobre as transformações das estatísticas oficiais no Canadá. De acordo com os autores, um regime estatístico específico pode ser definido de acordo com três dimensões principais: estruturas, normas e práticas. As estruturas dizem respeito às determinações políticas e econômicas que condicionam a organização das atividades do sistema estatístico. As normas se referem ao conjunto de regras, prescrições, padrões, discursos e ethos profissional que caracterizam o processo estatístico. As práticas englobam os procedimentos, métodos, fontes, ferramentas, tipos de informação e requisitos que incidem sobre a produção estatística. No caso canadense, os autores definem os seguintes regimes estatísticos: proto-estatístico (1800-1860), nacional (1860-1940), macrogestão (1940-1980) e neoliberal (1980-).

Partindo destes trabalhos canônicos, efetuemos uma livre reinterpretação e estratificação histórica das ferramentas e regimes estatísticos, que, embora instrumentalizem os esquemas de Desrosières (1998), Beau e Prévost (2011), adota periodizações e nomenclaturas que diferem daquelas utilizadas por estes autores. Neste sentido, construímos neste trabalho uma grade de leitura histórica e sociológica original para a análise da configuração e transição dos regimes estatísticos.

Nosso quadro analítico incorpora ainda uma contribuição póstuma de Desrosières (2014) na qual o autor inseriu uma nova categoria de análise à periodização histórica das estatísticas, quer seja, 'o impacto das crises econômicas sobre as ferramentas das estatísticas públicas'. De acordo com Desrosières (2014), as grandes crises econômicas:

são momentos em que as estatísticas são intensamente utilizadas para expressar a gravidade da situação. Mas são também momentos de grande debate, durante os quais o papel do Estado na regulação e condução da economia é profundamente repensado. Cada uma destas crises corresponde ao surgimento de novas formas de quantificar o mundo social. Novos modelos de ação envolvem novas variáveis e novos sistemas de observação (Desrosières, 2014, p.166)

À esta intuição original de Desrosières (2014), incorporamos os referenciais teóricos da escola francesa de regulação, para contextualizar as grandes crises

econômicas como resultado de desestabilizações no regime de acumulação ampliada do capital e em seu modo vigente de regulação (Boyer, 1990; Chesnais, 1996, Coriat, 2000; Anglietta, 2015). O regime de acumulação refere-se a um conjunto de regularidades que asseguram a progressão relativamente coerente da acumulação do capital em um determinado período histórico. Incluem-se aqui as características principais do sistema produtivo, como a divisão do trabalho, a tecnologia, o padrão de consumo e investimento. Já o modo de regulação, refere-se ao conjunto de regras, instituições e normas que regulam a dinâmica do regime de acumulação e que possibilitam manejar as distorções e desequilíbrios que surgem no processo. Trata-se aqui do quadro normativo do mercado, das políticas de governo e relações de trabalho. Para a Escola de Regulação, as crises do capitalismo não são eventos isolados, pelo contrário, são um elemento intrínseco ao funcionamento do sistema econômico, que manifestam as suas próprias contradições internas. Neste sentido, as crises do capitalismo são caracterizadas como eventos cíclicos e multideterminados que podem demarcar pontos de inflexão e de transição entre os diferentes regimes de acumulação e modos de regulação.

Finalmente, aproximamos a proposição de Desrosières (2014), de que as grandes crises econômicas corresponderiam à emergência de novas formas de quantificar o mundo social, ao que Foucault (2008) caracterizou como 'crises de governamentalidade', ou seja, crises político-epistêmicas que sucederiam as crises de acumulação do capital, na medida em que as últimas impõem uma rearticulação da intervenção governamental no campo social e econômico. Deste modo, conectamos os conceitos de modo de regulação e governamentalidade, afim de compreender mudanças que operam no nível da racionalidade, estratégias e práticas de governo. Na medida em que a estatística é uma ferramenta essencial à governamentalidade, sua mobilização nos períodos de crise do regime de acumulação tende igualmente a provocar mudanças no regime estatístico.

Nosso quadro analítico, portanto, fundamenta-se na noção de que as crises que transformam o regime de acumulação do capital, implicam em crises de governamentalidade, que, por sua vez, resultam em reconfigurações dos regimes estatísticos. Este processo ocorre na medida em que a estatística, enquanto ferramenta de prova e de governo, é mobilizada para assegurar a legitimidade do Estado e a reprodução do capital nos contextos de crise. Finalmente, o modo de

estruturação do campo estatístico se relaciona diretamente com os processos gerativos prevalentes em cada regime estatístico.

Figura 2 - Esquema Analítico da Investigação Histórica



Fonte: do autor.

Para a compreensão das características históricas próprias da transição de regimes estatísticos no Brasil e da estruturação do campo estatístico nacional, traçamos um paralelo entre as matrizes estatísticas do Norte-Global e o seu desenvolvimento desigual e combinado no contexto brasileiro. Para tanto, nos referenciamos no trabalho de pesquisadores brasileiros que investigaram o desenvolvimento histórico dos sistemas estatísticos no país (Penha, 1993; Botelho, 1998; Senra, 2006, 2008, 2009, 2016; Camargo, 2009, 2009a, 2016, 2018; Grabois, 2021; Gil, 2023; Pereira, 2023). Destacamos as contribuições de Camargo (2016), que nos ajudaram a melhor compreender o desenvolvimento da estatística em países periféricos como um jogo de influência recíproca entre o internacionalismo e o nacionalismo estatístico. Este enquadramento comparativo, nos possibilitou ter uma melhor compreensão das questões geopolíticas envolvidas no processo de dataficação das práticas estatísticas.

Finalmente, dois outros conceitos caros à pesquisa são as noções de ‘cercamento dos comuns’ de Marx (2013) e de ‘duplo movimento’ de Polanyi (2021).

Para Marx (2013), a propriedade privada surge da apropriação que exige a exclusão do acesso. Refletindo sobre a ‘acumulação primitiva’ que possibilitou a emergência de um novo modo de produção e organização social do capital, Marx (2013) argumentou que o capitalismo moderno emergiu como resultado de processos massivos de expropriação e de ‘cercamentos de bens comuns’ que possibilitaram a revolução industrial. Esta ‘forma original de acumulação’ é atualizada no presente em desapropriações que negam formas de riqueza coletivas, públicas e não mercantis – seja de terras, recursos naturais, conhecimentos ou dados. Em linha com releituras

do conceito marxiano, utilizamos a noção de ‘cercamentos de dados’ (Dean, 2016; Verdegen, 2022) para nos referir a etapa de acumulação que precede os processos de comoditização, mercantilização e capitalização dos dados digitais no contexto do capitalismo de plataforma.

Os cercamentos dos comuns foram igualmente instrumentalizados na análise de Polanyi (2001), acerca dos esforços do mercado em converter a terra, o trabalho e o dinheiro em ‘mercadorias fictícias’, nas origens do capitalismo industrial. De acordo com Polanyi (2001), a sociedade moderna seria governada por dois princípios opostos: por um lado, movimentos para expandir a extensão do mercado autorregulado ou para separar a economia da sociedade, e, por outro, contramovimentos protetivos que visam a conservação do homem, da natureza, e da organização produtiva. O autor identificou como o Estado tende a agir tanto para acomodar os contramovimentos protetivos, contra os efeitos destrutivos da lógica mercantil, como para atender aos interesses da economia de mercado. Utilizamos esta perspectiva analítica, para caracterizar o duplo movimento do campo estatístico frente aos esforços do mercado em converter os dados em mercadorias.

A segunda etapa da pesquisa consistiu na coleta de dados secundários e primários abordando especificamente a introdução de novas fontes de dados e métodos no IBGE. Esta investigação envolveu a análise de casos de uso de big data no IBGE e a implantação do Hub regional de Big Data. A necessidade de uma melhor compreensão sobre o significado e objetivos deste projeto, para o Brasil e América Latina, nos levou necessariamente à investigar a origem global da iniciativa, o Projeto ONU Big Data e a Plataforma Global das Nações Unidas.

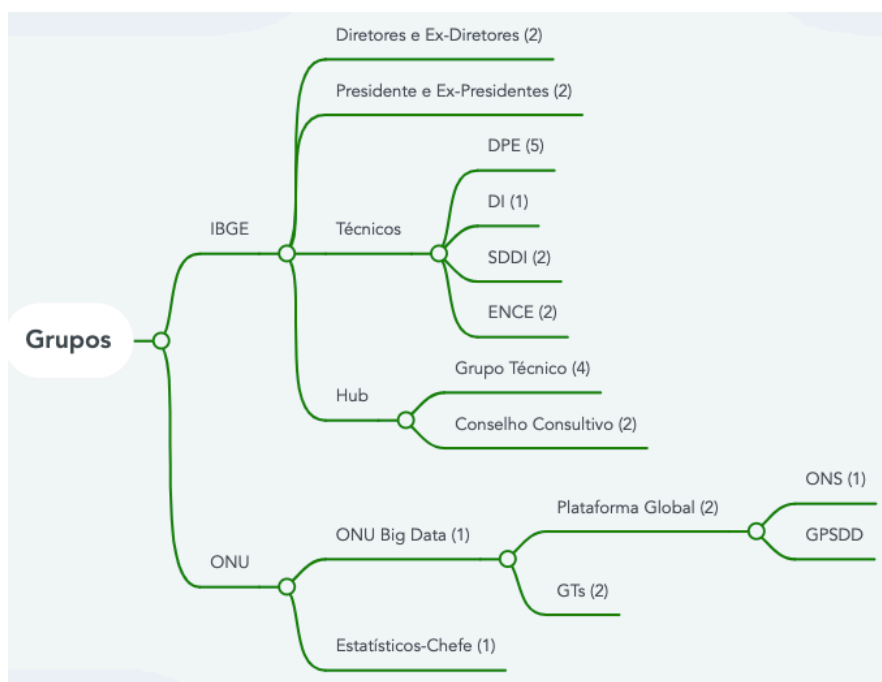
A coleta de dados secundários compreendeu o levantamento e análise de documentos públicos como planos, projetos, relatórios, vídeos, atas de reuniões, memorandos, discursos, postagens em redes sociais, balanços financeiros, registros contábeis, registros cartoriais, contratos, entre outros. Os documentos públicos necessários à esta etapa da pesquisa foram encontrados principalmente na internet. Adicionalmente, também foram levantadas e analisadas informações produzidas pela mídia, por meio de matérias jornalísticas, entrevistas e vídeos disponíveis na internet.

A coleta de dados primários envolveu a realização de entrevistas semiestruturadas e a observação não participante de atividades do projeto ONU Big Data e do hub regional de Big Data no Brasil. Esta etapa da pesquisa foi de certo modo facilitada pela minha vinculação institucional como servidor de carreira do IBGE,

apesar de licenciado para a investigação. Esta dupla vinculação com o objeto, por outro lado, requereu uma constante vigilância epistemológica (Bourdieu; Chamboredon, Passeron; 2015) para lidar com os obstáculos inerentes ao meu relacionamento pessoal, profissional, político e afetivo com o IBGE. Esta pesquisa é, portanto, em si mesma um campo de tensões entre o meu próprio habitus ibgeano e a busca pelo exercício autêntico da prática sociológica, o que não raro conduziu a crises pessoais ao longo do processo.

A realização das entrevistas foi precedida de uma fase exploratória em que buscamos: (i) a composição de um mapa do campo com um detalhamento dos grupos de interesse, buscando identificar as diferentes categorias, posicionamentos, capitais e diferenciais dos agentes que os compõe; (ii) a produção de roteiros com os principais tópicos da coleta de dados a partir do levantamento das principais dimensões, categorias e variáveis pertinentes à pesquisa; (iii) a identificação dos informantes-chave (sementes) da pesquisa.

Figura 3 - Mapa de Campo



Fonte: do autor.

Para a composição da amostra da pesquisa foi utilizada a técnica de amostragem não probabilística *snowball sampling* (bola de neve). A técnica consistiu em identificar informantes-chave ou sementes que indicaram outros informantes de

sua rede de contatos, que detivessem informações pertinentes e que estivessem potencialmente dispostos a participar da pesquisa. Desta forma, a própria coleta de dados compôs uma cadeia referencial em que uma entrevista levou às próximas, conforme indicações dos sujeitos entrevistados.

De acordo com Bernard (2005), a bola de neve é uma forma de amostragem adequada quando se está pesquisando populações relativamente pequenas que provavelmente estão em contato umas com as outras, como, por exemplo, grupos desviantes e grupos de elite. A principal vantagem da técnica é que os entrevistados são abordados a partir de uma rede interpessoal de relacionamentos, o que ao mesmo tempo em que contribui para o mapeamento da rede dos grupos pesquisados, pode emprestar confiabilidade ao entrevistador. De acordo com Becker (1993) que utilizou a técnica junto a grupos desviantes:

Essa estratégia resolve o problema de acesso de forma conveniente: pelo menos se conhece alguém que pode ser observado ou entrevistado, e pode-se tentar fazer com que este indivíduo o apresente a outros e seja seu fiador, desse modo deflagrando uma espécie de amostragem em bola de neve. Uma vez que você seja conhecido, e uma vez que se saiba a seu respeito em sua condição anterior, surgem poucas dúvidas quanto à sua confiabilidade (Becker, 1993, p.155).

Vinuto (2014) chama atenção para o risco de viés na amostra em virtude da possibilidade de se acessar apenas argumentações semelhantes, “já que os indivíduos necessariamente indicarão pessoas de sua rede pessoal, o que pode limitar a variabilidade de narrativas possíveis” (Vinuto, 2014, p. 207). A pesquisa buscou contornar esta limitação com o mapeamento inicial do campo, buscando identificar informantes chave que ocupavam posicionamentos distintos nos grupos pesquisados, aumentando a possibilidade de acessar uma maior variabilidade de discursos.

Junto aos informantes recrutados foram realizadas entrevistas semiestruturadas, em modo virtual. Os entrevistados foram sempre informados que estavam participando de uma atividade de pesquisa acadêmica. A identificação ou o anonimato dos dados obtidos no relatório final da tese foi facultada aos entrevistados e constava do termo de consentimento livre e esclarecido da pesquisa. Sempre que houve concordância do entrevistado a entrevista foi gravada. Recebemos autorização prévia do Comitê de Ética da UFC e Comitê de Ética da Goldsmiths University of London (onde realizamos um estágio doutoral) para a condução da pesquisa de campo. Entre maio de 2022 e dezembro de 2023 foram realizadas 21 entrevistas,

sendo: 18 entrevistas com profissionais ativos, ex-servidores e ex-presidentes do IBGE, 2 entrevistas com profissionais sêniores da Organização das Nações Unidas e 1 entrevista com uma estatística do ONS do Reino Unido.

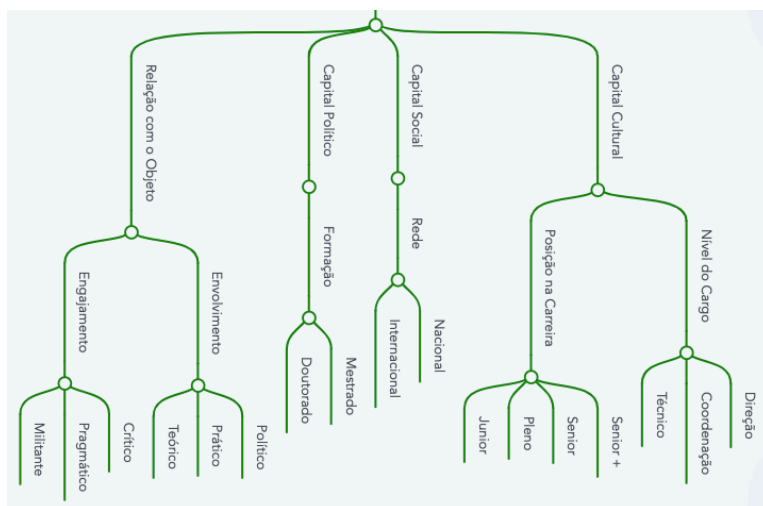
No caso dos dados obtidos por entrevistas, optamos por anonimizar os dados dos profissionais ativos no IBGE, identificando-os por pseudônimos (alfabeto grego) e conferindo a todos a identificação pelo gênero masculino. Isto se justifica para evitar que dados sensíveis possam comprometer de alguma forma os informantes em suas atividades profissionais atuais. Neste sentido, não sentimos a necessidade de fazê-lo para as entrevistas com quadros internacionais, aposentados e externos à instituição, bem como para os quadros ativos do IBGE, quando os dados foram obtidos por meio da observação não participante de atividades públicas.

O método de análise qualitativo consistiu no cotejamento dos dados primários e secundários e na identificação de padrões, consistências e inconsistências relevantes nas observações de campo. Finalmente, as informações relevantes e consistentes resultantes do processo de análise dos dados foram utilizadas para a produção dos resultados e conclusões da investigação. Como suporte à análise qualitativa e exposição dos dados primários, utilizamos a técnica estatística de Análise de Correspondências Múltiplas (ACM), desenvolvida pelo estatístico Jean-Paul Benzécri (1977) e amplamente utilizada por Bourdieu (2017), como uma forma de materialização operacional da teoria dos campos. O método possibilita a construção de modelos análogos do mundo social, por meio da representação geométrica em um plano cartesiano. No modelo, os agentes são situados em um espaço estruturado, no qual as distâncias demarcam diferenças de propriedades sociais, possibilitando a visualização dos posicionamentos dos agentes no espaço social (Kluger, 2018).

Com base em informações do perfil dos informantes e do tipo de relação estabelecidas com o objeto da pesquisa, efetuamos uma análise de correspondências múltiplas que pode ser visualizada no gráfico a seguir. Para a caracterização do perfil dos informantes selecionamos variáveis relativas aos capitais culturais, políticos e sociais dos agentes no campo estatístico. O capital político foi captado a partir do nível do cargo ocupado (direção, coordenação ou técnico) e posição na carreira (júnior, pleno, sênior e sênior+). O capital social foi abstraído da cadeia referencial do informante no âmbito nacional e(ou) internacional. O capital cultural foi medido com base no nível de formação acadêmica (mestrado e doutorado) e da distinção entre estatísticos e não estatísticos. Finalmente, a partir das entrevistas identificamos o tipo

de envolvimento dos informantes com o objeto da pesquisa (político, prático, teórico) e o modo de engajamento com o objeto (crítico, pragmático ou militante). Nos gráficos e tabela a seguir apresentamos o mapa e tabela de variáveis utilizados para compor a ACM dos informantes da pesquisa.

Figura 4 - Mapa de Variáveis dos Informantes



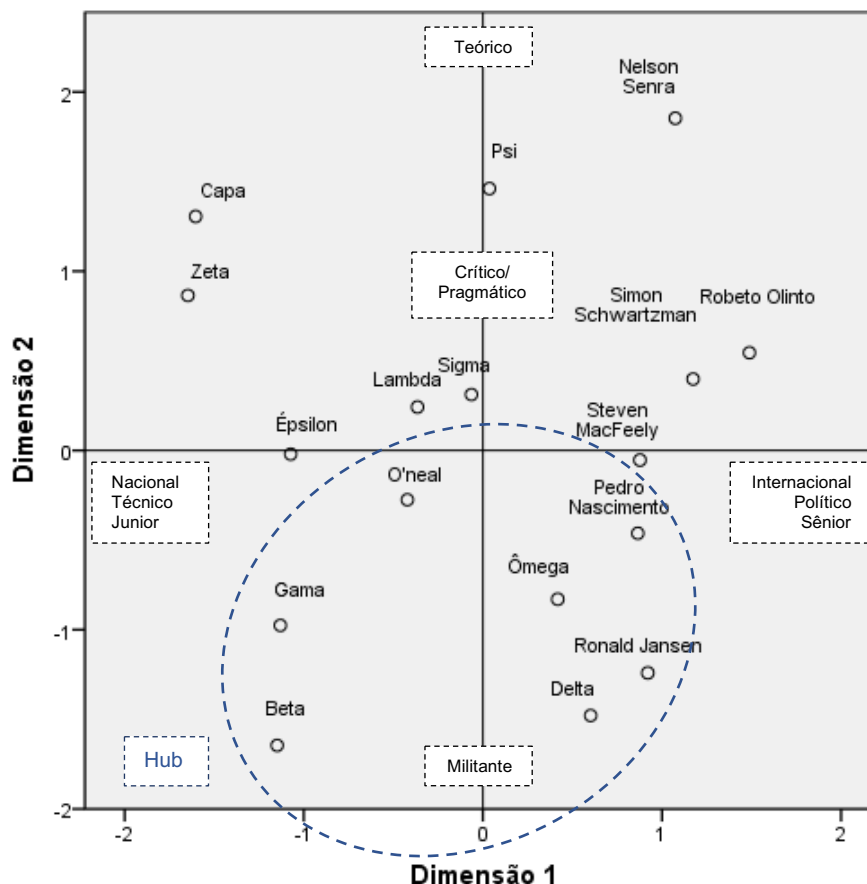
Fonte: do autor.

Tabela 1 - Variáveis dos Informantes

Informante	Hub	Cargo	Carreira	Formação	Nível	Rede Social	Envolvimento	Posicionamento
Beta	Sim	Coordenação	Pleno	Estatística	Mestrado	Internacional	Prático	Militante
Capa	Não	Técnico	Junior	Estatística	Mestrado	Nacional	Prático	Pragmático
Delta	Sim	Direção	Sênior	Estatística	Doutorado	Internacional	Político	Militante
Épsilon	Não	Coordenação	Pleno	Estatística	Doutorado	Nacional	Prático	Pragmático
Gama	Sim	Coordenação	Sênior	Estatística	Mestrado	Nacional	Prático	Militante
Lambda	Sim	Técnico	Sênior	Outra	Doutorado	Nacional	Prático	Pragmático
Nelson Senra	Não	Direção	Sênior +	Outra	Doutorado	Nacional	Teórico	Crítico
Ômega	Sim	Coordenação	Sênior	Outra	Doutorado	Internacional	Político	Pragmático
Pedro Nascimento	Sim	Direção	Aposentado	Estatística	Doutorado	Internacional	Político	Pragmático
Psi	Não	Direção	Sênior	Outra	Mestrado	Nacional	Teórico	Pragmático
Roberto Olinto	Não	Direção	Sênior +	Outra	Doutorado	Internacional	Político	Crítico
Ronald Jansen	Sim	Direção	Sênior	Outra	Doutorado	Internacional	Político	Militante
Steven MacFeely	Não	Direção	Sênior	Estatística	Doutorado	Internacional	Político	Crítico
Sigma	Não	Técnico	Sênior	Estatística	Doutorado	Internacional	Prático	Crítico
Simon Schwartzman	Não	Direção	Sênior +	Outra	Doutorado	Internacional	Político	Pragmático
Solange Onel	Sim	Coordenação	Sênior	Estatística	Doutorado	Internacional	Prático	Pragmático
Zeta	Não	Coordenação	Junior	Estatística	Mestrado	Nacional	Prático	Pragmático

Fonte: do autor.

Gráfico 1 - Análise de Correspondências Múltiplas - Informantes



Fonte: do autor.

O método de exposição dos dados consistiu na descrição textual dos achados da pesquisa e na utilização, sempre que conveniente, de citações diretas dos informantes e apresentação de dados em tabelas, gráficos, quadros e figuras. Na primeira parte da tese apresentamos os resultados da investigação teórica e histórica. Na segunda parte, discutimos os dados empíricos à luz da análise histórica.

No primeiro capítulo, buscamos responder à pergunta: o que são a estatística oficial e o campo estatístico, e que papéis desempenham na organização social da produção do capital e nos modos de ação do Estado? Evidenciamos como a estruturação do campo estatístico está intimamente relacionada com os processos de consolidação do Estado Nação e de transmutação da estatística em um bem público. Apontamos para o papel da estatística oficial como um instrumento de articulação do duplo movimento do Estado, enquanto meio geral de cognição do mercado e ferramenta de governo biopolítico. Evidenciamos como a estruturação do campo estatístico se efetua por meio de um conjunto de lutas, caracterizadas por tensões

que decorrem de sua dupla vinculação à soberania nacional e ao internacionalismo estatístico. Destacamos como a incorporação de construções cognitivas da matemática probabilística nas práticas de Estado fortaleceram a interpretação frequentista da estatística e se associaram à fatores políticos e econômicos que suscitaram mudanças históricas nos regimes estatísticos, da contabilização à enumeração, e da enumeração à precisão. Demonstramos como em resposta às crises de acumulação do capital e da governamentalidade liberal o fortalecimento da intervenção do Estado consolidou o caráter público e nacional das estatísticas oficiais no contexto do capitalismo administrado. Apontamos como no caso brasileiro em que pese a influência e incidência do internacionalismo estatístico, foram os contramovimentos nacionais que deram uma organicidade própria ao campo estatístico. Finalmente, sistemas estatísticos nacionais, centralizados por INEs, compostas por estatísticos de Estado especializados e inseridos em uma complexa rede internacional de instituições e profissionais caracterizam o campo estatístico em sua maturidade.

No segundo capítulo, discutimos como a crise do regime de acumulação fordista-keynesiano deu margem a uma nova crise de governamentalidade que abriu caminho para o neoliberalismo sob um novo regime de acumulação financeirizada. Neste contexto, apontamos um conjunto de transformações que conduziram à configuração de um regime estatístico de comensuração. Discutimos como estas transformações abriram flancos que tornaram os sistemas estatísticos nacionais mais permeáveis ao mercado, mais responsivos às inovações tecnológicas e mais susceptíveis à influência de agentes internacionais e privados. Por outro lado, apontamos para contramovimentos que possibilitaram a preservação da autonomia dos INEs e o caráter público e nacional das estatísticas oficiais. Finalmente, analisamos como a crise do regime de acumulação financeirizada aponta para uma crise contemporânea da governamentalidade neoliberal e para uma nova reconfiguração do regime estatístico. Tal reconfiguração se dá no contexto de uma nova economia orientada aos dados e de um regime de acumulação caracterizado pelo advento das finanças digitalizadas e do capitalismo de plataforma. Este novo contexto, implica em dois obstáculos sem precedentes para o campo estatístico, por um lado, uma nova economia política orientada pela comoditização dos dados, e por outro, uma nova epistemologia estatística associada ao big data. As injunções entre os campos de práticas algorítmicas e estatísticas dão margem a uma transição para

o regime estatístico-algorítmico de dataficação, marcada por um duplo movimento do campo estatístico de aproximação e reação ao mercado.

No terceiro capítulo, discutimos os resultados de nossa investigação empírica sobre o projeto da Plataforma Global da ONU. Demonstramos como a política de indicadores de desenvolvimento sustentável impulsionou o uso de big data e novos métodos algorítmicos pelos sistemas estatísticos nacionais e global, contribuindo assim para mudanças nas estruturas, normas e práticas do campo estatístico e para a disseminação do novo regime de dataficação. Evidenciamos como a articulação de alto nível entre agentes do campo estatístico e algorítmico transnacional abre um espaço político para a ingerência direta de Big Techs e outras corporações privadas nos sistemas estatísticos nacionais e global. O Projeto ONU Big Data e a plataformização das estatísticas oficiais que resulta desta articulação ressoam uma concepção mercantil e neocolonialista de dados, conferindo aos países do Sul-Global um status de provedores de dados e receptores passivos de métodos e tecnologias desenvolvidas por países e corporações do Norte-Global. Demonstramos como o capital simbólico da confiança pública na ONU e nas estatísticas oficiais é instrumentalizado para ações direcionadas ao mercado, conduzindo a uma série de conflitos de interesses e riscos aos sistemas estatísticos dos países do Sul-Global. Evidenciamos controvérsias entre agentes de alto nível do campo estatístico transnacional, que corroboram a tese do duplo movimento entre a adesão aos mercados de dados e a proteção dos dados para fins estatísticos como bens públicos. Finalmente, discutimos como o alcance global da Plataforma por meio dos hubs regionais pode potencialmente habilitar novas cadeias de valor. Neste sentido, concluímos que os hubs regionais no Sul-Global fazem parte de uma estratégia de governança política e econômica das Big Tech que visa vencer resistências nacionais aos mercados globais de dados e à nova estrutura de dominação.

No quarto capítulo, apresentamos os resultados da investigação empírica sobre a implantação do Hub Regional de Big Data no Brasil e os processos de dataficação no IBGE. No contexto de desestabilização institucional do Governo de extrema direita de Bolsonaro, verifica-se a desarticulação de iniciativas que visavam o fortalecimento da governança estatística nacional. A introdução de ações disruptivas de ‘modernização’, voltadas à promoção de big data e ciência de dados, e orientadas por acordos internacionais, caracterizam uma inserção subordinada do IBGE no regime de dataficação. Desdobrando controvérsias, verificamos diferentes

entendimentos e posicionamentos entre os agentes do IBGE sobre o processo de dataficação. Posicionamentos altivos, críticos e protetivos contrastam com posturas de subordinação e maior abertura às parcerias público-privadas. Finalmente, verificamos junto à maioria dos quadros ativos a incorporação de disposições características dos estatísticos de Estado, que se manifestam em um conservadorismo relativo à introdução de novas fontes de dados e em uma resistência ao novo perfil profissional do ‘cientista de dados’. Destacamos como a predominância do habitus estatístico e da lógica pública dos INEs constitui um obstáculo à difusão da lógica de mercado do big data e à introdução de novos métodos disruptivos. Neste sentido, estas disposições constituem um alvo das estratégia de governança política e econômica global das corporações privadas. Neste contexto, a transição para um regime de dataficação configura-se como uma disputa velada pelo controle dos sistemas estatísticos nacionais, com o cultivo no campo estatístico de um novo conjunto de disposições mais permeáveis aos interesses de mercado.

Na conclusão sumarizamos os principais resultados da investigação à luz do problema, hipótese e tese da pesquisa.

PARTE I

1. A ESTATÍSTICA OFICIAL E O CAMPO DE PRÁTICAS ESTATÍSTICAS

O problema de saber o que é a ciência deve ser posto. Não é a ciência, em si mesma, atividade política e pensamento político, na medida em que transforma os homens, torna-os diferentes do que eram antes? [...] Se a ciência for 'descoberta' da realidade antes ignorada, não será essa realidade, em certo sentido, concebida como transcendente? E o conceito de ciência como criação, afinal, não equivale à política? tudo consiste em ver se se trata de criação arbitrária ou racional, isto é, útil aos homens para ampliar o seu conceito da vida, para tornar superior (desenvolver) a própria vida (Gramsci, 202, p.331-332).

O objetivo deste capítulo é compreender o que são a estatística oficial⁵ e o campo de práticas estatísticas⁶, como se constituem historicamente e que papel desempenham na organização social da produção capitalista e no modo de ação do Estado. Esta compreensão é fundamental para que possamos posteriormente abordar o significado de transformações contemporâneas na produção, distribuição e consumo de estatísticas oficiais com a emergência de novas fontes de dados, métodos, estruturas e agentes no contexto da economia orientada aos dados e do capitalismo de plataforma (Srnicek, 2017).

Do ponto de vista político-administrativo, destacamos que a estruturação do campo estatístico se vincula historicamente ao processo de formação e consolidação dos Estados Nação. Disto resultam dois atributos ontológicos que são constitutivos da estrutura e da lógica de produção e distribuição de estatísticas oficiais, nomeadamente o seu caráter público e sua vinculação à soberania nacional. Com Desrosières (1998), destacamos o papel do Estado e da ciência na construção de espaços nacionais de convenções de equivalências, por meio de investimentos econômicos e cognitivos, que possibilitam o uso de medidas comuns. Com Bourdieu (2014), evidenciamos que a construção destes espaços é o resultado de uma série de despossamentos em favor da acumulação e controle de capital informacional pelo Estado, processo sobre o qual se constitui o campo de práticas estatísticas.

O processo histórico da estruturação do campo estatístico, que abordamos neste capítulo, é relevante para posteriormente demonstrarmos como os processos contemporâneos de dataficação e, em especial, a nova economia política do big data,

⁵ Compreendemos a estatística oficial como uma informação pública produzida e distribuída por instituições de Estado, de acordo com critérios e parâmetros geralmente definidos em lei.

⁶ Quando nos referimos ao campo de práticas estatísticas ou apenas campo estatístico, tratamos especificamente do campo constituído por agentes de Estado envolvidos na produção e distribuição de estatísticas oficiais.

orientada por princípios de mercado e regida por empresas privadas de porte global, desafiam aspectos ontológicos e estruturais das estatísticas oficiais. Nomeadamente, a concepção dos dados estatísticos como bens públicos e a soberania nacional sobre a produção de estatísticas oficiais.

Do ponto de vista epistemológico, argumentamos que a ciência estatística apropriada na produção de estatísticas oficiais, ao longo dos Sécs. XIX e XX, desenvolveu-se sob as premissas da interpretação frequentista do cálculo de probabilidades. De acordo com esta interpretação, a probabilidade é uma medida objetiva, baseada em dados empíricos e na observação do ‘mundo real’, determinada por frequências relativas de eventos observados junto a uma população conhecida. Estas premissas constituem os fundamentos das noções de representatividade e inferência da teoria da amostragem, que foram determinantes para o desenvolvimento das práticas estatísticas de Estado. Neste sentido, destacamos o papel de construções cognitivas como a ‘urna de bernoulli’ e a ‘curva normal’ para as aplicações, ferramentas e analogias historicamente construídas pelo campo de práticas estatísticas.

A análise do desenvolvimento histórico e cognitivo da ciência estatística e de sua incorporação nas práticas estatísticas de Estado são aspectos relevantes para posteriormente abordarmos transformações vinculadas ao uso de fontes de big data no processo de produção de estatísticas oficiais no Séc. XXI. O caráter desestruturado e não representativo destas fontes constitui um obstáculo epistemológico para as práticas tradicionais dos estatísticos de Estado, fundamentadas na interpretação frequentista da teoria da amostragem e na abordagem dedutiva do desenho de inquéritos. Conforme veremos, a incorporação de métodos baseados na modelagem indutiva e na abordagem bayesiana produz efeitos relevantes na estruturação contemporânea do campo estatístico, na medida em que estes métodos emergem associados ao perfil profissional do cientista de dados.

Finalmente, abordamos ao longo da investigação, como, a partir de sua matriz europeia e estadunidense se dá o desenvolvimento desigual e combinado da estatística no Brasil. Para tanto, utilizamos a abordagem crítica da teoria da dependência e nos referenciamos no trabalho de pesquisadores brasileiros que investigaram o desenvolvimento histórico dos sistemas estatísticos no país.

Apoiados neste quadro analítico, discutimos a seguir a estruturação do campo estatístico e as transformações históricas dos regimes estatísticos. Ao longo do

capítulo, as interfaces entre a economia política e a epistemologia da estatística serão abordadas, bem como os demais aportes teóricos e metodológicos que sustentam nossa análise.

1.1 O Estado Territorial e o regime estatístico de contabilização

A palavra estatística deriva do termo alemão *statistik*, que significa ciência do Estado. Essa origem etimológica não é fortuita. De acordo com as análises sobre o Estado realizadas por Pierre Bourdieu e Michel Foucault, podemos situar as origens da estatística administrativa em um conjunto de saberes e práticas implicados no processo de centralização dos Estados Territoriais na Europa ocidental entre os Sécs. XVI e XVIII. Ambos autores, sob óticas distintas, destacam o papel essencial desempenhado pela estatística neste contexto. Seja na racionalização do exercício do poder soberano, de acordo com uma nova razão de Estado (Foucault, 2007). Seja no processo de concentração de capitais, que conduz à constituição dos Estados Territoriais (Bourdieu, 2014)⁷.

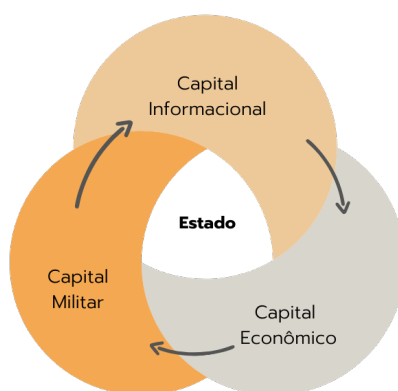
A análise do Estado de Bourdieu (2014) parte de uma revisão crítica do modelo sociológico de formação e transição do Estado de Norbert Elias. De acordo com Elias (2001), o processo de formação do Estado Territorial compreende um duplo monopólio do soberano: monopólio sobre a violência legítima, a partir da formação de um exército regular, e monopólio econômico, com a centralização da cobrança de impostos. Bourdieu (2014), por sua vez, enfatiza a concentração de capitais informacionais, possibilitada pela estatística, como um elo que articula os vetores de concentração dos capitais militar e econômico no Estado Territorial, de acordo com uma relação de causalidade circular.

⁷ Foucault e Bourdieu utilizam diferentes nomenclaturas em seus modelos de formação e transição do Estado. Neste trabalho, adotamos as noções de Estado Territorial e Estado Nação. A primeira designa monarquias absolutas europeias, que se consolidaram entre os Sécs. XVI e XVIII, centralizadas sobre o poder de um soberano. Aqui, a dominação baseava-se em valores tradicionais e teocráticos e a sucessão do poder era regida por uma lógica dinástica. A segunda refere-se à noção moderna de Estado, que se consolidou entre os Sécs. XIX e XX, na forma de monarquias constitucionais ou repúblicas, fundadas sob uma estrutura jurídica que impõe a soberania sobre um povo num dado território. Enquanto Bourdieu (2014) refere-se ao Estado Territorial como Estado Dinástico, destacando a distinção entre o monopólio privado de capitais da dinastia e o monopólio público do Estado Nação, Foucault (2008) refere-se à Razão de Estado (Estado Territorial) e Governamentalidade (Estado Nação), enfatizando a mudança nos modos de governar implicada nesta transição.

O nascimento do imposto é simultâneo à uma acumulação de capital detido pelos profissionais de gestão burocrática e à acumulação de um imenso capital informacional. É o vínculo entre o Estado e a estatística: O Estado está associado a um conhecimento racional do mundo social. Temos aí relações de causalidade circular – A causa B que causa A – entre a construção de um exército e a construção do imposto, a construção do imposto e a acumulação de capital informacional [estatística] (Bourdieu, 2014, p.273).

Para Elias (2001) e Bourdieu (2014), o aparelho administrativo do Estado Territorial, no topo do qual figurava o soberano, permanecia privado em muitos aspectos, constituindo um corpo de administração patrimonial que operava como “uma extensão da casa do rei” (Bourdieu, 2014, p.186). Enquanto segredos de Estado manuseados por uma elite patrimonialista, a estatística constituía um dos instrumentos que participavam da construção do monopólio privado do soberano, e voltava-se principalmente à catalogação, contabilização e ampliação das forças do Estado. Temos assim a constituição do que intitulamos de um ‘regime estatístico de contabilização’. De acordo com Desrosières (1998, p.28), “fossem destinadas ao rei ou à sua administração, [as estatísticas] eram secretas e vinculadas à prerrogativa real. Não pretendiam informar uma sociedade civil distinta do Estado, nem uma opinião pública autônoma”.

Figura 5 - Modelo de Formação do Estado



Fonte: do autor a partir de Bourdieu (2014).

Em sua análise genealógica, Foucault (2008, p.133) identifica na estatística um dos elementos que contribuíram para o desenvolvimento de uma teoria da arte de governo. Segundo esta nova razão de Estado — vinculada à soberania dos Estados

Territoriais, que se desenvolvem sob o quadro econômico e político do mercantilismo — a sabedoria requerida do soberano passava a englobar “o conhecimento das coisas, dos objetos que podem e devem ser obtidos e a disposição que deve ser empregada para alcança-los”.

Por um lado, a partir do século XVI a teoria da arte do governo estava ligada a todos os desenvolvimentos do aparelho administrativo das monarquias territoriais (a emergência dos aparelhos de governo, dos representantes do governo, etc.) esteve ligada também a todo um conjunto de análises e de saberes que se desenvolveram desde o fim do século XVI e que adquiriram toda a sua amplitude no século XVII; essencialmente esse conhecimento do estado em seus diferentes dados, em suas diferentes dimensões, nos diferentes fatores de seu poder, e foi isso que se chamou precisamente de ‘estatística’, como ciência do estado. Enfim, essa busca de uma arte de governar não pode não ser posta em correlação com o mercantilismo e o cameralismo, que são ao mesmo tempo esforços para racionalizar o exercício do poder, em função precisamente dos conhecimentos adquiridos pela estatística, e que também foram uma doutrina, ou melhor, um conjunto de princípios doutrinários quanto à maneira de aumentar o poder e a riqueza do Estado (Foucault, 2008, p. 133 – 134).

Do ponto de vista da estrutura econômica e produtiva, a centralização de poderes e capitais dos Estados Territoriais na Europa Ocidental está historicamente implicada em um período de transição, em que, segundo Marx (2013, p.515), “a estrutura econômica da sociedade capitalista surge da estrutura econômica da sociedade feudal”. De acordo com Marx (2013), a dissolução dos séquitos feudais, acelerada pela centralização da soberania absoluta, desencadeou dois processos fundamentais para o desenvolvimento das forças capitalistas. Por um lado, viabilizou a acumulação primitiva do capital, por meio do cercamento das terras comuns — processo de expropriação e expulsão de uma parte da população rural, que liberou trabalhadores e seus meios de produção para o nascente capital industrial. Por outro lado, por meio do avanço ultramarino e da instituição do sistema colonial, conduziu à expansão do mercado mundial. Estes fatores integram as condições gerais para o desenvolvimento da manufatura, a qual cria gradualmente uma nova organização do trabalho social e uma forma especificamente capitalista do processo de produção social, que conforma a estrutura social para o posterior advento do capitalismo industrial⁸.

⁸ O período manufatureiro (metade do século XVI até o último terço do século XVIII) caracteriza-se pela transição gradual do trabalho artesanal para o trabalho em fábricas e pelo uso de ferramentas e máquinas simples para a produção de mercadorias. A produção manufatureira reúne um conjunto maior de trabalhadores sob o comando de um mesmo capital, instaura a divisão do trabalho e deste modo

No contexto de expansão do mercado mundial sob a égide do mercantilismo, as colônias dos Estados Territoriais europeus foram igualmente objeto da produção de estatísticas que visavam sobretudo assegurar a soberania da metrópole sobre os territórios ultramarinos. No caso do Brasil colônia, de acordo com Gil (2023, p.1) “nunca houve, de fato, um censo geral de toda a América portuguesa. Eles sempre foram feitos aos ‘pedaços’, seja por capitânias ou estados”. Notadamente, após dois séculos de levantamentos parciais baseados em relatos e listagens que forneciam uma estimativa vaga do total de habitantes e das atividades econômicas, é somente a partir da segunda metade do século XVIII que métodos pré-definidos para estimar a população passaram a ser aplicados. A finalidade original desta estatística administrativa colonial era principalmente listar os homens disponíveis para a guerra, com o tempo esta função também passou a se relacionar com o controle do Estado metropolitano sobre o território e a economia coloniais, por meio de levantamentos populacionais “mais próximos do que chamamos hoje de censo”

Esta mudança não diz respeito apenas ao incremento do interesse lusitano em suas colônias [...] mas, também, à progressiva mudança de visão de mundo sobre soberania, controle e política. Cada vez mais a soberania territorial estava associada à presença de populações nos espaços. Os Estados europeus disputavam suas fronteiras recorrendo às guerras, mas a presença de súditos de um rei – e em grandes quantidades – era cada vez mais um sinal claro de posse territorial. Enumerar os súditos significa saber o tamanho do reino (Gil, 2023, p.2).

Mais do que apenas uma doutrina econômica, para Foucault (2008a, p.8) o mercantilismo corresponde a um determinado modo de organização da produção e do comércio, que é regido por três princípios: “primeiro, o Estado deve enriquecer pela acumulação monetária, segundo, deve se fortalecer pelo crescimento da população, terceiro, deve estar e se manter em um estado de concorrência permanente com as potências estrangeiras”. Analisamos a seguir estes três princípios e como se articulam ao desenvolvimento da estatística administrativa no âmbito dos Estados Territoriais.

Em primeiro lugar, o princípio mercantilista de acumulação monetária pressupõe um papel ativo do Estado na organização direta da economia. Deste quadro, emerge uma concepção planificadora da estatística que constitui o seu cerne. Este traço genético da estatística como instrumento de planificação, perdurará e aflorará em outros períodos históricos, caracterizados por um modo de ação típico

converte a mercadoria em um produto social.

que Desrosières (2008) intitula de ‘Estado Engenheiro’.

Nessa perspectiva, os volumes produzidos e consumidos eram as quantidades essenciais para um Estado engenheiro, que se julgava o responsável direto pela satisfação das necessidades, assim como o diretor técnico de uma empresa se preocupa com os insumos necessários à sua produção. Deste exemplo, resulta a especificidade histórica das estatísticas necessárias ao engenheiro do Estado, comparável à informação essencial ao general de um exército: quantidades produzidas e consumidas, equipamentos e, finalmente, não menos importante, a dos homens. As variáveis demográficas, a taxa de natalidade, a imigração, fazem parte das preocupações desse Estado (Desrosières, 2008, p. 43-44, tradução nossa).

Em segundo lugar, o princípio mercantilista do aumento da população, que de acordo com Porter (1986, p.20) consistia no “critério supremo de um país próspero e bem governado”, fez dos levantamentos populacionais um dos principais objetivos da estatística administrativa que se desenvolveu sob os Estados Territoriais. Conforme vimos, estes levantamentos relacionavam-se à acumulação de capitais informacionais necessários aos processos de monopolização da economia e da força pelo Estado. As informações sobre os súditos e seus bens atendiam à centralização da cobrança de impostos, utilizados, entre outros, para composição de um exército regular. O aparato militar, por sua vez, dependia das estatísticas para assegurar o provimento regular de homens em idade suficiente para fazer a guerra. Finalmente, o modo de organização do governo de acordo com a razão de Estado pressupunha o conhecimento e controle das forças internas do Estado para o exercício do poder de polícia.

Camargo (2016, p.55) chama atenção que não se pode falar propriamente em ‘estatística de população’ no contexto do Estado Territorial, mas de ‘pulosidade’, que “aponta para a natureza e o alcance da planificação na razão de Estado, para a qual o grau de pulosidade é o indicador de riqueza e a medida da política”. Na sociedade de estamentos do Antigo Regime, a única equivalência entre os indivíduos se dava em sua subordinação comum a um soberano que extraia utilidade justamente das posições diferenciais dos súditos. As classificações estatísticas neste contexto,

reforçavam papéis e posições herdados, encontrando sua inteligibilidade na relação entre a estabilidade política de uma ordem internamente diferenciada e a estabilidade cognitiva dos objetos de representação. Por tal razão, o Estado Territorial não conhecerá uma estatística de população propriamente, e sim um levantamento de seus volumes e superfícies, sobre a medida de sua mortalidade, natalidade, sobre as categorias de indivíduos que geram riqueza; elementos que, não por acaso, aparecerão ao lado das estimativas sobre clima, minas, florestas, comércio e finanças (Camargo, 2016, p.53).

Finalmente, o princípio mercantilista de concorrência entre os Estados, de acordo com Foucault (2008a), implica em uma nova arte de governar que agora se manifesta em um 'campo relacional de forças'. Foucault (2008) se refere ao processo histórico de construção de uma política de equilíbrio europeu, cuja "primeira manifestação completa" se dá no tratado de Westfália em 1648⁹. De acordo com Foucault (2008), na nova configuração implicada na construção do equilíbrio europeu, o principal objeto da razão política passa a ser o cálculo das forças do Estado e o problema da nova racionalidade governamental, que daí emerge, diz respeito ao desenvolvimento e conservação de uma dinâmica de forças entre os Estados.

Dois grandes conjuntos de tecnologia política são construídos para por em ação esta nova razão política: o conjunto militar-diplomático e o dispositivo de polícia. O primeiro caracteriza-se por uma dupla instrumentação, "uma diplomacia permanente e multilateral, de um lado e, de outro, organização de um exército profissional" (Foucault, 2008, p.420). O segundo refere-se ao "conjunto dos meios [cálculo e técnica] pelos quais é possível fazer as forças do Estado crescerem, mantendo ao mesmo tempo a boa ordem desse Estado" (Foucault, 2008, p.423). Para Foucault (2008), há uma relação de instrumentação entre os dois conjuntos tecnológicos, militar-diplomático e polícia, que é possibilitada precisamente pela estatística, enquanto ferramenta de racionalização da força interna de cada Estado, de um lado, e, de outro, instrumento para o cálculo da dinâmica de forças entre os Estados.

Esse instrumento comum ao equilíbrio europeu e a organização da polícia é o quê? É a estatística. Para que o equilíbrio seja efetivamente mantido na Europa, é preciso que cada Estado possa, primeiro conhecer suas próprias forças, segundo, conhecer, apreciar as forças dos outros e, por conseguinte, estabelecer uma comparação que possibilitará, justamente, acompanhar e manter o equilíbrio. É necessário portanto um princípio de decifração das forças constitutivas de um Estado [...] por esse domínio do conhecimento que se abre e se funda, se desenvolve nesse momento que é a estatística. Ora, como se pode estabelecer a estatística? Pode-se estabelecê-la justamente pela polícia, [...] porque é justamente o conjunto dos procedimentos instaurados para fazer as forças crescerem, para combiná-las, para

⁹ O acordo de paz inaugura um novo marco nas relações e enfrentamentos entre os Estados, não mais percebidos sob a forma de rivalidades e alianças religiosas e dinásticas, mas em termos de concorrência e composições políticas provisórias entre Estados soberanos. A nova razão governamental que é colocada em prática, entre os Sécs. XVI e XVII, passa a configurar um novo espaço político de relações entre Estados soberanos que, de acordo com Foucault (2008, p.389), implica em um tempo aberto e uma espacialidade múltipla, ou seja, de agora em diante o mundo será constituído de "uma pluralidade de Estados que tem sua lei e seu fim apenas em si mesmos".

desenvolvê-las, é todo esse conjunto, numa palavra, administrativo, que vai permitir que se identifique em cada Estado em que consistem suas forças, onde estão as possibilidades de desenvolvimento; polícia e estatística se condicionam mutuamente, e a estatística é, entre a polícia e o equilíbrio europeu, um instrumento comum. A estatística é o saber do Estado sobre o Estado, entendido como saber de si do Estado, mas também saber dos outros Estados. E é nessa medida que a estatística vai se encontrar na articulação dos dois conjuntos tecnológicos (Foucault, 2008, p.424).

A análise genealógica de Foucault nos permite compreender uma característica originária e estrutural da estatística de Estado. Enquanto medida absoluta da força interna de cada Estado, a estatística emerge vinculada ao princípio de soberania. Enquanto medida relativa da dinâmica de forças entre os Estados, a mesma estatística será tensionada pelo princípio da concorrência. A relação dialética entre a força centrípeta da soberania e a força centrífuga do espaço concorrencial global resulta em uma dinâmica de lutas que, conforme veremos, é uma das características constitutivas do campo de práticas estatísticas que emergirá no Séc. XIX.

Finalmente, da análise dos três princípios do mercantilismo identificados por Foucault (2008a) — acumulação monetária, aumento da população e quadro concorrencial — podemos isolar três traços genéticos da configuração do regime estatístico de contabilização. Em primeiro lugar, a estatística nasce como um instrumento de planificação voltado à intervenção direta do Estado na economia. Em segundo lugar, constitui uma ferramenta de medição e controle das forças internas do Estado vinculada ao aparato policial e ao princípio da soberania. Em terceiro lugar, estabelece um meio de comensuração e comparação das forças relativas entre Estados posicionados em um campo de concorrência permanente.

No que diz respeito às construções cognitivas que contribuíram para a constituição de um corpo de práticas estatísticas nos Estados Territoriais, a estatística administrativa resulta da combinação dos principais elementos de duas tradições distintas. Por um lado, a quantificação e matematização de fenômenos sociais (em especial nascimentos e óbitos) por meio de registros numéricos introduzido pela aritmética política inglesa¹⁰. Por outro, o uso da forma tabular para a exposição de

¹⁰ Na Inglaterra, verifica-se a emergência no Séc. XVII de um princípio de descrição quantitativa aplicada aos domínios da administração e política de Estado. A compilação e análise de registros de batismos, casamentos e óbitos por John Graunt (1620-1674), deu margem ao desenvolvimento deste novo método de raciocínio que William Petty (1623-1687) batizou de aritmética política (Desrosières, 1998). Petty compreendia a aritmética política como a aplicação do método científico de Francis Bacon (1561-1626) à arte de governo. Na filosofia empirista de Bacon, ele encontrara um paralelo entre o

dados qualitativos relativos à administração do Estado, de acordo uma taxonomia geral, concebida pela *statistik* alemã¹¹.

Desrosières (1998, p.24) distingue as tradições estatísticas inglesa e alemã tanto por suas diferenças cognitivas — por um lado, o aspecto quantitativo da abordagem inglesa e, por outro, a abordagem qualitativa da abordagem alemã — como pelo caráter prático dos aritméticos políticos contra o perfil generalista dos acadêmicos alemães. Os primeiros são pioneiros de um novo papel social que então tomava forma, o dos especialistas que “ofereciam uma linguagem articulada com precisão, enquanto os estatísticos alemães, que se identificavam com o Estado, ofereciam uma linguagem geral e abrangente”. A articulação deste duplo perfil científico/burocrático, do qual é possível identificar traços indiciais nas tradições originárias da estatística administrativa, constituirá uma marca distintiva do habitus do estatístico de Estado.

Conforme veremos na próxima seção, sob os mesmos condicionantes históricos, políticos e econômicos em que se originou a estatística administrativa na Europa Ocidental, desenvolvia-se a partir do Séc. XVII uma nova forma de raciocínio para lidar com problemas relacionados à incerteza e ao acaso. Em uma palavra, paralelamente à razão de Estado, emergia nos campos filosófico e científico uma nova razão probabilística que ensejará um novo significado cognitivo e científico da estatística.

1.2 A Razão probabilística e os fundamentos cognitivos da estatística

O desenvolvimento epistemológico da ciência estatística tem suas origens históricas no cálculo de probabilidades. Podemos dizer que a ciência estatística é uma

corpo natural e o corpo político. À semelhança do método científico, que se valia do uso de números, pesos e medidas para o estudo da natureza, Pety defendia a sua utilização pela ciência política, de forma a orientar os governantes em seu conhecimento sobre o território e seus habitantes (Porter, 1986).

¹¹ A *statistik* termo cunhado por Achenwall (1719-1772) no Séc. XVIII, designava um método originalmente desenvolvido por Conring (1606-1681) no Séc. XVII. Tal método consistia no uso de tabelas para sistematizar e apresentar uma série de informações qualitativas e descritiva sobre os Estados alemães. Concebida como uma ciência do Estado, a *statistik* dedicava-se a reunir de forma objetiva um conjunto de informações essenciais para orientar os príncipes soberanos dos Estados alemães. O sentido empregado por Achenwall e pela escola de estatística da Universidade de Göttingen, que se disseminará na Europa ao longo do Séc. XVIII, concebia a estatística como um saber puramente descritivo e qualitativo, um método de classificação baseado em uma lógica de taxonomia aristotélica (Desrosières, 1998).

filha da matemática probabilística. Entre os historiadores da ciência há certo consenso de que o cálculo matemático de probabilidades nasceu na Europa em meados do séc. XVII com as contribuições iniciais de Pascal e Fermat (Hacking, 2007; Daston, 1988)¹². Seu desenvolvimento estava ligado à resolução de problemas de equidade que se originaram no comércio e nos jogos aleatórios, por um lado e por outro, nas teorias do direito e das ciências morais, que relacionavam o conceito de probabilidade com a racionalidade de uma crença, conhecimento, julgamento ou decisão.

De acordo com Hacking (2007, p.169), ao longo dos Sécs. XVI e XVII os fenômenos aleatórios passaram a fazer parte de um novo *corpus* de problemas de divisão e equidade que surgiram no comércio, dos quais “ninguém era capaz de resolver aqueles que estavam relacionados ao acaso”. Para Coumet (1970), a emergência de novas modalidades comerciais e financeiras no Séc. XVII demandou o desenvolvimento de cálculos matemáticos que lidassem com a questão do risco e da incerteza. Essas modalidades relacionavam-se, por exemplo, a uma nova concepção de contratos aleatórios, que tratavam da troca de um valor presente certo, contra um valor futuro incerto: como os seguros, as rendas vitalícias, a especulação sobre safras agrícolas e determinadas condições de empréstimos a juros¹³.

O raciocínio sobre a probabilidade se desenvolveu quase como um sinônimo de uma nova concepção de racionalidade¹⁴. No séc. XVII, muitos pensadores

¹² De acordo com Franklin (2016), apesar dos primeiros cálculos matemáticos de probabilidade que lidaram com fenômenos estocásticos situarem-se no Séc. XVII, as origens filosóficas do conceito de probabilidade, entretanto, são anteriores e remontam a antiguidade greco-romana e medieval. A palavra probabilidade origina-se do Latim, *probabilitas* que, por sua vez, relaciona-se ao verbo *probare*, provar. Originalmente, portanto, uma probabilidade é um argumento, opinião ou crença para a qual existem boas provas (Shafer, 1990). Essa origem qualitativa é defendida por Franklin (2016), para quem a probabilidade tem raízes na oratória grega e direito romano, sendo transmitida pela teologia medieval, para finalmente ser formalizada em cálculo por matemáticos que eram ao mesmo tempo fortemente influenciados pela teoria do direito. De acordo com Franklin (2016, p.36): “Os criadores da probabilidade matemática foram todos advogados profissionais (Fermat, Huygens, de Witt) ou filhos de advogados (Cardano e Pascal), e assim tiveram algum contato com os conceitos mais amplos do pensamento jurídico”.

¹³ O surgimento de novos problemas no comércio e nas finanças, refletem transformações na base material da economia europeia que, conforme analisamos anteriormente, caracterizam uma transição do modo de produção feudal para o modo de produção capitalista. De acordo com Marx (2017), algumas características desta transição referem-se especificamente a movimentos do capital financeiro e comercial. Por um lado, o capital usurário que operara de forma restrita no medieval passava a funcionar em condições modificadas na forma de capital portador de juros, o que antecede a constituição do moderno sistema de crédito (Marx, 2017, p.705). Por outro, há um forte incremento no capital comercial com a expansão ultramarina e ‘descobrimientos’ geográficos (Marx, 2017, p.402) e com o desenvolvimento de um mercado interno vinculado à produção manufatureira.

¹⁴ De acordo com Hacking (2007), na idade média a escolástica referia-se à probabilidade exclusivamente como uma opinião ou crença comprovada pela autoridade das escrituras. Essa noção

buscavam encontrar um caminho do meio entre o dogmatismo e o ceticismo, renunciando à certeza absoluta e reconhecendo no conhecimento provável o conhecimento autêntico. Neste sentido, o pensamento probabilístico se tornou central na aliança entre a matemática e a racionalidade, expressa na ideia de um cálculo razoável (Daston, 1988).

No âmbito das teorias do direito e ciências morais, o pensamento probabilístico referia-se aos processos de tomada de decisão por indivíduos racionais com base em provas e indícios (Daston, 1988). Notadamente, a interpretação desenvolvida pelo filósofo e matemático alemão Gottfried Wilhelm Leibniz (1646-1716) concebia a probabilidade como uma espécie de jurisprudência natural. Diferente das concepções fundadas nos jogos aleatórios, sua teoria abordava a questão do acaso do ponto de vista de um conhecimento racional apoiado em evidências e propunha a noção de ‘graus de probabilidade’ como sinônimo de ‘graus de crença’ (Hacking, 2007).

De acordo com Hacking (2007), a probabilidade que emerge no tempo de Pascal tem a dupla face de Jano¹⁵. De um lado é objetiva, ocupa-se com leis

de probabilidade como testemunho bíblico, foi desafiada por Blaise Pascal (1623-1662) na *Lógica de Port-Royal* (1662), obra que segundo Foucault (1999) marca uma ruptura epistêmica no pensamento ocidental. Para Hacking (2007), a importância fundamental da nova lógica para a emergência da probabilidade, foi o desenvolvimento do conceito de evidência. Uma proposição passou a ser provável desde que houvessem evidências para sustentá-la, não mais recorrendo-se à autoridade das escrituras mas ao testemunho da natureza. A autoridade passou a ser lida em sinais da natureza, os quais deveriam ser frequentes para serem confiáveis. Deste modo a probabilidade, enquanto uma opinião ou crença provável, passaria a ser relacionada com regularidades e frequências.

¹⁵ As contribuições de Pascal para o desenvolvimento do pensamento e cálculo probabilístico são relevantes para as origens do caráter ambíguo do conceito de probabilidade, enquanto crença e frequência. Dois casos ilustrativos desta dicotomia são o ‘problema das partidas’ e a ‘aposta de Pascal’. O problema das partidas, proposto a Pascal por Chevalier de Méré (1607-1684) consistia em um problema matemático oriundo da idade média e que até então não encontrara uma solução adequada. O problema tratava de como dividir de forma justa entre dois jogadores, o valor da aposta de um jogo de várias partidas, que por algum motivo tivesse que ser interrompido antes do final. Em sua resolução do problema das partidas, Pascal estabeleceu um cálculo de expectativas baseado em chances iguais para cada jogador. A solução consistia no cálculo de um valor justo para a divisão da aposta, de forma que a alternativa entre ficar ou sair do jogo fosse indiferente para os jogadores. O cálculo de Pascal, fundado nas noções de chances iguais e valor justo, consistia na divisão da expectativa pela soma total da aposta. Tornando público nas cartas entre Pascal e Fermat, o feito foi consagrado como o evento inaugural do cálculo de probabilidades. Uma segunda contribuição de Pascal para o pensamento probabilístico, conhecida como a “Aposta de Pascal”, relaciona probabilidade, crença e decisão racional. Na passagem intitulada *infini-rien* na obra *Pensées*, Pascal propõe um jogo sobre a existência de Deus. A aposta implicaria em decidir entre levar uma vida ascética ou de prazeres na terra, contra a expectativa de uma vida eterna no céu ou no inferno, caso Deus exista. Pascal conclui que o homem racional tem mais a ganhar vivendo de acordo com a suposição de que Deus exista, mesmo que seja impossível à razão humana comprovar tal suposição. A polêmica aposta de Pascal demonstrou que a probabilidade poderia ser aplicada a outros modos de conjectura e contribuiu para popularizar a compreensão de que a estrutura dos jogos aleatórios poderia ser transferida para problemas de tomada de decisão sob incerteza. (Hacking, 2007)

estocásticas e frequências estáveis. De outro lado, é subjetiva, dedica-se a acessar graus razoáveis de crença para a tomada de uma decisão. Para Desrosières (1998), a abordagem do acaso, que nasce do pensamento sobre a estrutura dos jogos aleatórios, tenderá a relacionar o cálculo de probabilidades com frequências de eventos observáveis, propriedades físicas da natureza e estados do mundo. Por outro, a abordagem epistêmica que deriva da teoria do direito e ciências morais, tenderá a equiparar graus de probabilidade, que decorrem da observação de evidências, com graus de crença ou estados do espírito.

Na perspectiva epistêmica, probabilidade é um grau de crença. A incerteza do futuro, ou a incompletude de nosso conhecimento do universo, leva a apostas no futuro e no universo, e as probabilidades fornecem a uma pessoa razoável regras de comportamento quando falta informação. Mas na visão frequentista, a diversidade e o risco fazem parte da própria natureza, e não se refere simplesmente ao resultado de um conhecimento incompleto. Eles são externos à humanidade e fazem parte da essência das coisas. Cabe à ciência descrever as frequências observadas (Desrosières, 1998, p.7, tradução nossa).

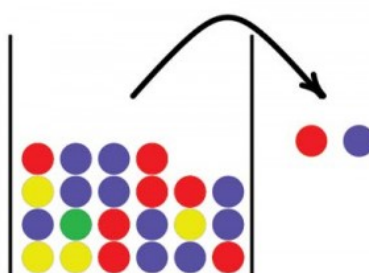
De acordo com Daston (1988), a interpretação clássica da probabilidade, que se desenvolve no Séc. XVIII, ignorava a distinção entre probabilidades objetivas (frequências) e probabilidades subjetivas (crenças). Esta distinção viria a ser evidenciada apenas no Séc. XIX. Em seus trabalhos, os primeiros probabilistas acomodaram as duas abordagens, produzindo importantes sínteses teóricas.

O teorema de ouro de Jakob Bernoulli (1655-1705), que relaciona graus de crença à frequências, é o principal exemplo desta indistinção da interpretação clássica da probabilidade. Posteriormente conhecido como 'Primeiro Teorema Fundamental da Probabilidade' ou 'Lei dos Grandes Números', a proposição de Bernoulli vale-se do ponto de vista frequentista implícito nos jogos aleatórios para demonstrar que a frequência de ocorrência de um evento, com uma dada probabilidade, tende a essa probabilidade quanto maior for o número de repetições das tentativas para obter o evento (Daston, 1988).

A Lei dos Grandes Números de Bernoulli possibilitava estimar a probabilidade de ocorrência de um determinado resultado a partir de uma causa conhecida. No exemplo clássico da 'urna de Bernoulli' — retirar uma bola preta de uma urna com uma composição conhecida de bolas brancas e pretas — a probabilidade é obtida por meio da razão das frequências entre casos favoráveis e casos possíveis, quando o número de casos tende ao infinito. A urna de Bernoulli é uma construção cognitiva

fundamental não apenas para a matemática probabilística, mas especialmente para a estatística de Estado. A futura construção de espaços nacionais unificados por meio da completa enumeração de seus elementos, pode ser compreendida como uma analogia político-administrativa da urna probabilística. É a construção do Estado como uma 'urna probabilística nacional', por meio de operações censitárias, que possibilitará o posterior desenvolvimento da inferência estatística aplicada a inquéritos amostrais (Desrosières, 2008).

Figura 6 - A Urna de Bernoulli



Fonte: <https://chance.amstat.org/2015/02/urns-and-forensics/>

O teorema de Bernoulli, entretanto, não apresentava uma solução para como estimar a probabilidade de eventos cujas causas eram desconhecidas. Retornando ao exemplo da urna, por exemplo, como calcular a probabilidade de retirar uma bola preta de uma urna da qual não se conhece a composição. Foi a este tipo de problema que se dedicou Thomas Bayes (1702-1761). Sua formulação implicou em uma inversão do cálculo de probabilidades.

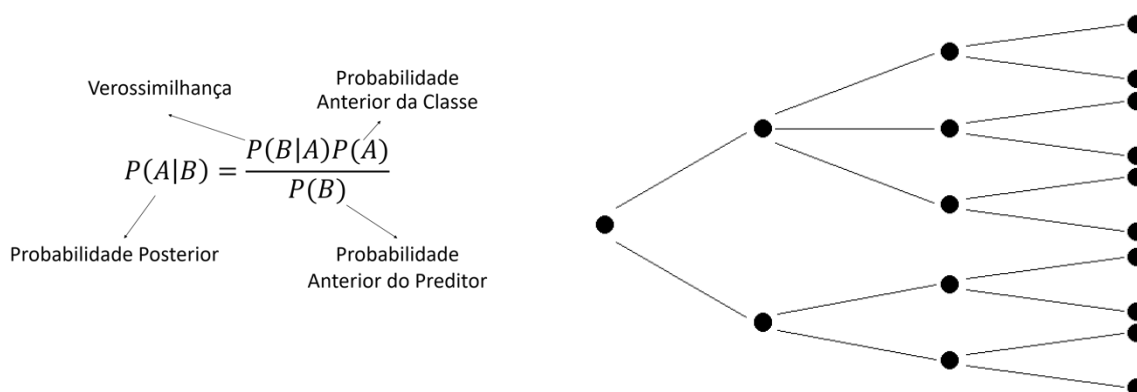
O teorema de Bayes se baseia na ideia de que a probabilidade da ocorrência de um evento de causas desconhecidas é determinada pela sua probabilidade condicional e pela probabilidade a priori do evento. A probabilidade condicional ou inversa é a probabilidade de um evento ocorrer dado que outro evento já ocorreu. A probabilidade a priori é a probabilidade de um evento ocorrer antes de se observar quaisquer dados. O método proposto por Bayes consistia no registro de informações parciais sobre eventos conhecidos para estimar a probabilidade a priori de uma causa desconhecida. A probabilidade à priori é, portanto, subjetiva e incorpora crenças, conhecimentos prévios e dados parciais que o probabilista utiliza para inferir uma

probabilidade condicional à posteriori, que permite aprimorar a racionalidade de uma decisão em cenário de incerteza e baseada em conhecimento incompleto¹⁶.

O Teorema de Bayes marca um importante momento no desenvolvimento do pensamento probabilístico, ao possibilitar sua aplicação a um novo conjunto de problemas. Para Desrosières (1998, p.45-60) esta formulação estabeleceu uma distinção entre decisões baseadas em avaliações não quantificáveis e decisões baseadas em observações repetidas. A formulação de Bayes teve, entretanto, uma repercussão muito limitada entre seus contemporâneos, tendo sido retomada apenas em meados do século XX por uma nova geração de estatísticos bayesianos (Stiegler, 1986).

Um dos aspectos que confere potência e versatilidade ao teorema de Bayes é a possibilidade de atualizar as crenças e probabilidades sobre uma hipótese com base em novos dados e evidências na medida em que surgem. Esse processo de atualizações sucessivas, que pode envolver cálculos complexos em modelos que envolvem muitas variáveis, foi impulsionado pelo desenvolvimento dos computadores eletrônicos no Séc. XX e por construções cognitivas como a árvore ou rede bayesiana, que constituem a base cognitiva de novos métodos de análise de dados não representativos como as fontes de big data.

Figura 7 - Teorema de Bayes e Árvore Bayesiana



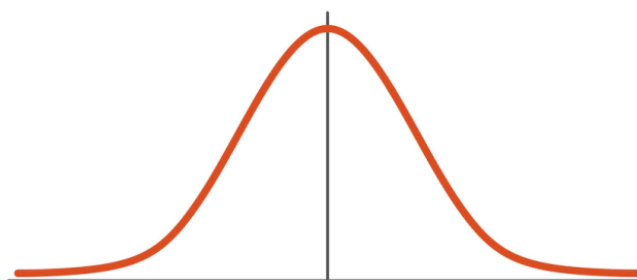
Fonte: Santos e Dias (2021)

¹⁶ De acordo com Desrosières (1998, p.59) a inferência bayesiana postula uma ideia de causalidade que tem aderência na filosofia de David Hume (1711-1776)¹⁶: “De fato, a ideia de causalidade postulada implicava que causas semelhantes podem levar a consequências semelhantes e que eventos futuros podem ser conectados a observações passadas de modo a circunscrever sua incerteza [...] Nesse aspecto, aderiu à filosofia de Hume”. Conforme veremos a relação da inferência bayesiana com o problema da indução é relevante para compreender a sua retomada no Séc. XXI na estatística aplicada a novos métodos de análise de dados baseados em modelos estatísticos.

A despeito da consagração da inferência bayesiana mais de dois séculos após à sua formalização, o conceito de probabilidade inversa foi difundido principalmente pela Lei dos Erros de Pierre Simon Laplace (1749-1827). Com Laplace, entretanto, o cálculo de probabilidade volta a ser associado à lei dos grandes números e à distribuições de frequências. Laplace utilizou a probabilidade inversa para abordar um problema que afligia o mundo da física, o de como combinar observações astronômicas imprecisas e realizadas em diferentes condições para obter a melhor estimativa possível. A inversão aqui consistia em partir de observações astronômicas distintas (resultado conhecido) para estimar o ponto astronômico efetivamente observado (causa desconhecida).

Concebendo as frequências de erros de observações astronômicas como distribuições aleatórias e simétricas, Laplace utilizou a Lei dos Grandes Números de Bernoulli e o método dos mínimos quadrados de Carl Friedrich Gauss (1777-1855) para formalizar uma Lei dos Erros ou Teorema do Limite Central. De acordo com a formalização de Laplace, a média das amostras de uma população tenderia a uma distribuição gaussiana ou distribuição normal. Esta distribuição corresponderia, portanto, ao comportamento normal do efeito agregado de experiências aleatórias, independentes e semelhantes quando o número de experiências é muito alto. A curva gaussiana ou curva normal constitui o fundamento sobre o qual se desenvolveu a estatística descritiva do Séc. XIX e a estatística inferencial do Séc. XX. Esta construção cognitiva exercerá considerável influência e fascínio não apenas sobre o pensamento social e científico, mas sobretudo em práticas de governo inspiradas na noção de 'normalidade', que emergem no Séc. XIX, e em especial no campo estatístico.

Figura 8 - A Curva Normal



De acordo com Daston (1988, p.371), a interpretação clássica da probabilidade é o produto de mentes formadas sobre os ideais do iluminismo. Neste contexto, o cálculo de probabilidades apoiava-se na noção de 'bom senso' do homem esclarecido, capaz de orientar racionalmente seus julgamentos e ações com base nos dados da realidade: "a crença de que os estados mentais subjetivos foram formados à imagem da experiência objetiva, permitiu aos probabilistas igualar indiferença epistemológica e física". A exemplo do teorema de Bayes, utilizar critérios subjetivos para estimar probabilidades, ou nas formulações de Bernoulli e Laplace, tratar indistintamente graus de crença e frequências. Conforme veremos, no Séc. XIX, a distinção entre probabilidades objetivas e subjetivas será estabelecida, relegando às últimas um status de não cientificidade, até sua retomada no Séc. XX pelos estatísticos bayesianos.

A aplicação do raciocínio matemático probabilístico a problemas relativos à moral, ao direito, à política, à economia e ao governo era uma pretensão comum entre os probabilistas clássicos. Para Daston (1988), as ciências morais do Séc. XVIII concebiam a racionalidade como um cálculo interno de operações mentais que a probabilidade poderia quantificar e tornar explícito. Neste sentido, o cálculo de probabilidades aplicado aos problemas humanos constituía uma tentativa de orientar o comportamento de homens racionais e prudentes, alternando descrição e prescrição. De um lado, pretendiam revelar a ordem do pensamento e ação humanas e de outro buscavam interferir em seu funcionamento.

Bernoulli foi um dos pioneiros na concepção de uma matemática social que ensaiava a aplicação do cálculo de probabilidades aos problemas da vida civil. Laplace construiu uma elaborada analogia entre os reinos natural e social e defendia que ambos eram governados pelas mesmas causas e deveriam ser estudados pelos mesmos métodos (Daston, 1988). Condorcet (1743-1794) defendeu a aplicação do cálculo de probabilidades aos problemas políticos e jurídicos e dedicou-se a uma álgebra da sociedade que expressasse em termos probabilísticos as decisões de júris e assembleias (Desrosières, 1998). Abraham de Moivre (1667-1754) aplicou o cálculo de probabilidades à estatística de mortalidade para atender à problemas originados nos ramos financeiros dos seguros e rendas vitalícias (Daston, 1988).

As iniciativas dos probabilistas clássicos nos campos do direito, da política e da economia somam-se a uma tradição de descrição social que emergiu em meados do Séc. XVIII entre homens movidos pela filosofia do Iluminismo e pela doutrina

política e econômica do liberalismo. Por um lado, a realização de estudos políticos, econômicos e sociais à partir dos registros de nascimentos e tábuas de mortalidade generalizou-se a partir da tradição dos aritméticos políticos, e, por outro, a produção de inquéritos independentes, difundida especialmente pelos médicos sanitaristas, passou a utilizar dados estatísticos em estudos que associavam patologias físicas e morais às características naturais, econômicas e sociais (Desrosières, 1998).

Nestes estudos e levantamentos, realizados fora do aparato estatal, a palavra estatística foi gradualmente assumindo uma nova conotação, relacionada a um sistema de descrição quantificada, à medida em que reunia adeptos do campo científico. Ao mesmo tempo, o cálculo de probabilidades passaria a revelar gradualmente novas possibilidades de aplicações empíricas, principalmente quando associado às frequências estatísticas. De acordo com Porter (1986), de uma prática relacionada estritamente ao aparato administrativo do Estado, a estatística passaria gradualmente a ser concebida como um empreendimento perseguido por cientistas, homens de negócios e profissionais liberais (Porter, 1986, p.17).

Como vimos, no âmbito do regime estatístico de contabilização das monarquias territoriais, as informações estatísticas constituíam segredos de Estado, manuseados por uma elite a serviço do monopólio privado de um soberano. É apenas com a derrocada das monarquias absolutas, a partir das revoluções burguesas, revoluções coloniais e contrarrevoluções, ao longo do Séc. XIX, que os dados estatísticos passarão a constituir informações públicas, cada vez mais acessíveis em meios impressos. Por outro lado, a construção do Estado Nação como um espaço social homogêneo, em contraposição à sociedade de estamentos do Estado Territorial, ensejará a construção de um novo espaço de medidas comuns, por meio de uma série de padronizações, classificações e convenções sustentadas pela autoridade da lei e da ciência.

É precisamente no contexto de formação do Estado Nação que duas novas concepções da estatística poderão emergir. Por um lado, a estatística oficial, concebida como um bem público e um espelho da nação, produzida e distribuída por novas burocracias, que emergem e se diferenciam de acordo com um modo de reprodução burocrático, e por outro lado, a concepção cognitiva da estatística social, por meio da aplicação de ferramentas da matemática probabilística na análise de frequências de dados populacionais. Nas próximas seções, analisaremos como a confluência destas duas concepções da estatística, que se interceptam na construção

do Estado Nação, são determinantes para a estruturação do campo de práticas estatísticas.

1.3 O Estado Nação e a transmutação da estatística em bem público

Conforme vimos, há uma homologia histórica entre a origem da estatística e a origem do Estado, processo que situamos no contexto de centralização das monarquias territoriais europeias dos Séc. XVI ao XVIII. Ao discutir estas origens históricas, identificamos como a estatística administrativa veio a constituir um instrumento fundamental para a formação do Estado Territorial, possibilitando a concentração de recursos informacionais necessários ao exercício do poder soberano. Nesta seção, analisaremos a transformação das práticas estatísticas em sua transição para o Estado Nação.

A formação dos Estados Nação ocidentais, na Europa e nas Américas, se inicia no final do Séc. XVIII e se estende ao longo do Séc. XIX, em meio a um contexto atravessado por uma série de revoluções: revolução industrial, revoluções burguesas e nacionais europeias, e revoluções pela independência nas Américas. Este processo histórico corresponde ao mesmo tempo à emancipação da estrutura econômica da sociedade capitalista, com o advento do capitalismo industrial e com a crise das tradicionais legitimações das monarquias territoriais (dinastia, ordenação divina e direito tradicional). Do ponto de vista da racionalidade do Estado, verifica-se a reformulação dos pressupostos teóricos mercantilistas com a emergência do liberalismo econômico dos fisiocratas franceses e utilitaristas ingleses. Enquanto o Estado Territorial constituiu uma modalidade de organização social de transição para o capitalismo, o Estado Nação corresponderá à formação de um Estado capitalista que funciona como nação (Poulantzas, 2015).

De acordo com o modelo de transição do Estado de Bourdieu (2014), a progressiva acumulação de diferentes espécies de capitais — econômico, físico, militar, informacional — que na centralização do Estado Territorial, realizava-se em benefício do rei e da casa real, conduzirá, em um segundo momento, a uma transformação qualitativa de uma série de recursos em capitais públicos no Estado Nação: força pública, finanças públicas, estatísticas públicas, etc. Para Bourdieu (2014, p.268), “esse processo de concentração é ao mesmo tempo um processo de

separação, constituindo um monopólio na base de um desapossamento”, daqueles que não participam ou não estão do lado do Estado.

De acordo com Bourdieu (2014), um dos elementos centrais da transição para o Estado Nação é a passagem de um modo de reprodução de base dinástico/familiar, cuja legitimidade sustentava-se na linhagem hereditária, para um modo de reprodução burocrático, cuja autoridade é baseada na competência dos funcionários do Estado e na aquisição de capitais culturais no sistema escolar. Essa passagem conduz a uma substituição da antiga nobreza por uma ‘nobreza do Estado’, processo que se dá por meio da distribuição do poder central em cadeias de autoridade. Esta distribuição, por sua vez, é acompanhada pela diferenciação de distintos campos de práticas (político, administrativo, jurídico, intelectual, etc.), que lutam entre si para que seu poder seja o poder legítimo. Neste processo, o Estado, “se constitui progressivamente como uma espécie de metacampo de um campo em que se produz, se conserva, se reproduz um capital que dá poder sobre as outras espécies de capital” (Bourdieu, 2014, p.267).

Para Bourdieu (2014), a constituição do Estado como um campo relativamente autônomo requer a constituição do público como um espaço político legítimo, que confere autoridade ao Estado em sentido restrito — compreendido como o seu corpo político e burocrático. No Estado Nação, este processo se faz ao mesmo tempo acompanhar por um trabalho de unificação do espaço social do território e da nação. A nação é assim compreendida, uma construção e manifestação ampliada do Estado.

O Estado no sentido restrito, Estado 1 (administração, forma de governo, conjunto de instituições burocráticas etc.), se faz fazendo o Estado em sentido amplo, Estado 2 (território nacional, conjunto de cidadãos unidos por relações de conhecimento, que falam a mesma língua, portanto, aquilo que se põe a noção de nação). Portanto, o Estado 1 se faz fazendo o Estado 2. Eis a fórmula simplificada. De modo mais rigoroso, a construção do Estado como campo relativamente autônomo exercendo um poder de centralização da força física e da força simbólica, e constituído assim em objeto de lutas, está inseparavelmente acompanhada pela construção do espaço social unificado que é de sua alçada (Bourdieu, 2014, p.176).

De acordo com Bourdieu (2014, p.38), para que o Estado Nação se constitua em um princípio da ordem e do poder público são necessárias não apenas as suas formas físicas evidentes mas sobretudo as suas formas simbólicas inconscientes. Assim, “uma das funções mais gerais do Estado é a produção e canonização das classificações sociais, não por acaso há um vínculo entre o Estado e a estatística”. Os princípios de classificação do Estado constituem a essência do trabalho estatístico,

funcionando como estruturas estruturantes “capazes de serem aplicadas a todas as coisas do mundo, e em especial às coisas sociais” (Bourdieu, 2014, p.227). É por meio deste trabalho de classificação e universalização que o Estado “tem o poder extraordinário de produzir um mundo social ordenado sem necessariamente dar ordens, sem exercer coerção permanente”. (Bourdieu, 2014, p.228-229)

Seguindo o modelo de Bourdieu, podemos compreender como a transição para um modo de reprodução burocrática do Estado, implicou em uma transformação qualitativa da estatística enquanto ferramenta de governo. Conforme vimos, a estatística administrativa do Estado Territorial limitava-se a uma catalogação das forças do Estado em benefício do monopólio privado da soberania. Já a estatística oficial do Estado Nação se vincula a um novo papel de produção do espaço social unificado da nação, em que agora se sustenta a legitimidade do Estado, o que faz por meio da ‘produção e canonização de classificações oficiais’. Este novo papel implica necessariamente em uma transmutação da estatística em um ‘bem público’, pois para que as classificações oficiais tenham eficácia social é necessário que se constituam em uma referência comum e legítima, na qual o conjunto de pessoas unificadas em um mesmo território nacional se reconheçam.

Figura 9 - Modelo de Formação do Estado-Nação



Fonte: do autor a partir de Bourdieu (2014).

Em sua definição de público, Bourdieu (2014) destaca dois aspectos fundamentais: a oposição ao particular/singular e a oposição ao escondido/invisível. Em oposição ao particular, o público se refere ao universal e às ações atribuídas aos representantes da coletividade, personalidades oficiais que agem oficialmente em

nome do universal. Em sua oposição ao invisível, o público é aquilo que se opera à vista de todos, diante de uma audiência universal.

Esta dupla definição nos ajuda a compreender como a transmutação qualitativa da estatística no Estado Nação se faz acompanhar de dois processos coerentes com a construção do público como um espaço político legítimo para a produção e disseminação de classificações sociais universais. Em primeiro lugar, verifica-se a gradual diferenciação e autonomização de um campo estatístico público, ao qual o metacampo do Estado atribui oficialmente a ação legítima de produção de estatísticas oficiais. Em segundo lugar, as estatísticas oficiais, passam a ser distribuídas, publicadas e acessíveis à uma audiência pretensamente universal, formada pelo conjunto dos cidadãos. Deste modo, as estatísticas do Estado Nação deixam de ser o ‘espelho secreto do príncipe’, para se constituir em um ‘espelho público da nação’ (Desrosières, 1998). A primeira república francesa fundada após a revolução de 1789, constitui um caso ilustrativo deste processo:

A unificação da nação foi acompanhada por uma ampla difusão de informações [...] as estatísticas passaram de manuscritos enclausurados em arquivos administrativos para impressos teoricamente destinados a um grande público. Essa mudança estava ligada ao fato de que o Estado republicano, tendo se tornado literalmente a coisa pública, passou a representar a sociedade em sua totalidade, por meio da representação eleitoral, e também por meio das estatísticas, que passaram a ser o espelho da nação’ e não mais simplesmente o ‘espelho do príncipe’. O objetivo de oferecer à sociedade um espelho de si mesma por meio de uma rede de pesquisas foi o primeiro objetivo do novo ‘bureau de estatísticas da República’, criado em 1800 (Desrosières, 1998, tradução nossa, p.34).

Para Hobsbawm (2004, p.100), “em seu sentido moderno e basicamente político, o conceito de nação é historicamente muito recente”, tendo emergido na era das revoluções burguesas. Doravante a legitimidade do Estado baseava-se na identificação com um povo e um território, “assim considerada, a nação era o corpo de cidadãos cuja soberania coletiva os constituía como um Estado concebido como sua expressão política [...] A equação nação = Estado = povo, povo soberano, vinculou indubitavelmente a nação ao território”. A necessidade de refundação nacional do Estado não se limitou aos Estados que emergiram de revoluções, mas se estendeu rapidamente mesmo entre regimes longamente estabelecidos e “seguros contra a revolução” (Hobsbawm, 2004, p. 105). Hobsbawm (2004) destaca a ‘máquina dos censos’ como um dos instrumentos implicados no processo de ‘refundação nacional’,

que conduziu a uma expansão do domínio e da agência do Estado sobre o território nacional:

Ao longo do Séc. XIX, essas intervenções se tornaram tão universais e rotinizadas nos Estados 'modernos' que uma família teria que viver em um lugar muito inacessível se um de seus membros não quisesse entrar em contato regular com o Estado nacional e seus agentes [...] Cada vez mais o Estado detinha informações sobre cada um dos indivíduos e cidadãos através do instrumento representado por seus censos periódicos regulares [...] Nos Estados burocráticos e bem policiados, um sistema de documentação e registro pessoal trouxe os habitantes para um contato ainda mais direto com a máquina de administração e dominação [...] pela maquinaria que registrava os nascimentos, casamentos e mortes, que complementavam a máquina dos censos. (Hobsbawm, 2004, p.102)

A tarefa de realização dos primeiros censos demográficos nacionais, que proliferaram ao longo do Séc. XIX, tanto na Europa quanto nos Estados pós-coloniais nas Américas, foi um dos principais fatores que conduziram a um processo de diferenciação de estruturas burocráticas especializadas, seja na forma de escritórios estatísticos nacionais ou unidades estatísticas departamentais. Os censos populacionais correspondiam a um esforço de refundação nacional, que visava forjar um espaço de unidade social e uma identidade nacional comum. Neste sentido, visavam a completa enumeração da população compreendida no território nacional, distinguindo-se dos levantamentos populacionais anteriores, que, no contexto dos Estados Territoriais e das colônias, limitavam-se a catalogar uma fração da população para atender a interesses específicos da soberania, como o recrutamento militar e a cobrança de impostos.

De acordo com Desrosières (1998), do ponto de vista das práticas estatísticas, a unificação do espaço social que caracterizou a formação dos Estados Nação correspondeu a um processo de construção política e cognitiva de 'espaços nacionais de convenções de equivalências', espaços de medida e representação mental comuns, nascidos de uma linguagem comum e demarcados e constituídos principalmente pelo Estado e pela lei. A França revolucionária constitui um tipo ideal do processo de construção do espaço comum de equivalências que caracterizou a refundação do Estado em bases nacionais. Desrosières (1998) se refere ao termo 'adunação' para tratar da deliberada homogeneização do território nacional francês que se fez acompanhar pela padronização, centralização e universalização de uma série de 'medidas comuns' no período que vai de 1789 a 1815:

O processo de criação de equivalências foi, durante o quarto de século que abrangeu a Revolução e o Império, particularmente espetacular. Em poucos momentos da história do mundo esse processo foi desejado, pensado e iniciado tão sistematicamente em tão pouco tempo e em tantas áreas: o sistema métrico foi instituído, assim como a unificação de pesos e medidas (o mesmo em todos os lugares, e logicamente expresso em relação ao metro); generalizou-se a língua francesa e reduziu-se o número de dialetos provinciais (através do exército e das escolas); foi proclamada a universalidade dos direitos do homem ('Todos os homens nascem e permanecem livres e iguais'); os privilégios nobiliárquicos e as guildas profissionais foram abolidos; foi instituído o Código Civil (inspirado em uma lei natural do homem em geral, ao invés de estar vinculado a uma sociedade particular); e o território nacional, homogeneizado pela supressão de direitos próprios de certas províncias (Desrosières, 1998, p.31, tradução nossa).

De acordo com Desrosières (1998, p.31), esse processo de unificação e universalização na escala da nação, acompanhado por “monumentais levantamentos estatísticos” — que na França se deu a reboque de uma ruptura revolucionária e que assumirá contornos diferentes em outros contextos nacionais — é o que caracteriza, a construção do espaço de convenções de equivalências no contexto do Estado Nação. Para Desrosières (1998), a construção deste ‘espaço de medida comum’ (Camargo, 2016) implica em uma série de investimentos econômicos e cognitivos que sustentam uma rede complexa de agentes, instituições e construtos mentais envolvidos no processo estatístico. Trata-se, portanto, de um espaço político e cognitivo que assume características comuns e que confere um caráter nacional às ferramentas que constituem a essência do trabalho estatístico.

Bourdieu (2014, p.146), por outro lado, destaca que o processo de concentração e universalização de diferentes formas de capitais informacionais na gênese do Estado Nação, “vai de par com um processo de desapossamento” e de “monopolização do universal”. Na medida em que o Estado estabelece uma classificação oficial, um padrão universal, uma língua nacional, ou seja, na medida em que “a cultura legítima é a cultura garantida pelo Estado”, os diversos sistemas locais — de referência, classificação, pesos, medidas, línguas, etc. — são desqualificados e conseqüentemente,

o próprio processo pelo qual se ganha em universalidade é acompanhado por uma concentração da universalidade [ou seja] o próprio processo de constituição de recursos comuns [espaço de medida comum] é inseparável da constituição desses recursos comuns em capital monopolizado pelos que têm o monopólio da luta pelo monopólio do universal. Todo esse processo — constituição de um campo; autonomização deste campo em relação a outras necessidades, constituição de uma necessidade específica em relação à necessidade econômica e doméstica [...] constituição de uma reprodução específica do tipo burocrática [...] é inseparável de um processo de

concentração e de constituição de uma nova forma de recursos que são do universal [...] A gênese do Estado é, no fundo, inseparável da constituição de um monopólio do universal (Bourdieu, 2014, p.147).

Para Bourdieu (2014), portanto, o trabalho de unificação envolvido na construção do espaço público e nacional pelo Estado Nação foi necessariamente acompanhado por um processo de desapossamento de saberes e competências para que uma série de recursos pudessem ser qualitativamente transformados em capitais públicos e universais. A luta pelo controle e monopolização destes capitais constitui uma característica fundamental do modo de reprodução burocrático do Estado Nação, o qual implica em um processo de diferenciação de campos de práticas vinculados ao Estado, de onde deriva o poder associado aos capitais monopolizados por cada campo.

Articulando as contribuições de Bourdieu (2014) e Desrosières (1998), podemos compreender, portanto, que tudo se passa como se a construção do espaço nacional de convenções de equivalências, ou espaço de medida comum, constituísse ao mesmo tempo um espaço de lutas pelo controle e monopolização legítimos da medida comum. É justamente este espaço de lutas que denominamos por 'campo estatístico'.

Finalmente, para resumir nossa discussão nesta seção, podemos concluir que no contexto de formação do Estado Nação, a constituição do 'público' como espaço político legítimo para o exercício do poder do Estado se fez necessariamente acompanhar pela construção de um espaço social unificado que corresponde à 'nação'. Este processo de unificação nacional implicou na construção de uma representação comum da nação, por meio da centralização e universalização de classificações e medidas que constituem um espaço nacional de convenções de equivalências. A construção deste espaço implicou, por um lado, na quebra do monopólio tradicional da soberania sobre as estatísticas, e por outro, em desapossamentos de saberes locais em benefício da universalização de referências comuns, monopolizados pelo Estado, e, em tese, acessíveis ao público universal da nação. Este processo corresponde a uma transformação qualitativa de uma série de recursos informacionais, e em particular da estatística, em bens públicos, o que implica na configuração do campo estatístico como um espaço de lutas pelo controle e monopolização legítimos destes capitais, e em última instância, pelo controle do poder simbólico mobilizado pelas estatísticas oficiais.

Ao nos referirmos ao campo estatístico como um espaço de lutas faz-se necessário compreender que, no contexto das transformações políticas e econômicas em que se dá a formação do Estado Nação, as disputas em torno da construção do espaço de medida comum e pelo controle da estatística enquanto ferramenta de governo estão inseridas em um conjunto ampliado de lutas que demarcam a constituição do Estado no contexto de emancipação da burguesia e da consolidação do modo de produção capitalista. Estas lutas compreendem, portanto, a acomodação pelo Estado dos interesses dos capitais industriais e comerciais que se expandem no contexto da revolução industrial. Conforme observado por Porter (1995):

[...] a expansão do capitalismo foi uma fonte importante do ímpeto para unificar e simplificar as medidas. O outro, é claro, era o Estado — às vezes colaborando com grandes interesses industriais ou comerciais e às vezes agindo por seus próprios motivos. Medidas padrão e classificações uniformes eram pelo menos tão úteis para a atividade governamental centralizada quanto para o comércio e as manufaturas em larga escala [...] Medidas precisas e uniformes ajudaram a mover a economia de uma ordem baseada no privilégio para o domínio da lei. Elas também melhoraram o controle administrativo sobre questões de tributação e desenvolvimento econômico (Porter, 1995. p.25, tradução nossa).

Para Poulantzas (2015), o processo de unificação do Estado na escala do território nacional, vincula-se essencialmente a modificações nas matrizes do tempo e do espaço que têm seu fundamento nas novas relações de produção e na divisão social do trabalho, com o advento do capitalismo em sua fase industrial. O modo de organização social da produção sob o capitalismo implica em uma matriz espacial totalmente diferente, que surge como um pressuposto da divisão social do trabalho na fábrica e que se expressa como um espaço matricial no qual “se inscrevem os movimentos do capital e sua reprodução ampliada, a generalização das trocas e os fluxos monetários” (Poulantzas, 2015, p.103).

Para compreendermos como os interesses do Estado e do mercado se interceptam na constituição do campo estatístico, abordaremos na próxima seção o papel que as estatísticas oficiais passam a desempenhar em um contexto histórico marcado ao mesmo tempo por uma profunda transformação nas relações sociais de produção com o advento do capitalismo industrial (Marx, 2017; Poulantzas, 2015), e por uma ruptura no quadro da teoria e prática de governo com o advento do liberalismo (Foucault, 2008a).

1.4 Governamentalidade liberal e mais-valia: a estatística oficial e o mercado

Após a revolução francesa e no decurso da revolução industrial, o liberalismo que vinha sendo gestado ao longo do Séc. XVIII, enquanto oposição ao poder absoluto da soberania e crítica aos dogmas do mercantilismo, torna-se uma doutrina política e econômica hegemônica que passa a orientar a ação dos governos tanto na Europa quanto na América pós-colonial ao longo do Séc. XIX. De acordo com Foucault (2008a), o desenvolvimento neste período de uma nova 'arte liberal de governar', a qual intitula de governamentalidade, difere fundamentalmente da razão de Estado da soberania, na medida em que, visa "não tanto assegurar o crescimento do Estado em força, riqueza e poder, crescimento indefinido do Estado, mas sim limitar do interior o exercício do poder de governar" (Foucault, 2008a, p.39).

Para Foucault (2008a), o desenvolvimento da nova governamentalidade liberal caracteriza-se fundamentalmente pela articulação de dois princípios de autolimitação da prática governamental. Por um lado, a economia política clássica introduz na razão de Estado a noção do mercado como um novo espaço de verificação-falsificação da prática governamental. Por outro, a constituição do poder público na forma da lei passa a ser pautada pela noção de liberdade dos indivíduos governados, impondo uma limitação ao governo no campo do direito.

Deriva da economia política clássica ou liberalismo econômico (fisiocratas e utilitaristas) a concepção do mercado como um espaço autorregulado e regido por mecanismos naturais e espontâneos, que quando são deixados funcionar livremente e sem intervenções, permitiriam a formação de um preço natural que coincide com o valor justo das mercadorias. Deste modo, em sua 'verdade natural', o mercado possibilitaria alcançar, por meio da troca, um equilíbrio perfeito entre preço e valor, entre produção e demanda. Com a introdução da ideia de mercado autorregulado pelo liberalismo econômico, a prática governamental passa a ser precedida por leis naturais, as quais os governantes deveriam respeitar. De acordo com Foucault (2008a), a economia política estabelece, deste modo, uma separação entre mercado e governo, e eleva o mercado à instância reveladora do êxito ou fracasso da ação governamental, impondo uma autolimitação interna na razão de Estado. Doravante, os governantes devem conhecer os mecanismos de mercado e deixá-los funcionar livremente ou com um mínimo de intervenção.

Podemos depreender uma mudança significativa no papel desempenhado pela

estatística no contexto da nova razão liberal de governo. De um instrumento secreto e destinado exclusivamente ao planejamento da intervenção monopolista do Estado na economia, conforme sua formatação mercantilista, a estatística agora se constitui em uma ferramenta de governo voltada a revelar os mecanismos econômicos do mercado, para a burocracia do Estado e sobretudo para os agentes econômicos. Doravante, os dados sobre o comércio internacional, movimento dos preços, tarifas aduaneiras, taxas de câmbio, safras agrícolas, produção industrial, etc., tornam-se fundamentais para fornecer um quadro cognitivo comum, que viabiliza a competição capitalista e, ao mesmo tempo, fornece os critérios para a verificação da ação governamental, de acordo com a governamentalidade liberal.

Este novo papel da estatística oficial, como um meio de cognição do mercado, implica em uma aparente contradição daquilo que Foucault (2008a) define como princípio de 'autolimitação da prática governamental' e que em teoria tenderia a um 'Estado mínimo', uma vez que decorre do próprio modo liberal de governar, o desenvolvimento de instrumentos, instituições e práticas de Estado, como as estatísticas oficiais, para garantir a eficácia deste princípio. De acordo com Foucault (2008a), esta aparente contradição decorre da própria limitação interna do poder público, de acordo com os princípios dos direitos fundamentais e da utilidade da ação governamental, que no governo liberal passa a ser pautada pelo binômio liberdade e segurança¹⁷. Na medida em que o governo liberal só pode funcionar se houver determinadas liberdades (de mercado, de expressão, da propriedade, etc.), a prática governamental liberal constitui-se ao mesmo tempo em uma produtora e gestora destas liberdades, o que faz precisamente por meio de uma série de intervenções governamentais, que constituem um conjunto de técnicas, instrumentos e aparatos

¹⁷ De acordo com Foucault (2008a), no Estado Nação, o problema essencial do direito, que nos Estados Territoriais era o de fundamentar a soberania e as condições de legitimidade para o exercício do poder soberano, desloca-se para o de como estabelecer limites jurídicos ao exercício do poder público. Esta limitação na elaboração do direito público se dá de acordo com uma dupla noção de liberdade individual. Por um lado, a via revolucionária francesa partirá de uma concepção jurídica da liberdade, compreendida na forma de direitos naturais, originais e fundamentais do indivíduo a partir dos quais se delimitam os poderes e as fronteiras da competência do governo. Por outro, a via utilitarista inglesa compreenderá a liberdade, simplesmente como a independência dos governados em relação aos governantes, e definirá os limites da intervenção governamental na forma de um cálculo de sua utilidade. Embora o problema da utilidade tenha se constituído como o principal critério de elaboração dos limites do poder público e de formação do direito público e administrativo a partir do Séc. XIX, as duas concepções da lei e da liberdade (jurídica e utilitarista) introduzem uma ambiguidade que é característica do modo liberal de governar. Esta ambiguidade reside no fato de que esta prática governamental emergente não se limita a respeitar as liberdades, mas se constitui em uma gestora de liberdades por meio de uma série de intervenções governamentais.

que Foucault (2008a) denomina de mecanismos de segurança.

[...] esse liberalismo que podemos caracterizar como a nova arte de governar formada no século XVIII, implica no seu cerne uma relação de produção/destruição com a liberdade. É necessário, de um lado, produzir liberdade, mas, esse gesto mesmo implica que, de outro lado, se estabeleçam limitações, controles, coerções, obrigações apoiadas em ameaças, etc. Temos evidentemente exemplos disso. É preciso haver liberdade do comércio, mas como poderá esta exercer-se efetivamente se não se controla, se não se limita, se não se organiza uma série de coisas, de medidas, de prevenções, etc. [...] Temos aí uma espécie de estímulo para uma formidável legislação, a uma formidável quantidade de intervenções governamentais que serão a garantia da produção de liberdade de que se necessita, justamente, para governar (Foucault, 2008a, p.87-88).

Do ponto de vista das práticas estatísticas, para Desrosières (2008, p.45), a existência da estatística oficial chega a ser "quase paradoxal para o Estado liberal puro, se é que este é sequer imaginável". A economia de mercado e o 'liberalismo real', portanto, passam a demandar do Estado o exercício de um papel de 'inteligência' por meio de instituições e operações estatísticas:

Em sua formulação mais pura e abstrata, a teoria do mercado não implica necessariamente informação estatística. Os preços, revelados pelo jogo das trocas de mercado e da competição entre produtores, carregam em si todas as informações essenciais ao desdobramento das supostas virtudes dessa forma de organização econômica, sem que uma autoridade central, estatal ou não, precise estar envolvida, nomeadamente através da produção de informação estatística [...] No entanto, muitas instituições e operações estatísticas foram justificadas diretamente pelas necessidades de uma economia de mercado. Os primeiros exemplos disso foram fornecidos pela existência de comércio internacional, tarifas alfandegárias, taxas de câmbio e gestão da moeda nacional. O Statistical Office of the English Board of Trade foi criado em 1833, justamente quando ocorria uma série de reformas políticas e econômicas destinadas a libertar completamente o mercado capitalista de vários grilhões herdados do passado. A discussão dessas questões exigia documentação estatística ad hoc sobre preços e salários. Assim, o liberalismo "real" (e não apenas teórico) implicava um Estado desempenhando o papel de órgão de inteligência econômica, reunindo e disseminando as informações de que os agentes precisam para atuar no mercado (Desrosières, 2008, p.45)

Para Poulantzas (2015), a aparente separação entre mercado e governo, que Foucault (2008a) depreende de um efeito de teoria da economia política, constitui em realidade uma característica própria do modo de organização social da produção capitalista. Por meio de um "conjunto de determinações econômicas, políticas e ideológicas", a lógica do capital delimita novos espaços e campos relativos ao Estado e à economia, transformando os seus elementos. Esta separação para Poulantzas

(2015, p.17) é apenas relativa e decorre da própria estrutura das relações de produção capitalista, que na medida em que separa os trabalhadores de seus meios de trabalho, transforma a força de trabalho em mercadoria e condiciona sua introdução no processo de trabalho pela forma jurídica de um contrato. Deste modo, essa estrutura “dá lugar igualmente nas relações do Estado e da economia a uma separação relativa (acumulação do capital e produção de mais-valia), base da ossatura institucional própria ao Estado capitalista”. De acordo com Poulantzas (2015, p.16), embora sob formas diferentes, o Estado “sempre esteve constitutivamente presente nas relações de produção e assim em sua reprodução” e

esta separação não nos deve levar a crer em real exterioridade do Estado e da economia, como se o Estado só, do exterior, interviesse na economia. Esta separação é a forma precisa que encobre sob o capitalismo, a presença constitutiva do político nas relações de produção e, dessa maneira, em sua produção. A separação do Estado e da economia e a presença-ação do Estado na economia, que não passa de uma única e mesma figura das relações do Estado e da economia sob o capitalismo, atravessam, embora modificadas toda a história do capitalismo (Poulantzas, 2015, p.18).

As estatísticas oficiais no contexto do Estado capitalista participam, portanto, do que Poulantzas (2015, p.30) denomina de eficácia positiva do Estado e que se refere, entre outros, ao papel de organização das classes dominantes, por meio da formulação e declaração aberta, sem encobrimentos ideológicos, das próprias táticas de reprodução de seu poder. Para Poulantzas (2015, p.31) o revelar da tática — por exemplo, por meio de estatísticas oficiais que revelam a estrutura e o funcionamento do mercado — é parte “do espaço cênico do Estado em seu papel de representação dessas classes”, e neste sentido, o Estado

declama a verdade de seu poder, como também assume os meios de elaboração e formulação das táticas políticas. Produz o saber e as técnicas de saber, que, imbricadas na ideologia, de muito a superam. As estatísticas ‘burguesas’, por exemplo, elementos do saber do Estado com fins de estratégia política, não são mera mistificação. (Poulantzas, 2015, p.31)

O novo papel que a estatística oficial passa a assumir no campo econômico relaciona-se sobretudo com mudanças radicais na estrutura de organização da produção, com o advento da revolução industrial. De acordo com Marx (2017, p.310), “a revolução no modo de produção da indústria e da agricultura provocou também uma revolução nas condições gerais do processo de produção social”. Por ‘condições gerais de produção’, Marx se refere às condições que beneficiam ou limitam todo o

desenvolvimento do processo de produção de mercadorias, como por exemplo a infraestrutura (estradas, eletricidade, canais, ferrovias, meios de comunicação, etc.). A criação e manutenção das condições gerais correspondem a uma das funções que Marx (2017) atribuiu ao Estado, sendo o controle político sobre estas condições, um dos instrumentos mobilizado pelos capitais individuais para a extração de mais-valia.

Na medida em que as condições gerais de produção são compartilhadas por todas os agentes econômicos em uma determinada sociedade, constituem uma base comum para a produção de mercadorias. Assim, os capitais individuais não precisam investir em todas as condições necessárias para a produção de mercadorias, mas podem economizar e compartilhar essas condições. A disputa pelo controle político sobre as condições gerais de produção, constitui assim, parte das lutas constitutivas do Estado, na medida em que estas condições podem ser utilizadas para ampliar a extração de mais-valia relativa do trabalho, ampliando as taxas de lucro do capital. Por exemplo, viabilizando investimento em tecnologias mais eficientes que aumentam a produtividade do trabalho, por meio de sua intensificação ou controle, mas que não aumentam os salários na mesma proporção.

Partindo de uma definição mais expansiva do conceito de condições gerais de produção social, como um conjunto de “tecnologias, instituições e práticas que formam o ambiente para a produção capitalista em um dado tempo e local” (Dyer-Witford et al., 2019, p.30), podemos compreender que no contexto de revolução das condições gerais de produção, as estatísticas oficiais tornam-se parte constitutiva destas condições. Isto por que passam a fornecer informações comuns que beneficiam o conjunto de capitais individuais.

Finalmente, articulando as contribuições de Foucault (2008a), Poulantzas (2015) e Marx (2017), podemos compreender que, no contexto do desenvolvimento do capitalismo industrial e Estado liberal, a estatística oficial, assume três novos papéis no campo econômico. Em primeiro lugar, enquanto instrumento da governamentalidade liberal (Foucault, 2008a), a estatística constitui-se em um meio de cognição do mercado e de verificação e autolimitação da prática governamental. Em segundo lugar, enquanto ação positiva do Estado na economia (Poulantzas, 2015), a estatística oficial passa a operar como um instrumento de coordenação e organização estratégica dos agentes econômicos. Estes dois papéis reforçam, por sua vez, a característica da estatística oficial como um bem público, conforme vimos na seção anterior, na medida em que se destina a uma audiência pretensamente

universal e constitui uma função estratégica para a separação relativa entre Estado e mercado. Finalmente, depreendemos da teoria do valor de Marx (2017) um terceiro papel das estatísticas oficiais, enquanto ‘meio geral de cognição’ e parte das ‘condições gerais de produção’ sob responsabilidade do Estado, implicadas no processo geral de extração da mais-valia.

Tendo analisado os novos papéis que a estatística oficial passa a desempenhar no contexto da organização social do capital e modo de ação do Estado, nos dedicaremos na próxima seção a compreender suas interfaces com a estatística cognitiva. Trata-se de investigar a relação entre o Estado e a ciência, a partir de uma nova concepção do conceito de população. Esta relação, constitutiva do processo de estruturação do campo de práticas estatísticas no Séc. XIX, se materializa na estatística social e efetua uma transição para um novo regime estatístico de enumeração.

1.5 População e normalização: a estatística social e o regime de enumeração

Atravessado por uma série de revoluções, o Séc. XIX assistiu à emergência do fenômeno da sociedade moderna, vista ao mesmo tempo como fonte de progressos e como causa de grandes instabilidades. Neste contexto, os ensaios de Thomas Malthus (1766-1834) sobre a população exerceram forte influência no pensamento político e científico do século. Se no quadro do mercantilismo a população era percebida como um fator de fortalecimento dos Estados, sob a racionalidade liberal será associada às principais causas dos problemas políticos e sociais da época. O estudo da sociedade por meio dos números, que conforme vimos já vinha se manifestando desde o Séc. XVIII em levantamentos realizados fora do aparato estatal, se consolida agora como uma prática em sintonia com um novo espírito científico empirista-positivista e como uma resposta ao aumento da ansiedade sobre a mudança social. Nestes estudos, uma nova categoria de problemas associados ao fenômeno da população — como a pobreza, a saúde, a higiene e a criminalidade — corresponderá a construção de novos objetos, técnicas e métodos voltados à sua classificação, mensuração, análise e controle, dando margem ao desenvolvimento de uma ‘estatística social’, essencialmente descritiva.

De acordo com Foucault (2008), a ‘descoberta’ do fenômeno da população pela estatística é um dos principais fatores técnicos que contribuiram para o desbloqueio

da arte liberal de governar e para o desenvolvimento de uma nova estratégia geral de poder, a qual denomina de biopolítica. Essa estratégia consiste em fazer dos traços biológicos básicos da espécie humana um objeto de intervenção política. Intervenção que se dá por meio de instrumentos, técnicas e aparatos que constituem o mecanismo de segurança da governamentalidade liberal, no interior do qual a estatística passa a ocupar um lugar e papel estratégicos.

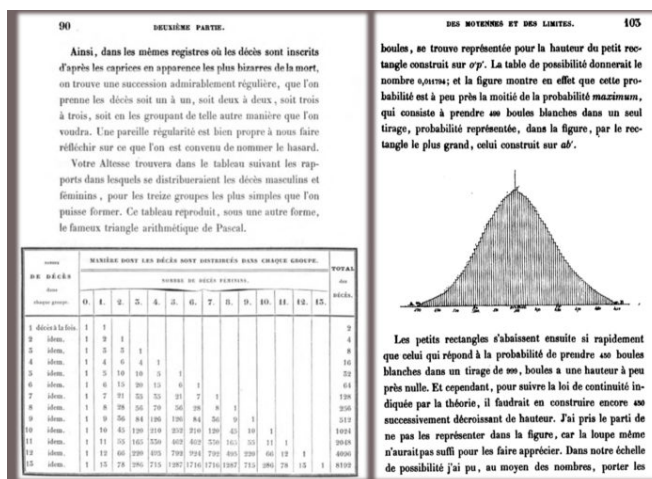
Para Foucault (2008), o que permitiu à estatística desbloquear a arte liberal de governo da rígida estrutura da razão de Estado da soberania foi principalmente a eliminação de um modelo até então baseado na estrutura da família. Isto possibilitou reorientar o foco da economia política e do governo para a escala da população, o que só foi possível na medida em que a estatística passou a quantificar os fenômenos populacionais e a revelar que a população possuía suas próprias regularidades (taxa de mortalidade, incidência de doenças, regularidade de acidentes, etc.) e seus próprios efeitos agregados (taxa de crescimento, epidemias, força de trabalho, etc.). Do ponto de vista dos seus efeitos econômicos, a população passou a constituir o principal objeto da economia política, e do ponto de vista das estratégias biopolíticas de poder, tornou-se o alvo dos mecanismos de segurança do Estado Nação.

Foucault (2008) identifica, a partir de meados do Séc. XVIII, alguns indícios desta nova concepção do fenômeno da população. Por exemplo, na doutrina do *laissez faire* dos fisiocratas e em sua abordagem ao problema da produção e escassez de grãos, ou na emergência de uma nova concepção de medicina preventiva, associada ao controle da epidemia de varíola. De acordo com Foucault, essas novas leituras passam a identificar uma 'naturalidade' nos movimentos e flutuações da população, agora compreendidos como um conjunto de processos que poderiam ser governados por meio de um cálculo refletido. Emergem assim técnicas de 'normalização' que passam a constituir uma característica dos mecanismos de segurança. Essas técnicas diferem fundamentalmente da 'normatização' dos mecanismos disciplinares, na medida em que a norma social passaria a ser deduzida da interação de diferentes curvas de normalidade.

A formalização da noção de que seria possível conhecer, quantificar e controlar os fenômenos populacionais por meio dos grandes números e da distribuição normal, se consolida na primeira metade do Séc. XIX por meio dos trabalhos do astrônomo belga Adolph Quetelet (1796-1874), que cunha o termo 'estatística social'. Amparado na teoria da lei dos erros de Laplace e no construto cognitivo da curva normal, Quetelet

propõe a existência de uma ‘lei estatística’ que regeria os fenômenos sociais. A ideia de que o acaso poderia ser domado pelos grandes números e emulado nos fenômenos sociais, teve um efeito político fenomenal na busca pela ordem social que caracterizou o Séc. XIX (Porter, 1986).

Figura 10 - Quetelet e a Curva Normal



Fonte: <https://archive.org/>

De inspiração positivista, a estatística social de Quetelet se equiparava a uma física social e defendia a utilização das leis da probabilidade para o estudo da população e dos fenômenos sociais, por meio da análise da distribuição de suas frequências. À frente do Bureau estatístico dos países baixos a partir de 1827, Quetelet se valeu de seus conhecimentos astronômicos e probabilísticos para a análise de dados estatísticos populacionais. Sua análise revelou uma série de regularidades nas distribuições de fenômenos sociais (nascimentos, mortes, casamentos, crimes, suicídios, etc.) e atributos físicos humanos (altura, peso, estrutura corpórea, etc.). Estas regularidades o levaram a difundir a noção de um ‘homem médio’ ideal que combinaria características médias de uma dada população a partir das quais os desvios físicos, morais e comportamentais poderiam ser conhecidos e controláveis (Porter, 1986).

Quetelet é um personagem relevante para o desenvolvimento da estatística no Séc. XIX sob vários aspectos. Aqui nos interessa destacar principalmente a aproximação que promoveu entre a estatística oficial, enquanto ferramenta de governo de caráter político e administrativo, e a estatística cognitiva, enquanto ferramenta de prova vinculada à matemática probabilística. Este primeiro encontro

entre os dois ramos da estatística contou com a decisiva participação de Quetelet, tanto em sua elaboração teórica, quanto em sua construção técnica e política. A materialização da ideia de que seria possível compreender e controlar os fenômenos sociais por meio da estatística teve grandes consequências epistemológicas, sociais e políticas.

Em primeiro lugar, do ponto de vista epistemológico, a estatística social ajudou a consolidar a interpretação frequentista e objetiva do cálculo de probabilidades, vinculando-o definitivamente à análise de frequências estatísticas. Em 1835, a publicação do ensaio de Quetelet em que apresentou os fundamentos da estatística social, associando a análise de frequências populacionais à distribuição normal, teve grande repercussão internacional tanto nos meios científicos como para além destes (Stiegler, 1986). A distinção entre as interpretações objetiva e subjetiva de probabilidade será apresentada pouco tempo depois em 1837 por Poisson (1781-1840). A partir daí, são formuladas uma série de críticas às concepções 'subjetivas' da probabilidade em defesa de uma aceção estritamente objetiva¹⁸ (Daston, 1988). Para Shafer (1990, p.437), tais críticas se deveram à influência das correntes empiristas e positivistas na filosofia e ciência do Séc. XIX, para as quais a probabilidade "só fazia sentido se fosse definida como frequência".

De acordo com Daston (1988), o declínio da interpretação clássica da probabilidade, que conduzirá a primazia da interpretação frequentista é marcado pela transição entre os ideais representados pelo 'homem esclarecido' do iluminismo e o 'homem médio' da sociedade moderna. Enquanto o primeiro é um exemplo do critério de crença e ação racional codificados na definição e comparação de expectativas probabilísticas, o segundo personifica as leis uniformes da sociedade, reveladas pela curva normal, em oposição as condutas variáveis dos indivíduos. Ambos, entretanto, personificam padrões sociais a serem matematicamente codificados e reforçados.

Para Desrosières (1998), a oposição entre probabilidades objetivas e subjetivas decorre principalmente de concepções distintas acerca do papel descritivo e prescritivo da estatística e do cálculo de probabilidades, tendo a primeira ganhado proeminência a partir do Séc. XIX:

¹⁸ Entre os mais destacados críticos constam Cournout (1801-1877), John Stuart Mill (1806-1873), George Boole (1815-1864), Jakob Fries (1773-1843), Auguste Comte (1798-1857), entre outros (Daston, 1988).

Numerosas construções conectaram esses dois conceitos de probabilidade. Em primeiro lugar, estão as diversas formulações da “lei dos grandes números”, começando com Jacques Bernoulli (1713) [e] o teorema de Bayes (1765) - que liga a informação parcial fornecida pela ocorrência de alguns eventos com a hipótese de “probabilidade a priori” levando a uma “probabilidade a posteriori” mais certa, melhorando a racionalidade de uma decisão baseada em conhecimento incompleto. Este raciocínio, plausível do ponto de vista da racionalização do comportamento (epistêmico), não é mais plausível do ponto de vista descritivo (frequentista) no qual a “probabilidade a priori” não tem base. Esta tensão permeia toda a história da estatística. É crucial a oposição entre dois pontos de vista: num caso, é aceitável porque é preciso tomar uma decisão; no outro, não se satisfaz com uma hipótese injustificada, destinada apenas a orientar ações (Desrosières, 1998, p.7, tradução nossa).

Em segundo lugar, do ponto de vista administrativo, a intervenção do Estado torna-se um imperativo para garantir a completa enumeração dos fenômenos populacionais, de forma a prover os dados que, por meio da objetividade científica e das novas ferramentas cognitivas da estatística, poderiam ser analisados, estabilizados e controlados. A busca por regularidades estatísticas pressupunha o levantamento exaustivo de dados, neste sentido, do ponto de vista do método, a estatística social é puramente descritiva e não inferencial. Sob a sua influência, os escritórios estatísticos da Europa e das Américas no Séc. XIX, adotarão a prática dos levantamentos censitários regulares e exaustivos e serão por muito tempo refratários ao uso de amostras probabilísticas (Stiegler, 1986).

Desrosières (2008) observa que a intuição original de Quetelet e seus seguidores, na promoção dos censos exaustivos nas primeiras décadas do séc. XIX, operava a materialização da ‘urna de Bernoulli na escala da nação’. Apenas após a construção da urna nacional, e de conhecidos todos os seus elementos populacionais, é que seria possível conceber a introdução de procedimentos amostrais na produção de estatísticas oficiais, o que se dará apenas nas primeiras décadas do Séc. XX, com o desenvolvimento da estatística inferencial.

Finalmente, do ponto de vista político, a construção da legitimidade da prática estatística, necessária à eficácia das táticas de normalização, passaria doravante a requerer uma dupla autorização, do Estado e da ciência. Essa dupla autoridade ganhou materialidade objetiva nos escritórios estatísticos e foi decisiva na estruturação do campo estatístico ao longo dos Séculos XIX e XX. Desrosières (1998) destaca que a combinação dos dois tipos de autoridade, do Estado e da ciência, no processo de construção da legitimidade das estatísticas oficiais, foi o que possibilitou que novos objetos, classificações e técnicas de formalização — resumos, médias,

curvas, codificações, totalizações, gráficos e tabelas — fossem apreendidos pelos indivíduos como ‘coisas’ reais sem serem constantemente questionados.

Essa solidez pode resultar ela mesma do arbítrio da força, ou de uma legitimidade cuidadosamente criada, nos Estados de direito que se fundaram no século XIX, e que assumiram diversas formas. Essa legitimidade não caiu do céu por decreto. Foi moldada e tecida dia após dia, esquecida, ameaçada, questionada e reconstruída com custos adicionais. Dentro da legitimidade das instituições estatais, as estatísticas ocupavam um lugar particular: o da referência comum, duplamente garantida pelo Estado e pela ciência, cuja sutil combinação, constituía a originalidade e a particular credibilidade das estatísticas oficiais (Desrosières, 1998, p.148, tradução nossa).

De acordo com Hacking (2004), a associação entre fenômenos sociais e leis estatísticas só se tornou possível pelos trabalhos prévios de classificação, enumeração e tabulação executados pela estatística oficial, que então emergia como uma nova tecnologia implantada por ‘novas burocracias’. Esta mesma estatística, que conforme analisamos anteriormente, se transmutava em um bem público voltado à uma audiência universal e nacional, concorreu nas primeiras décadas do Séc. XIX, para uma ‘avalanche de números impressos’ e para a sedimentação de um nova forma de raciocínio estatístico.

A impressão de números foi um efeito de superfície. Atrás dele estavam novas tecnologias para classificar e enumerar, e novas burocracias com a autoridade e continuidade para implantar a tecnologia. Em certo sentido, muitos dos fatos apresentados pelas burocracias nem mesmo existiram antes do tempo. Foi preciso inventar categorias nas quais as pessoas pudessem convenientemente se enquadrar para serem contadas [...] Leis estatísticas que se parecem com fatos brutos e irreduzíveis foram encontradas pela primeira vez nos assuntos humanos, mas só puderam ser notadas depois que os fenômenos sociais foram enumerados, tabulados e tornados públicos. Esse papel foi bem servido pela avalanche de números impressos no início do século XIX. A maioria das regularidades semelhantes a leis foram percebidas pela primeira vez em conexão com desvio: suicídio, crime, vadiagem, loucura, prostituição, doença (Hacking, 2004, p. 2-3, tradução nossa).

Porter (1986) destaca o papel dos ‘estatistas’ — como eram chamados no início do Séc. XIX os reformistas, cientistas e burocratas dedicados às práticas estatísticas — na difusão de uma nova forma de raciocínio que será futuramente associada ao novo ramo científico da matemática estatística.

Não é por acaso que uma ciência social quantitativa e empírica do século XIX deu seu nome a um ramo da matemática aplicada, pois a contribuição dos "estatistas" para o estilo de raciocínio associado à nova matemática foi

fundamental. Para começar, os estatistas familiarizaram o mundo científico e o público instruído com o uso de números agregados e valores médios para estudar um objeto inerentemente variável. Os escritores estatísticos persuadiram seus contemporâneos de que os sistemas que consistem em numerosos indivíduos autônomos podem ser estudados em um nível mais alto do que os diversos constituintes atômicos. Ensinaram-lhes que se poderia presumir que tais sistemas gerassem ordem e regularidade em larga escala, que seriam virtualmente inalteradas pelo capricho que parecia prevalecer nas ações dos indivíduos. Uma vez que mudanças significativas no estado do sistema apareceriam apenas como consequência de causas proporcionalmente grandes, uma ciência poderia ser formulada usando frequências relativas como seus dados elementares (Porter, 1986, p.5, tradução nossa).

Finalmente, articulando as contribuições de Desrosières (1998), Hacking (2004) e Porter (1986) podemos compreender que, tudo se passa como se houvesse uma relação intrínseca entre a estatística cognitiva que começava a despontar como um novo campo de conhecimentos científicos e a estatística pública e oficial produzida pelos escritórios de estatísticas, que surgiam no contexto de refundação nacional dos Estados.

Do ponto de vista cognitivo, a estatística social abriu caminho para a adoção da matemática probabilística aplicada à análise de dados estatísticos, o que contribuiu para consolidar o frequentismo como uma concepção dominante na ciência estatística e nos escritórios estatísticos. A compreensão limitada de Quetelet sobre a lei dos erros e sua insistência nas médias, foi objeto de críticas dos matemáticos estatísticos que o sucederam. De fato, a maioria das regularidades identificadas por Quetelet não correspondiam verdadeiramente à distribuição normal (Stiegler, 1986). O desenvolvimento da estatística inferencial — a partir de construtos como a correlação e a regressão por Francis Galton (1822-1911) e Karl Pearson (1857-1936) e o teste de hipóteses por Ronald Fisher (1890-1962), Egon Pearson (1895-1980) e Jerzy Neyman (1894-1981) — elevará a matemática estatística ao estatuto de ciência experimental no Séc. XX, construída, entretanto, sob as mesmas bases frequentistas da estatística social de Quetelet.

Do ponto de vista político, o trabalho dos 'estatistas' tanto dentro como fora do aparato estatal concorrerá para a conformação de um novo campo de práticas, que se caracterizará pela conflagração de disputas tanto no interior dos espaços nacionais, entre os campos burocrático e científico, como em novos espaços transnacionais que passam a ser constituídos a partir da segunda metade do Séc. XIX, configurando um movimento que dá origem ao 'internacionalismo estatístico'. De

fato, é a partir da realização de uma série de congressos internacionais de estatística, entre 1853 e 1878, promovidos por Quetelet e seus seguidores, que o campo de práticas estatísticas ganha materialidade, constituído por um novo corpo técnico-científico-burocrático estatal e nacional em um espaço transnacional de disputas. Processo que passamos a analisar na próxima seção.

1.6 Entre a nação e o internacionalismo: a construção política do campo estatístico

A primeira metade do Séc. XIX correspondeu a ‘era de entusiasmo’ da estatística em que, de acordo com Porter (1986), a prática estatística passou a refletir um temperamento liberal e uma busca por reformas. Para Porter (1986, p. 27), por meio da observação empírica cuidadosa, os estatísticos “acreditavam que a confusão na política poderia ser substituída por um reino ordenado dos fatos” (Porter, 1986, p.27). De forma semelhante, Randeraad (2010, p.189) observa que esta época foi caracterizada pela busca incansável por uma ‘ciência da sociedade’ que também servisse de guia para a ação governamental. A estatística representava a possibilidade de entender e resolver os problemas que assolavam a sociedade e durante algum tempo “houve esperança de que as estatísticas tornariam possíveis políticas objetivas e uma intervenção estatal incontestada”. Fazer da política um reino da técnica e dos fatos, amparada na objetividade e neutralidade dos números, eis o sonho e a ilusão dos primeiros estatísticos, que, como veremos, ainda reverbera em formulações contemporâneas.

O sonho era da utilidade progressiva do conhecimento estatístico e foi compartilhado por muitos estatísticos do século XIX. Seu sonho seria realizado em três fases. Primeiro, os dados sobre a sociedade seriam coletados em todos os países, empregando métodos e categorias uniformes. Em seguida, os dados seriam comparados e os governos baseariam suas políticas no conhecimento assim adquirido. E, finalmente, toda a humanidade experimentaria maior felicidade e prosperidade. A crença no progresso não teve proponente mais verdadeiro, mais fiel ou mais ambicioso do que o estatístico. Ele calculou, classificou e concluiu, até que todas as leis que regiam a sociedade pareciam materializar-se dos números espontaneamente. Por mais óbvio que seja para nós que isso era uma ilusão, o estatístico não tinha dúvidas de que seu ideal era alcançável. (Randeraad, 2010, p.1, tradução nossa)

Esta fase de entusiasmo deu impulso a um trabalho político capitaneado por Quetelet e seus contemporâneos, que resultará na configuração de um campo

transnacional de práticas estatísticas¹⁹. Este campo passava a ser caracterizado por um lado, por novas estruturas burocráticas nacionais voltadas à produção e disseminação de estatísticas oficiais²⁰, e, por outro, pela articulação internacional de burocratas, cientistas e entusiastas da estatística social. Estas interfaces resultarão na criação de estruturas transnacionais voltadas à padronização de classificações, medidas e métodos estatísticos que tinham por finalidade estabelecer um quadro objetivo de comparação entre as diferentes nações. Um marco inicial e decisivo deste processo foi a realização dos Congressos Internacionais de Estatística (CIE) entre 1853 e 1878²¹.

Conforme analisamos anteriormente, de acordo com Foucault (2008), a política de equilíbrio europeu sob a doutrina do mercantilismo lançara as bases, a partir do Séc. XVII, para uma nova relação entre os Estados soberanos baseada em uma lógica concorrencial. Esta lógica impunha pela via da concorrência uma limitação externa à soberania. Entre o Séc. XVIII e XIX, manifesta-se, de acordo com Foucault (2008a, p.80), a emergência de “uma nova forma de cálculo político em escala internacional”, coerente com uma estratégia de expansão ilimitada do mercado em uma escala global sob o domínio Europeu. Para Foucault (2008, p.83), doravante, a garantia da paz estaria atrelada à mundialização do mercado. Assim, o equilíbrio europeu, especialmente após o Tratado de Viena de 1815, passaria a se assentar no princípio da Europa como uma região econômica perante “um mundo que, para ela, deve constituir um mercado”. Verifica-se a partir daí “um novo tipo de cálculo planetário na prática governamental europeia” (Foucault 2008, p.86) e podem ser identificados

¹⁹ Quetelet foi um dos principais incentivadores do intercâmbio de ideias entre os primeiros estatísticos “compartilhando os resultados de suas pesquisas, bem como suas ideias sobre a organização da ciência estatística e seus objetos” (Randeraad, 2010,p.5). Já mundialmente conhecido por sua produção teórica, o seu sucesso na condução do primeiro censo nacional da Bélgica como Estado independente em 1846, havia contribuído para projetá-lo ainda mais como o principal líder do movimento estatístico internacional.

²⁰ Até 1850 serão criados bureaux estatísticos nos principais países europeus, ao que se seguirá na segunda metade do século à criação de estruturas semelhantes nos Estados pós coloniais da América Latina.

²¹ Os Congressos Internacionais de Estatística foram realizados entre 1853 e 1878 nas cidades de Bruxelas (1853), Paris (1855), Viena (1857), Londres (1860), Berlim (1863), Florença (1867), Haia (1869), São Petersburgo (1872) e Budapeste (1876). Um comissão permanente foi instalada após o Congresso de São Petersburgo e realizou reuniões em Viena (1873), Estocolmo (1874), São Petersburgo (1876) e Paris (1878). Com a morte de Quetelet, em 1874 e em um ambiente de conflitos envolvendo disputas nacionais, uma solução de continuidade para os CIE é encontrada com a fundação do Instituto Internacional de Estatística, em 1885 (Randeraad, 2010).

sinais do aparecimento de uma nova forma de ‘racionalidade planetária’, que ganha materialidade nos projetos de organização internacional. O desenvolvimento do internacionalismo científico e estatístico, a partir do Séc. XIX, é coerente com esta nova racionalidade política.

Foucault (2008, p.80) destaca, por outro lado, que o desenvolvimento deste novo tipo de cálculo não significou a “mundialização pacífica da política”, pelo contrário, “com o Séc. XIX inicia-se a pior época de guerras, das taxas alfandegárias, dos protecionismos econômicos, das economias nacionais, dos nacionalismos políticos”. Neste sentido, a contradição entre o ideal estatístico de um cálculo político na escala global e as disputas concorrenciais entre os Estados Nação, demarcadas pela lógica da soberania nacional, constitui o pano de fundo dos principais embates travados nos Congressos Internacionais de Estatística (CIE).

Os CIE reuniam chefes de escritórios nacionais de estatística e membros de sociedades nacionais de estatística (associações constituídas por notáveis reformadores, médicos, higienistas, leigos, etc.) que, de acordo com Desrosières (2008, p.28), “constituíam uma espécie de lobby pelo desenvolvimento das instituições estatísticas”. Constituído majoritariamente por representantes dos países europeus, é digna de nota a participação oficial de países latino americanos como o Brasil, que se fez presente nos congressos de 1860 em Londres, 1872 em São Petersburgo e 1876 em Budapeste (Camargo, 2016). De acordo com Camargo (2016), a participação das ex-colônias “esteve longe de ser meramente passiva ou receptiva em relação aos argumentos de força dos sábios europeus”. Destacam-se a aprovação de emendas às resoluções do congresso pelas delegações das jovens nações, a exemplo da proposta do delegado brasileiro no encontro de 1872, para a flexibilização da data universal de referência para a realização dos censos, levando em conta as particularidades do hemisfério sul.

Para Desrosières (2008, p.27), os congressos e a crise política que conduziu ao seu encerramento são ilustrativos de uma tensão que é própria das estatísticas oficiais, “entre um papel eminentemente político e um objetivo de conhecimento ‘objetivo e neutro’, repetidamente reivindicado pelos estatísticos”. Neste sentido, a historiografia dos CIE revela que os congressos foram marcados por uma série de disputas e embates, que caracterizam a formação do campo de práticas estatísticas “como um campo de tensões entre as reivindicações científicas de neutralidade e

universalidade, por um lado, e a realidade política e econômica dos interesses conflitantes dos Estados Nação, por outro” (Randeraad, 2010, p.6).

Havia tensão entre a ‘comunidade imaginada’ de estatísticos de orientação internacional e a ‘comunidade imaginada’ da nação. Os congressos internacionais de estatística rapidamente tiveram de abandonar o seu caráter cosmopolita e, para confusão e aborrecimento dos estatísticos, em si, tornaram-se o campo de batalha dos interesses nacionais [...] Esses interesses conflitantes manifestaram-se de diversas maneiras. Por vezes, as batalhas constituíram verdadeiros testes de força na política internacional, enquanto outras giravam em torno de laboriosas comparações de economias divergentes ou de avaliações pragmáticas de vários métodos de registo (Randeraad, 2010, p.6, tradução nossa).

Para Desrosières (2008, p.27), na medida em que as estatísticas oficiais “estão de muitas formas ligadas ao Estado-Nação, através de esquemas e categorias indissociavelmente cognitivas e políticas”, as relações entre a estatística e a política “vão muito além do exercício quotidiano do poder, e incluem as várias dimensões que mantêm uma unidade nacional”. Notadamente, a busca dos CIE por uniformização e padronização internacional de pesos, medidas, moedas, procedimentos censitários, classificações estatísticas, etc., constantemente esbarrava em uma diversidade de métodos e conceitos, que denunciavam tradições e princípios nacionais que não podiam ser facilmente harmonizados.

De acordo com Randeraad (2010, p.50), embora a busca pela uniformidade unisse todos os estatísticos e importantes consensos tenham sido alcançados nos CIE, a exemplo da universalização do sistema métrico e da periodicidade decenal dos censos demográficos, “o desacordo ocorreu sempre que foi feita uma tentativa de definir esse esforço em termos específicos”. Por exemplo, a resolução que estabelecia a adoção da população real para efeito de contagem nos censos demográficos era defendida por estatísticos de perfil mais científico como Quetelet e chegou a perdurar por alguns congressos. A sua posterior flexibilização, no congresso de 1872, atendia a pressões e objetivos nacionais específicos, em flagrante contraste com os objetivos iniciais de uniformização do internacionalismo estatístico.

De acordo com Desrosières (2008), a crise do internacionalismo estatístico foi deflagrada quando o recém-criado Império Alemão instituiu um escritório estatístico central e decidiu não mais participar dos congressos, onde os estatísticos franceses desempenhavam um papel relevante. Embora este não tenha sido o único fator que levou à interrupção dos CIE, cujo modelo já vinha sendo criticado em face dos poucos

resultados alcançados em termos de comparabilidade estatística (Randeraad, 2010), o conflito político quebrou a dinâmica do congresso, que se reuniu pela última vez em 1878.

Embora o caráter oficial da estatística e seus objetivos nacionais tenham se sobreposto às aspirações universalizantes dos estatísticos, construiu-se nos congressos um consenso sobre a necessidade de aplicar uma 'abordagem científica' à prática estatística (Randeraad, 2010). Em meio a disputas, os estatísticos buscaram acomodar tanto interesses científicos e profissionais, como políticos na construção do novo campo. Por um lado, a noção de que a prática estatística no âmbito da administração pública envolvia a necessidade de conhecimentos científicos especializados, atendeu a interesses comuns dos campos burocrático e científico, e contribuiu para a diferenciação e legitimidade dos escritórios estatísticos, conferindo-lhes um status de cientificidade. Por outro lado, a defesa de um status de neutralidade e autonomia para o desenvolvimento da ciência estatística, em seu esforço de harmonização internacional de métodos e práticas, implicou na demarcação de um novo espaço de articulação do campo estatístico a partir da criação do Instituto Internacional de Estatística (ISI).

Fundado em 1885 na Inglaterra como uma instituição 'puramente científica', o Instituto Internacional de Estatística (ISI), em atividade até o presente, foi uma solução de compromisso encontrada para dar seguimento à articulação internacional dos estatísticos e ao trabalho de harmonização de suas práticas e métodos, sem comprometer politicamente os governos com suas recomendações. Os representantes dos escritórios estatísticos nacionais agora participavam do ISI na condição de especialistas, e as harmonizações e padronizações ali discutidas não precisariam necessariamente ser implementadas nacionalmente. Entretanto, na medida em que a produção técnica do ISI estabelece padrões internacionalmente aceitos e validados cientificamente, o instituto passou a exercer considerável influência sobre os órgãos estatísticos nacionais. (Desrosières, 2008).

A descontinuidade dos CIE e a criação do ISI marca o início de um trabalho de separação relativa entre os campos da ciência estatística e da estatística oficial que se enfatizará no Séc. XX. Essa separação é resultado tanto do desenvolvimento científico da matemática estatística inferencial como da emergência de novas instituições internacionais de coordenação estatística no contexto de construção de uma governança global no pós guerra. Neste sentido, após a primeira guerra mundial

será constituída uma primeira comissão estatística internacional vinculada à Liga das Nações (1920–1946) e após a segunda guerra mundial, juntamente com a constituição do Sistema da Organização das Nações Unidas, será criada a Comissão Estatística da ONU.

Finalmente, o episódio dos CIE é relevante para compreendermos como o processo de estruturação do campo estatístico é caracterizado por um conjunto de lutas e tensões que decorrem de seu caráter ao mesmo tempo científico/político e nacional/global. A interrupção dos CIE revela que em detrimento do ideal estatístico de uma governança planetária ‘apolítica’, prevaleceu o papel da estatística enquanto ferramenta de governo do Estado Nação e instrumento de acumulação de capital informacional fundamental ao exercício da soberania sobre a população, o território e o mercado. Na próxima seção investigamos o papel do internacionalismo científico na importação de práticas estatísticas por Estados Pós-Coloniais como o Brasil.

1.7 A estatística desigual e combinada dos Estados Pós-Coloniais

Embora a Europa tenha sido a matriz intelectual do processo de refundação nacional dos Estados Territoriais sob a forma moderna do Estado Nação, os censos nacionais e as estatísticas oficiais serão igualmente utilizadas pelos países latino americanos pós-coloniais, para legitimar sua existência como nações independentes ao longo do Séc. XIX (Otero, 2018). De acordo com Camargo (2016), as novas formas institucionais dos Estados Nação europeus seriam ‘desejadas’ pelos novos países pós-coloniais, como o Brasil, que passavam a ocupar posições periféricas no sistema de Estados que se constituía na segunda metade do Séc. XIX. Notadamente, o internacionalismo científico e estatístico que emergiu neste período foi decisivo para a importação de tecnologias ‘liberais’ de governo, típicas do Estado Nação, por países de passado colonial. As novas instituições (e.g. imprensa livre, sistema escolar) e tecnologias liberais (e.g. estatísticas, censos, mapas) passariam, por sua vez, a coexistir com as estruturas sociais, políticas e administrativas herdadas do Estado Territorial — escravocrata, estamental e patrimonialista — concorrendo para uma hibridação que “não leva à superação da forma antiga mas a uma conformação nova e original” (Camargo, 2016, p.83).

Esta conformação ‘nova e original’ do Estado Pós-Colonial, pode ser compreendida do ponto de vista do processo histórico de desenvolvimento

econômico, político e cultural do capitalismo mundial. De acordo com a lei do desenvolvimento desigual e combinado de Trotsky (1988, p.226), essa hibridação ou “amálgama das formas arcaicas com as mais modernas” é uma característica intrínseca da própria natureza do desenvolvimento capitalista, que tende a combinar “duas tendências fundamentais, centrípeta e centrífuga, nivelamento e desigualdade”. Assim, em face da desigualdade no ritmo de desenvolvimento das forças capitalistas no plano internacional, o desenvolvimento interno dos países de passado colonial possui uma historicidade própria que é marcada pela incorporação dos elementos mais modernos das nações desenvolvidas e sua adaptação às condições culturais e materiais arcaicas. A essência do desenvolvimento capitalista nos países pós-coloniais, é, portanto, caracterizada por uma combinação dialética das inovações políticas, culturais e tecnológicas dos países avançados com relações sociais pré-capitalistas.

A lei de desenvolvimento desigual e combinado exerceu considerável influência sobre teóricos latino-americanos que se dedicaram à compreensão do desenvolvimento capitalista nos países periféricos. Destaca-se, neste sentido, o conceito de ‘capitalismo dependente’ de Florestan Fernandes e o seu posterior desenvolvimento pelos teóricos da dependência (Chilcote, 2012). Para Fernandes (2006, p. 209), “só o capitalismo dependente permite e requer tal combinação do ‘moderno’ com o ‘arcaico’, uma descolonização mínima, com uma modernização máxima”. Assim, de acordo com a teoria da dependência (Bambirra, 2013; Cardoso; Faletto, 1970; Frank, 1969; Furtado, 1961, 1968; Marini, 2008; Prebish, 1949; Rangel, 2005; Santos, 2011, Tavares, 1982, 1992), o subdesenvolvimento é compreendido como produção histórica da expansão do capitalismo, que vinculou a um mesmo sistema economias com graus variados de diferenciação produtiva e que passaram a ocupar posições distintas na estrutura global do sistema capitalista.

Nas nações subdesenvolvidas o processo histórico de ‘independência’ colonial gerou uma ambiguidade: uma busca por autonomia do centro político da ação das forças sociais, enquanto a economia continua sendo definida objetivamente em função do mercado externo, limitando a ação autônoma. A dependência se expressa, portanto, como um tipo específico de relação entre as classes, que implica uma situação de domínio que mantém estruturalmente a vinculação econômica com o exterior. Neste sentido, o desenvolvimento do Estado de uma maneira geral — e o

desenvolvimento da estatística oficial e campo estatístico em particular — nos países pós-coloniais será igualmente marcado por esta ambiguidade.

Para Cardoso e Faletto (1970), o problema da organização nacional pós-colonial na América Latina consistiria de uma maneira geral em manter sob controle local o sistema produtivo exportador herdado do sistema colonial por meio de um mecanismo interno de alianças políticas que garantisse a sua dominação. No caso brasileiro, tratava-se de organizar a economia interna como uma base agrícola da economia metropolitana, em particular do sistema importador europeu. As condições de êxito deste modelo derivaram tanto da dinâmica de expansão industrial inglesa, interessada principalmente em assegurar o abastecimento de produtos primários sem que fosse necessária a substituição dos grupos produtores nacionais, como da capacidade dos grupos exportadores, em particular dos cafeicultores, em organizar um sistema de alianças com as oligarquias locais o que envolvia a representação e mediação de seus interesses pelo Estado, que no Brasil pós-colonial organizou-se como uma monarquia constitucional.

De acordo com Fernandes (2006), no Brasil Império, a correlação entre o ‘velho’ e o ‘novo’ se manifesta no preenchimento de funções sociais distintas pelo Estado. Por um lado, contra ‘os princípios liberais’ absorvidos de forma indecisa pelas elites dominantes, o Estado se converte em fator de preservação da escravidão, da dominação senhorial, da sociedade estamental e da estrutura patrimonialista. Por outro lado, este mesmo Estado se vincula “de modo profundo e dinâmico com as inovações requeridas institucionalmente, a curto e a longo prazo, para que ele próprio fosse possível e para que viesse a contar com uma sociedade nacional que lhe servisse de base econômica, social e cultural” (Fernandes, 2016, p.66).

As práticas estatísticas inserem-se portanto no rol de inovações institucionais necessárias à legitimação dos Estados Pós Coloniais. Para Camargo (2016), o princípio de comparabilidade que emergiu dos Congressos Internacionais de Estatística no Séc. XIX — e que ganhou materialidade, por exemplo, em rankings que classificavam os países de acordo com posições objetivas em um ‘espaço de medida comum’ — foram determinantes para fomentar nos países de passado colonial o ‘desejo’ que levou à importação das modernas práticas estatísticas que emergiam nos países europeus. Conforme veremos, ecos desse ‘desejo’ modernizador continuam a se manifestar no presente, por exemplo, na busca pelo big data.

Estas formas, liberais e modernas, ao mesmo tempo que produziam e performavam a hegemonia dos países europeus e dos Estados Unidos, se inscreverão no horizonte do desejo das jovens nações de passado colonial, justamente porque a medida comum tornava objetiva e superável a distância, antes abissal e intransponível, que existia em relação às suas antigas metrópoles. Este movimento convidava à comparação sistemática entre os países e instalava os indicadores de desenvolvimento que encarnavam o progresso moral e material - como os de letramento e de superfície viária construída – no horizonte político de sociedades cuja dinâmica interna não comportava o uso produtivo de dispositivos liberais (Camargo, 2016, p.83-84).

A visualização destas ‘posições objetivas’ entre as nações, por outro lado, passou a evidenciar a situação subordinada das ex-colônias, produzindo efeitos políticos relevantes. No caso brasileiro, Camargo (2016) observa como a publicação de estatísticas comparativas internacionais contribuíram para orientar o discurso estatístico em relação à perspectiva internacional, em manifestações que oscilavam “entre o lamento pelo atraso e o sonho do futuro promissor, uma vez superado o legado colonial” (Camargo, 2016, p.126). Notadamente, a comparação internacional será internalizada menos como um instrumento para o monitoramento da ação governamental e mais “como revelação de uma fraqueza, que deve ser combatida em nome da aspiração progressista partilhada pelas elites” (Camargo, 2016, p.127). Assim, o interesse no discurso e na produção estatística será “muito mais determinado pelas pretensões internacionais do Brasil, do que pela possibilidade de criação de um referente capaz de justificar o aumento da intervenção do Estado” (Camargo, 2016, p.129).

De acordo com Senra (2006), as estatísticas já eram ‘desejadas’ pelo Brasil Império desde sua independência em 1822, o que se manifestava, por exemplo, em calorosos discursos parlamentares, mas em poucas ações efetivas. Durante a maior parte do Séc. XIX, os registros estatísticos do Brasil Império limitavam-se aos alistamentos eleitorais organizados por autoridades eclesiais, comissões que lidavam com questões de limites territoriais e estimativas populacionais por iniciativa de particulares. As instabilidades políticas na primeira metade do Séc. XIX, marcadas por revoluções separatistas regionais, explicam em parte as iniciativas frustradas de organização centralizada das estatísticas neste período (Senra, 2006).

A partir de 1850, entretanto, verificam-se esforços legislativos para centralizar a produção estatística nacional. A centralização estatística se insere em um conjunto de medidas que visavam, por um lado, consolidar o Estado Nacional e, por outro, lançar as bases para a emancipação de uma economia capitalista de mercado no

Brasil. Tratava-se de impulsionar a partir do Estado um processo de ‘modernização’, que, por um lado, se espelhava nas transformações econômicas e políticas dos países europeus recém industrializados e, por outro, buscava acomodar os interesses da classe senhorial, latifundiária e escravagista nacional. Assim, são promulgadas a partir de 1850: a Lei de Terras que regulamentou a propriedade fundiária e instituiu um mercado de terras privilegiando o sistema latifundiário; a Lei Eusébio de Queiroz que suspendeu o tráfico negreiro abrindo o caminho para a criação de um mercado de trabalho; a Lei do código comercial que instituiu um quadro normativo para disciplinar as relações das empresas privadas no mercado nacional; e finalmente, as leis que determinaram a realização de um recenseamento geral e a criação de um serviço de registro civil em 1852, e, posteriormente, a lei que impôs a adoção do sistema métrico e decimal em 1862 (Sodré, 1987; Faoro, 1989).

É curioso observar que dentre o pacote de leis ‘modernizadoras’, foram justamente a criação do registro civil, o recenseamento geral e a imposição do sistema métrico e decimal que enfrentaram as mais violentas reações populares (Senra, 2006). Os eventos históricos conhecidos como a ‘revolta dos marimbondos’ e ‘revolta dos quebra quilos’ são emblemáticos de como a construção do monopólio público sobre os capitais informacionais, implicou em desapossamentos e, não raro, em resistências. Em 1852, os levantes populares levaram ao cancelamento do recenseamento geral. Já a imposição legal do sistema métrico e decimal enfrentou aguerrida resistência popular em diversas províncias contra a extinção das medidas locais, em eventos que se caracterizaram pela “destruição de pesos, balanças e medidas; queima de documentos relativos às coletorias fiscais; destruição de prédios e de objetos pertencentes a repartições do governo” (Senra, 2006, p.22).

É apenas vinte anos após a experiência frustrada do recenseamento geral, e já sob a decisiva influência do internacionalismo estatístico, que, em 1872, o Brasil realizará o seu primeiro censo nacional. Notadamente, o envio de representações oficiais do Brasil para os Congressos Internacionais de Estatística, conforme vimos na seção anterior, fazia parte de uma estratégia que visava posicionar o país “no seletivo círculo das nações civilizadas que haviam produzido um censo moderno, chancelando as pretensões internacionais do Império” (Camargo, 2018, p.3). A influência dos congressos na primeira experiência estatística nacional do Brasil pode ser identificada, por exemplo, na adoção do modelo de um escritório central — com a criação da Diretoria Geral de Estatística (DGE) em 1871 — bem como na incorporação

de critérios técnicos no censo de 1872, que correspondiam ao estado da arte da tecnologia censitária moderna, como a previsão legal, a definição de um período e território de referência e a enumeração exaustiva e simultânea da população (Botelho, 1998; Camargo, 2018).

De acordo com Camargo (2016), a partir do censo de 1872, a população irrompe como fonte de construção utópica da nação brasileira, embora a representação hierárquica e homogênea da população, que daí emerge, corresponda à uma representação idealizada pelas elites imperiais. Era a própria realização do censo, de acordo com o receituário internacional, e não necessariamente o uso de suas informações para orientar a ação governamental, que constituía o símbolo de modernidade desejado por estas elites. De todo modo, a criação de uma estrutura estatística na burocracia central do Estado e a gradual incorporação de tecnologias de governo como os censos, tabelas, gráficos e mapas, serão fundamentais para que, em um momento posterior, movimentos nacionais sejam capazes de dar um sentido político a esta estrutura e às equivalências construídas pelas práticas estatísticas. No Brasil, conforme veremos, estes serão os casos dos movimentos sanitarista e educacional na Primeira República, e do nacional desenvolvimentismo na era Vargas.

Para Camargo (2016, p.86), enquanto fenômeno endógeno das sociedades coloniais e pós-coloniais, o nacionalismo quando integrado à racionalidade governamental “redimensiona o papel até então desempenhado pelas dinâmicas internas na mudança social”. O trabalho de Camargo (2016) sobre a construção da estatística no Brasil Império e Primeira República nos ajuda, portanto, a compreender a estruturação do campo estatístico nos países pós-coloniais, como “um jogo de influência recíproca entre o internacionalismo e o nacionalismo”.

[...] o internacionalismo científico e estatístico, gerado no seio do governo liberal, impõe o modelo normativo de desenvolvimento a ser aspirado por todas as nações que se querem civilizadas e que integram o sistema de Estados. Os projetos de construção do Estado-nação, no Brasil e na América Latina, foram em larga medida moldados pela emergência e consolidação daquele sistema. Quer dizer que tanto o repertório institucional quanto as tecnologias de governo característicos do Estado-nação são constituídos em uma base relacional. Concebidos pela racionalidade liberal, saberes (como a demografia e a saúde pública) e dispositivos (como censos e seguros) serão desejados e recomendados em acordos internacionais, representando a modernidade política e cultural das potências nacionais a que aspiravam os Estados de passado colonial. A adoção de certas tendências, como a contagem exaustiva e periódica da população, pode ser explicada pela dinâmica supranacional, ainda que a sua produtividade política dependa, em última análise, da construção de equivalências alavancadas internamente

através do movimento nacional. Trata-se de um jogo de influência recíproca entre o internacionalismo e o nacionalismo (Camargo, 2016, p.86-87).

De acordo com Camargo (2016, p.109-111), enquanto nos países centrais o nacionalismo nasceu inscrito na mesma lógica da governamentalidade liberal e de seus mecanismos de segurança — referenciando-se nos princípios da cidadania, soberania e universalidade dos direitos — nos países de passado colonial, a conformação dos movimentos nacionais e anticoloniais se deu sob determinações próprias e radicalmente distintas. Aqui, os próprios programas operacionalizados pela governamentalidade colonial, baseados na diferença racial e na estigmatização, bem como a transplantação de parte do repertório moderno da governamentalidade liberal para estes países, “contribuíram, para produzir de forma não planejada um horizonte de superação”. Deste modo, o próprio espaço de comparação criado pela ordem internacional contribui para inflar o movimento nacional a partir da “tomada de consciência de um ‘déficit civilizatório’ [...] mas que só será acionado quando o projeto ético-político de letrados e setores médios em ascensão lograr traduzir seus interesses em interesse nacional”.

Podemos abstrair da contribuição de Camargo (2016) a existência de uma relação dialética entre internacionalismo e nacionalismo estatístico nos países pós coloniais. Assim, no caso do Brasil, tudo parece funcionar como se a influência do internacionalismo estatístico na conformação da estatística oficial nacional operasse como um artefato simbólico para a construção de uma imagem desejada de modernidade. Por outro lado, conforme veremos, é a instrumentalização a posteriori dos números públicos e equivalências construídas por movimentos nacionais, que poderá dar significado e consequência política às práticas estatísticas, conferindo legitimidade e autoridade às estruturas estatísticas nacionais e, assim, impulsionando a estruturação de um campo estatístico nacional.

Finalmente, a decisiva influência do internacionalismo nas primeiras experiências estatísticas dos países pós-coloniais na segunda metade do Séc. XIX, pode ser evidenciada pelos próprios efeitos de sua crise. Verifica-se com o encerramento dos congressos em 1878 a dissolução de diversos escritórios nacionais de estatística nestes países. No caso brasileiro, os efeitos foram imediatos com o rebaixamento da Diretoria Geral de Estatística ao status de repartição em 1879, logo após a dissolução dos CIE (Camargo, 2016).

Conforme veremos na próxima seção, a crise do internacionalismo estatístico está inserida em um contexto ampliado de crises do capitalismo e da governamentalidade liberal a partir do final do Séc. XIX. Estas crises conduziram a mudanças profundas no modo de ação do Estado, no papel desempenhado pelas estatísticas oficiais e no desenvolvimento de novas ferramentas cognitivas da estatística inferencial, resultando em um processo prolongado de transição de regime estatístico.

1.8 Crise de governamentalidade liberal e transição de regime estatístico

O último quarto do Séc. XIX e as primeiras décadas do Séc. XX, inauguram um período prolongado de crises sistêmicas do capitalismo. Nomeadamente, verificam-se a longa depressão do último quarto do Séc. XIX e a crise financeira mundial de 1929²². De acordo com Anglietta (2015), as crises financeiras são os elementos visíveis de uma ruptura mais profunda do processo de acumulação do capital que revelam as limitações do modo de regulação fordista e abrem caminho para uma nova etapa de acumulação/regulação, o capitalismo monopolista de Estado.

De acordo com Foucault (2008a), as crises do capitalismo corresponderão por sua vez à crises da governamentalidade liberal, na medida em que passam a impor a necessidade da intervenção governamental no campo social e econômico, colocando em xeque a justificação doutrinal do liberalismo. Para Dardot e Laval (2016, p.38), o período de 1880 a 1930, corresponde a uma ‘longa crise do liberalismo’, em que se verificam a revisão dos dogmas liberais em todos os países industrializados. Neste sentido, o aumento da intervenção do Estado passaria a “fazer frente às mutações

²² A primeira grande crise de 1873, se manifesta como uma duradoura depressão econômica que se estende até o início do Séc. XX, a qual corresponde aos efeitos da segunda revolução industrial e da entrada do capitalismo em uma etapa monopolista. Verifica-se neste período uma elevada concentração de capitais e uma superprodução de mercadorias, acompanhada pelo aumento do desemprego, retração dos salários e empobrecimento da classe trabalhadora. O descompasso entre o aumento da oferta e a queda da demanda, resultou em um longo período de deflação e contração geral da atividade econômica. A longa depressão do Séc. XIX desencadeou uma série de turbulências sociais internas e deflagrou uma competição imperialista entre os países Europeus em busca de novos mercados e colônias. As disputas imperialistas conduzirão à conflagração da Primeira Guerra Mundial, após a qual se seguirá um período de reconstrução e forte crescimento econômico nos Estados Unidos e Europa. Os anos dourados da década de 1920 conduzirão a um novo cenário de superprodução mundial agravado pela especulação no mercado de ações. A queda das taxas de lucros, retração da produção industrial e paralisação do comércio conduz à uma queda generalizada nos valores das ações o que culmina na crise mundial de 1929 e na grande depressão dos anos 1930 (Curvo, 2021).

organizacionais do capitalismo, aos conflitos de classe e às novas relações de força internacionais”.

De acordo com Polanyi (2021), o elemento central destas crises residia nos esforços do liberalismo econômico em desenraizar a economia da sociedade, constituindo um mercado autorregulado que passava a desestruturar aspectos fundamentais da vida humana — nomeadamente o trabalho, a terra e o dinheiro — convertendo-os em ‘mercadorias fictícias’. Na medida em que se evidenciavam os efeitos e consequências destrutivas da subordinação da sociedade ao mercado, o princípio do liberalismo econômico entrou em choque com um princípio de proteção social que visava a conservação do homem, da natureza e da organização produtiva, valendo-se de legislações protetivas e outros instrumentos de intervenção. A ação destes dois princípios contraditórios, que Polanyi (2021) intitula de ‘duplo movimento’, levou a uma profunda tensão institucional nos Estados Nação, agravada por conflitos internos de classe e disputas imperialistas, que conduziram em última instância à deflagração dos conflitos mundiais do Séc. XX.

Para Polanyi (2021), na medida em que os dois princípios organizadores da sociedade, o princípio do mercado autorregulado e o princípio da proteção social, entravam em choque, mobilizando forças sociais e objetivos institucionais contraditórios, a ação do Estado correspondeu igualmente a um duplo movimento. Por um lado, o Estado atuou na criação de mecanismos que apoiavam o movimento em direção a uma sociedade de mercado, e por outro, implantou instrumentos que o limitavam, reforçando um contramovimento de proteção da sociedade contra os efeitos nocivos do mercado autorregulado.

Neste contexto, a estatística oficial passou a constituir um elo de articulação dos dois princípios da ação do Estado. Por um lado provendo os dados necessários ao funcionamento do mercado e à extração de mais-valor, e por outro, produzindo as informações demandadas para as intervenções biopolíticas norteadas pelo princípio de proteção social. Neste duplo movimento, as estatísticas passaram igualmente a fornecer subsídios que fundamentavam o reformismo social e a crítica ao capital que emerge em meados do Séc. XIX²³.

²³ O uso crítico de dados estatísticos destaca-se, entre outros, na obra de Karl Marx. Em sua crítica à economia política e análise do funcionamento da estrutura do capital, Marx se valeu de uma série de dados estatísticos e conceitos analíticos da estatística descritiva de Quetelet, em particular a aplicação das médias para a construção e análise de categorias abstratas, a exemplo do trabalho médio e dos preços médios. Em *O Capital*, Marx faz uso de estatísticas de expectativa de vida dos

De acordo com Dardot e Laval (2016, p.37), as primeiras linhas de fratura do programa liberal, podem ser identificadas a partir de meados do séc. XIX, na tensão “entre dois tipos de liberalismo, o dos reformistas sociais que defendem um ideal de bem comum e o dos partidários da liberdade individual como um fim absoluto”, incluídos no segundo grupo os darwinistas sociais. Estas duas correntes de pensamento e ação no âmbito do liberalismo, por sua vez, influenciarão importantes desenvolvimentos no campo estatístico.

As diversas iniciativas de reforma social que emergiram nos países industrializados no final do Séc. XIX passaram a enquadrar o problema da pobreza do ponto de vista de seus fatores econômicos e sociais. Pela via legislativa e da intervenção estatal, os contramovimentos reformistas buscaram estabelecer mecanismos de proteção dos trabalhadores assalariados contra os efeitos da lógica de mercado sobre o mundo do trabalho. De acordo com Polanyi (2021, p.218), diferentemente do *laissez faire* que resultou de uma ação deliberada e planejada do Estado, este contramovimento que se opunha a lógica do mercado autorregulado revelou-se espontâneo e “acionado por um espírito puramente pragmático”²⁴. Esta reorientação política, que passaria a caracterizar o ideal de ‘bem estar social’, concorreu para a construção de sistemas de proteção social baseados na lógica da seguridade (seguro desemprego, seguro saúde, pensões, etc.). Neste contexto, o desenvolvimento das estatísticas do trabalho promoveu uma profunda renovação nos objetos, práticas e métodos das estatísticas oficiais, destacando-se a realização dos primeiros inquéritos por amostragem.

Nesse período, das décadas de 1880 a 1930, as estatísticas do trabalho foram o motor da renovação das estatísticas oficiais, tanto pelos objetos estudados quanto pelos métodos de investigação. Os salários, o emprego e o desemprego, a atividade profissional, os orçamentos dos trabalhadores, o

trabalhadores de diferentes ramos da indústria para denunciar as condições insalubres a que eram submetidos, analisa séries históricas para desvendar o movimento dos preços, utiliza estatísticas demográficas para demonstrar problemas distributivos nas sociedades capitalistas e finalmente a partir da análise de dados das flutuações do setor têxtil da Inglaterra, constrói a teorização dos ciclos econômicos, que exercerá forte influência na teoria econômica e em particular na econometria do Séc. XX (Rubin, 1968).

²⁴ A tese de Polanyi (2021, p.223-224) de que o ‘laissez-faire foi planejado e o planejamento não’ apoia-se em evidências que demonstram “a espantosa diversidade das questões sobre as quais se tomaram medidas”, dentre as quais um levantamento realizado por Herbert Spencer para sustentar a ideia de uma ‘conspiração antiliberal’, quando de fato a diversidade de leis e ações que lidavam “apenas com um problema advindo das condições industriais modernas e visava salvaguardar algum interesse público contra os perigos inerentes a essas condições [...] revelavam a natureza puramente prática e pragmática do contramovimento”.

índice de preços ao consumidor [...] foram integrados à agenda dos bureaux de estatística. Para mensurá-los, novas formas de pesquisas, baseadas em amostras representativas foram concebidas (Desrosières, 2008, p.48, tradução nossa).

A crescente intervenção do Estado no campo econômico e social será objeto de uma violenta reação dos defensores do livre mercado. Sob a influência do darwinismo social de Herbert Spencer (1820-1903), o dogmatismo do *laissez faire* é renovado no final do século XIX por uma ideologia concorrencialista que “institui a competição como norma geral da existência individual e coletiva, tanto da vida nacional como internacional” (Dardot e Laval, 2016, p.54). Segundo esta nova doutrina, o progresso da sociedade passava a pressupor a destruição de seus componentes menos aptos. Para Dardot e Laval (2016, p.45), o ‘spencerismo’ introduziu alguns dos temas que serão posteriormente renovados em meados do Séc. XX pelo neoliberalismo, em especial “a primazia da concorrência nas relações sociais”.

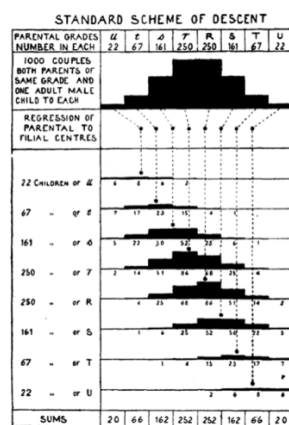
Esta corrente concorrencialista encontrou na eugenia de Francis Galton, Karl Pearson e Ronald Fisher, a materialização científica da ideia de progresso social baseado na seleção dos mais aptos e fortes (Galton, 1907). A pobreza para os eugenistas não decorria de questões sociais e econômicas, mas constituía um problema de natureza biológica a ser equacionado por meio da esterilização dos pobres e deficientes físicos e mentais, em nome da melhoria genética da população. Curiosamente, decorrem dos trabalhos controversos de Galton, Pearson e Fisher, no campo da eugenia, as principais elaborações da matemática estatística, em particular, a correlação, a regressão e os testes estatísticos (Desrosières, 1998).

Na formalização estatística da eugenia, a construção cognitiva da curva normal volta a ser utilizada para postular uma distribuição ‘normal’ e hereditária das capacidades humanas de acordo com as classes sociais. Em virtude da menor fertilidade das classes altas em comparação à classe trabalhadora, o tratado eugenista de Galton (1909) apontava para o risco de declínio qualitativo das aptidões e talentos das nações, concluindo que a ‘melhoria da raça humana’ implicava na esterilização das classes mais baixas²⁵.

²⁵ Neste quadro extraído dos ensaios eugenista de Galton (1909), as letras maiúsculas (R, S, T, U) indicam as classes de indivíduos acima da mediocridade e as letras minúsculas (r, s, t, u) àquelas abaixo da mediocridade. As curvas de normalidade indicam a distribuição das aptidões nas distintas classes sociais. Para Desrosières (2008a) as classificações estatísticas de habilidades hereditárias propostas por Galton, antecipam, em certo sentido, a ideia subjacente aos rankings de desempenho

Figura 11 - A Eugenia de Galton e a Curva Normal

THE POSSIBLE IMPROVEMENT OF
THE HUMAN BREED,
UNDER THE EXISTING CONDITIONS OF
LAW AND SENTIMENT.*



Fonte: Galton (1909, p.2)

De acordo com Desrosières (2008a), foi para defender as ideias eugenistas de Galton, que Karl Pearson moldou as noções de correlação, regressão e teste qui-quadrado, “assim, esta configuração desacreditada deixará, no entanto, as premissas da matemática estatística, que serão herdadas pela estatística inferencial (Desrosières, 2008a, p.18). Construindo suas formulações sobre o legado controverso de Galton e Pearson, os estatísticos inferenciais — Wilhelm Lexis (1837-1914), George Udny Yule (1871-1951), Ronald Fisher (1890-1962), Jerzy Neyman (1894-1981) e Egon Pearson (1895-1980) — apesar de críticos às formulações matematicamente limitadas da estatística social, adotarão os mesmos pressupostos frequentistas de Quetelet, compreendendo a probabilidade como um cálculo estritamente objetivo e baseado em frequências. Estes matemáticos estatísticos, entretanto, estavam muito mais empenhados na análise das variabilidades, correlações e dispersões de diferentes frequências de dados, do que em suas médias estáveis. Notadamente, sua abordagem mais complexa implicava em maior rigor na aplicação das leis da probabilidade. Suas novas teorias, métodos e construtos — amostragem, teste de hipóteses, significância, intervalo de confiança, margem de erro, regressão, etc. — provocarão um profundo impacto na ciência experimental e também afetarão substancialmente as práticas dos escritórios nacionais de estatística ao longo

que surgirão um século depois com o desenvolvimento das políticas de indicadores e benchmarking típicas do neoliberalismo.

do Séc. XX.

Conforme vimos, o estilo das estatísticas oficiais produzidas de acordo com as noções preconizadas pela estatística social de Quetelet, ligava-se à ideia de exaustividade. Os levantamentos censitários cobrindo o universo da população compreendida no território nacional e o cálculo das médias, que possibilitava transformar ações individuais aleatórias em agregados estáveis, passaram a constituir o núcleo central dos instrumentos estatísticos públicos desenvolvidos em meados do Séc. XIX. Formaram-se assim os espaços nacionais de convenções de equivalências à semelhança de grandes ‘urnas de Bernoulli’, das quais, entretanto, ainda não se extraíam amostras. Apesar do ponto de vista de Quetelet ter feito quase desaparecer a especulação probabilística no meio estatístico por cerca de um século — na medida em que os procedimentos amostrais eram rejeitados pela lógica da enumeração exaustiva — para Desrosières (2008), o Estado do bem estar social caracteriza-se como um ‘Estado probabilístico’ que se fundamenta na intuição original de Quetelet:

O estado de bem-estar está associado à ideia de seguro. Um procedimento baseado em um cálculo estatístico garante a cobertura dos riscos, com base nas probabilidades (medidas por frequências) dos diversos eventos descritos pelas novas estatísticas do trabalho. Assim, o estado de bem-estar é um estado probabilístico. Baseia-se na intuição central de Adolphe Quetelet: as quantidades agregadas, resumidas por médias, possuem propriedades de estabilidade e previsibilidade que faltam aos indivíduos. Esta é a justificativa para o mecanismo de seguro. Além disso, tratando-se a população de maneira uniforme segundo as mesmas leis em nível nacional, é coerente considerá-la como uma urna probabilística da qual se extraem amostras [...] Assim, as filosofias políticas e os esquemas cognitivos do estado de bem-estar estão intimamente interligados. Ao mesmo tempo, construiu-se uma nova forma de governo e uma nova forma de fazer estatística (Desrosières, 2008, p.49, tradução nossa).

O novo conceito de ‘representatividade’, que possibilitará o desenvolvimento dos inquéritos amostrais, emerge na virada do século vinculado aos contramovimentos reformistas e aos estudos sobre pobreza, renda e trabalho, que subsidiam novas iniciativas legislativas e governamentais de proteção social nos países industrializados. Na Inglaterra, os debates sobre a pobreza que conduziram a uma reforma da *Poor law* em 1908, foram subsidiados por um conjunto de inquéritos sobre as condições de vida da classe trabalhadora realizados Arthur Bowley (1869-1957). Na Noruega, em 1895, a realização do primeiro survey amostral por um Bureau estatístico nacional, então comandado por Anders Nicolai Kiaer (1838-1919), atendia, por sua vez, ao objetivo de criação de um sistema de previdência social (Desrosières,

1998).

Na medida em que a pobreza deixava de ser tratada como um problema de moral individual, para constituir uma questão nacional objeto de políticas sociais, dados estatísticos periódicos e comparáveis passaram a ser demandados pelos governos que adotavam a intervenção protetiva. Diferente da enumeração exaustiva dos censos e dos estudos descritivos e monográficos que caracterizaram o Séc. XIX, o ‘método representativo’ introduzido nos inquéritos realizados por Bowley e Kiaer, atendia justamente a uma nova abordagem estatística voltada a subsidiar ações específicas de intervenção do Estado no campo social.

Outra importante mudança no processo produtivo das estatísticas oficiais, que se verifica entre o final do Séc. XIX e início do Séc. XX, foi a introdução de máquinas de tabulação para o processamento de dados dos censos e demais inquéritos estatísticos (Anderson, 2015). O processamento mecânico dos dados representou um importante salto tecnológico nos escritórios estatísticos, que possibilitou encurtar o tempo entre a coleta de dados e a disponibilização pública das informações. Desenvolvidas especificamente para atender a necessidade de processamento dos grandes volumes de dados das operações censitárias, as máquinas tabuladores constam como precursores mecânicos dos computadores eletrônicos²⁶.

Estas inovações possibilitam, por um lado, maior agilidade nos levantamentos estatísticos por enumeração, e, por outro, concorrem para a construção de uma nova noção de representatividade, na medida em que os inquéritos amostrais permitem expandir para a escala nacional, resultados parciais obtidos em operações mais rápidas, mais específicas e menos onerosas que os censos. A introdução da amostragem nos bureaux estatísticos, entretanto, foi um processo gradual e permeado por resistências internas. A experiência norueguesa, por exemplo, foi objeto dos primeiros debates sobre o tema no Instituto Internacional de Estatística (ISI), nos quais de acordo com Beaud e Prévost (1998, p.694), “o método representativo encontrou uma oposição feroz daqueles que estavam apegados às virtudes do censo exaustivo”²⁷.

²⁶ A máquina de tabulação desenvolvida pelo funcionário do US Census Herman Hollerith (1860-1929), para o processamento dos dados do Censo dos EUA em 1890, consta como o evento de fundação da Tabulating Machine Company (TMC), posteriormente renomeada para International Business Machines (IBM). O computador mecânico de Hollerith foi posteriormente disseminado para os escritórios estatísticos do Canadá, Europa e América Latina. No Brasil consta a sua introdução no Censo de 1920.

²⁷ Para Beaud e Prévost (1998, p.694) a história internalista sobre o desenvolvimento da amostragem

Para Beaud e Prévost (1998), a introdução da amostragem nos escritórios nacionais, não dependeu apenas do desenvolvimento cognitivo das ferramentas estatísticas, mas sobretudo da característica do modo de ação do Estado. Os métodos de enumeração exaustiva, apesar de demandarem uma grande estrutura administrativa de coleta de dados, não exigiam muito mais dos governos além de oferecer aos agentes um conhecimento geral que orientava as suas ações. Já os métodos de amostragem passavam a oferecer a possibilidade de ajustar a precisão dos dados de acordo com a demanda específica da informação a ser obtida e com os recursos disponíveis. Este novo tipo de levantamento, portanto, pressupõe que os usos para os quais os dados são destinados sejam bem especificados a priori: “a nova concepção de precisão das técnicas de amostragem só fazia sentido para um governo inclinado à uma intervenção ativa, um governo para o qual os dados significassem ação” (Beaud e Prévost, 1998, p.699).

As condições para a incorporação dos surveys amostrais à prática estatística de Estado se deram efetivamente nos Estados Unidos na década de 1930 sob o governo de Franklin D. Roosevelt (1882-1945). Verifica-se neste contexto, a coincidência de três fatores decisivos. Em primeiro lugar, com o desemprego atingindo níveis críticos uma forte pressão foi exercida sobre o US Census Bureau para fornecer dados rápidos para a organização de uma política nacional. Em segundo lugar, a partir de 1934, o método de amostragem estratificada de Neyman passava a oferecer uma solução tecnicamente viável e politicamente respaldada pelo campo estatístico para a questão da representatividade dos estudos amostrais. Finalmente, a política do *New Deal* de Roosevelt correspondia exatamente a um novo modo de intervenção governamental ativa, que passava a demandar instrumentos de precisão para orientar a ação do Estado.

probabilística, tende a apresentá-la normalmente como um processo evolutivo em três etapas. A primeira fase corresponderia ao estudo norueguês e sua repercussão no ISI entre 1897 a 1903, tendo os primeiros debates girado em torno da oposição entre o ainda pouco conhecido método de amostragem e o já institucionalizado sistema de enumeração completa. A segunda etapa corresponderia a retomada da discussão sobre o método representativo no ISI em 1925, quando se verifica uma mudança nos termos do debate. A questão aqui já não se trataria mais sobre a validade da amostragem, mas sobre qual método de amostragem seria o mais adequado, o aleatório ou o intencional, este último defendido pelo estatístico italiano Conrado Gini (1884-1965). Finalmente a terceira etapa, seria marcada pelo triunfo da amostragem probabilística com a publicação do artigo de Jerzy Neyman (1894-1980) em 1934, no qual o estatístico polonês e discípulo de Karl Pearson apresenta a teoria da estratificação e o conceito de ‘intervalo de confiança’, demonstrando definitivamente a superioridade do método aleatório sobre o intencional.

Sob esta nova lógica governamental, verifica-se não apenas a incorporação dos inquéritos amostrais à prática estatística de Estado, mas uma completa reestruturação dos escritórios estatísticos nos EUA, com o recrutamento de uma nova geração de estatísticos, matemáticos e economistas. Estes centros viriam a se transformar em locais de inovação e experimentação dos novos ramos do conhecimento derivados da matematização da estatística e da economia. Para Desrosières (1998, p.199), este período marca o encontro entre as duas tradições estatísticas, administrativa e matemática, as quais “convergiam em uma construção doravante dotada de dupla legitimidade: a do Estado e a da ciência” (Desrosières, 1998, p.199)

No espaço de alguns anos, entre 1933 e 1940, os termos do debate social e as ferramentas estatísticas que o informavam foram radicalmente alterados. As duas transformações estiveram intimamente ligadas, pois o mesmo período testemunhou tanto o nascimento de uma nova maneira de conceber e administrar os desequilíbrios políticos, econômicos e sociais do estado federal americano, quanto uma linguagem para expressar essa ação. Desemprego, visto em escala nacional; desigualdade entre classes, raças ou regiões; o próprio fato de equipar esses objetos com ferramentas estatísticas para debatê-los – tudo isso entrou na nova linguagem, que se tornou comum em todos os países ocidentais após 1945 (Desrosières, 1998, p.199).

A transição de um ‘regime estatístico de enumeração’, amparado na exaustividade dos censos, para um novo ‘regime estatístico de precisão’, fundamentado na estatística inferencial e na econometria, conduzirá a uma mudança estrutural no campo estatístico a partir da reformulação das práticas e do perfil profissional dos agentes do campo. Deste modo, as gerações de estatistas e administradores generalistas, formados sob os ideais de exaustividade da estatística social, perderão gradualmente espaço para os modernos estatísticos de Estado, profissionais especializados e versados em métodos quantitativos avançados.

Para Beaud e Prévost (1998), o processo de transição de regime estatístico será muito mais dilatado cronologicamente em países como o Canadá, onde os autores realizaram sua pesquisa empírica, e sobretudo em países periféricos e de industrialização tardia como o Brasil. No caso canadense, as novas técnicas de precisão não encontraram aderência no modo de ação do Estado, ainda pautado pela não intervenção do liberalismo clássico, o que, somado a resistências internas no Bureau estatístico do Canadá, retardaram a sua incorporação. A pesquisa empírica de Beaud e Prévost (1998) nos permite compreender como o descompasso entre os

âmbitos cognitivo e político-administrativo, tende a se refletir em disputas no campo estatístico, em torno de mudanças que possam afetar a lógica e a estrutura do campo.

A amostragem, enquanto avanço teórico e técnico, não conseguiu obter o acordo dos principais estatísticos do governo do Canadá quanto aos seus próprios méritos científicos, apesar das suas promessas como multiplicador económico de inquéritos estatísticos. A tradição do censo [...] baseava-se numa máquina administrativa na qual muito esforço tinha sido investido e que exigia um exército de funcionários, mas um núcleo muito pequeno de pessoal altamente qualificado. Além disso, a concepção que poderia tornar a amostragem atrativa — a de um governo intervencionista cuja tomada de decisões precisava ser sustentada em um grande volume de dados recentes — ainda estava longe de ser amplamente partilhada na década de 1930 (Beaud; Prévost, 1998, p.724, tradução nossa).

Em resumo, assinalamos quatro elementos relevantes desta seção para a composição de nosso modelo de transição de regimes estatísticos e seus efeitos sobre o modo de estruturação do campo estatístico. Em primeiro lugar, a crise de governamentalidade que deriva da crises do regime de acumulação do capital, demanda respostas e soluções técnicas e políticas, abrindo um espaço de possibilidades para novos investimentos económicos, políticos e cognitivos em novas ferramentas e práticas estatísticas. Em segundo lugar, o campo estatístico tende a se reconfigurar de acordo com os princípios contraditórios do duplo movimento (pró-mercado e protetivo), se conformando ao modo predominante de ação do Estado. Em terceiro lugar, a incorporação de novas práticas pelo campo estatístico relaciona-se, por um lado, com o nível de desenvolvimento cognitivo destas práticas e, por outro, com a sua aderência ao modo de ação do Estado, o que em última instância determina a lógica dos investimentos e a transição para um novo regime estatístico.

Na próxima seção analisamos como no caso brasileiro, sob o contexto de crise da governamentalidade liberal, a emergência de contramovimentos de carácter nacional concorrem para uma profunda reformulação do Estado e das práticas estatísticas. É sob um Estado interventor orientado por um projeto de desenvolvimento nacional em bases capitalistas que se verifica a construção de um sistema estatístico nacional no Brasil. A transição, no caso brasileiro, consolida o regime estatístico de enumeração, revelando uma 'dilatação cronológica' entre o desenvolvimento das práticas estatísticas nos países industrializados e periféricos.

1.9 Os contramovimentos nacionais e a consolidação do regime de enumeração no Brasil

No Brasil, a ‘longa crise do liberalismo’ corresponde de um modo geral ao período de transição compreendido entre a derrubada do Império, no final do séc. XIX, e a derrocada da Primeira República, no final da segunda década do século XX. Para Cardoso e Faletto (1970), este período é caracterizado por um processo histórico-estrutural de diferenciação da economia exportadora que, por um lado, criou as bases para o surgimento de setores médios na dinâmica social e política da sociedade e por outro, constituiu uma burguesia nacional, vinculada ao exterior e apoiada por um complexo sistema de alianças interno com oligarquias regionais e estamentos burocráticos. No jogo de alianças regionais, predominavam os interesses ligados ao açúcar e ao café. A substituição da mão de obra escrava pelos cafeicultores de São Paulo, a partir de 1870, marca a cristalização de uma situação capitalista burguesa, que leva gradualmente a uma maior divisão social do trabalho e fortalecimento da economia urbana. A abolição da escravatura e a derrubada do império resultam deste processo, com o surgimento de novos grupos de pressão política (burocracia civil e militar).

A adesão a determinados aspectos da doutrina liberal pela elite política da república — como o individualismo político e econômico, o federalismo, o metalismo, o laissez faire, o mercado autorregulado, a teoria das vantagens comparativas e a vocação natural do Brasil para a especialização primária-exportadora — em uma estrutura econômica baseada na agro exportação, conjuga-se com as modalidades arcaicas de poder, baseadas no clientelismo e coronelismo, para a formação de um ‘liberalismo oligárquico’ (Resende, 2018). A estrutura de Estado que deriva deste arranjo e a forma federativa republicana passam a expressar o sistema de alianças locais, sob predomínio do centro-sul, sem excluir as oligarquias regionais e a presença de novos grupos anti-oligárquicos (Cardoso e Faletto, 1970). A ação do Estado se caracterizará no campo econômico pela proteção ao latifúndio e à agro exportação por um lado, e, por outro, à política fiscal necessária à própria expansão da estrutura do Estado.

A restauração da Diretoria Geral de Estatísticas (DGE), que se dá sob a Primeira República, reflete o apreço pelas fórmulas políticas liberais europeias e estadunidenses. Neste sentido, verifica-se a adoção de recomendações do

internacionalismo estatístico, mesmo sem as condições materiais e institucionais necessárias à sua implantação. Assim, são estabelecidas na constituição de 1891 a regularidade censitária decenal e a proporcionalidade da representação política dos estados de acordo com a população recenseada (Camargo, 2016). No que pesem as relevantes tarefas constitucionais atribuídas à DGE, o órgão caracterizou-se por uma notória precariedade material, técnica e política. Com um quadro pessoal mínimo e sem autoridade sobre os governos estaduais, tanto as operações censitárias conduzidas em 1890 e 1900 se revelaram “incríveis fracassos” (Senra, 2006a, p.56), quanto o trabalho de compilação estatística a partir de registros administrativos produziu resultados débeis.

De acordo com Camargo (2016), é apenas a partir do desenvolvimento de uma ideologia nacionalista no Brasil que a autoridade das estatísticas oficiais ganhará um novo impulso. Destaca-se, no início do Séc. XX, a emergência do movimento médico sanitaria e sua capacidade de tradução de interesses dos setores médios urbanos e segmentos da burocracia estatal em torno de um novo sentido de consciência nacional. O projeto nacional se volta para a interdependência da saúde e educação no território, para a solidarização com os grupos populacionais dispersos pelo país e para a necessidade de integração das ‘regiões’ nacionais por meio da presença do Estado. Do ponto de vista político e administrativo, a organização de serviços de demografia sanitária e a demanda por dados estatísticos confiáveis pela rede de médicos e demógrafos sanitaristas exercerá forte pressão sobre a DGE. Esta pressão se materializará na reformulação da Diretoria com a nomeação do médico Bulhões Carvalho, principal autoridade da demografia sanitária, para a sua direção em 1907²⁸.

Após a reformulação promovida por Bulhões Carvalho, verificam-se consideráveis avanços e realizações pela DGE, com destaque para a publicação do primeiro anuário estatístico do Brasil em 1916 e para a realização do censo de 1920, única operação censitária bem sucedida da primeira república (Senra, 2006a). Algumas inovações são dignas de nota nestas iniciativas, como a introdução do

²⁸ Considerado o fundador da estatística geral brasileira, deve-se a Bulhões Carvalho a tentativa pioneira de construção de um sistema estatístico nacional. O decreto de reformulação da DGE em 1907 atribuiu à Diretoria a competência de execução dos levantamentos estatísticos e criou o Conselho Superior de Estatística como um órgão central com função de orientação dos demais órgãos da administração federal, estadual e municipal responsáveis pela execução dos serviços estatísticos. Apesar deste ‘sistema’ não ter conseguido atingir seus objetivos em termos de coordenação das atividades estatísticas, sob o quadro federativo do Estado Oligárquico, será a base para o desenvolvimento posterior da estrutura estatística nacional no governo de Getúlio Vargas (Penha, 1993).

primeiro computador mecânico no Brasil, utilizado na tabulação de dados do censo, e os significativos avanços em termos de visualização dos dados. Destacam-se a apresentação das estatísticas nacionais em mapas e séries históricas desagregadas por regiões, unidades da federação e municípios. Isto passa a possibilitar a visualização das desigualdades econômicas e sociais no território nacional, ao mesmo tempo em que, a construção de novos espaços de equivalência, como a região, abre “novas possibilidades para a imaginação da nação” (Camargo, 2016, p.239).

De acordo com Camargo (2016), a produção estatística que deriva da reformulação da DGE atuará para reforçar os principais temas da agenda nacional que vinha sendo formulada por novos atores como sanitaristas e educadores, colocando em evidência os limites da ordem oligárquica para a gestão dos problemas públicos. Para Camargo (2016), a prática e discurso estatísticos que emergem associados ao nascente movimento nacional, constituem o embrião da nova mentalidade de governo que emergirá sob as novas condições políticas do projeto nacional desenvolvimentista da era Vargas.

As mudanças na produção estatística brasileira nos anos 1920 estão inseridas em um conjunto de profundas transformações da sociedade brasileira após a Primeira Guerra Mundial. Estas transformações resultam em uma crise da dominação oligárquica, que se manifestará plenamente na Revolução de 1930. Do ponto de vista econômico, o processo de industrialização por substituição de importações e as fortes oscilações dos preços internacionais do café neste período geraram um conjunto de efeitos relevantes. Por um lado, verifica-se uma maior complexificação da estrutura econômica, por outro lado, crises fiscais e inflacionárias provocaram importantes desestabilizações políticas e sociais. Este processo é acompanhado pela intensificação da urbanização e crescimento das camadas médias. Os setores médios urbanos (profissionais liberais, jovens militares, artistas e intelectuais) passam a constituir uma reação aos padrões culturais e políticos das oligarquias tradicionais mobilizando-se em torno da ideia de construção de uma identidade e projeto nacionais. O nacionalismo emerge em uma série de contramovimentos que se opõem à dominação oligárquica e aos dogmas liberais. Destacam-se, neste sentido, movimentos de conformações e origens distintas como o tenentismo, o modernismo, o sanitarismo e o movimento educacional. Finalmente, o conjunto de transformações econômicas, políticas e sociais da década de 1920 contribui para minar as bases de sustentação do sistema oligárquico, o que, juntamente com a crise financeira mundial

de 1929, estabelece as condições objetivas para a mudança radical do quadro político pela via da força, com a ascensão de Getúlio Vargas ao poder, sustentado por uma aliança entre oligarquias e movimentos militares dissidentes (Fausto, 1997).

Do ponto de vista ideológico, o 'Estado de compromisso' que emerge da Revolução de 1930 está inserido no contexto da crise internacional da governamentalidade liberal. Assim, verificam-se na era Vargas tanto a adoção de fórmulas do 'novo liberalismo', que justificam o aumento do intervencionismo econômico e social e a proposta de cidadania centrada nos direitos do trabalho, como a adesão à matrizes fundamentalmente antiliberais com a incorporação de postulados autoritários e projetos de organização corporativista da sociedade e do Estado (Gomes, 2018). A conformação deste 'nacionalismo antiliberal' será reforçada no Estado Novo com a implantação de uma política centralista, que, sem desrespeitar os interesses locais agroexportadores, aumentou o controle governamental e o investimento na infraestrutura, formando uma base econômica urbana que permitiu o surgimento de uma burguesia industrial e mercantil, e de setores da nova classe média e operário-popular (Cardoso; Faletto, 1970).

A criação do Instituto Nacional de Estatística (INE) em 1936, rebatizado para Instituto Nacional de Geografia e Estatística (IBGE) em 1938, sob a liderança intelectual de Teixeira de Freitas, se dá neste contexto de centralização do poder político no Estado Novo. Tal centralização está imbricada a uma política de integração nacional pautada em mecanismos para o controle de esferas estratégicas da economia pelo Estado e na unificação político-administrativa que amplia o poder central sobre os governos estaduais e território nacional. As exigências da política econômica estatal, por sua vez, provocam a expansão da tecnoestrutura com a criação de novas instituições públicas voltadas ao planejamento e orientação da ação do Estado (Ianni, 1971). Deste modo, o IBGE nasce como um instrumento técnico, político e científico estratégico para a integração nacional e interiorização do Estado (Penha, 1993).

Diretamente subordinado à Presidência da República, o IBGE é concebido como um órgão responsável pela coordenação nacional das atividades estatísticas e geográficas da União, dos estados e municípios. Neste sentido, a criação do Conselho Nacional de Estatística (CNE) e o estabelecimento de um pacto de cooperação intergovernamental, consolidou a estrutura do Sistema Estatístico Nacional, presidida pelo IBGE a quem cabia a orientação e coordenação das atividades estatísticas dos

órgãos federais e serviços estatísticos estaduais e municipais responsáveis pela sua execução. Até 1940, ano de realização bem sucedida do primeiro censo sob a nova estrutura, todos os estados da federação e os 1574 municípios brasileiros passaram a dispor de um órgão exclusivamente dedicado à produção estatística (Penha, 1993). A partir de 1942, no contexto de ‘Estado de Guerra’, a rede estatística municipal será nacionalizada, concorrendo para a interiorização da administração federal por meio do IBGE, em um movimento que visava assegurar a soberania nacional sobre o território. Sob um Estado interventor e orientado por um modelo de desenvolvimento nacional, o Brasil consolida definitivamente sua entrada em um regime estatístico de enumeração, constituindo uma estrutura estatística de abrangência nacional, coordenada por um órgão central, processo que na maioria dos países europeus já havia se consolidado desde o final do Séc. XIX.

A constituição do IBGE como um ente coordenador de serviços estaduais com autonomia relativa, entretanto, impôs diversas limitações à produção estatística. Notadamente, para além da produção censitária, havia muita deficiência na produção de informações conjunturais, notadamente no campo econômico. Apesar da era Vargas ter marcado o início da intervenção direta do Estado nas questões vinculadas ao mundo trabalho — com a criação do Ministério do Trabalho em 1930, promulgação da Consolidação das Leis Trabalhistas (CLT) em 1943 e criação de uma rede de seguridade social — as informações sobre o mercado de trabalho limitavam-se aos levantamentos censitários e compilações de registros administrativos nos anuários estatísticos. Neste sentido, não se verifica neste período a consolidação de um sistema de estatísticas econômicas e estatísticas do trabalho apoiado em levantamentos conjunturais e amostrais como será o caso nos Estados Unidos e países europeus.

Embora a introdução da estatística inferencial no Brasil se verifique já na década de 1920 em trabalhos científicos de médicos e demógrafos sanitaristas, influenciados pelas correntes eugenistas (Camargo, 2016), a sua introdução nas práticas estatísticas de Estado será muito mais dilatada e se dará apenas partir do final da década de 1960, sob um novo conjunto de determinações nacionais e internacionais. Conforme veremos a seguir, nos países industrializados, a transição para um regime estatístico de precisão — que efetivamente incorpora a estatística inferencial ao processo da produção estatística nacional — se consolida no pós-guerra sob o quadro do capitalismo administrado, pautado pelo planejamento estatal e

macroeconomia keynesiana.

1.10 O capitalismo administrado e o regime estatístico de precisão

As crises econômicas sucessivas, entre o final do Séc. XIX e as primeiras décadas do Séc. XX, somadas à consolidação do regime soviético e a perspectiva real de novas revoluções socialistas, colocavam em questão a própria sobrevivência do sistema capitalista. Neste contexto, se verifica a emergência de um “novo liberalismo” sob a liderança intelectual de John Maynard Keynes (1883-1946). Keynes efetua uma profunda revisão dos fundamentos do liberalismo clássico, de modo a reposicionar o problema da intervenção governamental na economia (Dardot e Laval, 2016). Para Polanyi (2021), a ‘grande transformação’ dos anos 1930 e 1940 marca o fim do capitalismo liberal construído no Séc. XIX. A ressocialização das economias, as disputas neocoloniais e a emergência do nazi-fascismo, denunciavam os sintomas de uma sociedade de mercado disfuncional, processo que culminou com a deflagração da segunda guerra mundial.

Após 1945, em um nova dinâmica geopolítica polarizada entre dois blocos de poder, capitalista e socialista, verifica-se a emergência de uma nova ordem no capitalismo global caracterizada pelo modo de regulação keynesiano-fordista (Chesnais, 1996; Harvey, 2013; Anglietta, 2015). Por um lado, um conjunto de práticas de controle do trabalho, tecnologias e hábitos de consumo baseados na produção industrial em massa, e por outro, uma configuração de poder político-econômico fundamentada na intervenção ativa do Estado na economia. Sob este ‘capitalismo administrado’, há uma profunda reformulação das práticas estatísticas, na medida em que novos instrumentos são demandados para operacionalizar a gestão das economias capitalistas ocidentais sob a macroeconomia keynesiana.

A busca pelo equilíbrio macroeconômico entre a oferta e a demanda agregadas, conforme a formalização teórica de Keynes, passou a ser operacionalizada pelos Estados por meio da contabilidade nacional. O sistema de contas nacionais subsidiado por uma série de levantamentos estatísticos (pesquisas de produção industrial e agrícola, pesquisas de serviços, pesquisas de preços,

pesquisas de orçamentos familiares, etc.) expandiu fortemente o papel e a estrutura de produção das estatísticas oficiais (Desrosières, 1998)²⁹.

A síntese de elementos previamente desconectados (registros estatísticos, renda nacional, pesquisas amostrais, estatísticas inferenciais), associada a uma situação socioeconômica e política que possibilitou a intervenção governamental para regular o equilíbrio econômico, deu margem a construção de ferramentas e objetos estatísticos que foram dominantes por pelo menos três décadas, e ainda se mantém até o presente. Construtos como o Produto Interno Bruto (PIB), calculado a partir de matrizes de insumos e produtos, a taxa de inflação, calculada a partir de índices de preços ao consumidor, e a taxa de desocupação, obtida por meio de inquéritos amostrais conjunturais, entre outros, são o resultado da unificação do debate social e econômico em torno da linguagem comum do planejamento e da macroeconomia Keynesiana (Desrosières, 1998).

Para Duncan e Shelton (1978), verifica-se nos Estados Unidos a partir dos anos 1940 uma 'revolução nas estatísticas oficiais' que tem como principais elementos os inquéritos por amostragem, as contas nacionais e a introdução dos computadores eletrônicos. A era dos 'mainframes' programáveis, computadores eletrônicos de grande porte, capazes de realizar operações complexas em grande velocidade sobre um grande volume de dados, representa um importante salto tecnológico e organizacional no processo estatístico. Amplia-se imensamente a capacidade de processamento e análise dos dados, bem como a tempestividade na publicação das estatísticas oficiais, favorecendo-se, ao mesmo tempo a centralização da coordenação estatística nacional.

Esta profunda reformulação das práticas estatísticas será posteriormente seguida pelos países europeus e latino-americanos. Neste contexto, os investimentos públicos na produção de estatísticas oficiais conduzem à implantação de um novo modelo estatístico nos países industriais e periféricos. Este modelo baseia-se na centralização dos sistemas estatísticos nacionais por Instituições Nacionais de Estatística (INEs) dotadas de maior autonomia técnica e política, que passam a assumir a maior parte da produção de estatísticas oficiais, bem como a coordenação

²⁹ Do ponto de vista dos investimentos na formulação e circulação de registros estatísticos, Desrosières (1998) compara o esforço de constituição dos sistemas de contas nacionais às grandes iniciativas científico-militares com financiamento público que surgiram a partir dos anos 1930 como a pesquisa nuclear, aeronáutica, espacial e eletrônica.

dos demais produtores oficiais. Estas novas instituições nascem muito mais robustas em termos de mandato, instrumentos, quadros e recursos do que os escritórios estatísticos do Séc. XIX.

Pequenas oficinas, dispersas e pouco legítimas, com estatísticas administrativas muitas vezes datadas das décadas de 1830 a 1850, mudaram completamente a sua natureza um século depois, a partir da década de 1930 nos Estados Unidos, e das décadas de 1940 e 1950 na Europa [...] e finalmente, em outros continentes [...] Inquéritos, contabilidade nacional, estatísticas matemáticas, econometria, informática, coordenação de métodos de registo e nomenclaturas entre as diversas administrações, todas estas transformações tendem a constituir grandes instituições, em parte comparáveis, reconhecidas como produtoras de linguagens comuns legítimas para apoiar decisões económicas e políticas e dotar o debate social de um vasto leque de pontos de referência coerentes, à primeira vista inconcebíveis no século XIX e mesmo no início do século XX (Desrosières, 2008a, p.50, tradução nossa).

A disseminação das novas práticas estatísticas para os demais países capitalistas industrializados e periféricos será uma tarefa exercida sobretudo pelas novas organizações do sistema de governança internacional construído no pós guerra, em um contexto de expansão da hegemonia dos Estados Unidos sobre o bloco de países capitalistas e periféricos. As instituições multilaterais nascidas dos acordos de Bretton Woods em 1944 (Banco Mundial e Fundo Monetário Internacional) e a Organização das Nações Unidas constituída em outubro de 1945, em substituição à Liga das Nações, passam a conformar um novo sistema de governança internacional em que a comparabilidade estatística entre as nações assume uma nova dimensão, a partir da construção de um sistema estatístico internacional. A criação da Comissão Estatística das Nações Unidas (UN-StatCom) que passa a congrega os chefes estatísticos dos Estados Membros, apenas quatro meses após a fundação da ONU, em fevereiro de 1946, é um indício da relevância da organização estatística para o novo sistema de governança global (Holly, 2000)³⁰.

A introdução e combinação de novas técnicas e ferramentas de origens cognitivas diversas nos INEs, impulsionada pelos esforços de padronização

³⁰ A Comissão Estatística da ONU (StatCom) nasce com um mandato bastante amplo que envolve: (i) a promoção do desenvolvimento das estatísticas em nível nacional e a melhoria da sua comparabilidade em nível global, (ii) a coordenação do trabalho estatístico das agências especializadas das Nações Unidas e (iii) a criação de uma divisão central de estatísticas no âmbito do Secretariado das Nações Unidas com a tarefa de atender às necessidades estatísticas de todos os departamentos e agências das Nações Unidas (Holly, 2000).

internacional das práticas estatísticas, conduzem a uma diversificação dos perfis profissionais no campo estatístico. Por um lado, a disseminação das pesquisas amostrais fortalece a posição dos matemáticos estatísticos. Por outro lado, a contabilidade nacional introduz um novo tipo profissional, o economista de Estado, para quem, a necessidade de integração de uma série de dados em tabelas e matrizes, não se refere a noção de intervalo de confiança dos estatísticos mas à sua consistência contabilística. Além disso, entram em cena outros campos profissionais como a sociologia, a antropologia, a geografia e a demografia. A conciliação destes diferentes perfis no âmbito do processo produtivo das estatísticas oficiais é um processo permeado por uma série de tensionamentos e disputas (Desrosières, 2008a)³¹.

Neste contexto, a concepção holística dos estatistas, administradores e reformadores públicos que prevaleceu nos escritórios estatísticos é reposicionada com a introdução das novas competências científicas dos profissionais especializados nos INEs. Para Desrosières (2008a, p.48), o perfil do 'estatístico de Estado' combinará "em formas e proporções que variam de um país para outro", os tipos sociais do servidor público que "gere circuitos administrativos regulados por leis, regulamentos e rotinas inscritas no Estado" e do cientista/acadêmico detentor de conhecimentos especializados.

O fortalecimento das funções do Estado interventor, no contexto do capitalismo administrado, é um aspecto relevante para o modo de estruturação do campo estatístico no Séc. XX, o que vem a consolidar o seu caráter público e nacional. Por um lado, a eficácia e legitimidade da intervenção governamental do Estado, requer o fortalecimento de instituições públicas relativamente autônomas ao mercado. Por outro, o novo sistema de governança internacional, construído sob o fundamento da

³¹ De acordo com Desrosières (2008a), podemos analisar as disputas entre os diferentes tipos profissionais envolvidos no processo estatístico do ponto de vista de suas articulações com processos e redes que estão à montante (construção de classificações e equivalências, codificação das variáveis, construção dos métodos e ferramentas estatísticas, registros administrativos, etc.) e à jusante (usuários das estatísticas, mídia, pesquisadores, governos) do processo estatístico. Além disso, os conflitos envolvem diferentes interpretações e prioridades dos aspectos sintático (consistência interna), semântico (interpretação dos dados) e pragmático (uso dos dados) do processo estatístico. Embora o 'estatístico de Estado' combine de forma complexa em sua prática estas categorias, enquanto tipos ideais, os estatísticos matemáticos tendem a relacionar-se com redes envolvidas na produção dos métodos e ferramentas estatísticas e a priorizar aspectos sintáticos, os economistas ocupam-se com a consistência dos registros administrativos e usos pragmáticos da contabilidade nacional, enquanto os sociólogos priorizam os aspectos semânticos dos processos de construção das classificações e a interpretação das informações produzidas.

soberania, concorre para enfatizar o carácter ‘nacional’ do Estado. Neste contexto, há um notável incremento da produção teórica e crítica sobre o papel, domínios, objetos e práticas do campo público e estatal no sistema capitalista³². A construção dos sistemas estatísticos nacionais coordenados por Instituições Nacionais de Estatística tecnicamente robustas e politicamente legítimas, por uma lado, reafirma a vinculação das estatísticas oficiais à soberania nacional, e por outro, consagra a concepção das informações estatísticas como um bem público e um monopólio do Estado.

De acordo com Desrosières (1998, 2008a), um aspecto importante da transição do pós-guerra foi o fato de as estatísticas “terem sido cada vez mais pensadas, construídas e utilizadas para um território nacional, considerado como um espaço estatístico único, tanto para a definição e padronização das variáveis, quanto para o domínio abrangido pelas medidas” (Desrosières 2008a, p.50). No que diz respeito ao fortalecimento do carácter público das estatísticas oficiais neste contexto, Desrosières (1998, p.324) destaca como a linguagem estatística contribuiu para ‘moldar uma esfera pública’. Isto foi possível na medida em que o sistema estatístico conseguiu articular, por um lado, as informações necessárias para orientar a ação e o investimento público do Estado e, por outro, os seus usos no ‘debate público’:

durante o período 1950-1975 foi feito um esforço – pelo menos provisório – para unificar o debate económico e social em torno de uma linguagem comum. Esta foi a linguagem do planeamento e da macroeconomia keynesiana, do crescimento e das contas nacionais, da sociologia das desigualdades sociais e dos seus indicadores estatísticos. Foi também a linguagem da negociação coletiva apoiada pelo Estado entre patrões e sindicatos sobre salários e sobre um sistema de bem-estar social igualitário e redistributivo. (Desrosières, 1998, p.333)

Concluimos, que sob a racionalidade governamental que emerge do pós-guerra, fundamentada na intervenção planejada do Estado na economia sob o ideário keynesiano, verifica-se uma reestruturação do campo estatístico em novas bases. Os escritórios estatísticos comandados por administradores e estatistas generalistas são substituídos pelos modernos Institutos Nacionais de Estatística (INEs) integrados por estatísticos de Estado com competências técnicas diversificadas e especializadas. Sob a égide de uma nova governança internacional fundamentada na soberania

³² Destacam-se o desenvolvimento das teorias do desenvolvimento económico por Roy Harrod (1900-1978) e Evsey Domar (1914-1997), do nacional-desenvolvimentismo por Raúl Prebisch (1901-1986) e Celso Furtado (1920-2004), dos bens públicos por Paul Samuelson (1915-2009) e da esfera pública por Jurgens Habermas (1929-).

nacional e em um contexto de expansão das funções públicas do Estado, esta reestruturação do campo estatístico terá como fundamento o caráter público e nacional das estatísticas oficiais.

Finalmente, na esteira de uma nova crise do capitalismo global entre as décadas de 1970-1980, o espaço público ‘relativamente coerente’ sobre o qual operavam os sistemas nacionais de estatísticas entra igualmente em crise e a ‘linguagem comum’ que unificou o sistema de bem estar social, a política macroeconômica keynesiana e a produção estatística orientada para o planejamento público será “distorcida e gradualmente destruída” (Desrosières, 1998, p.334). Trataremos desta crise e processo de transição sob o modelo neoliberal no próximo capítulo, antes analisaremos na próxima seção como se deu a consolidação do regime de precisão no Brasil.

1.11 Entre a soberania e a dependência: a consolidação do regime de precisão no Brasil

A América Latina, e em particular o Brasil, experimentaram intensas transformações na primeira metade do século XX. A grande depressão e a Segunda Guerra desorganizaram o sistema de divisão internacional do trabalho e possibilitaram um processo de industrialização por substituição de importações na região. Este processo, ao mesmo tempo em que promoveu a diversificação das economias nacionais, impactou profundamente suas sociedades, expondo contradições com as estruturas sociais oriundas do período colonial. A problematização da subordinação dos interesses nacionais à determinações externas dos centros de poder geopolítico, bem como, a possibilidade de superação da condição de subalternidade por meio do desenvolvimento, constituem o cerne do pensamento crítico latino-americano que se desenvolveu paralelamente a estas transformações.

A constituição de centros estratégicos de irradiação da produção intelectual latino-americana, como a Comissão Econômica para a América Latina e o Caribe (CEPAL) e o Instituto Superior de Estudos Brasileiros (ISEB), lançou as bases para uma concepção de desenvolvimento de caráter nacional, efetuando uma translação do modelo keynesiano para a realidade regional, em oposição à ortodoxia liberal neoclássica. Neste contexto, a ‘teoria do subdesenvolvimento’ latino-americano passou a imputar os problemas econômicos da região à divisão internacional do

trabalho, introduzindo, em oposição à tese clássica das vantagens comparativas, o conceito da deterioração dos termos de troca (Prebisch, 1949). Em linhas gerais, o modelo nacional desenvolvimentista propõe a implantação de uma política de industrialização autônoma que pressupõe um Estado dotado de alto grau de autonomia e a liderança de um classe burguesa industrial nacional (Furtado, 1968). As tentativas de implantação deste modelo no Brasil, na década de 1950 e até o início da década de 1960, levaram a uma intensificação da ação Estado na defesa do mercado interno, na transferência de rendas do setor exportador para o industrial e na criação da infraestrutura para apoiar a industrialização (Cardoso e Faletto, 1970).

A prática do planejamento estatal, que já vinha se esboçando em diversas iniciativas governamentais desde a década 1930, ganhou relevância no contexto do nacional desenvolvimentismo dos anos 1950 (Ianni, 1971). Neste sentido, a demanda por informações para suprir as necessidades da planificação, especialmente no campo econômico, passaram a revelar a fragilidade da produção estatística nacional e a inadequação do modelo de produção descentralizada adotado pelo IBGE (COPLAN, 1962). De fato, a desarticulação institucional do aparato administrativo do Estado Novo e a reorganização do pacto federativo na Terceira República, implicaram na gradual desestruturação do sistema estatístico construído na era Vargas. Assim, enquanto novas instituições como a Fundação Getúlio Vargas emergem como produtoras de estatísticas econômicas, verifica-se a fragilização do IBGE, com a perda de recursos, quadros, autoridade e capacidade de coordenação dos serviços estatísticos dos órgãos federais, estaduais e municipais.

A crise no IBGE, que se verifica ao longo de todo o período da Terceira República, se manifesta em tensões e conflitos entre grupos que disputavam a sua direção. A instabilidade institucional somada à inadequação do sistema estatístico descentralizado ao novo contexto federativo resultam em uma disfuncionalidade que se revela na incipiência do programa estatístico, na quebra de séries históricas e finalmente no fracasso do censo de 1960, ocasionando uma série de críticas públicas ao órgão (Senra, 2008). A necessidade de informações estatísticas para o planejamento estatal e a demanda por levantamentos contínuos com uso de procedimentos amostrais passaram a constituir o pano de fundo destas críticas (Pereira, 2023). O diagnóstico elaborado pelo ensaísta ibgeano Alberto Passos Guimarães (1962) oferece uma boa síntese desta situação:

Nos últimos tempos, as críticas às estatísticas brasileiras se vêm tornando cada vez mais frequentes. Incidem estas críticas, principalmente, sobre a sua limitada extensão, sobre as falhas em sua continuidade e sobre a demora na divulgação dos resultados [...] grande parte das críticas, possam ou não ser consideradas justas de todos os pontos de vista, resultam da aspiração, bastante generalizada entre os consumidores, por colocar o sistema estatístico brasileiro no nível exigido pelas necessidades atuais do desenvolvimento econômico nacional. No Brasil, nada haverá de estranho que, em face das significativas modificações surgidas no panorama nacional durante os mais recentes decênios, o sistema estatístico nacional, implantado desde cerca de trinta anos, já não esteja em condições de atender plena e satisfatoriamente à crescente e intensa procura de informações indispensáveis à fase atual de desenvolvimento do país [...] Desde a década de cinquenta, nosso país deixou de realizar em extensão nacional, as estatísticas de comércio por vias internas, desde 1958 interrompeu a série anual de estatísticas industriais, desde 1960 interrompeu a série mensal de estatísticas da indústria e do comércio atacadista [...] o Brasil não fornece dados sobre numerosos aspectos econômicos e sociais cujos levantamentos são realizados em outros países, mesmo nos de grau de desenvolvimento inferior ao nosso. Não dispomos de boas estatísticas de preços atacadistas, de salários, de custo de vida, não efetuamos estatísticas de desemprego [...] nem dispomos de cadastros de empresas econômicas, atualizados nos intervalos de operações censitárias (Guimarães, 1962, p.1-2).

O diagnóstico de Guimarães (1962) foi um dos documentos de base que subsidiaram a proposta de reestruturação do sistema estatístico nacional. Esta proposta foi elaborada em 1962 durante o governo de João Goulart, pela Comissão Nacional de Planejamento (COPLAN). O grupo de trabalho criado com esta finalidade por Tancredo Neves, presidente do conselho de ministros, tinha por objetivo ampliar e atualizar os levantamentos estatísticos “em função das necessidades do planejamento econômico e social do país” (COPLAN, 1962, p.3). O relatório final propunha a reformulação das estatísticas nacionais com a centralização das atividades de planejamento, execução e divulgação à cargo do IBGE, a quem também caberia a coordenação do sistema estatístico nacional. A proposta previa a ampliação da capacidade técnica, financeira e institucional do IBGE para a execução de um Plano Nacional de Estatística supervisionado pelo Conselho Nacional de Estatística, que deveria abranger um conjunto de estatísticas censitárias, contínuas e secundárias.

A proposta de reformulação do IBGE pela COPLAN (1962) visava reposicionar o órgão como instrumento central do planejamento governamental. Para Grabois (2021, p.238), a ideia de planejamento no Governo João Goulart transcendia a concepção do ‘plano escrito’ por especialistas e assumia o caráter de “uma ação permanente de governo em uma lógica incremental de construção nacional”. Neste sentido, a nova concepção do IBGE e do sistema estatístico nacional passava a

compreender muito mais do que um papel meramente técnico, mas de relevância política e estratégica no debate nacional, o que, de acordo com Grabois (2021), é evidenciado na publicação do livro “Ação Governamental de João Goulart” pelo IBGE em janeiro de 1964. Para o autor, a nova concepção de planejamento que emergia neste contexto, caracterizava sobretudo um governo que

almejava fortemente agir de acordo com uma racionalidade que expressava e significava que ele – o Governo – agia considerando o conjunto da população e de seus problemas, que compreendia a economia em uma dimensão fortemente social e que essa racionalidade deveria ser, portanto, tecida e estruturada como bem-estar social”. Em outros termos, o Governo deveria agir de modo calculado e coordenado e não simplesmente deixar as coisas por conta do mercado ou da ação espontânea e dos interesses das ‘forças econômicas’ (Grabois, 2021, p.238).

A reestruturação formal do IBGE e do sistema estatístico nacional, entretanto, viria a ocorrer apenas sob a ditadura militar que, com o golpe civil-militar de 1964, encerrou a breve experiência democrática da Terceira República. A centralização estatística que se segue, se dará no âmbito de supressão da autonomia dos entes federativos, concentração de poderes no âmbito federal e reorganização política e administrativa do Estado brasileiro, com a outorga da constituição de 1967. Constituído como uma fundação pública de direito privado subordinado ao Ministério do Planejamento Econômico, o IBGE incorporou em sua estrutura a rede de serviços municipais de estatística e assumiu a função de órgão produtor e centralizador das estatísticas oficiais destinadas aos usuários do governo, e principalmente ao Instituto de Pesquisa Econômica Aplicada (IPEA), que assumia o papel de centro de inteligência do planejamento nacional. A derrota do nacional desenvolvimentismo democrático pelo autoritarismo burocrático dará lugar a adoção de um modelo liberal-desenvolvimentista, comandado pelo economista Roberto Campos e por uma equipe de jovens economistas formados em universidades norte americanas (Silva, 2021). A implantação deste modelo aprofunda a inserção dependente do país no capitalismo mundial, sob a tutela militar e a hegemonia dos Estados Unidos.

Para Cardoso e Faletto (1970), a sustentação política do modelo nacional desenvolvimentista dependia da criação de uma base econômica de sustentação de novos grupos hegemônicos e da inserção econômico-social dos grupos populares, de modo a alterar o sistema de dominação política. No Brasil, o equilíbrio da ‘aliança desenvolvimentista’ esteve condicionado a uma conjuntura propícia de manutenção

dos preços de exportação que permitissem remunerar os setores exportadores e transferir rendas para financiar os setores urbano-industriais. O término dessa conjuntura revelou a inconsistência das alternativas de desenvolvimento capitalista nacional e autônomo, restando como viável a opção de desenvolvimento dependente com a abertura do mercado interno para o capital estrangeiro nos anos 1950. Na fase de substituição de importações, a penetração de capitais estrangeiros não é percebida como problema essencial para o desenvolvimento. Dado o dinamismo dessa fase, existiu uma coincidência transitória que permitiu conciliar os objetivos protecionistas, a pressão das massas e os investimentos estrangeiros. Esse processo entrou em crise quando, com o esgotamento do modelo de substituição de importações, se evidenciam problemas mais complexos como os limites dos setores tecnológicos e da indústria de bens de capital e a cisão das classes industriais e proletárias.

A intensificação da industrialização através dos capitais, da técnica e da organização transferidos do exterior inaugura um novo eixo de ordenamento da economia, implicando em restrições à autonomia dos centros de decisão nacional. Estabelece-se assim um novo tipo de dependência, em que as empresas monopolistas e o setor financeiro internacional procuram exercer influência sobre as decisões nacionais. Nesta nova etapa, o desenvolvimento dependente realiza-se intensificando a exclusão social por meio de um esquema político de sustentação que implica em conter as demandas reivindicatórias das massas. Pela via do golpe civil-militar, os grupos dominantes efetuam a reorganização do regime político para permitir a centralização autoritária (Cardoso e Faletto, 1970). Assim, por um lado, a burocracia estatal, se reorganiza para a implantação do regime autoritário, o que envolve a 'depuração' dos órgãos públicos, com a perseguição aos quadros identificados com a esquerda e com o projeto nacional desenvolvimentista, enquanto, por outro, a dinâmica das relações internacionais no contexto da guerra fria e a ideologia de segurança nacional legitimam as mudanças econômicas e políticas.

Do ponto de vista geopolítico, os golpes civil-militares que se sucedem nas décadas de 1960 e 1970 na América Latina marcam a expansão e consolidação da hegemonia dos Estados Unidos na região, cujo apoio político e militar foi fundamental para a implantação dos regimes autoritários. Respalhada pelo contexto geopolítico da guerra fria, a deposição de governos alinhados com o projeto nacional desenvolvimentista atendia aos interesses de instituições financeiras e empresas

estadunidenses e multinacionais, que intensificaram sua presença na região no processo de industrialização (Bandeira, 1978).

Pereira (2023, p.8) destaca a importância dos mecanismos de *soft power* utilizados pelos Estados Unidos para a construção de sua hegemonia no Brasil, a partir da década de 1940. Este processo envolveu o estabelecimento de acordos bilaterais de cooperação técnica e a mobilização de uma rede complexa formada por fundações filantrópicas, universidades e serviços estatais estadunidenses, bem como organismos multilaterais. De acordo com Pereira (2023, p.8), esta estratégia visava, por um lado, a formação nas instituições americanas de indivíduos que ocupassem cargos chave no Estado e em empresas privadas brasileiras, e, por outro lado, “trazer ao Brasil um grande número de especialistas com o objetivo de orientar a condução de políticas públicas, capacitar um público mais amplo e intervir no debate intelectual nacional”. É neste contexto que sob a influência direta dos Estados Unidos, o método da amostragem probabilística é introduzido nas práticas estatísticas do IBGE.

No caso dos quadros do IBGE, é nomeadamente através da intervenção de especialistas em áreas muito específicas e da realização de formações de curta ou média duração que, numa primeira fase, se concretiza esta aproximação com as EUA. Foi nessas condições que foi realizada a primeira pesquisa amostral do IBGE, em 1957-1958 e os trabalhos preparatórios do censo de 1960. Este censo refletiu bem este processo: foi o primeiro a ser realizado por amostragem representativa a nível nacional, entrevistando 25% dos domicílios em cada município, concebido com a ajuda de especialistas do Escritório do Census dos Estados Unidos (Pereira, 2023, p.9, tradução nossa).

O estatístico do US Census, Thomas B. Jabine, que atuou como assessor do Núcleo de Planejamento Censitário do IBGE entre 1957 e 1960, produziu neste período uma série de propostas para a introdução da amostragem probabilística nas pesquisas do órgão (Senra, 2008). Apesar do plano amostral proposto por Jabine ter sido implantado no Censo de 1960, os relatórios do estatístico apontam para resistências internas no IBGE à adoção dos procedimentos amostrais em outras pesquisas. De acordo com Jabine (1969, apud Senra 2008, p.625), no caso do censo agrícola, “o grupo não foi favorável à ideia de coletar informações para itens selecionados em base amostral”, fundamentalmente porque havia a expectativa de divulgação de todos resultados na escala do município e, neste sentido, os procedimentos amostrais não implicariam em redução de custos ou tempo das operações. Tampouco no caso dos censos econômicos suas propostas foram

implementadas (Senra, 2008). De acordo com Senra (2008, p.625), para o chefe do grupo técnico de coordenação do IBGE “os textos [de Jabine] embora bons no trato das questões técnicas, pecavam por desatenção às exigências financeiras e administrativas”. De acordo com Pereira (2023, p.11), “a realização dos levantamentos amostrais entrava em conflito direto com o sistema de coordenação de serviços estatísticos relativamente autônomos”, na medida em que o IBGE dependia da adesão dos agentes estaduais e municipais ao procedimento amostral, o que, para além do censo demográfico não foi alcançado.

Foi apenas no contexto da reestruturação do IBGE e centralização da produção estatística, efetuada pela ditadura militar, que se verifica a efetiva implementação de uma pesquisa amostral em escala nacional no Brasil, o que se deu sob intervenção direta dos Estados Unidos. De acordo com Pereira (2023), a produção de informações para o monitoramento da economia no curto prazo pelos órgãos governamentais brasileiros e financiadores internacionais, por meio de pesquisas amostrais, foi um dos eixos da reestruturação das estatísticas nacionais após 1964. Assim, no contexto do projeto Aliança para o Progresso³³ e por intermédio da Comissão Estatística da ONU, a Agência dos Estados Unidos para o Desenvolvimento Internacional (USAID) e o US Census desenvolveram o projeto ‘Atlântida’, que visava a implantação de inquéritos amostrais nos países latino-americanos. A ação incluía a formação intensiva dos estatísticos nacionais, a introdução de uma pesquisa padrão elaborada pelo US Census e o apoio para a adaptação do plano amostral e análise dos resultados nos INEs latino-americanos por especialistas da instituição estadunidense (Pereira, 2023).

A participação do IBGE no projeto Atlântida, a partir de 1965, resultou na criação da Pesquisa Nacional por Amostra de Domicílios (PNAD) voltada à produção de estatísticas sociodemográficas (educação, emprego, rendimentos e características habitacionais) e informações adicionais em inquéritos suplementares com temas específicos. Produzindo resultados trimestrais entre 1967 e 1969 e, a partir de 1971, reformulada para uma periodicidade anual, a recepção dos dados da PNAD pelos órgãos públicos nacionais e instituições internacionais foi muito positiva. Assim, em

³³ O Projeto Aliança para o Progresso foi um amplo programa de financiamento estadunidense para países latino-americanos lançado por John F. Kennedy em 1961. Moldado nos objetivos da Guerra Fria, entre 1961 e 1969, o programa de financiamento e cooperação técnica visava se contrapor à influência soviética na região tendo contribuído para desestabilizar sistemas políticos democráticos na região e facilitar o advento e sustentação dos regimes autoritários (Loureiro,2023).

um curto espaço de tempo, a pesquisa se tornou a principal fonte de informações socioeconômicas conjunturais do país. De acordo com Pereira (2023, p.13), no que pese a sua utilização pelos órgãos de planejamento econômico nacionais, “acima de tudo a PNAD foi concebida como fonte de informação para o Aliança para o Progresso, que era então o principal financiador de projetos econômicos no Brasil”.

Pereira (2023) destaca como a adaptação para o contexto brasileiro do questionário básico do US Census — construído para o país genérico ‘Atlântida’, que teria as características médias de um país latino-americano — compreendeu uma tradução literária deste instrumento. Apesar da coleta de informações que atendiam aos objetivos nacionais terem sido realizadas por meio de questionários adicionais e módulos especiais, a observância a conceitos e classificações para fins de comparabilidade internacional se deu em detrimento de sua relevância para a realidade nacional. Assim, “apesar da entrada muito precoce dos indivíduos no mercado de trabalho, inicialmente a PNAD interrogaria apenas indivíduos com 15 ou mais anos. Da mesma forma, num país onde, em 1970, 43,7% da população era rural, apenas questionaríamos as formas de emprego comerciais e bastante urbanas” (Merrick, 1972, apud Pereira 2023, p.13). Para Pereira (2023), apesar de sua importância histórica como instrumento que passou a representar o conjunto da população brasileira, esta representação estatística foi obtida por meio ‘de uma grade de leitura estrangeira’:

Se por um lado a PNAD foi o primeiro inquérito a ter como universo (no sentido estatístico) toda a população brasileira e, assim, poder expressar um todo que não é a simples soma das partes que o compõem e, por isso, representou o resultado do sistema estatístico nacional; por outro lado, este resultado foi alcançado através de uma grade de leitura estrangeira. Ou seja, apesar das suas qualidades e da sua importância, a PNAD não produziu a informação mais relevante para compreender as diferentes facetas da sociedade brasileira, mas a informação que interessava principalmente aos Estados Unidos, aos organismos internacionais e àqueles que queriam aplicar técnicas econométricas neoclássicas (Pereira, 2023, p.13-14, tradução nossa).

Apesar do projeto da PNAD ter contribuído para reposicionar o IBGE como órgão produtor no processo de centralização do sistema estatístico nacional, conformando-o às necessidades da planificação econômica do Estado, a efetiva reestruturação do órgão nestes novos moldes se consolidou a partir de 1970 sob a gestão de Isaac Kerstenetzky. Economista pragmático, Kerstenetzky era crítico das visões ortodoxas neoclássicas, tendo sido influenciado tanto pelas teorias keynesiana e pós

keynesiana, como pela econometria de Jan Tinbergen (RBE, 1991). Na área pública, destacou-se nos anos 1950 e 1960 na implantação do projeto de contas nacionais na Fundação Getúlio Vargas. Em sua longa gestão de 11 anos à frente do IBGE, foi responsável pelo reposicionamento do órgão como principal produtor nacional de estatísticas oficiais para o planejamento econômico, destacando-se sua contribuição fundamental nos Planos Nacionais de Desenvolvimento (PND I e II), ao mesmo tempo em que expandiu as ações do órgão na área de pesquisas sociais. O compromisso de sua gestão com as estatísticas sociais contribuiu para expor a extrema desigualdade social, regional e econômica do ‘milagre econômico brasileiro’ entre 1969 e 1975, período caracterizado por altas taxas de crescimento do Produto Interno Bruto.

Durante os anos 1970, verifica-se a renovação dos quadros técnicos do IBGE com a contratação de profissionais qualificados nas áreas de estatística, geografia, economia, demografia, antropologia e sociologia. De acordo com Camargo (2009, p.154), este período corresponde a uma transição em que “o intelectual universal, livre pensador e grande escritor [...] tão bem representado na geração de intelectuais da era Vargas, cede lugar ao intelectual específico [...] [que] desempenharia um papel de cientista perito”. Do ponto de vista das disposições que passariam a ser incorporadas pelos agentes do IBGE, essa fase consolida a transição dos administradores e ‘estatistas’ generalistas que caracteriza o regime estatístico de enumeração para a conformação do novo perfil dos ‘estatísticos de Estado’ característico do regime de precisão. De acordo com Camargo (2009, p.160), esta transição deu margem a conflitos geracionais que foram habilmente equacionados por Kerstenetzky. Isso se deu por meio da mobilização da memória institucional no processo de refundação do órgão e valorização da integração da estatística e da geografia em um só corpo técnico, um diferencial do IBGE que foi instrumentalizado para fortalecer “a identidade coletiva da comunidade ibgeana como fator de mobilização e engajamento no projeto de reestruturação do instituto”³⁴.

³⁴ De acordo com Camargo (2009) a estratégia de Kerstenetzky envolveu por exemplo a transferência da data comemorativa de fundação do IBGE de 1936, ano de criação do INE associado “ao heroísmo que envolveu o trio Teixeira de Freitas, Macedo Soares e Juarez Távora” para 1938, ano que marca a integração do Conselho Brasileiro de Geografia ao Conselho Nacional de Estatística e a criação do IBGE. Neste mesmo sentido figura a adoção do 29 de maio como dia do ‘ibgeano’ em que simultaneamente se comemoram no Brasil o dia do estatístico e do geógrafo.

O programa estatístico implementado durante a gestão de Kerstenetzky envolveu a implantação de projetos estratégicos que consolidaram o novo papel institucional do IBGE. Destacam-se o desenvolvimento da matriz de insumos e produtos que possibilitou a implantação de um novo sistema de contas nacionais a partir de 1979 (Neto; Forte, 2016), a consolidação das pesquisas econômicas, sociais e agropecuárias, a implantação de um sistema de indicadores econômicos e sociais, além de consideráveis avanços nas áreas de cartografia e geodesia. O fortalecimento da produção e divulgação estatística, somado à autoridade técnica e burocrática de seus novos quadros possibilitou a superação do período de crises ancorando-se, ao mesmo tempo, na legitimidade herdada do passado institucional.

De acordo com Nascimento (2017), a realização do Estudo Nacional de Despesa Familiar (ENDEF), precursor das pesquisas de orçamento familiar, em 1974-75, marcou uma etapa de consolidação no IBGE do emprego da amostragem. O ENDEF também se notabilizou por outros motivos. Em primeiro lugar, apesar de atender aos objetivos técnicos de composição de um índice de preços, a pesquisa foi a primeira a levantar dados representativos na escala nacional sobre o tema da nutrição. Em segundo lugar, o estudo adotou uma metodologia inovadora que consistiu em conciliar a aplicação do questionário fechado com o levantamento de dados qualitativos em uma folha aberta, destinada a colher as impressões dos agentes de pesquisa sobre a situação dos domicílios e famílias selecionadas, que eram visitados pelo agente ao longo de uma semana (Camargo, 2009a). Apesar de parcialmente censurada pela ditadura militar, que determinou o encerramento da pesquisa qualitativa, a divulgação dos dados do ENDEF, a partir de 1977, deu margem a importantes debates nacionais sobre a pobreza, a desigualdade e principalmente sobre a prevalência da fome no país (Suplicy, 1977, 1978; IBGE, 1979; Simonsen, 1978), no contexto de gradual abertura política e 'liberalização' da ditadura.

A partir do final dos anos 1970, o IBGE ampliou bastante o uso da amostragem em suas pesquisas. Inicialmente restrito às pesquisas domiciliares, neste período o procedimento se estende às pesquisas econômicas — Índices de preço ao consumidor em 1979, Pesquisa Mensal de Emprego (PME) em 1980, Pesquisa Industrial Anual em 1981. Além disso, no Censo de 1980, a amostragem foi empregada de forma mais ampla que na operação de 1970, sendo incluída no controle de qualidade durante as etapas de coleta e processamento (Nascimento, 2017). De

acordo com Nascimento (2017), verifica-se neste período uma ‘maior autonomia técnica’ e o desenvolvimento de ‘uma cultura própria’ de amostragem no IBGE:

Essa fase de consolidação da amostragem como instrumento básico para a realização de pesquisas foi caracterizada pela maior independência e autonomia técnica. O planejamento amostral passou a ser realizado por pessoal do próprio IBGE, incorporando inovações técnicas e refletindo o desenvolvimento de uma cultura própria. Data ainda dessa fase o reconhecimento da necessidade de documentar e disseminar entre os usuários as metodologias de pesquisa, em particular detalhes dos procedimentos amostrais adotados, tendo sido publicados relatórios descrevendo as metodologias adotadas nas principais pesquisas do IBGE (Nascimento, 2017, p.48).

Finalmente, verificamos que no final dos anos 1970 está consolidado o regime estatístico de precisão no Brasil. Conforme vimos, foi em um contexto autoritário e subordinado à hegemonia dos Estados Unidos que as práticas amostrais foram introduzidas no campo estatístico brasileiro, configurando uma espécie de ‘dependência estatística’ do Brasil nos primeiros anos da ditadura. A concepção política do planejamento como elemento de construção nacional inclusiva da classe trabalhadora, em que caberia um novo papel do órgão estatístico foi radicalmente abortada. Em que pese o seu caráter tecnocrático e autoritário, a centralização e vinculação técnica do sistema estatístico nacional ao sistema de planificação do Estado, a profissionalização e diversificação dos quadros técnicos do IBGE, e a racionalização e ampliação do programa estatístico, lograram consolidar a legitimidade e autonomia do órgão. Adicionalmente, a divulgação de estatísticas econômicas e sociais, mesmo que sob o controle da ditadura, contribuíram para subsidiar o debate público e as críticas ao regime, que ganharam espaço e maior dimensão a partir do processo de abertura, inserindo assim o campo estatístico brasileiro nos contramovimentos que conduzem ao processo de redemocratização.

Conclusão: A constituição da estatística oficial como um bem público e a estruturação do campo estatístico como um espaço de lutas entre a dinâmica internacional e a soberania nacional

Neste capítulo discutimos o processo histórico e sociológico de estruturação do campo estatístico. Conduzimos esta investigação para responder algumas questões fundamentais: o que são as estatísticas oficiais, o que é o campo estatístico e qual o

seu papel na organização social do modo de produção capitalista e na ação do Estado.

Nossa análise norteou-se pela ideia de que as crises que fundam e transformam o regime de acumulação do capital implicam em crises de governamentalidade que resultam em uma reconfiguração dos regimes estatísticos. Este processo ocorre na medida em que a estatística, enquanto ferramenta de prova e de governo, é mobilizada para assegurar a legitimidade do Estado e a reprodução do capital nos contextos de crise.

Vimos como a estatística emergiu como uma prática vinculada à razão de Estado operando em favor da concentração de capitais que conduziram à centralização dos Estados Territoriais europeus. Neste contexto, as estatísticas voltavam-se principalmente à catalogação dos bens e forças do Estado, constituindo um monopólio privado do soberano. Neste 'regime estatístico de contabilização' orientado pelos princípios mercantilistas, a estatística cumpria três funções principais: i) planificação da intervenção direta do Estado na economia; ii) medição e controle das forças internas do Estado vinculada ao aparato policial e ao princípio da soberania; iii) medição relativa de forças em comparação com outros Estados.

Paralelamente à razão de Estado, analisamos a emergência de uma razão probabilística, caracterizada pela aliança entre a matemática e uma nova racionalidade fundamentada na ideia de conhecimento provável. Desta aliança resultaram novas formas de cálculo para lidar com o acaso e a incerteza. Vimos como as formalizações clássicas combinavam interpretações objetivas (frequências) e subjetivas (crenças) do cálculo de probabilidades e como suas construções cognitivas (urna de Bernoulli e curva normal) resultaram em aplicações e analogias determinantes para o desenvolvimento posterior da estatística de Estado. Este processo foi tão mais acentuado na medida em que as estatísticas superaram a condição de segredos de Estado para se constituírem em referentes públicos.

Analisamos como a formação do Estado Nação, na esteira da emancipação da estrutura social do capitalismo, foi acompanhada pela construção de espaços nacionais de equivalências que implicaram em uma transmutação da estatística em um bem público e nacional. Estes espaços foram o resultado de processos de unificação e universalização constituídos por desapossamentos de sistemas locais de referência e classificação. Concluimos que o 'espaço de medida comum' constitui-se ao mesmo tempo como o espaço de lutas pelo controle e monopolização legítimos da

medida comum, ou seja, dos capitais informacionais e simbólicos assim constituídos. Este espaço de lutas, por sua vez, corresponde ao que denominamos de campo estatístico.

Ao tratarmos do campo estatístico como um espaço de lutas, buscamos compreender o seu modo de estruturação a partir do papel desempenhado pela estatística em um conjunto ampliado de disputas, que interceptam o Estado, a ciência e o capital, bem como suas contradições entre as dinâmicas nacionais e globais. Assim, no contexto da governamentalidade liberal, identificamos que as estatísticas oficiais, passam a assumir um novo conjunto de papéis políticos e econômicos.

No campo econômico, as estatísticas oficiais passam a operar como instrumentos de coordenação estratégica dos agentes e de verificação da prática governamental pelo mercado. Concluímos que este novo papel da estatística é fundamental para sustentar a legitimidade do Estado em sua separação relativa do mercado. Coordenada pelo Estado a estatística oficial cumpre o papel de 'meio geral de cognição' do mercado, estando assim diretamente implicada no processo de extração da mais-valia.

No campo político, a estatística passa a constituir o cerne de uma nova forma de intervenção que visa o controle das características biológicas e sociais da população por meio do cálculo. Vimos como a estatística social passou a articular ferramentas cognitivas e político-administrativas coerentes com estratégias biopolíticas e táticas de normalização. Do ponto de vista cognitivo, a estatística social promove uma primeira aproximação da estatística com a matemática probabilística, aplicando a distribuição normal na análise de dados populacionais, consolidando assim a interpretação frequentista do cálculo de probabilidades. Do ponto de vista político-administrativo, a intervenção do Estado torna-se um imperativo para garantir a estrutura necessária à enumeração exaustiva da coleta de dados por meio de operações censitárias e registros administrativos na escala do território nacional.

No campo internacional, identificamos como o desenvolvimento do internacionalismo estatístico, coerente com a nova racionalidade política do liberalismo e orientado pelo sonho de governança técnica dos estatísticos, resultou em disputas que derivavam de interesses conflitantes dos Estados Nacionais. Estas disputas enfatizam uma tensão intrínseca no campo estatístico, a contradição entre o papel da estatística como instrumento de soberania nacional e de mediação da competição internacional.

Na esteira das primeiras depressões econômicas do capitalismo, a crise de governamentalidade que se segue conduz a uma revisão dos dogmas liberais e em contramovimentos protetivos que resultam na proteção do trabalho e na intervenção do Estado na economia. Analisamos como neste contexto a estatística oficial passou a constituir um instrumento de articulação do duplo movimento do Estado. Por um lado provendo os dados necessários ao funcionamento do mercado, e por outro, produzindo as informações demandadas pelo sistema de proteção social. Além disso, analisamos como ferramentas estatísticas inicialmente desenvolvidas no arcabouço político da eugenia puderam ser aplicadas e reorientadas para os objetivos da proteção social. Verifica-se, neste contexto, o início da transição de um 'regime estatístico de enumeração', amparado na exaustividade dos censos, para um novo 'regime estatístico de precisão', com a incorporação da estatística inferencial em surveys amostrais.

Sob o capitalismo administrado que emerge do pós-guerra, apontamos para uma nova ênfase no caráter público e nacional das estatísticas oficiais. Isto decorre, por um lado, da intervenção planejada do Estado na economia, que implica no fortalecimento das instituições públicas e no papel da estatística como 'meio geral de cognição', e, por outro lado, da construção de um sistema de governança internacional sob o fundamento da soberania nacional. A consolidação do regime estatístico de precisão, fundamentado na estatística inferencial, nas contas nacionais e nos computadores eletrônicos, conduz a uma mudança estrutural no campo estatístico. Este processo é caracterizado pela sucessão dos administradores e estatistas generalistas dos escritórios estatísticos, por 'estatísticos de Estado' especializados no contexto de fortalecimento das Instituições Nacionais de Estatística.

Finalmente, vimos como no caso do Brasil verifica-se um desenvolvimento desigual e combinado das estatísticas oficiais que deriva da função periférica da economia brasileira no mercado global e do desenvolvimento contraditório do Estado, que historicamente incorporou inovações institucionais e tecnológicas dos países centrais a um quadro interno de relações políticas e sociais arcaicas. Neste sentido, o desenvolvimento do campo estatístico no Brasil é o resultado de uma relação dialética entre a sua vinculação dependente ao internacionalismo estatístico e o desenvolvimento autônomo de uma organicidade própria que resultou de contramovimentos de caráter nacional e de rupturas do sistema político. Apesar de subordinado às macro determinações que derivam das crises do regime de

acumulação capitalista e das crises de governamentalidade, o desenvolvimento da capacidade estatística no Brasil dependeu em larga medida do grau de dependência e/ou autonomia da vinculação do sistema político e econômico aos centros de decisão internacional. Este desenvolvimento histórico particular, implicou tanto em uma dilatação cronológica na transição e consolidação dos regimes estatísticos, como no desenvolvimento de características próprias que resultam do jogo entre a subordinação internacional e a defesa da soberania nacional.

Embora a adoção de novas tecnologias estatísticas no Brasil, como o censo demográfico em 1872 e os inquéritos amostrais nos anos 1960, tenham dependido fundamentalmente da importação de inovações desenvolvidas no Norte-Global, é apenas quando os movimentos de caráter nacional são capazes de se apropriar destas tecnologias e lhes conferir sentido e consequência política, que verificamos um efetivo processo de estruturação do campo estatístico e de transição de regime estatístico no Brasil.

No primeiro caso, embora o censo demográfico tenha sido adotado no Brasil Império, apenas podemos conceber a transição para um regime de enumeração no país, quando os movimentos sanitaristas dos anos 1910/20 efetuam uma transformação das estatísticas e mapas em instrumentos políticos de um projeto nacional (Camargo, 2016). Esta instrumentação tem seu apogeu com a fundação do IBGE e articulação de um sistema estatístico e cartográfico nacional, enquanto elementos estratégicos do projeto desenvolvimentista da era Vargas (Penha, 1993, Grabois, 2021).

No segundo caso, embora as técnicas de amostragem tenham sido inicialmente introduzidas no IBGE pelo US Census nos anos 1960 (Pereira, 2023), é apenas a partir da apropriação da estatística inferencial e econometria às necessidades do planejamento nacional nos anos 1970, que o Brasil efetivamente adentra em um regime de precisão, com a centralização do IBGE e a profissionalização dos estatísticos de Estado.

Finalmente, tudo se passa como se, embora sob a influência do internacionalismo estatístico, o nacionalismo estatístico se impusesse como a principal dinâmica gerativa tanto da estruturação do campo estatístico, como da transição de regimes estatísticos. Como veremos, este resultado da investigação histórica é uma chave de leitura relevante para a compreensão do contexto contemporâneo de dataficação.

2. A TRANSIÇÃO PARA UM REGIME DE DATAFICAÇÃO

As forças produtivas e as relações sociais – ambos aspectos diferentes do desenvolvimento do indivíduo social – aparecem somente como meios para o capital, e para ele são exclusivamente meios para poder produzir a partir de seu fundamento acanhado. De fato, porém, elas constituem as condições materiais para fazê-lo voar pelos ares [...] A natureza não constrói máquinas nem locomotivas, ferrovias, telégrafos elétricos, máquinas de fiar automáticas, etc. Elas são produtos da indústria humana; material natural transformado em órgãos da vontade humana sobre a natureza ou de sua atividade na natureza. Elas são órgãos do cérebro humano criados pela mão humana; força do saber objetivada. O desenvolvimento do capital fixo indica até que ponto o saber social geral, conhecimento, deveio força produtiva imediata e, em consequência, até que ponto as próprias condições do processo vital da sociedade ficaram sob o controle do intelecto geral e foram reorganizadas em conformidade com ele. Até que ponto as forças produtivas da sociedade são produzidas, não só na forma do saber, mas como órgãos imediatos da práxis social; do processo real da vida. (Marx, 2011, p. 943-944)

O objetivo deste capítulo é compreender um conjunto de transformações que conduzem a reconfigurações no regime estatístico durante o período de hegemonia neoliberal e em seu contexto contemporâneo de crise.

Discutimos inicialmente como a crise do regime de acumulação keynesiano-fordista implicou em uma reconfiguração dos sistemas estatísticos nacionais, a qual conduziu a um reposicionamento dos INEs sob a governamentalidade neoliberal. Neste contexto, a perda de relevância do planejamento e dos modelos macroeconômicos e macrossociais no âmbito da ação pública é acompanhada pela disseminação de novas modalidades de quantificação, vinculadas aos processos de financeirização. A política de indicadores, a cultura de auditoria e as avaliações microeconômicas caracterizam a emergência de um novo regime estatístico de comensuração orientado pela lógica da concorrência social.

Sob o modelo neoliberal, as reformas gerenciais do Estado no contexto de disseminação da New Public Management (NPM) ameaçam os sistemas estatísticos nacionais e promovem uma maior permeabilidade dos INEs à linguagem e lógica de mercado. A manutenção da autonomia e caráter público e nacional das estatísticas oficiais esteve vinculada à adaptação de sua relevância ao novo contexto, bem como a contramovimentos protetivos do campo estatístico, que buscou preservar os seus princípios fundamentais contra os ataques neoliberais.

O desenvolvimento das tecnologias de informação e comunicação (TICs) ampliam a capacidade de produção e disseminação de informações dos sistemas estatísticos nacionais, ao mesmo tempo em que conduzem a uma maior

descentralização da produção estatística e à emergência de novos agentes quantificadores privados, subnacionais e internacionais. A generalização dos processos de comensuração, impulsionados pelas TICs, constituem-se como uma importante etapa de transição, que estabelece as condições materiais e cognitivas para a emergência de um novo regime estatístico-algorítmico de dataficação no Séc. XXI.

Apontamos como o processo de transição para um regime de dataficação vincula-se aos desdobramentos da crise financeira de 2007/2008 e à emergência de uma nova economia orientada aos dados. Neste contexto, Big Techs e um conjunto variado de agentes associados, que intitulamos de campo de práticas algorítmicas, atuam na extração e cercamento de novas fontes de dados com vistas à transformá-las em mercadorias e ativos em novos mercados, de acordo com o modelo de plataformação.

O big data e as novas modalidades de informação a ele vinculadas impõe considerável pressão sobre os sistemas estatísticos nacionais dos Estados. O acesso às novas fontes de dados se impõe ao campo estatístico como um elemento essencial para a manutenção de sua relevância e as estratégias que emergem neste contexto são traduzidas em um discurso de ‘modernização’ das estatísticas oficiais. As estratégias de ‘modernização’ passam a implicar no estabelecimento de intersecções entre os campos de práticas estatística e algorítmica, as quais implicam em tensões, controvérsias e disputas que caracterizam o atual contexto de transição para um novo regime de dataficação.

Do ponto de vista da ontologia e economia política dos dados, apontamos que as disputas em curso envolvem, por um lado, tentativas de criação de novos mercados de dados para estatísticas oficiais que partem da premissa de que os dados guardados por privados são mercadorias e, por outro, contramovimentos protetivos que visam pela via da regulação assegurar o acesso às novas fontes de dados como bens públicos.

Do ponto de vista da epistemologia estatística, destacamos como a introdução das novas fontes de dados nos INEs implica em um obstáculo à interpretação frequentista da teoria da amostragem e à abordagem dedutiva de desenho de inquéritos. Isto conduz a adoção de novos métodos baseados na estatística bayesiana e na abordagem indutiva da modelagem. Finalmente, apontamos como este conjunto de mudanças produz efeitos na estruturação do campo estatístico e no habitus

incorporado pelos estatísticos de Estado, especialmente a partir da introdução da ciência de dados nos INEs, acompanhada pelo novo perfil profissional dos cientistas de dados, vinculado às noções de ‘inovação’ e ‘empreendedorismo’ típicas da racionalidade neoliberal.

De forma similar à adotada no capítulo anterior, buscamos contrapor o processo global de desenvolvimento do campo estatístico no norte-global, especialmente o europeu, com o caso brasileiro. Neste sentido, discutimos como se processou no país a crise de acumulação e governamentalidade que conduziu a adoção do modelo neoliberal e quais efeitos produziu no campo estatístico e em particular no IBGE. Finalmente, destacamos os processos de democratização e organização dos trabalhadores do IBGE como fatores relevantes na atenuação dos efeitos nocivos da governamentalidade neoliberal sobre o sistema estatístico nacional.

2.1 A crise da governamentalidade keynesiana e a emergência do neoliberalismo

A partir de meados dos anos 1960, o modelo keynesiano-fordista dá sinais de sua incapacidade de “conter as contradições inerentes ao capitalismo” (Harvey, 2013, p.135). A rigidez do modelo — baseado no pleno emprego e em investimentos de larga escala e de longo prazo que presumiam crescimento estável — e a resistência da classe trabalhadora às tentativas de flexibilização, somadas ao aumento dos gastos públicos com programas de assistência para manter a legitimidade do Estado conduzem a uma política monetária inflacionária como meio de sustentação do crescimento econômico (Chesnais, 1996, Harvey, 2008, 2013). Em um cenário de estagnação econômica, as dificuldades de valorização do capital no processo produtivo ampliam a massa de capitais que busca valorizar-se a partir da esfera financeira (Chesnais, 1996; Paraná, 2024, Mollo et. al., 2022). As medidas para frear a inflação, somadas ao choque do petróleo em 1973/1978 e ao abandono da paridade ouro-dólar, implicam em uma nova crise econômica mundial.

De acordo com Harvey (2013, p.140), “as décadas de 70 e 80 foram um conturbado período de reestruturação econômica e reajustamento social” que marca a passagem para um novo regime de acumulação caracterizado pela expansão e centralidade do capital financeiro e pela fragmentação e flexibilização da produção. Neste contexto, verifica-se uma nova crise da governamentalidade liberal, agora caracterizada pelo esgotamento da política econômica keynesiana, do planejamento

estatal e do compromisso político baseado na ideia do pleno emprego (Dardot e Laval, 2016).

A teoria econômica neoliberal, que já vinha sendo gestada desde 1940 pela escola austríaca capitaneada por Friedrich Hayek (1899-1992), encontrou terreno fértil para sua implementação, inicialmente junto a governos autoritários e conservadores³⁵. Posteriormente, nos anos 1980-1990, a crise do Estado de bem estar social, a derrocada do bloco soviético e os problemas generalizados de endividamento nos países periféricos pavimentaram o caminho para a disseminação global do neoliberalismo, com a decisiva participação de organizações internacionais e multilaterais, sob o novo regime de acumulação financeirizado e flexível³⁶ (Chesnais, 1996; Harvey, 2013).

Do ponto de vista do campo estatístico, a crise de governamentalidade desarticula o espaço político e cognitivo que fundamentava as ferramentas estatísticas do modelo keynesiano construído no pós-guerra, baseadas na intervenção estatal e no modelo de negociação coletiva entre sindicatos e empresários. De acordo com a racionalidade neoliberal, este espaço deve doravante corresponder ao próprio mercado, onde os agentes individuais dotados de informações orientam suas expectativas e ações com total liberdade para a maximização de suas utilidades. Desrosières (1998) destaca a perda de relevância do planejamento e dos modelos macroeconômicos e macrossociais no âmbito da ação pública, deslegitimados, entre outros, pela teoria das expectativas racionais³⁷.

Durante cerca de trinta anos, existiu um espaço cognitivo de negociação e cálculo, dotado da legitimidade da ciência e do Estado. Dentro deste espaço um grande número de debates e estudos técnicos foram realizados,

³⁵ As primeiras experiências neoliberais se dão nos anos 1970 sob a ditadura militar de Augusto Pinochet (1915-2006) no Chile e nos anos 1980 nos governos conservadores de Ronald Reagan (1911-2004) nos EUA e Margaret Thatcher (1925-2013) no Reino Unido.

³⁶ De acordo com Harvey (2013, p.140) o regime de acumulação flexível “se apoia na flexibilidade dos processos de trabalho, dos mercados de trabalho, dos produtos e padrões de consumo. Caracteriza-se pelo surgimento de setores de produção inteiramente novos, novas maneiras de fornecimento de serviços financeiros, novos mercados e, sobretudo, taxas altamente intensificadas de inovação comercial, tecnológica e organizacional” (Harvey, 2013, p.140).

³⁷ De acordo com a teoria das expectativas racionais, os agentes utilizam a informação disponível sobre o atual comportamento e as previsões para o futuro da economia e com base em sua experiência antecipam de forma racional as atitudes e políticas futuras do governo, reagindo no presente a partir destas expectativas, resultando na minoração ou anulação da efetividade das políticas macroeconômicas (Muth, 1961; Lucas; Rapping, 1969).

antecedendo ou acompanhando decisões de política econômica. Entretanto esse espaço relativamente coerente, composto por instituições, objetos sociais e palavras para debatê-los, ele próprio entrou em crise no final dos anos 1970. As redes de equivalência que conduziam às políticas e estatísticas foram parcialmente desmanteladas. O plano tem importância menor como lugar para a concertação e para realizar previsões de médio-prazo para decisões públicas importantes. Os modelos econométricos que simulavam o desenvolvimento de relações entre os objetos macroeconômicos e macrossociais mais centrais deste sistema estatístico são com frequência considerados incapazes de prever crises e tensões. Debates sobre a medição efetiva de alguns destes objetos e sua significância ficaram mais amargas: população ativa, desemprego, produto interno, pobreza, economia informal (Desrosières, p. 334, tradução nossa).

Do ponto de vista da teoria econômica e de sua aplicação na prática governamental, o neoliberalismo é caracterizado pelo retorno da ortodoxia liberal sob uma nova configuração. Enfatiza-se agora a promoção da concorrência em todas as esferas da economia e sociedade e um novo papel do Estado. Para Dardot e Laval (2016), ‘economia livre, Estado forte’ é o slogan que resume o sentido da ‘grande virada’ neoliberal que se verifica a partir dos anos 1980. Embora as políticas neoliberais caracterizem-se doravante pelos princípios de auto-regulamentação da economia pelos mercados e privatização de determinadas funções públicas do Estado, “não estamos lidando com uma simples retirada de cena do Estado, mas com um reengajamento político do Estado sobre novas bases, novos métodos e novos objetivos” (Dardot e Laval, 2016, p.190).

No neoliberalismo não se trata, portanto, de um ‘Estado mínimo’, mas sobretudo de um ‘Estado ativo’ na criação de novos mercados e na instauração e supervisão de um quadro jurídico que estabelece a concorrência como a norma geral das práticas sociais e econômicas. Sob este quadro concorrencial, o próprio Estado deve reconstruir-se de acordo com as normas do mercado, como se fora uma empresa. O empreendedorismo e a empresa privada tornam-se assim o modelo central da governamentalidade neoliberal, não apenas para o funcionamento do mercado e do Estado, mas sobretudo como fundamento de um novo modo de subjetivação, no qual “cada indivíduo é uma empresa que deve se gerir e um capital que deve se fazer frutificar” (Dardot e Laval, 2016, p.378).

A concepção empresarial do Estado implicará em uma série de consequências para o campo estatístico em todo o mundo, principalmente na medida em que esta noção ganhou materialidade nos programas de ‘modernização’ do setor público, introduzidos pela *New Public Management* (NPM), a partir dos anos 1980 (Hood, 1995). A NPM consistiu na reestruturação gerencial do Estado de acordo com o

corolário neoliberal, por meio da adoção de práticas de gestão do setor privado na administração pública, sob os alegados objetivos de aumentar a eficiência, eficácia e transparência, com redução de custos. Desenvolvida inicialmente no Reino Unido, sob a administração conservadora de Margareth Thatcher (Oliver e Drewry, 1996), a NPM será disseminada a partir dos anos 1990 para os países do norte (McLaughlin et al., 2002) e sul-global (McCourt e Minogue, 2001).

De acordo com Harvey (2008, p.86), a implementação de mudanças estruturais na natureza da governança sob o neoliberalismo visava sobretudo “integrar a tomada de decisões do Estado à dinâmica de acumulação do capital”. Neste sentido, a reconfiguração das instituições e práticas do Estado implicou, ao mesmo tempo, na redução de sua autonomia relativa, em grande medida vinculada à tradicional independência do serviço público. Harvey (2008, p.88) destaca como neste processo “a fronteira entre o Estado e o poder corporativo se tornou cada vez mais porosa” por meio das agendas de privatizações e parcerias público-privadas (PPPs).

No Reino Unido, onde a doutrina da NPM adquiriu maior radicalidade, é simbólico que as estatísticas oficiais tenham sido o primeiro alvo da política de redução do Estado durante o governo Thatcher (Radermacher, 2020). Neste caso, a revisão do sistema estatístico nacional partiu do pressuposto de que “os dados não são recursos gratuitos” e de que “não há mais razões para o governo ser um fornecedor universal no domínio estatístico” (Beaud; Prévost, 2011, p.40). De fato, embora o Central Statistics Office (CSO) não tenha sido incluído no programa de privatizações do Reino Unido, opção cogitada pelos membros mais conservadores do governo Thatcher (Jones et al., 1990), a reestruturação do sistema estatístico implicou em uma drástica redução no orçamento e nos quadros do CSO, levando à interrupção de uma série de inquéritos estatísticos³⁸.

De acordo com Beaud e Prévost (2011), a reforma do setor estatístico no Reino Unido implicou na introdução de uma nova modalidade de contratos comerciais entre o serviço de estatística e seus ‘clientes’, formados por outros departamentos governamentais e pelo setor privado. Para Beaud e Prévost (2011, p.40), verifica-se “ao longo de uma década uma redução espetacular dos serviços

³⁸ De acordo com Jones et al. (1990), a privatização do CSO não ocorreu, entre outros motivos, em face do risco de perda de controle político sobre informações estratégicas para o governo. A reestruturação do sistema estatístico ficou a cargo do *ex-chairman* da Mark & Spencer, Derek Rayner (1926-1998) responsável pela condução de uma revisão do setor estatístico, que ficou conhecida como “relatório Rayner”.

estatísticos que levou a uma grave crise de confiança, da qual o sistema estatístico Britânico ainda não se recuperou totalmente”.

A imunidade dos INEs às iniciativas generalizadas de privatização de serviços públicos em todo o mundo, a partir dos anos 1990, é um indício de que determinadas funções de Estado mantiveram importância estratégica sob a governamentalidade neoliberal. No caso das estatísticas oficiais, suas tradicionais funções na produção de informações públicas que sustentam a legitimidade dos governos perante a opinião pública e orientam a ação dos agentes econômicos no mercado, se sobrepuseram a agenda privatista. Deste modo, a resiliência dos INEs às reformas neoliberais vinculou-se sobretudo à capacidade de adaptação das estatísticas oficiais à nova agenda e à emergência de contramovimentos protetivos no campo estatístico contra a investida disruptiva do neoliberalismo.

2.2 O regime estatístico de comensuração: financeirização e política de indicadores

Os indicadores estatísticos têm uma longa história e não nasceram sob o neoliberalismo. Foi sob esta nova racionalidade, entretanto, que adquiriram proeminência como ferramentas de governo. De acordo com Porter (2015, p.53), neste novo contexto, os indicadores passam a “assumir um papel dentro de sistemas de incentivos que mimetizam o mercado, funcionando como mercados artificiais”. A governamentalidade neoliberal demanda uma comensuração generalizada de objetos e processos nos mais variados campos de práticas e esferas do conhecimento, onde os indicadores passam a reger a lógica das ações, orientando-as de acordo com os princípios da concorrência.

Do ponto de vista estatístico, os indicadores são “proposições quantitativas simples e facilmente comparáveis” (Rottenburg e Merry, 2015, p.3). Tratam-se fundamentalmente de instrumentos de prescrição, adequados a orientar a ação, mas dotados de baixo poder explicativo e descritivo. Porter (2015) destaca a aderência destas ferramentas superficiais à economia intelectual do neoliberalismo:

Etimologicamente, um indicador, assim como um índice, tem a ver com apontar. Anatomicamente, o músculo indicador (extensor do indicador) endireita o dedo indicador. Logicamente, os indicadores detectam, apontam ou medem, mas não explicam [...] Um índice ou indicador quantitativo normalmente não pode medir exatamente o que interessa, mas em seu lugar

algo cujos movimentos mostram uma relação consistente com esse objeto. Dado que o seu objetivo é meramente indicar como um guia para a ação, a facilidade de medição é preferível ao significado ou à profundidade. O indicador está entre as variedades de informação cuja ascensão tem sido tão acentuada na economia intelectual dos tempos modernos, tendendo talvez a excluir aquelas formas epistêmicas mais exigentes, como o conhecimento e a sabedoria (Porter, 2015, p.34).

Muito embora sob diferentes denominações, os indicadores quantitativos “apareceram em cena mais ou menos simultaneamente com as estatísticas públicas regulares” nas primeiras décadas do Séc. XIX, na forma de números ou taxas que sintetizavam e representavam algum fenômeno ou objeto, como por exemplo as taxas de mortalidade (Porter, 2015, p.41). Já o uso explícito da palavra ‘indicador’, como referência a uma medida quantitativa, surgiu no âmbito do mercado nas primeiras décadas do Séc. XX, quando “grandes corporações, especialmente nos Estados Unidos, desenvolveram indicadores de negócios como parte de seu esforço para reduzir a sua vulnerabilidade às flutuações da economia” (Porter, 2015, p.48). A referência a ‘indicadores’ passou posteriormente a ser empregada no campo estatístico, a partir dos anos 1940, especialmente em relação às estatísticas econômicas (taxas de desocupação, produto interno bruto, preços, inflação, etc.). A partir de meados dos anos 1960, entretanto, verifica-se uma considerável expansão no uso de indicadores no campo estatístico, desencadeada pelo ‘movimento dos indicadores sociais’ (Duncan, 1969).

A formulação dos ‘indicadores sociais’ tem origem nos trabalhos de sociólogos quantitativos da escola de Chicago, no campo da teoria da mudança social (Land, 1983)³⁹. De acordo com Land (1983, p.3), o interesse dos sociólogos nos indicadores se deveu em grande medida ao sucesso dos economistas na utilização do PIB para a proposição e avaliação de políticas econômicas. O êxito do sistema de contas nacionais “sugeriu aos cientistas sociais que um conjunto análogo de indicadores sociais poderia ser utilizado de forma semelhante na manipulação da política social”. A partir daí, ganhou força em setores da academia e do governo estadunidense a

³⁹ Um precedente do movimento dos indicadores sociais foi o trabalho de William F. Ogburn na Universidade de Chicago nas décadas de 1930 e 1940 sobre a teoria e medição da mudança social. Como presidente do Comitê Presidencial de Pesquisa sobre Tendências Sociais, Ogburn supervisionou a produção do relatório *Recent Social Trends* (1933), contribuição que estabeleceu um precedente para os relatórios sociais. As ideias de Ogburn sobre a medição da mudança social influenciaram vários dos seus alunos - nomeadamente Albert D. Biderman, Otis Dudley Duncan, Albert J. Reiss e Eleanor Bernert Sheldon - que desempenharam papéis importantes no surgimento e desenvolvimento do campo dos indicadores sociais

ideia de um 'sistema de contas sociais', que avaliasse o 'bem estar da Nação' e facilitasse uma análise de custo-benefício de outros aspectos da sociedade além do econômico.

Na década de 1970, há uma consolidação da agenda dos indicadores sociais nos Estados Unidos e sua disseminação para a Europa e América Latina. No âmbito dos sistemas estatísticos nacionais, verificam-se investimentos direcionados a coordenação da produção de relatórios sociais nacionais. Na década de 1980, em meio à crise do Estado Keynesiano, há uma descontinuidade nos investimentos e esforços de coordenação dos indicadores sociais, na medida em que, a ideia de construção de um modelo de contas sociais é abandonada (Desrosières, 2014).

A eficácia das contas nacionais resultou, em particular, do consenso reunido, nas décadas do pós-guerra, em torno de uma teoria econômica descritiva e prescritiva, aceita nos Estados Unidos e na Europa. As chamadas políticas 'sociais', por outro lado, baseavam-se em diferentes tradições nacionais, e a teoria 'sociológica' nunca forneceu uma linguagem comum comparável à da economia. Os 'indicadores sociais' foram, portanto, propostos de forma dispersa e não foram integrados num quadro conceitual comum. A integração foi, portanto, na melhor das hipóteses, principalmente técnica e foi obra de estatísticos e não de teóricos, economistas ou sociólogos (Desrosières, 2014, p.248-249, tradução nossa).

A partir dos anos 1990, a quantificação por meio de indicadores será intensificada sob uma nova orientação teórica e política, conduzindo ao que Desrosières (2014) intitula de uma 'política de indicadores'. Por um lado, a lógica normativa e macro sistêmica do planejamento estatal, que orientou as primeiras iniciativas de coordenação dos indicadores sociais, será reorientada pela perspectiva 'micro teórica' que passa a prevalecer na economia, sociologia e ciência política. De acordo com esta abordagem, a relação adequada entre os indicadores sociais e a política social ocorre ao nível gerencial e operacional das organizações e agentes. Por outro lado, o incremento no uso de indicadores está diretamente relacionado à financeirização da economia e à emergência de uma nova convenção de quantificação no campo das práticas financeiras a partir dos anos 1980 (Chiapello e Walter, 2016)⁴⁰.

⁴⁰ De acordo com Chiapello e Walter (2016), os métodos de quantificação da valoração financeira estão historicamente associados a três estágios de desenvolvimento das ferramentas de cálculo da 'taxa de retorno do investimento': i) a convenção atuarial, baseada no 'fluxo de caixa descontado'; ii) a convenção da 'variância média', fundamentada em modelos probabilísticos aplicados à gestão de portfólios de investimentos; iii) e a convenção de 'consistência mercadológica', associada à emergência dos derivativos financeiros, orientada por indicadores de mercado e modelos preditivos.

No contexto de expansão da financeirização do capital em um mercado supostamente autorregulado, as opções de investimentos passaram a ser avaliadas de acordo com os indicadores do mercado sobre os preços dos ativos. Neste novo estágio das convenções de quantificação no campo financeiro, “o objetivo já não é descontar os fluxos de caixa futuros esperados para obter um valor presente, mas sim o contrário: tomar os valores presentes observados nos mercados e deduzir deles todos os retornos esperados de um possível investimento” (Chiapello; Walter, 2016, p.5). Esta inversão na lógica de valoração dos ativos, que dá margem ao desenvolvimento dos ‘derivativos’ financeiros, é fundamentada na presunção de um mercado autorregulado e marca a introdução da estatística bayesiana nas práticas financeiras (Delbaen; Schachermayer, 1994; MacKenzie; Spears, 2014; Chiapello; Walter, 2016)⁴¹.

Para Chiapello e Walter (2016), a noção de que os valores dos ativos devem ser consistentes com os indicadores de mercado, passou a ser igualmente utilizada para avaliar ‘ativos’ que não têm necessariamente um preço de mercado. Essa permeabilidade das convenções de quantificação financeira para outros campos é uma importante característica do processo de financeirização, que ganha impulso nesta nova etapa do capitalismo, à medida em que a reorganização da lógica geral de acumulação do capital submete o processo produtivo e outras relações sociais à lógica da valorização financeira (Chesnais, 1996, 2016; Mollo et. al., 2022, Paraná, 2024). Assim, para Chiapello e Walter (2016):

O processo de financeirização, que está redefinindo setores inteiros da economia e transformando as lógicas de funcionamento das empresas, bem como as políticas públicas, traz consigo concepções de mundo, métodos de análise de problemas, técnicas de cálculo e princípios de tomada de decisão que foram originalmente forjados para um número limitado de casos especiais, mas tendem agora a alastrar-se a todas as questões e atividades

⁴¹ Sob a hipótese da “oportunidade de ausência de arbitragem” em um mercado autorregulado (Delbaen; Schachermayer, 1994), a nova matemática financeira passa a considerar os preços de mercado como a medida perfeita para calcular a expectativa de retorno dos investimentos, sendo a taxa de retorno definida a partir de uma função de probabilidade à priori, denominada “probabilidade de risco neutro” (Chiapello; Walter, 2016). Esta probabilidade, que tornaria um participante do mercado indiferente à compra ou venda de um ativo, não se refere à uma ‘probabilidade real’ calculada a partir das frequências de eventos observáveis, mas é sobretudo, uma medida da crença e expectativas do mercado sobre a probabilidade de um evento futuro. A construção desta probabilidade à priori baseia-se em um série de informações empíricas (indicadores de mercado) e pode ser interpretada como uma crença hipotética sobre o comportamento dos investidores em um mercado ficcional onde o risco é irrelevante. Essa crença é atualizada à medida que novas informações sobre o mercado e o comportamento dos investidores são disponibilizadas por meio de modelos estatísticos bayesianos. (MacKenzie; Spears, 2014).

humanas. Estruturas de raciocínio, representação e cálculo extraídas das finanças podem aparentemente ser aplicadas e redefinir todas as esferas da existência (Chiapello; Walter, 2016, p.2)

A auditoria financeira foi um dos principais campos de disseminação das convenções de quantificação e valoração das finanças para outras esferas de conhecimento e ação. De acordo com Power (2000), o crescimento e circulação das práticas que conduziram a uma “explosão da auditoria”, inicialmente no Reino Unido e posteriormente no resto do mundo, relacionam-se à crise e restrição financeira dos anos 1980 e a um novo compromisso em torno de reformas nos setores público e privado em torno das ideias de controle, governança e qualidade. O fundamento destas reformas é uma ‘reafirmação da ordem capitalista’ que visa fortalecer o controle gerencial por meio da internalização de esquemas de autocontrole pelas organizações e agentes. Para Bruno e Didier (2013), a incorporação das práticas de auditoria financeira na gestão pública, por meio da NPM, deu margem à disseminação de sistemas de incentivos e sanções, medidos por meio de indicadores que governam os agentes à distância, introjetando sistemas de classificação por meio de rankings e benchmarkings (Bruno e Didier, 2013).

De acordo com Desrosières (2014), a disseminação da ‘política de indicadores’ provocou uma descontinuidade na utilização tradicional das estatísticas oficiais pelos governos e causou uma perturbação no seu ‘estado de realidade’ que passou a afetar o ‘ethos’ dos estatísticos de Estado e a confiança pública nas estatísticas oficiais. Isto decorre de dois fatores principais. Em primeiro lugar, a lógica política e cognitiva dos indicadores e rankings, que se originam no meio corporativo privado, difere substancialmente da tradição estatística. Em segundo lugar, na medida em que agem diretamente sobre aquilo que é medido, passam a introduzir importantes efeitos de feedback desafiando “a epistemologia realista que está subjacente à metrologia das estatísticas públicas” (Desrosières, 2014, p.73).

Apesar de reconhecer que as estatísticas oficiais tradicionais também produziram historicamente efeitos sobre a realidade, para Desrosières (2014) estes efeitos, entretanto, possuíam até então um caráter macrossocial que embora não revelado, era tolerado pelos estatísticos⁴². Na medida em que os indicadores passam

⁴² De modo similar ao que Bourdieu (2008, p.82) se refere como ‘efeito de teoria’, na medida em que a ciência “ao contribuir para impor uma maneira mais ou menos autorizada de ver o mundo social, contribui para fazer a realidade desse mundo”.

a ser associados a metas, entretanto, tanto os efeitos de feedback são ampliados, em virtude do comportamento e ações dos agentes serem orientados em direção às metas, quanto aumentam os riscos de manipulação das métricas para alcançá-las⁴³. Isto se torna especialmente relevante nos sistemas de avaliação introduzidos pela NPM, que passam a conferir um papel fundamental aos indicadores estatísticos, por exemplo, atrelando decisões orçamentárias à avaliações de desempenhos e estimulando a competição por meio de benchmarkings e ‘melhores práticas’ (Bruno e Didier, 2013).

Os efeitos de feedback dos indicadores, quando atrelados a metas e sistemas de recompensas e punições, entretanto, não se tratam de ‘efeitos colaterais’ destes sistemas, como a interpretação de Desrosières (2014) pode levar a crer, pelo contrário, correspondem ao resultado desejado da governamentalidade neoliberal. O objetivo desta tecnologia de governo é justamente o de fazer com que os próprios agentes se internalizem, ajustando ‘automaticamente’ seus comportamentos e condutas por meio da auto verificação. De acordo com Merry (2009, p.8), efetua-se assim “uma transferência de responsabilidade que mascara a dinâmica de poder subjacente: o próprio indicador faz o trabalho de crítica e o agente governado procura conformar-se com os seus termos”, enquanto o agente governante pode ater-se à simples verificação dos indicadores.

Os efeitos diretos de feedback, entretanto, não são o único problema introduzido pelos indicadores no campo estatístico. Os processos de tomada de decisão baseados em evidência, se por um lado reforçam a importância das estatísticas, por outro, passam a criar demandas irrealizáveis ou simplesmente contraproducentes, produzindo uma série de vieses observacionais⁴⁴. Para Radermacher (2020, p.89), novas ameaças ao campo estatístico que surgem neste contexto, estão relacionadas com a cultura generalizada de comensuração: “negar que existe uma lacuna entre o que se deseja medir e o que se pode realmente medir, faz com que esta lacuna seja preenchida de uma forma ou de outra, seja por estimativas erradas, por um esforço inapropriadamente elevado ou simplesmente por

⁴³ Este risco associado à qualidade estatística dos indicadores ficou consagrado pela ‘lei de Goodhart’ que afirma que, “qualquer regularidade estatística tende a entrar em colapso quando pressão é colocada sobre ela para fins de controle” (Chrystal e Mizen, 2001).

⁴⁴ Um dos vieses de observação relacionados à ênfase nos indicadores é conhecido como o “efeito da luz de poste”, que consiste em desconsiderar, ignorar, julgar menos importante ou mesmo inexistentes aquilo que não pode ser facilmente medido (Radermacher, 2020).

fatos falsos”. Conforme veremos, esta lacuna também motivará a busca por novas tecnologias e fontes de dados para povoar sistemas contraproducentes de indicadores, como os ODS, justificando parcerias público-privadas para uso de big data em estatísticas oficiais.

Desrosières (2014) aponta duas críticas emergentes e de naturezas opostas que envolvem o problema da comensuração. Por um lado, identifica uma crescente resistência social aos aspectos perversos dos sistemas de avaliação por indicadores. Esta resistência se apresenta como uma crítica generalizada aos processos de comensuração que se estende às estatísticas oficiais. Por outro lado, as iniciativas de medição de conceitos complexos e multidimensionais como ‘pobreza’, ‘qualidade de vida’ e ‘sustentabilidade’ e a demanda de novos movimentos sociais pela inclusão de aspectos sociais e ambientais no cálculo do PIB, reivindicam uma intensificação das práticas de comensuração, mesmo quando alegadamente orientadas por uma concepção não economicista.

A comensuração, ou seja, a transformação de qualidades em quantidades, por meio da expressão de características normalmente representadas por diferentes unidades, de acordo com uma métrica comum (Espeland; Stevens, 1998), constitui um etapa fundamental da prática estatística. Os processos de comensuração enquanto práticas sociais, pressupõem, portanto, a negociação e o acordo em torno das classificações e convenções de equivalências que tornam um determinado objeto comensurável (Desrosières, 1998). Sob o neoliberalismo, entretanto, este espaço de negociação passou a ser cada vez mais submetido às regras da concorrência e do mercado, tornando-o permeável à lógica da financeirização (Porter, 2015).

Finalmente, a intensificação do uso de indicadores em processos de controle e tomada de decisão implicou em uma comensuração generalizada nas esferas da vida econômica e social. Neste sentido, o que aqui caracterizamos como um novo regime de comensuração implica, por um lado, em mudanças internas no processo de produção das estatísticas oficiais pelos INEs, na medida em que interfere na agenda e nas práticas estatísticas, e por outro lado, em mudanças externas, ao promover a emergência de novos centros de quantificação na iniciativa privada e nas organizações internacionais.

2.3 Virada empresarial dos Institutos Nacionais de Estatística e contramovimentos protetivos

No contexto neoliberal, verifica-se uma maior permeabilidade dos INEs à lógica empresarial e competitiva. Um dos elementos que caracteriza esse processo é a introdução de uma nova linguagem e de um novo conjunto de práticas oriundas do setor corporativo privado no processo produtivo das estatísticas oficiais. De acordo com Beaud e Prévost (2011), como resultado das reformas gerenciais do Estado nos anos 1990 verifica-se uma ‘virada empresarial’ nos INEs, que contrasta fortemente com o perfil profissional público e científico dos estatísticos de Estado.

Em muitos países o instituto estatístico é agora descrito como uma agência que fornece serviços e produtos num mercado onde os clientes são sensíveis à qualidade e aos preços: essas são algumas das palavras-chave que comprovam – em um nível simbólico pelo menos – a virada empresarial que foi tomada e que contrasta fortemente com a autoimagem científica do último período (Beaud; Prévost, 2011, p.39)

Ao mesmo tempo em que há uma maior permeabilidade à linguagem e práticas empresariais em determinados países e segmentos do campo estatístico, verifica-se, por outro lado, uma relativa resistência às reformas neoliberais. Isto se deve à emergência de contramovimentos que buscam salvaguardar a relevância e o caráter público das estatísticas oficiais, protegendo ao mesmo tempo a autonomia relativa dos INEs. Um dos mais importantes contramovimentos defensivos do campo estatístico à ofensiva neoliberal foi a formalização de um conjunto de princípios fundamentais que regem as normas e práticas das estatísticas oficiais. Inicialmente adotado em 1992 no âmbito da Conferência dos Estatísticos Europeus articulada pela UNECE (1992), os Princípios Fundamentais das Estatísticas Oficiais (FPOS) foram adotados em nível global pela Comissão Estatística da ONU em 1994 (UN-StatCom, 1994) e ratificados sem alterações pela Assembleia Geral das Nações Unidas em 2014. Os dez princípios estabelecem: i) relevância, imparcialidade e igualdade de acesso; ii) Independência profissional, princípios científicos e ética profissional; iii) responsabilidade e transparência; iv) prevenção do uso indevido; v) qualidade das fontes; vi) confidencialidade; vii) leis públicas; viii) coordenação nacional; ix) uso de padrões internacionais e x) cooperação internacional.

Entre as evidências empíricas da virada empresarial nos INEs a partir da década de 1990, Beaud e Prévost (2011) destacam os casos do Canadá, Austrália,

Nova Zelândia, Suécia e Finlândia. Nestes países constam a adoção de novos planos e estratégias de marketing, bem como o estabelecimento de metas de captação de receitas financeiras por meio, entre outros, da realização de surveys para ‘clientes’ e do desenvolvimento de produtos impressos e eletrônicos. Beaud e Prévost (2011, p.41) também apontam que, no caso do Statistics Canadá, “esta virada empresarial significou uma série de mudanças nas práticas da instituição, tais como a subcontratação de algumas das suas operações para empresas privadas”.

No âmbito das práticas estatísticas, um dos meios de penetração da lógica empresarial nos INEs foram os processos gerenciais de controle de qualidade. De acordo com Desrosières (2014, p.237) verifica-se, sob o contexto da produção flexível, uma ‘metamorfose’ do conceito de qualidade: “anteriormente associada a uma interpretação probabilística da confiabilidade [...] a ‘qualidade’ está agora inserida numa cadeia performativa muito mais diretamente orientada para o mercado”⁴⁵. No campo estatístico, segundo Beaud e Prévost (2011, p.41), os “conceitos de excelência científica, neutralidade e integridade, que constituíram o núcleo normativo das estatísticas durante mais de um século, são igualmente reformulados de acordo com a linguagem da qualidade total”. Não obstante, de acordo com Beaud e Prévost (2011), os INEs lograram utilizar a nova dinâmica da qualidade em estratégias voltadas ao fortalecimento dos programas estatísticos e da autonomia institucional.

Muito embora os critérios de qualidade sejam predominantemente abordados do ponto de vista de seu caráter técnico — incluindo aspectos como relevância, precisão, tempestividade, acessibilidade, interpretabilidade e coerência — Beaud e Prévost (2011, p.42) destacam que as questões de qualidade também podem ser compreendidas como “o espaço e as posições em torno das quais são travadas uma série de lutas político-administrativas”. Estas lutas no interior do campo estatístico passam a envolver, entre outros, o tipo de informações produzidas e a publicidade dos dados. Notadamente, embora a adoção dos processos de qualidade tenha implicado em uma maior permeabilidade dos INEs à lógica empresarial, Beaud e Prévost (2011) destacam a utilização estratégica dos critérios de relevância e

⁴⁵ Desrosières (2014) observa que na década de 1930, o estatístico americano W. Edwards Deming que foi um dos responsáveis pela introdução do cálculo de probabilidades nas estatísticas públicas, utilizou a partir de 1940 as ferramentas probabilísticas para desenvolver o conceito de “controle de qualidade” na produção industrial, por sorteio aleatório e verificação de lotes de peças na produção em massa. Esta orientação fez muito sucesso no Japão, onde Deming foi um dos iniciadores do “movimento da qualidade” que ganhou projeção mundial a partir de 1980.

tempestividade para preservar os programas estatísticos de cortes orçamentários e para fortalecer a autonomia técnica e política dos INEs.

De acordo com Beaud e Prévost (2011, p.35), o princípio de ‘relevância’ foi utilizado pelos INEs para preservar as estatísticas oficiais de cortes orçamentários, em um contexto de aparente ineficácia das políticas macroeconômicas e do questionamento da adequação das intervenções governamentais. Esta estratégia, implicou na introdução de mudanças nos programas estatísticos para atender às novas demandas de informações dos governos. Neste sentido, verificam-se investimentos para a “obtenção de dados sobre o comportamento microeconômico e social a um nível mais preciso do que os fornecidos por macro agregados como o PIB ou a taxa de desemprego”, de forma a produzir informações relevantes para o planejamento e avaliação das políticas públicas setoriais.

O movimento para o desenvolvimento de indicadores de desempenho pode ser visto como uma resposta a essa situação, tal como o foi o desenvolvimento de programas estatísticos mais direcionados para a compreensão dos problemas microeconômicos e do comportamento social de grupos específicos (no que diz respeito a problemas de saúde, desemprego, criminalidade, etc.). As estatísticas relevantes são, portanto, definidas como estatísticas que fornecem informações para identificar, medir e avaliar os efeitos de políticas e programas específicos ou das próprias variáveis suscetíveis de afetar o sucesso ou o fracasso desses programas. Quando apresentam a relevância como pedra de toque, os institutos de estatística apresentam-se como a fonte privilegiada de inteligência graças à qual o governo pode prestar contas das suas políticas e da utilização que faz do dinheiro dos contribuintes (Beaud e Prévost, 2011, p.43).

Já o princípio de ‘tempestividade’ — ou seja, a disponibilização da informação no momento em que é demandada e relevante — se por um lado implicou em uma maior pressão por agilidade no processo estatístico, por outro lado, foi estrategicamente utilizado em benefício da ampliação da autonomia dos INEs em um contramovimento de democratização da informação pública e aproximação com setores da mídia e do público mais amplo. De acordo com Beaud e Prévost (2011, p.44), a adoção de um calendário público de divulgações estatísticas se tornou uma prática coerente com os princípios de qualidade para garantir a tempestividade das estatísticas oficiais. Entretanto, na medida em que “a informação é divulgada ao governo e ao público ao mesmo tempo – ou quase ao mesmo tempo, qualquer intervenção governamental para impedir a divulgação de dados torna-se evidente”. Há diversos casos que comprovam que o adiamento de divulgações estatísticas

resultou em controvérsias afetando a confiança pública sobre os governos⁴⁶. O custo político, portanto, fez dos calendários de divulgações um instrumento de proteção e ampliação da autonomia dos INEs.

Finalmente, concluímos que sob a governamentalidade neoliberal verifica-se a introdução de transformações tendentes a uma maior permeabilidade do campo estatístico aos princípios do mercado e das práticas corporativas do setor privado. Por outro lado, há ao mesmo tempo uma relativa resiliência do campo estatístico a esta incidência, o que se traduz em sua capacidade de adaptação e em contramovimentos que visam resguardar o caráter público das estatísticas oficiais e a autonomia relativa dos INEs.

2.4 Espaços transnacionais de quantificação e governança por indicadores

Na esteira dos processos de globalização que se intensificam no final do Séc. XX, verifica-se o fortalecimento de espaços internacionais e transnacionais de quantificação de pessoas, objetos e fluxos. De acordo com Beaud e Prévost (2011, p.35-36), no que diz respeito à estrutura estatística, “o estatuto do ‘nacional’ como o nível mais significativo de governança” e “o princípio da centralização estatística, que personificava a dimensão nacional durante o século XX” são colocados em questão na medida em que a harmonização e a decisão sobre programas e classificações estatísticas são parcialmente transferidas para o nível supranacional. Este processo é mais evidente na Europa, onde, a partir do tratado da União Europeia em 1992, o Eurostat passa a assumir um novo papel na coordenação estatística europeia, mas também se verifica nos esforços de coordenação estatística na América do Norte a partir do acordo do NAFTA em 1994 e em menor medida na América do Sul com a criação do Mercosul em 1991.

Do ponto de vista das Organizações Internacionais, verifica-se, a partir dos anos 1990, a intensificação de um modelo de ‘governança’ global que tem como fundamento o estabelecimento de metas nacionais calculáveis e comparáveis por meio de indicadores estatísticos e rankings globais. Evidentemente, a comparação

⁴⁶ Beaud e Prévost (2011) citam como exemplos a decisão tomada em Março de 2007 pelo Institut National de la Statistique et des Études Économiques (INSEE), o organismo estatístico francês, de adiar para Novembro a divulgação do seu inquérito anual sobre o emprego, alegadamente por razões metodológicas e uma série de problemas relativos ao momento da divulgação de estatísticas na Grã Bretanha. Conforme veremos situações semelhantes também podem ser verificadas no Brasil.

estatística internacional não surge neste contexto, mas tem raízes em uma longa tradição histórica que, como vimos, é subjacente às primeiras iniciativas do internacionalismo estatístico e ao sonho de governança técnica e apolítica dos primeiros estatistas do Séc. XIX. No contexto neoliberal, entretanto, os indicadores, metas e rankings alcançaram o status de instrumentos políticos de governo definidos por meio de acordos internacionais e negociações diplomáticas.

Este novo modelo de governança global por indicadores implicou em importantes repercussões, na medida em que questões técnicas e métricas passaram a se sobrepor ao processo político e aos aspectos normativos dos acordos, conferências e resoluções pactuadas internacionalmente sob o quadro dos direitos humanos. De acordo com Merry (2009):

A expansão da utilização de indicadores na governança global significa que as lutas políticas sobre o que significam os direitos humanos e o que constitui o seu cumprimento são reformuladas por questões técnicas de medição, critérios e acessibilidade aos dados. Os debates políticos sobre a conformidade mudam para argumentos sobre como formar um indicador, o que deve ser medido e o que cada medição deve representar. Estes debates normalmente contam com especialistas na área de medição e estatística, geralmente em consulta com especialistas no tema substantivo no terreno nacional e internacional. Eles se baseiam em pesquisas anteriores e no conhecimento gerado por acadêmicos. Os resultados aparecem como formas de conhecimento e não como representações particulares de uma metodologia e de decisões políticas específicas sobre o que medir e nomear. Um indicador fornece uma transição da ambiguidade para a certeza, da teoria para o fato, da variação complexa e do contexto para números verdadeiros e comparáveis. Em outras palavras, o processo político de julgar e avaliar se transforma em uma questão técnica de medição e contagem (Merry, 2009, p.9-10, tradução nossa).

Do ponto de vista do campo estatístico transnacional, esta ‘virada estatística’ na ‘governança global’ implicou em uma participação mais pronunciada dos órgãos técnicos da Comissão Estatística da ONU (UN-StatCom) e Divisão Estatística da ONU (UNSD) nos processos de negociação das políticas de desenvolvimento e na coordenação estatística global. O principal antecedente deste processo, entretanto, não se origina na UN-StatCom mas no Programa das Nações Unidas para o Desenvolvimento (UNDP), a partir do lançamento do Relatório do Desenvolvimento Humano (RDH) em 1990 e de um ranking de países classificados com base em sua pontuação no Índice de Desenvolvimento Humano (IDH)⁴⁷.

⁴⁷ De acordo com um de seus principais promotores, o objetivo explícito do RDH foi “mudar o foco do desenvolvimento econômico, da contabilidade da renda nacional para as políticas centradas nas pessoas” (UI Haq, 1995). Fundamentado no esquema conceitual do economista Amartya Sen (1999),

A nova abordagem do ‘desenvolvimento humano’, que Sandbrook (2000) intitula de ‘neoliberalismo pragmático’, dá margem a adoção de um novo modelo de intervenção pelas organizações internacionais e multilaterais com reflexos nos Estados Nacionais. Este modelo, mais sofisticado que as reformas conservadoras dos anos 1980, apresenta-se como uma crítica ao economicismo e uma defesa a uma abordagem multideterminada do desenvolvimento, que o iguala à ampliação das liberdades individuais em diferentes esferas (social, política, econômica, cultural). Fundamentado na concepção de que a participação livre das pessoas no mercado é “um padrão natural e intrinsecamente valioso” (Sandbrook 2000, p.1074), o modelo implica em uma subordinação das políticas de desenvolvimento à lógica do mercado autorregulado. Assim, os Estados Nacionais são convocados a ampliar as capacidade das pessoas, das empresas e da economia nacional para a competição em um mercado global idealizado, em que as relações assimétricas de poderes e capitais que determinam as condições desiguais de participação nos mercados, não são consideradas.

Esta abordagem orientada para o mercado está muito distante do reducionismo bruto da análise neoclássica ortodoxa. Embora as reformas macroeconômicas continuem a ser centrais, esta nova perspectiva adota o bem-estar humano em vez do mero crescimento como objetivo e alarga a agenda de desenvolvimento para incluir reformas políticas, sociais e institucionais [...] este modelo apresenta uma estratégia baseada no mercado que é holística, sinérgica e complexa. Primeiro, abrange objetivos políticos e sociais, para além dos objetivos convencionais macroeconômicos ou de mercado. Em seguida, estes objetivos são complementares e reforçam-se mutuamente. E terceiro, considera-se que os sistemas de mercado eficientes requerem a ação de apoio de Estados nacionais eficazes (Sandbrook, 2000, p. 1071-1073, tradução nossa).

De acordo com Fukuda-Parr (2003), o estabelecimento de ferramentas simples de medição como o IDH “foi fundamental para introduzir o desenvolvimento humano como um paradigma alternativo e para chamar a atenção dos atores políticos”. Do

a agenda do ‘desenvolvimento humano’ abre caminho para uma mudança global na orientação das políticas de desenvolvimento, da provisão de serviços públicos e atendimento às necessidades coletivas, para o ‘empoderamento’ das pessoas e ampliação das ‘escolhas’ individuais. O Índice de Desenvolvimento Humano (IDH) desenvolvido pelos economistas UI Haq e Sen é um índice estatístico composto pela expectativa de vida, educação (média de anos de escolaridade completados e anos esperados de escolaridade ao entrar no sistema educacional) e indicadores de renda per capita. De acordo com a sua pontuação os países são classificados em quatro níveis de desenvolvimento humano de acordo com os quartis da distribuição do índice.

ponto de vista das práticas estatísticas, de acordo com Ward (2004, p.202) “O RDH representou um afastamento ousado das tradições estatísticas da ONU e adentrou no que até então era considerado um território de medição proibido”. Ward (2004) destaca ‘desvios’ da prática estatística da ONU, tradicionalmente situada no âmbito da harmonização de estatísticas nacionais, para a efetiva produção de índices nacionais.

A publicação do IDH suscitou uma série de críticas no campo estatístico⁴⁸. De acordo com Ward (2004, p.203) a questão foi levada à Comissão Estatística da ONU e resultou na composição de um painel independente de especialistas para conduzir uma revisão do conteúdo estatístico do relatório e do índice, em face do que “a comunidade estatística profissional considerou um tratamento inadequado, seletivo e tendencioso dos dados para apoiar posições políticas e sua definição estatística pouco precisa”. Para Ward (2004, p.203), entretanto, em face da repercussão política global alcançada pelo índice, a despeito das críticas técnicas, “a conclusão foi de que o índice não pretende ser um instrumento estatisticamente preciso, mas uma ferramenta para advocacy”.

A abordagem do ‘desenvolvimento humano’, atrelada a estratégia de compilação de indicadores voltados à mobilização política, exerceu uma grande influência na redefinição da agenda de desenvolvimento das Nações Unidas ao longo da década de 1990. O ponto de culminância deste processo foi a expressão dos compromissos da ‘Declaração do Milênio’ adotada pela Assembleia Geral da ONU em 2000, na forma de metas e indicadores estatísticos no quadro dos Objetivos de Desenvolvimento do Milênio (ODMs). Para uma série de analistas, embora os indicadores tenham funcionado eficazmente como instrumentos de divulgação da agenda de desenvolvimento, a adoção dos ODM como um quadro normativo para as políticas nacionais, implicou em uma série de efeitos políticos e cognitivos problemáticos. Destacam-se a sua abordagem vertical e tecnocrática e a simplificação do desenvolvimento como um processo de obtenção de resultados concretos e mensuráveis (Fukuda-parr; Yamin; Greenstein, 2014).

⁴⁸ As críticas versam, entre outros, sobre a superficialidade e caráter simplista do índice para abordar questões complexas de desenvolvimento, sobre a escolha das variáveis e qualidade dos dados utilizados, sobre o método de ponderação do índice, sobre erros de medição e vieses implícitos nos dados e sobre a alta correlação do IDH com o PIB que o tornaria irrelevante para uma discussão séria sobre desenvolvimento (Kovacevic, 2010)

Do ponto de vista do campo estatístico, a despeito das diversas críticas de inconsistência técnica, os ODM representaram uma oportunidade de fortalecimento político e requalificação institucional da Comissão Estatística e da Divisão Estatística da ONU, enquanto órgãos responsáveis pelo monitoramento da agenda de desenvolvimento (ONU, 2013). Os interesses corporativistas são evidentes nas 'lições aprendidas' do grupo de experts da 'comunidade estatística' da ONU. Para os experts, a ampliação das capacidades estatísticas e o fortalecimento da parceria entre os sistemas estatísticos nacionais e internacionais despontam como pontos fortes do modelo de 'governança global'. Notadamente, o maior envolvimento da 'comunidade estatística' na agenda pós-2015 foi uma das demandas apresentadas pelo grupo:

O envolvimento precoce da comunidade estatística nacional e internacional no processo de formulação das metas e indicadores é necessário para garantir que o monitoramento esteja firmemente alicerçado em princípios estabelecidos e práticas estatísticas sólidas. O papel de autoridade e liderança desempenhado pela comunidade estatística global em questões técnicas deve ser reconhecido para garantir que sejam consideradas as melhores opções em termos de indicadores, fontes de dados relacionadas e compilação de dados. Além disso, um mecanismo interagências e de grupos de peritos – tal como foi estabelecido para o monitoramento dos ODM – é fundamental para a coordenação das atividades de monitoramento global dos novos objetivos e metas de desenvolvimento pós-2015, uma vez definidos. (ONU, 2013, p.7)

A despeito das críticas e do insucesso dos ODM como um quadro adequado para a coordenação das políticas de desenvolvimento, a ONU dobrará a aposta na 'governança global' por indicadores com a agenda 2030. A nova agenda intensifica a orientação ideológica do 'neoliberalismo pragmático' agora sob o slogan da 'sustentabilidade', incorporando a dimensão ambiental aos novos Objetivos de Desenvolvimento Sustentável (ODS). Conforme veremos no próximo capítulo, a concertação da nova agenda de desenvolvimento com ativa participação de corporações privadas resultará em um compromisso político em torno da 'revolução de dados', estratégia voltada à criação de novos mercados de dados para estatísticas oficiais, por meio da promoção de parcerias público-privadas para a 'modernização' dos INEs com fontes de big data. Antes disso porém, analisaremos nas próximas seções outros fatores que caracterizam o regime estatístico de comensuração sob o neoliberalismo, seus desdobramentos no Brasil e finalmente, como conduzem a um processo de transição para um novo regime de dataficação.

2.5 As Tecnologias de Informação e Comunicação e seus efeitos no campo estatístico

De acordo com Castells (2011), apesar das principais descobertas tecnológicas em eletrônica terem ocorrido durante a Segunda Guerra Mundial, a exemplo do computador programável e dos transistores, apenas “na década de 1970 as novas tecnologias da informação difundiram-se amplamente, acelerando seu desenvolvimento sinérgico e convergindo em um novo paradigma” (Castells, 2011, p.76). Para o autor os principais campos responsáveis por esta aceleração das TICs foram a microeletrônica, os computadores e as telecomunicações:

Esse sistema tecnológico, em que estamos totalmente imersos na aurora do século XXI, surgiu nos anos 70 [...] essas tecnologias representaram um salto qualitativo na difusão maciça da tecnologia em aplicações comerciais e civis, devido a sua acessibilidade e custo cada vez menor, com qualidade cada vez maior. Assim, o microprocessador, o principal dispositivo de difusão da microeletrônica, foi inventado em 1971 e começou a ser difundido em meados dos anos 70 [...] O primeiro comutador eletrônico industrial apareceu em 1969, e o comutador digital foi desenvolvido em meados dos anos 70 e distribuído no comércio em 1977. A fibra ótica foi produzida em escala industrial pela primeira vez pela Coming Glass, no início da década de 1970 [...] E, finalmente, mas não menos importante, foi em 1969 que a ARPA (Agência de Projetos de Pesquisa Avançada do Departamento de Defesa Norte-Americano) instalou uma nova e revolucionária rede eletrônica de comunicação que se desenvolveu durante os anos 70 e veio a se tornar a Internet (Castells, 2011, p.91).

De acordo com Paraná (2018, p.176), o desenvolvimento das Tecnologias de Informação e Comunicação, “acelera, intensifica e dá base tecno-material às dinâmicas de reestruturação e transnacionalização da produção, de financeirização e neoliberalização, que avançam com maior intensidade”. Assim, a partir dos anos 1990, no ‘frenesi’ dos interesses financeiros especulativos em torno das possibilidades comerciais da internet, as TICs foram consideravelmente impulsionadas (Srnicek, 2017).

[...] este entusiasmo com a nova indústria traduziu-se numa injeção massiva de capital nos ativos fixos da Internet [...] Concretamente, este investimento significou a instalação de milhões de quilômetros de cabos de fibra ótica e cabos submarinos, o estabelecimento de grandes avanços nos softwares e no design de redes e a realização de grandes investimentos em bases de dados e servidores (Srnicek, 2017, p.31-32)

Do ponto de vista dos Estados Nacionais, o desenvolvimento das TICs desencadeou, em um primeiro momento, um processo de descentralização dos agentes e redes de produção e distribuição de informações estatísticas. Esta dinâmica ‘policêntrica’, impulsionada pelas TICs, encontrou aderência na racionalidade política e econômica do neoliberalismo. De acordo com Desrosières (2008, p.55), o Estado sob a governamentalidade neoliberal se multiplicou em vários centros de gestão. Este processo foi acompanhado pela descentralização de ‘centros de computação’ e pela interligação de numerosos “lugares de ação, de decisão e, portanto de produção e uso da informação”.

Do ponto de vista dos sistemas estatísticos nacionais, as inovações promovidas pelas tecnologias de informação e comunicação (TICs) no final do Séc. XX são um dos principais fatores que — somados à pressão sobre os orçamentos públicos e o papel crescente do setor privado e organismos internacionais em atividades correlatas — passam a desafiar o tradicional ideal de centralização estatística dos Estados Nacionais. De acordo com Beaud e Prévost (2011, p.36), a inovação tecnológica “inverteu completamente as características do período anterior com o advento do microcomputador individual, o desenvolvimento de pacotes estatísticos fáceis de usar, o surgimento da Internet e o aumento das competências estatísticas dos usuários”. Como resultado deste processo, verifica-se uma tendência à descentralização da produção estatística.

Para Beaud e Prévost (2011, p.41), enquanto os INEs abdicam de seu papel centralizador buscam assumir um maior protagonismo na coordenação e harmonização das estatísticas oficiais, apelando à cooperação com outros agentes produtores no âmbito governamental. Impulsionados pelas TICs, os registros administrativos digitalizados ganham maior relevância na produção de estatísticas oficiais e as relações com outros produtores de dados públicos passam a se dar “no âmbito da parceria e, portanto, como resultado de um acordo entre iguais, em vez de um exercício de autoridade por parte de uma agência central”.

Do ponto de vista das práticas estatísticas dos INEs, o desenvolvimento das TICs expandiu enormemente as capacidades ao longo de todo o ciclo de produção das estatísticas oficiais, da coleta de dados à divulgação. A partir dos anos 1990 a transição da arquitetura dos sistemas de tecnologia informação — do modelo hierárquico baseado em mainframes e terminais, para modelos mais descentralizados com a adoção de microcomputadores pessoais e bancos de dados distribuídos

(UNECE, 1999) — conferiu ao corpo técnico dos INEs maior controle sobre as etapas do processo de produção e divulgação das estatísticas oficiais (concepção de questionários, coleta de dados, tabulação, publicação). A descentralização da informatização nos INEs também ampliou o acesso interno e externo a dados, metadados, parados, insumos cartográficos, softwares, entre outros importantes instrumentos de apoio às práticas estatísticas (DESA, 2007).

No que se refere à coleta e processamento dos dados a utilização da codificação assistida por computador e de leitores ópticos agilizou substancialmente a entrada de dados das operações censitárias e surveys amostrais nos anos 1990 (Dekker, 1994). Já na primeira década do Séc. XXI o aprimoramento dos protocolos de comunicação na internet e a adoção de computadores de mão, celulares e tablets revolucionou os processos de coleta de dados ao possibilitar a digitalização dos questionários e transmissão dos dados em tempo real (DESA, 2022). Finalmente, o avanço dos processos de digitalização da administração pública intensificou o uso de registros administrativos como instrumentos de apoio e fontes de dados para a produção de estatísticas oficiais pelos INEs.

Os processos de divulgação das estatísticas oficiais e a disponibilização dos dados, microdados e metadados dos censos e surveys amostrais também foram substancialmente transformados com o desenvolvimento das mídias digitais e da internet. Notadamente, a transição tecnológica contribuiu para um reposicionamento dos INEs junto aos setores da academia, mídia e sociedade em geral, na medida em que o acesso público às estatísticas oficiais e microdados foi substancialmente ampliado e facilitado (DESA, 2022).

Em resumo, o contexto de impulsionamento das TICs entre o final do Séc. XX e início do Séc. XXI, implica, por um lado, em uma maior descentralização e diversificação dos produtores de informações estatísticas, o que desafia o tradicional princípio de centralização estatística nacional que conduziu à consolidação dos INEs durante o Séc. XX (Beaud e Prévost, 2011). Há, por outro lado, um fortalecimento das capacidades e um reposicionamento dos INEs — enquanto produtores relevantes de informações públicas para o Estado, academia e sociedade em geral, e enquanto coordenadores de sistemas estatísticos nacionais — na medida em que a incorporação de novas tecnologias se somou a processos e métodos já longamente consolidados e possibilitou o desenvolvimento autônomo de soluções adaptadas às práticas estatísticas.

Na segunda década do Séc. XXI, entretanto, verifica-se uma mudança significativa nos contextos tecnológico, econômico e político globais, na medida em que os dados se tornam um elemento central da reconfiguração do regime de acumulação do capital em um novo cenário de crise do capitalismo global. Antes de adentrarmos no processo contemporâneo de dataficação, entretanto, analisaremos na próxima seção como se processaram no caso do sistema estatístico brasileiro as mudanças provocadas pela governamentalidade neoliberal.

2.6 Democratização e resiliência do IBGE às reformas neoliberais

Os efeitos da crise do modelo de acumulação fordista-keynesiano, especialmente a partir do choque do Petróleo de 1975-1978, afetaram profundamente os países periféricos da América Latina. A crise encerrou um período de expansão econômica promovida com alto grau de endividamento externo e desigualdade social, o que contribuiu para desestabilizar a sustentação de regimes autoritários na região. Assim, na década de 1980 verifica-se um período de redemocratização em vários países latino-americanos, processo que no Brasil se deu pela via da negociação e coalizão conservadora entre as elites políticas, em um contexto de crescente mobilização dos trabalhadores organizados.

Conforme vimos, os dados do IBGE, especialmente as estatísticas sociais, contribuíram para subsidiar e alimentar o debate público e a oposição à ditadura no contexto de liberalização e abertura do regime entre o final dos anos 1970 e início dos anos 1980. As estatísticas econômicas, por sua vez, passavam a revelar uma profunda crise caracterizada por dois componentes principais, por um lado, uma grave instabilidade cambial, devido à incapacidade de financiar o déficit da balança de pagamentos e pagar os serviços da dívida, e, por outro lado, uma elevada taxa de inflação pressionada pelo desequilíbrio fiscal. Neste contexto, duas questões de política econômica se impõem ao país, a estabilização da economia e a definição de um novo modelo de desenvolvimento (Prado; Leopoldi, 2018).

De acordo com Sallum (1999), o modelo desenvolvimentista amparado em um Estado interventor, que emergiu no Estado Novo e perdurou sob a ditadura militar, sustentou-se em um sistema de dominação que gradualmente se desagregou a partir de 1980. Neste sentido, dois fatores principais contribuíram para a perda de capacidade de comando do Estado sobre a economia e a sociedade: i) as

transformações econômicas internacionais, que marcam a transição do capitalismo mundial para a sua forma transnacional; e ii) a emergência de movimentos e formas de organização autônomas dos segmentos sociais com a democratização do país.

Para Sallum (1999, p.25), verifica-se neste contexto uma crise de hegemonia, que permaneceu sem resolução ao longo da década de 1980 e primeiros anos de 1990, em que os segmentos sociais oscilaram ideologicamente entre o neoliberalismo e o nacional-desenvolvimentismo. Por um lado, há um movimento do empresariado em defesa de uma orientação desestatizante e internacionalizante — desregulamentação, privatizações e abertura ao capital estrangeiro — endossada pelos meios de comunicação de massa e retransmitida para as classes médias. Por outro lado, verifica-se um contramovimento dos assalariados organizados, pessoal do Estado e movimentos sociais em defesa do ‘nacional’ e do ‘estatal’.

O embate entre as correntes neoliberais e nacionais desenvolvimentistas se expressa na Assembleia Nacional Constituinte e resulta em uma derrota parcial das organizações empresariais, incapazes de se fazer representar institucionalmente neste processo. Deste modo, a constituição de 1988 consolidou juridicamente o modelo nacional desenvolvimentista — expresso no aumento do controle do Estado sobre o mercado e em limitações ao capital estrangeiro — e assegurou uma série de conquistas democráticas aos trabalhadores (e.g. direito a sindicalização, direito à greve, regime jurídico único para os servidores públicos, ampliação de mecanismos de proteção social e trabalhista). A partir de então, a Constituição será ao mesmo tempo, um alvo de ataques das elites empresariais e uma trincheira de defesa dos movimentos sociais e sindicais (Sallum, 1999).

No IBGE, o período compreendido entre 1985 e 1994 foi marcado por intensa crise interna, caracterizada, por um lado, por tentativas de implementação de reformas administrativas e, por outro, pela forte resistência dos trabalhadores organizados às reformas. Orientada pelos princípios da NPM, a principal reforma foi implantada na gestão do economista liberal Edmar Bacha. Para Bacha, “o objetivo era fazer uma organização mais enxuta [...] Havia pessoas em excesso no IBGE, e instituímos um programa de aposentadoria voluntária, que foi muito bem sucedido. Entre a minha entrada e minha saída a redução foi significativa” (Senra; Fonseca; Milions, 2016, p.45-46). Verifica-se ao longo deste período, a demissão de funcionários, redução no número de agências municipais, cortes nas estruturas administrativas estaduais,

reformulação das diretorias e a tentativa frustrada de implantar um programa de marketing, visando a geração de receitas por meio da venda de produtos e serviços.

Sob forte oposição organizada dos trabalhadores do IBGE contra as reformas, o que incluiu a realização de uma série de greves, verifica-se uma grande rotatividade de presidentes do órgão. No intervalo de nove anos verificam-se seis diferentes gestões⁴⁹, o que contribuiu para comprometer a execução do plano estatístico e a imagem pública do órgão. A estes óbices, somou-se o adiamento do censo de 1990. Por outro lado, este período também se caracterizou por uma maior abertura do IBGE ao público externo no contexto de maior democratização da informação pública, com a criação do Centro de Documentação e Disseminação de Informações e realização da Conferência Nacional de Estatística reunindo produtores e usuários (Senra, 2009).

A crise de governamentalidade no Brasil será equacionada a partir de 1994 com o atrelamento da política econômica ao corolário neoliberal do 'Consenso de Washington'⁵⁰. De acordo com Sallum (1999), este processo se deu por meio de uma composição econômica e política que envolveu, por um lado, a estabilização da economia com o Plano Real — pautado essencialmente pelo controle da inflação por meio da sobrevalorização do câmbio, abertura aos capitais externos, elevação dos juros e ajuste fiscal com cortes de gastos públicos — e, por outro, uma aliança entre o empresariado, mídia corporativa e a classe política, que se expressou na formação de um bloco político-partidário hegemônico de centro-direita, em torno de um projeto de reconstrução do Estado pautado pelo modelo neoliberal.

O governo de Fernando Henrique Cardoso (1995-2002), que resulta desta composição, implementa o programa neoliberal tendo por principal objetivo encerrar definitivamente a 'Era Vargas' e estabelecer uma nova relação entre o Estado e o mercado, por meio da quebra “dos alicerces legais do Estado nacional-

⁴⁹ Entre 1985 e 1994 ocuparam a presidência do IBGE: Edmar Bacha Lisboa (1985-1986), Edson de Oliveira Nunes (1986-1988), Charles Curt Mueller (1988-1990), Eduardo Augusto Guimarães (1990-1992), Eurico de Andrade Neves Borba (1992-1993), Sílvio Augusto Minciott (1993-1994).

⁵⁰ O Consenso de Washington consiste em um conjunto de medidas propostas em 1989 pelo economista estadunidense John Williamson que passaram a ser adotadas como um programa por organizações como o FMI e o Banco Mundial para orientar a política econômica dos países latino-americanos. As medidas consolidam a ortodoxia neoliberal que passou a prevalecer no período, são elas: i) disciplina fiscal; ii) redirecionamento dos gastos públicos; iii) reforma tributária; iv) taxas de juros determinadas pelo mercado; v) taxas de câmbio competitivas; vi) livre comércio; vii) liberalização do investimento estrangeiro; viii) privatização de empresas estatais; ix) desregulamentação; x) segurança jurídica para direitos de propriedade privada (Williamson, 2005).

desenvolvimentista, parte dos quais fora constitucionalizado em 1988” (Sallum, 1999, p.31). Este programa é alcançado, de uma maneira geral, por meio da redução da participação do Estado nas atividades econômicas e na equiparação do tratamento conferindo às empresas nacionais e estrangeiras. A simpatia da grande maioria da população à estabilidade dos preços e à melhora temporária das condições de vida, somada a presença minoritária das forças de esquerda no Congresso Nacional, garantem uma boa margem de governabilidade para a implementação da agenda neoliberal. Do ponto de vista das reformas do Estado, verifica-se a concessão de diversos serviços públicos para a iniciativa privada, a execução de um enorme programa de privatizações, a criação de uma nova modalidade de organizações públicas não-estatais (Organizações Sociais) e a transferência do intervencionismo estatal para o campo normativo, com a criação das agências reguladoras.

A hegemonia liberal alcançada após o Plano Real se reflete em gestões relativamente estáveis no IBGE. A gestão de Simon Schwartzman (1994-1999), que correspondeu ao primeiro governo de FHC, caracterizou-se fundamentalmente por: (i) promover uma maior interface entre o IBGE e o internacionalismo estatístico — ampliando a cooperação com outros INEs e a presença na Comissão Estatística da ONU; (ii) implementar a transição tecnológica no órgão com a introdução de microcomputadores e disseminação de informações públicas pela internet e; (iii) introduzir os princípios gerenciais da qualidade no processo estatístico, com subsídios de uma missão de avaliação do Statistics Canadá. A gestão de Sérgio Besserman (1999-2003) por sua vez, notabilizou-se por aprofundar o processo de transição tecnológica com a adoção de leitores ópticos nos questionários do Censo 2000, e por ampliar consideravelmente a relação do IBGE com a mídia, profissionalizando a assessoria de imprensa, estabelecendo um calendário fixo de divulgações e criando o instrumento do ‘embargo’ que passou a anteciper aos órgãos de imprensa a divulgação das pesquisas com o compromisso de divulgação em datas pré-estabelecidas (Senra, 2009).

Cabe destacar que um conjunto de fatores políticos e administrativos, relacionados ao processo coincidente de consolidação da democratização e implementação das normas constitucionais, contribuiu para um novo ciclo de fortalecimento do papel do IBGE como órgão público de Estado à serviço da união federativa e da sociedade democrática. Destacam-se: i) os processos de descentralização administrativa que demandaram dados para diferentes níveis de

governo e inclusive para transferências de recursos com base em estatísticas populacionais; ii) a implantação de programas de combate à pobreza que passaram a exigir estatísticas mais precisas para medição de diferentes aspectos do fenômeno, o que implicou em maior complexidade dos censos e surveys amostrais; iii) a agenda de transparência pública que passou a demandar a melhoria dos processos de disseminação e acesso aos dados, microdados e metadados pelos usuários e a maior publicidade dos métodos.

É interessante observar como o fortalecimento do papel público e estatal do IBGE, neste contexto, contrastava com a agenda de reforma gerencial do Estado promovida no governo FHC. A proposta de ‘modernização’ do órgão defendida por Schwartzman (1995, p.9), por exemplo, envolvia a sua reestruturação em uma agência autônoma regida pelas “regras gerais do setor privado”, o que incluía a geração de “receitas próprias através da venda de produtos, licenciamentos e convênios com instituições governamentais e não governamentais”. Ao encaminhar a proposta de reestruturação do IBGE ao governo federal pouco antes do final de sua gestão, Schwartzman (1998, p.2) demandava a sua implantação de forma ágil: “transformações mais profundas na administração pública podem ocorrer, mas dependem de que exista uma decisão política de colocá-las em uma linha prioritária de ação, um *fast-track* que faça com que as decisões sejam tomadas com rapidez”. Em que pese a continuidade de uma série de ações legadas por Schwartzman após sua saída do órgão em 1999, a sua proposta de reestruturação não chegou a ser implementada.

A não adesão à reestruturação do IBGE em uma base ‘semiprivatizada’ se deve a alguns fatores. Em primeiro lugar, como já mencionado, deve se considerar às funções e obrigações do órgão que decorreram da implementação dos preceitos constitucionais de 1988. Em segundo lugar, houve intensa resistência dos trabalhadores do IBGE à proposta, obstáculo que foi apontado pelo próprio Schwartzman (1998, p.2) referindo-se à “pressão política do sindicato dos funcionários, que mantém na justiça uma ação popular contra a direção do Instituto e faz denúncias sucessivas ao Tribunal de Contas e a outras autoridades”. Em terceiro lugar, há que se levar em consideração que apesar da hegemonia liberal ter caracterizado a estrutura de dominação do Governo FHC, “o ideário neoliberal foi incorporado às políticas de Estado de forma atenuada e seletiva” (Sallum e Goulart, 2016, p.124), o que se deveu tanto à presença de quadros ‘liberais-

desenvolvimentistas' no governo, como às pressões de grupos sociais atingidos (tecnocracia estatal, empresários e trabalhadores).

De acordo com Sallum e Goulart (2016, p.122-123), ao longo do governo FHC “o neoliberalismo se efetivou no núcleo das políticas macroeconômicas, mas não conseguiu se converter em diretriz orientadora da totalidade das políticas de Estado”. Nestas condições, o governo oscilou entre o fundamentalismo neoliberal e o liberal-desenvolvimentismo, tendo esta segunda corrente contribuído para adotar políticas que, por exemplo, deram alguma proteção e estímulo às empresas nacionais. Após a crise cambial do real em 1999, entretanto, dobrou-se a aposta neoliberal e os efeitos nocivos da política macroeconômica se evidenciaram. Em síntese, esta política resultou em “crescimento medíocre, privilegiamento dos rendimentos financeiros, desnacionalização da economia, intercâmbio cronicamente deficitário com o exterior e, por fim, alto desemprego” (Salum, 1999, p.40).

As dificuldades econômicas do país na entrada do Séc. XXI e o aprofundamento de sua inserção dependente na economia mundial colocaram em xeque a hegemonia liberal. Neste contexto, o cenário sociopolítico se moveu “em direção a uma nova coalizão que pudesse empunhar de novo as bandeiras do desenvolvimento e da defesa dos interesses nacionais” (Sallum e Goulart, 2016, p.126). Esta coalização, que reuniu segmentos progressistas da classe média, parcelas de grupos empresariais e segmentos organizados dos trabalhadores, conduziu a eleição de Lula em 2002 e ao governo do Partido dos Trabalhadores.

Os governos de Lula (2003-2010) caracterizaram-se pela contenção do ímpeto reformista neoliberal, ao mesmo tempo em que mantiveram a orientação liberal na macropolítica econômica, “à qual foi agregada um maior impulso desenvolvimentista e forte ação redistributiva” (Sallum e Goulart, 2016, p.131). De uma maneira geral, a estabilidade e o crescimento da economia em um contexto favorável até 2008, possibilitaram uma convergência entre a política macroeconômica liberal e às posições originais do PT — maior presença do Estado, proteção da economia nacional, melhor distribuição de renda e proteção aos trabalhadores.

As políticas de proteção social e democratização do acesso aos bens e serviços públicos (e.g. transferência de renda, aumento do salário mínimo, acesso à educação superior) promoveram uma maior distribuição de renda e ascensão social, enquanto políticas afinadas com o liberal desenvolvimentismo (investimentos públicos na infraestrutura, financiamento público de empresas nacionais, parcerias público-

privadas) ganharam maior relevância. Neste sentido, a estabilidade política deste período foi alcançada por meio de um pacto “que permitiu, de um lado, preservar e mesmo ampliar os ganhos do capital e, de outro, tornar a democracia mais substantiva mediante um impulso distributivo”. (Sallum e Goulart, 2016, p.132).

No que se refere ao campo estatístico oficial, inaugura-se sob os governos do PT a indicação de quadros técnicos do IBGE para a presidência da instituição. Sob a presidência do ibgeano Eduardo Pereira Nunes (2003-2011) há um notável fortalecimento institucional do órgão, que se reflete: (i) na realização de concursos públicos, na garantia de recursos orçamentários para a realização da contagem populacional (2007), censo agropecuário (2007) e censo demográfico (2010); (ii) em investimentos expressivos no desenvolvimento de novas tecnologias de coleta e processamento dos dados (e.g.: computadores de mão, georreferenciamento dos dados, uso de registros administrativos); (iii) na ampliação do papel de coordenação do sistema estatístico e geocientífico nacional pelo IBGE com a reestruturação da CONCAR (Comissão Nacional de Cartografia), criação da INDE (Infraestrutura Nacional de Dados Espaciais), fortalecimento da CONCLA (Comissão Nacional de Classificação), instalação do Comitê de Estatísticas Oficiais (CES) e ampliação da atuação do IBGE junto aos órgãos estatísticos estaduais no Sistema de Contas Regionais.

Concluimos que a resiliência do IBGE às reformas neoliberais, se deveu em grande medida à resistência organizada de seus trabalhadores, ao novo posicionamento do órgão no contexto democrático e à atenuação da ortodoxia neoliberal pela prevalência de correntes desenvolvimentistas e grupos de pressão dentro e fora do governo. Alguns efeitos do regime de comensuração sobre o programa estatístico podem ser percebidos na medida em que se amplia a utilização de registros administrativos e a produção e coordenação de indicadores sociais em diferentes esferas do governo. O IBGE, entretanto, seguiu fundamentalmente estruturado sobre as práticas do regime de precisão (censos, inquéritos amostrais e contabilidade nacional), mantendo seu status de organização pública e estatal. Tais efeitos, entretanto, serão identificados na medida em que se verifica uma maior adesão à política internacional de indicadores de desenvolvimento sustentável, conforme veremos no capítulo quatro. Esta dilatação cronológica faz com que esta transição seja combinada com a emergência dos processos de dataficação que passamos a analisar na próxima seção.

2.7 Dataficação, governamentalidade algorítmica e capitalismo de plataforma

A significativa transferência de capitais financeiros nos anos 1990 para o setor de tecnologia, preparou o terreno para o desenvolvimento da 'economia digital', por meio do impulsionamento das TICs, da criação da infraestrutura necessária à expansão global da internet e desenvolvimento da web 2.0. Este processo implicou em uma crise especulativa, conhecida como a bolha das 'ponto com'. A crise resultou na venda, fusão, redução ou simplesmente desaparecimento de muitas empresas da nova economia digital. Verifica-se a partir daí uma expressiva concentração de capitais no setor de tecnologia da informação. Neste contexto, de acordo com Castells (2009, p.105), "o apetite corporativo pela comercialização da internet cresceu exponencialmente":

Quando a expansão da banda larga e o surgimento da Web 2.0 abriram novas oportunidades de lucro na primeira década do século XXI, um novo conjunto de políticas regulatórias foi introduzido, visando à apropriação privada não da própria Internet, mas da infraestrutura de rede da qual a Internet depende (Castells, 2009, p.105, tradução minha).

Um dos elementos centrais da nova visão empresarial que passou a prevalecer na economia da internet foi o conceito de uma nova 'era dirigida pelos dados'. De acordo com Musser e O'Relley (2007), que cunharam o termo Web 2.0, a competição no novo ambiente de negócios da internet passou a demandar das empresas o desenvolvimento de estratégias de negócios baseadas nos dados:

Na medida em que o mercado se afasta dos aplicativos de desktop para indivíduos e passa a adotar um modelo de serviços online compartilhados, está se tornando cada vez menos sobre funções isoladas, como processamento de texto, e mais sobre dados, incluindo catálogos de varejo, bancos de dados de pesquisa, bancos de dados de leilão, bases de conhecimento, enciclopédias, dados de mapeamento e localização, coleções de mídia e assim por diante. Para muitos serviços online, o valor são os dados [...] Novas vantagens competitivas estão nas camadas superior e inferior dos pacotes emergente da Internet. Na camada superior dos software, uma vantagem de cercamento agora vem dos efeitos de rede e do uso de software como serviço. Na camada inferior, os dados e o controle dos dados fornecem a vantagem competitiva (Musser e O'Relley, 2007, p.19).

Na medida em que as TICs passaram a possibilitar a conversão de diversos aspectos da vida humana, do mundo físico, da economia, da sociedade, dos bens e serviços em artefatos digitais, verifica-se uma transformação da economia e

sociedade por meio da utilização destes artefatos, processo que alguns autores intitulam de digitalização (Gradillas; Thomas, 2023). A digitalização caracteriza-se fundamentalmente pela automação de tarefas por meio de processos computacionais; pela ampliação da conectividade entre pessoas, empresas e máquinas; e pela maior convergência entre diferentes setores e áreas de conhecimento e produção.

Em linha com Paraná (2019, 2024), compreendemos criticamente a digitalização como um processo multifacetado que engloba uma série de transformações estruturais do capitalismo contemporâneo sob o regime neoliberal, sobredeterminadas principalmente pela tecnologia e financeirização. Conjugados, estes processos implicam em tendências globais como “a aceleração social ligada à compressão dos fluxos espaço-temporais [e] a crescente importância estratégica da propriedade sobre a produção [...] facilitando a utilização da descentralização técnica para a concentração econômica e política” (Paraná, 2024, p.20).

Um dos resultados do processo de digitalização foi a emergência de novas e massivas fontes de dados (Srnicek, 2017). Dados, microdados, metadados e nanodados passaram a ser gerados como produtos ou subprodutos de interações digitais entre humanos e máquinas (internet, redes sociais, comércio eletrônico, telefonia celular, scanners) e entre máquinas e máquinas (sensores, internet das coisas) em volumes, variedades e velocidades sem precedentes, constituindo o que se convencionou chamar de big data (Kitchin, 2014)⁵¹. Este processo é acompanhado

⁵¹ O termo ‘big data’ que se popularizou na segunda década do Séc. XXI refere-se em geral a um novo momento de geração de conjuntos massivos de dados que possuem ao mesmo tempo três atributos fundamentais, que ficaram conhecidos como os 3 Vs: volume, velocidade e variedade, a partir de um artigo publicado por um analista da empresa Meta em 2001 sobre novas tendências para a gestão de dados (Laney, 2001). Os 3V se tornaram a definição mais usual para os novos conjuntos de dados, de acordo com Kitchin (2014, p.99), “Os Big Data são: grandes em volume, consistindo em terabytes ou petabytes de dados; altos em velocidade, sendo criados em, ou quase, tempo real; diversos em variedade, sendo estruturados e não estruturados por natureza, e frequentemente referenciados temporal e espacialmente”. Os três atributos (volume, velocidade e variedade) representavam uma limitação tradicional dos bancos de dados, que em geral não conseguiam desempenhar, ao mesmo tempo, as três propriedades. A sua conjugação simultânea, tem se tornado possível graças ao aumento do poder computacional, novos designs de banco de dados e processamento distribuído, o que tem permitido novas formas de análise e visualização dos dados (Kitchin, 2014). Com o tempo novos “Vs” foram sendo adicionados à definição de Big Data. Um quarto V de ‘veracidade’ foi adicionado por pesquisadores da IBM para lidar com o problema da incerteza associada às novas fontes de dados, de acordo com Schroeck et al. (2012 et. al, p.5): “obter dados de alta qualidade é um desafio no big data, mesmo os melhores métodos de limpeza não podem remover a imprevisibilidade inerente a alguns dados”. Demchenko et. al. (2013, p.50) introduziram um quinto V para ‘valor’ definido pelo “valor agregado que os dados podem trazer ao processo, atividade ou análise/hipótese preditiva”. Para Gandomi e Haider (2014) que também utilizam a noção de valor, “os big data recebidos na forma original costumam ter um valor baixo em relação ao seu volume”. Khan et al. (2018) listaram mais cinco ‘Vs’ à lista de definições de Big Data: ‘volatilidade’, que diz respeito ao tempo de vida dos dados, ou seja por quanto tempo serão gerados e estarão disponíveis e armazenados; ‘variabilidade’, que tem a

pela diferenciação de uma nova ‘indústria’ voltada à extração, cercamento e comoditização dos dados, utilizados para a otimização de processos de produção, desenvolvimento de estratégias de mercado baseadas nas preferências dos consumidores, controle dos trabalhadores, desenvolvimento de novos produtos e serviços, direcionamento de publicidade, etc. (Srnicek, 2017).

O processo contemporâneo de dataficação que resulta da digitalização, pode, portanto, ser compreendido como um duplo processo de conversão da economia, sociedade, bens e serviços para o formato de dados digitais, e de geração de diferentes formas de valor a partir dos dados (Couldry; Mejias, 2019). Por um lado, um processo sociotécnico, na medida em que novas modalidades de processamento, quantificação e análise automatizada dos dados transmutam a cognição e ação humanas em uma forma analisável (Mayer-Schönberger; Cukier, 2013). Por outro lado, um novo regime político e econômico impulsionado pela lógica da acumulação do capital e caracterizado pela extração, cercamento e conversão dos dados em mercadorias, ativos e capitais (Sadowski, 2019).

A transição para a ‘nova economia orientada aos dados’ se dá no contexto de uma nova crise do capitalismo global, caracterizada, entre outros pela baixa produtividade do setor industrial, financeirização da economia e por um novo modelo de crescimento orientado pelo preço dos ativos financeiros (ações e imóveis) que conduziu ao choque imobiliário de 2007/2008 e à crise do Euro em 2011 (Anglietta, 2015; Chesnais, 2016). Em meio às crises, a narrativa dominante nos países de capitalismo avançado volta-se para as promessas e oportunidades representadas pela ‘ascensão da tecnologia’ e ‘indústria 4.0’. Neste contexto, as Big Tech (Google, Apple, Facebook, Amazon, Microsoft), grandes empresas globais de tecnologia, despontam como indutoras dos mercados, a partir da implantação de um novo e

ver com à inconsistências nos fluxos de dados; o que está vinculado à mudanças no ambiente de negócios em que são produzidos; ‘viscosidade’, que se refere ao grau de correlação e interdependência entre os dados, na medida em que mudanças em um tipo de dado podem provocar efeitos em outros conjuntos de dados; ‘viabilidade’, associada à identificação de atributos e fatores que satisfaçam as necessidades preditivas esperadas; e ‘validade’ da base de dados para as aplicações a que se destina. Uma avaliação dos dez vetores propostos por Kahn et. al. (2018) para caracterizar os big data revela possíveis sobreposições de algumas categorias, como por exemplo: variedade e variabilidade, validade e veracidade, viabilidade e valor. Neste sentido alguns analistas têm utilizado classificações intermediárias. MacFeely (2019, p.27) destaca que a inclusão de outros vetores, além de volume, velocidade e variedade, para a caracterização das novas fontes de dados é relevante para expressar uma perspectiva analítica mais equilibrada, que reflita alguns dos “riscos negativos” associados ao uso do Big Data, especialmente quando passam a ser abordadas como fontes potenciais para estatísticas oficiais. Neste sentido o autor sugere uma classificação com 6 categorias, assumindo além dos tradicionais 3V, as dimensões de veracidade, volatilidade e valor.

poderoso modelo de negócios: as plataformas digitais (Srnicek, 2017). Este modelo se baseia em três pilares fundamentais: **i)** desintermediação: eliminação de intermediários tradicionais e conexão direta entre consumidores e produtos/serviços; **ii)** capitalização dos dados: extração, cercamento e conversão dos dados em mercadorias, ativos e capitais; e **iii)** governança algorítmica: uso dos algoritmos para influenciar processos de tomada de decisão. No contexto das sucessivas crises do capitalismo em sua fase financeirizada, tudo parece funcionar como se a plataformização emergisse como um novo regime de acumulação do capital, caracterizado pela centralização de poder e extração de valor e rendas por meio de plataformas digitais, controladas por corporações privadas (Srnicek, 2017, 2021; Sadowski, 2019, 2020).

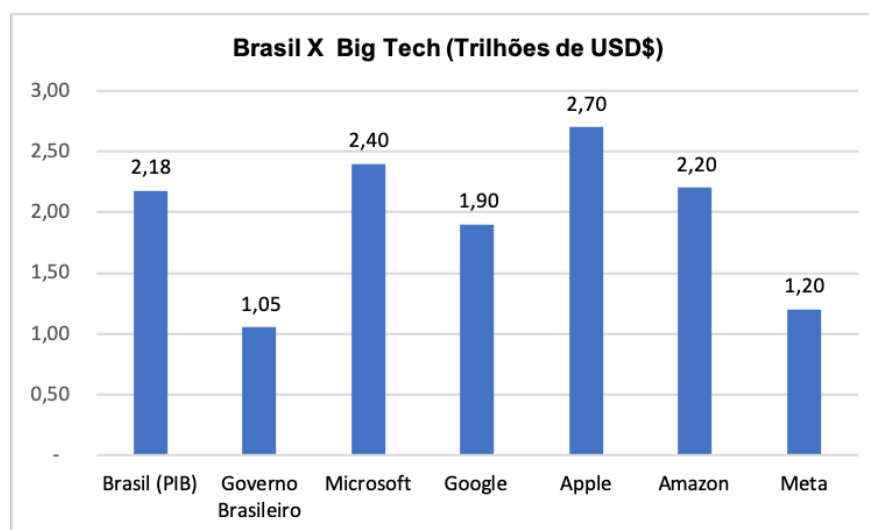
Para Srnicek (2021) e Sadowski (2020), o regime de propriedade das plataformas (intelectual e material) lhes possibilita a apropriação de valor produzido em outros espaços da economia global. O capitalismo de plataforma (Srnicek, 2017) pode, portanto, ser compreendido como um processo de atualização e expansão do rentismo e da monopolização no capitalismo contemporâneo, por meio das tecnologias digitais. Este processo passa a englobar a extração de rendas de conhecimento (Rotta e Teixeira, 2018; Rikap, 2021) e uma tendência de plataformização das finanças (Hendrikse et al., 2024) com a crescente interligação entre as grandes empresas globais de tecnologia, Big Techs e o setor financeiro digitalizado (Paraná, 2019, 2024).

A convergência entre plataformização e financeirização aponta para uma concentração sem precedentes de capitais pelas Big Tech, processo que ganha ainda maior impulso com o desenvolvimento da Indústria de Inteligência Artificial (Dyer-Witheyford et al., 2019). Uma série de analistas tem apontado para aspectos disfuncionais da nova economia orientada aos dados, em especial às suas características monopolísticas (Zuboff, 2019; Hendrikse et al. 2021; Birch; Bronson, 2022; Rikap, 2021; Rikap; Lundvall, 2022). Na medida em que estas corporações passam a deter cada vez mais influência na determinação da forma como o valor é capturado no sistema capitalista, ampliam o seu poder econômico e político. A expansão da dominação política das Big Tech se reflete em sua penetração em campos tradicionalmente públicos e estatais e em estruturas nacionais, supranacionais e globais de governo em que passam a exercer poderes análogos aos Estados (Morozov, 2020).

Para Couldry e Mejias (2019) as assimetrias econômicas e políticas que caracterizam a nova economia de dados implicam na emergência de novas formas de poder geopolítico, nas quais os dados são utilizados por empresas tecnológicas e governos do Norte Global para manter poder e influência sobre os países do Sul Global, configurando formas contemporâneas de 'colonialismo de dados'. Um dos aspectos deste processo é o avanço das Big Tech em vários setores tradicionalmente estatais, o que resulta numa extração ainda maior de dados dos usuários, aumentando o seu poder a nível global. Conforme veremos na parte empírica deste trabalho, no setor estatístico estatal este avanço vem ocorrendo por meio da agenda do 'big data', o que tem envolvido o incentivo à parcerias público-privadas e a transferência da lógica técnica, política e mercantil da plataformização para o campo estatístico.

No contexto da economia de dados e do capitalismo de plataforma, a acumulação de capital informacional pelos Estados Nacionais passa a ser desafiada pelas Big Tech. As grandes empresas globais de tecnologia detêm hoje maiores bases de dados e maiores capacidades computacionais do que diversos Estados Nacionais e seus tradicionais campos estatísticos (UNCTAD, 2021). Para efeitos de visualização destas assimetrias, podemos observar no gráfico a seguir em valores de 2023 (trilhões de dólares) a comparação entre o PIB Brasileiro, o orçamento do Governo Federal do Brasil e o valor de mercado das Big Techs.

Gráfico 2 - O Brasil e as Big Tech em 'Valores de Mercado'



Fonte: do autor. PIB em valores nominais (IBGE), Orçamento do Governo Federal do Brasil (Portal da Transparência), Valor de mercado das Big Tech em 2023 (Guerato, 2023).

Os processos de digitalização, dataficação e plataformização, são acompanhados pela crescente utilização de algoritmos para a extração de conhecimento dos dados. Algoritmos são procedimentos codificados para transformar dados em informações e resultados desejados, com base em cálculos especificados⁵². Os computadores são essencialmente máquinas algorítmicas. Na medida em que as ferramentas computacionais se tornam o principal meio de expressão e interação social e econômica, o conhecimento e as práticas humanas são cada vez mais submetidos à lógica e às normas algorítmicas (Gillespie, 2013). A crescente automatização de tarefas e processos de tomada de decisão baseados em regras predefinidas por algoritmos, conduz à emergência de novas dinâmicas de

⁵² Algoritmos são geralmente definidos como uma sequência finita de instruções passo a passo para a execução de uma tarefa ou processo. Sabe-se que a palavra “algoritmo” deriva da pronúncia latina *algorismi* em referência ao matemático árabe do Séc. IX Al Khwarizmi (Dantzig, 2005), cujos escritos constam como o primeiro trabalho de matemática aritmética - a descrever os métodos de adição, subtração, multiplicação e divisão utilizando o sistema numérico indo-arábico - introduzido na Europa Ocidental no Séc. XII (Miyazaki, 2012). Apesar de ser um conceito central para a matemática e para a ciência da computação, a noção de algoritmo pode referir-se a diferentes tipos de atividades. Tanto operações cotidianas como procedimentos mais complexos podem ser compreendidos como processos algorítmicos, desde que mediados por comandos descritos na forma de sequências de instruções. Do ponto de vista da cibernética, campo de estudos desenvolvido por Norbert Wiener (1894-1964), podemos compreender os algoritmos como elementos da teoria das mensagens em sua abordagem ao problema da comunicação e controle. Em meados do Séc. XX, com o desenvolvimento da ciência da computação e das primeiras linguagens de programação de alto nível, como o Algol, abreviação para linguagem algorítmica, o computador passou a assumir o papel de máquina processadora de algoritmos (Miyazaki, 2012). Expresso na forma de um programa (*software*), o algoritmo passou a ser compreendido como um conjunto de etapas definidas e estruturadas que especificam instruções e operações de processamento de dados (*input*) executadas pelo processador (*hardware*) para a produção de um resultado (*output*). Um elemento fundamental no estudo dos algoritmos são as bases de dados. Estas representam as entradas (*inputs*) que os alimentam e que possibilitam o processamento. Sem bases de dados em que possam ser pareados, os algoritmos são como “máquinas inertes e sem sentido” (Gillespie, 2013, p.3). Se não há entrada de dados, não há o que processar, nem resultados (*outputs*) para gerar, muito menos retroalimentação e aprendizagem. Para Manovich (1999), estruturas de dados e algoritmos tem uma relação simbiótica e juntas compõem “as duas metades da ontologia do mundo de acordo com um computador” (Manovich, 1999, p.83). A partir da segunda metade do Séc. XX, os desenvolvimentos na tecnologia de processamento dos computadores eletrônicos, nos bancos de dados e nas linguagens de programação passaram a “permitir a realização de tarefas extensas e complexas que seriam praticamente impossíveis à mão ou com máquina analógica” (Kitchin, 2017, p.1). Este processo resultou na abertura de novas aplicações para os algoritmos, em áreas antes consideradas inviáveis. No Séc. XXI os avanços nas tecnologias de informação e comunicação, em especial a partir da consolidação da infraestrutura global da rede mundial de computadores, resultaram na ampliação da digitalização de áreas cada vez mais diversas e extensas da vida social e econômica, o que gerou um volume crescente de dados. Na internet, “cada login, cada visualização de página, cada clique deixa um traço digital. Transformar esses rastros em bancos de dados envolve uma gama complexa de práticas de informação” (Gillespie, 2013, p.2, tradução minha). Conjuntos massivos de dados gerados por meio de interações na internet e em redes sociais consistem principalmente de dados não estruturados, cujo crescimento em volume tem superado em taxas muito superiores os dados estruturados. Isto tem requerido e impulsionado o desenvolvimento de novos modelos de bancos de dados e novos algoritmos baseados em técnicas de aprendizado de máquina, envolvidas na classificação, análise e armazenamento dos dados (Kitchin, 2014).

dominação, que alguns autores têm chamado de poder algorítmico, governança algorítmica, governamentalidade algorítmica ou, simplesmente, algoritmização (Beer, 2009; Musiani, 2013; Ruvroy; Berns, 2013, Kitchin, 2016). A algoritmização tem produzido efeitos disruptivos sobre o funcionamento de diversos processos e sistemas (Steiner, 2012; Pasquale, 2015; Lehdonvirta, 2016), especialmente a partir do desenvolvimento do modelo de plataformização (Gillespie, 2010; Srnicek, 2017, 2020; Sadowski, 2019, 2020; Duffy et al., 2019, Morozov, 2020).

Sob a liderança das grandes corporações de dados, verifica-se a emergência de uma variedade de agentes econômicos de grande, médio e pequeno porte que atuam subsidiariamente às Big Techs na construção de novos mercados de dados: empresas de tecnologia, empresas de telefonia, *startups*, laboratórios de pesquisas, consultorias e think-tanks (Hayashi et al.2021, Richter e Slowinski 2019, Liang et al. 2018, Schomm et al. 2013, Stahl et al. 2014). Uma característica comum destes agentes é o direcionamento estratégico de seus modelos de negócios para a extração e cercamento dos dados (Birch; Cochrane; Ward, 2021, Grabher, 2020, Wu e Taneja, 2021), com vista a transformá-los em mercadorias digitais (Rotta; Paraná, 2022), visando a extração de rendas de conhecimento (Ciuriak 2020, Rota; Teixeira, 2018, Foley 2013).

Em face da relevância que os métodos algorítmicos assumem nas práticas de dataficação, adotamos neste trabalho a denominação de ‘campo algorítmico’ para caracterizar o espaço social de relações objetivas entre os agentes que atuam na economia de dados. O campo algorítmico compreende um conjunto de práticas computacionais complexas de extração, classificação, codificação, programação, anonimização, automação, armazenamento, distribuição, compartilhamento e análise de dados. Seus agentes têm adotado a ciência de dados — híbrido que combina a ciência estatística e a ciência da computação — como um novo quadro de conhecimentos que articula sua linguagem e prática comuns⁵³.

⁵³ A literatura técnica (Aggarwal, 2015; Brammer, 2016; Skiena, 2017) concebe a ciência de dados como um campo interdisciplinar emergente, que gradualmente vem delimitando suas fronteiras na intersecção entre a estatística e a ciência da computação e abrindo-se para um vasto campo de aplicações. Uma das primeiras citações ao termo “cientista de dados” foi realizada no âmbito do *National Science Board (NSB)* dos Estados Unidos em relatório da *National Science Foundation (NSF)* de 2005, que propõe um conjunto de políticas e estratégias para o aprimoramento de acervos de dados digitais, e identifica entre as prioridades, o apoio a este grupo emergente de pesquisadores. A necessidade de gestação de uma nova geração de cientistas para lidar com os desafios da gestão e análise de quantidades crescentes de dados também foi objeto de considerações do *National Research Council* da *National Academy of Sciences* por meio do *Comitê on the Analysis of Massive Data*,

De acordo com Ruvroy e Berns (2013), na medida em que as novas modalidades informacionais que se originam das práticas de dataficação são desenraizadas das tradicionais convenções de equivalências que conformaram o espaço normativo e cognitivo da informação estatística pública (Desrosières, 1998), verifica-se a emergência de uma nova governamentalidade algorítmica. Este termo é utilizado por Ruvroy e Berns (2013, p.10) em referência a um novo tipo de racionalidade “(a)normativa ou (a)política fundada na coleta, agregação e análise automatizada de big data, de modo a modelar, antecipar e preventivamente afetar possíveis comportamentos”. Os autores apontam para a ação das corporações privadas como principais poderes gerativos da governamentalidade algorítmica. Para Ruvroy e Berns (2013, p.5), diferentemente da informação estatística, a dataficação não gera um espaço público, pelo contrário, “sob o disfarce de ‘personalizar’ informações, serviços e ofertas de produtos, na era da governamentalidade algorítmica estamos assistindo a uma colonização dos espaços públicos por uma esfera privada hipertrofiada”.

Neste novo contexto de crise do modo de acumulação financeirizada do capital e emergência do capitalismo de plataforma e governamentalidade algorítmica, identificamos indícios de mudanças nas normas, práticas e estrutura do regime estatístico. Conforme vimos, as transformações do campo estatístico ao longo do período neoliberal, por um lado, abrem flancos que o tornaram mais permeável ao mercado, mais responsivo às inovações tecnológicas e mais susceptível à participação de atores privados e transnacionais. Por outro lado, a despeito destas mudanças, os INEs se mostraram resilientes e capazes de se reposicionar neste contexto, preservando ao mesmo tempo sua autonomia relativa e o caráter público e

constituído em 2010. O relatório do comitê identificou a necessidade de novas abordagens para a ‘era do big data’ e defendeu o caráter interdisciplinar da nova geração de cientista de dados, destacando a ‘combinação íntima’ da estatística e da ciência da computação. Em 2012, o Big Data Research and Development Initiative, lançado pela administração Obama (2009-2017), financiou diversos programas federais para o desenvolvimento da pesquisa e infraestrutura para big data, principalmente nas áreas de defesa e saúde. A iniciativa Obama aportou considerável apoio financeiro na Defense Advanced Research Projects Agency (DARPA), a qual por sua vez apoiou a constituição de laboratórios e centros de ciência de dados em universidades norte americanas em parceria com Big Techs e outras empresas do Vale do Silício. A Universidade da Califórnia em Berkeley (UC Berkeley) foi uma das primeiras universidades estadunidenses a agrupar os departamentos de ciência da computação e ciência estatística a partir de 2018. Em fevereiro de 2020, a nova divisão comemorou o recebimento de recursos privados da ordem de US\$252 milhões para implantação de um Data Hub, o valor consta como a maior doação individual da história da UC Berkeley, sem dúvida uma demonstração de força e dos interesses comerciais envolvidos em torno da nova ciência de dados.

nacional das estatísticas oficiais. Entretanto, na medida em que as fontes de big data são introduzidas no campo estatístico e interfaces são estabelecidas com o campo algorítmico, verifica-se a emergência de novas disputas que caracterizamos como indícios de transição para um novo regime estatístico-algorítmico de dataficação. Estas disputas decorrem de dois fatores fundamentais.

Em primeiro lugar, há um conjunto de mudanças na economia política dos dados que afetam as normas do campo estatístico. Com o advento dos big data e sua introdução no processo estatístico, há um conflito entre a lógica pública do campo estatístico e a lógica de mercado do campo algorítmico. Para o primeiro, as estatísticas oficiais e os dados para fins estatísticos são bens públicos. Para o segundo, os dados são mercadorias comercializáveis e ativos capitalizáveis em mercados de dados. Neste contexto, o campo estatístico passa a ser fortemente assediado pelos interesses de agentes do mercado. Conforme veremos, isso implica na emergência de diferentes estratégias no interior do campo estatístico, caracterizando um duplo movimento. Verifica-se, por um lado, cooperação com os agentes de mercado e criação de estruturas estatísticas supranacionais público-privadas, e, por outro lado, contramovimentos protetivos que visam assegurar a autonomia relativa do campo e o caráter público e nacional das estatísticas oficiais.

Em segundo lugar, os cercamentos de dados por empresas privadas e os novos processos de computação em nuvem provocam mudanças na estrutura do campo estatístico. O tradicional modelo de centralização de dados em data centers dos INEs para a produção de estatísticas oficiais, que é intensificado com a expansão das TICs e utilização de registros administrativos, é desafiado pelos cercamentos privados de dados. Neste contexto, verifica-se a emergência de um novo modelo de acesso descentralizado e monetizado aos dados, defendido por Big Techs e outras corporações privadas. Neste novo modelo em construção, baseado na computação distribuída em nuvem e em tecnologias de ampliação de privacidade, os dados permaneceriam guardados pelas empresas e seriam acessados pelos INEs por meio de algoritmos de consulta, que gerariam dados agregados e anonimizados.

Em terceiro lugar, a introdução dos big data nos INEs é acompanhada por um obstáculo epistemológico que afeta as práticas do campo estatístico. Isto se dá na medida em que a estatística frequentista e a lógica dedutiva que tradicionalmente subsidiaram os métodos de produção e análise das estatísticas oficiais, passam a ser confrontadas pela estatística bayesiana e lógica indutiva, utilizadas nos novos

modelos de análise de big data. Neste contexto, os estímulos à gestação de uma nova facção de cientistas de dados no interior dos INEs, versada nos novos métodos e tecnologias e moldada pelos valores da inovação e empreendedorismo, típicos da racionalidade neoliberal, implicam em disputas internas que evidenciam mudanças nos capitais que estão em jogo no campo. Estas mudanças afetam o habitus dos estatísticos de Estado e induzem contramovimentos reativos e protetivos.

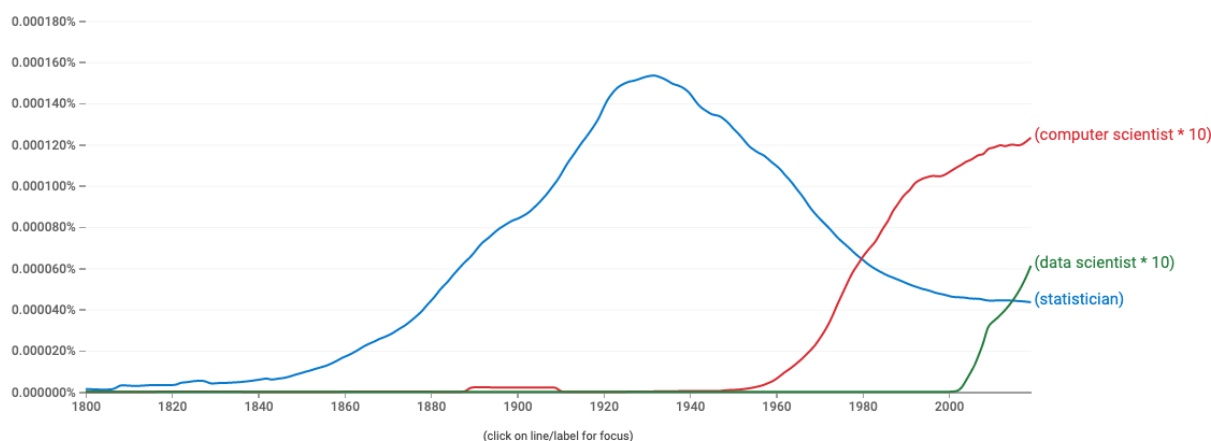
Um olhar panorâmico sobre o desenvolvimento da ciência estatística e ciência da computação, juntamente com os indícios levantados, ajudar a compreender, de um ponto de vista histórico, algumas controvérsias na interseção destes campos. Tanto a estatística quanto a computação são ciências que provocaram importantes rupturas e construíram novos objetos e agregados que transformaram as formas de conhecimento e ação no mundo social. Conforme vimos, a emergência e confluência da estatística cognitiva e administrativa, entre a segunda metade do século XIX e meados do Séc. XX, foram acompanhadas por profundas mudanças na economia, sociedade e modo de ação do Estado. A ciência da computação que por sua vez, se origina, entre outros, da confluência entre o pensamento estatístico e probabilístico com a física, a partir do desenvolvimento da cibernética (Wiener, 2000), vem desde meados do século XX ampliando as possibilidades e capacidades de quantificação e acumulação de informações em níveis inimagináveis, e em escala e velocidades crescentes.

O gráfico ⁵⁴ a seguir ajuda a ilustrar visualmente um panorama histórico destas ciências e perfis profissionais. Se a era dos estatísticos que iniciou no século XIX, teve seu pico por volta dos anos 1940, a era dos cientistas da computação que iniciou por volta dos anos 1950 parece viver hoje o seu apogeu. A ciência de dados se constitui como um híbrido entre a ciência estatística e a ciência da computação. Este novo agrupamento implica tanto em inovações epistemológicas, como em novos agenciamentos de atores privados e públicos na conformação de um novo campo de práticas. Neste sentido, os cientistas de dados que começam a despontar no século

⁵⁴ No gráfico as curvas representam a frequência de citações em fontes impressas dos termos *statistician*, *computer scientist* e *data scientist* entre 1800 e 2019. Os termos *data scientist* e *computer scientist* foram multiplicados por uma constante (10) para melhor visualização junto ao termo *statistician* que apresenta maior escala de frequência de citações. Sobre o uso do *Google Ngrams* para o estudo de narrativas, ver o artigo de Shiller (2017), em que o autor apresenta o conceito de economia narrativa e aplica um modelo clássico da epidemiologia (SIR) para analisar “epidemias narrativas” que seguem padrão semelhante à distribuição normal.

XXI, são um novo agrupamento que pode se apresentar tanto como uma oportunidade para os defensores dos processos de modernização das estatísticas oficiais, como um risco para grupos consolidados como os estatísticos de Estado.

Gráfico 3 - Estatísticos, Cientistas da Computação e Cientistas de Dados



Fonte: do autor, gerado com a ferramenta *Google Ngram*.

Nas próximas seções analisamos as disputas e obstáculos (ontológicos e epistemológicos) que caracterizam o atual processo de transição das estatísticas oficiais para um novo regime estatístico-algorítmico de dataficação.

2.8 A dataficação e as novas 'pressões' sobre o campo estatístico

No início da segunda década do Séc. XXI, verificam-se uma série de reações no campo estatístico à emergência de novas fontes de dados e agentes produtores de novas modalidades de informações estatísticas. Estas reações — que podem ser observadas em uma profusão de artigos, seminários, conferências, e grupos de trabalho sobre estes temas⁵⁵ — confluem para a articulação de um discurso em defesa

⁵⁵ Em 2010 a Conferência dos Estatísticos Europeus institui um Grupo de Alto Nível para o Desenvolvimento Estratégico de uma Arquitetura de Negócios em Estatística (HLG-BAS) coordenado pelo Netherlands Statistics e formado por 10 representantes de INEs europeus. Em 2012 este grupo promoveu um seminário de alto nível em St. Petersburg que encaminhou a elaboração de um documento estratégico sobre big data para estatísticas oficiais voltado aos chefes e gestores seniores dos INEs (UNECE, 2013). A partir de 2012 o grupo passou a denominar-se Grupo de Alto Nível para a Modernização das Estatísticas Oficiais (HLG-MOS). Os estatísticos deste grupo produziram uma série de artigos sobre os potenciais e desafios representados pelo big data e ciência de dados para as estatísticas oficiais (Strujis e Daas, 2013; Daas e Loo, 2013; Strujis et al. 2014; Daas et al. 2015). Em abril de 2013, a Comissão Econômica para a Europa (UNECE), o Serviço de Estatística da Comissão

da ‘modernização das estatísticas oficiais’ com o uso de fontes de big data. O discurso modernizador se apresenta, por um lado, como uma resposta ao que é percebido pelo campo estatístico como um novo conjunto de ‘pressões’ e ‘ameaças’ que poderiam vir a comprometer o papel, a relevância e mesmo a sobrevivência das Instituições Nacionais de Estatística. Por outro lado, a estratégia discursiva da modernização é utilizada para a mobilização do campo em torno de um conjunto de ações que visam a aproximação com o setor privado e campo algorítmico.

Em um artigo endereçado ao Sistema Estatístico Europeu (ESS), em que tratava das perspectivas futuras das estatísticas oficiais, o diretor geral do Eurostat Walter Radermacher (2009) lançou o seguinte questionamento aos seus pares: “Serão o Eurostat e o ESS capazes de cumprir a sua missão dentro de dez anos se continuarmos a trabalhar do mesmo modo? ou será que outros fornecedores de dados assumirão o controle?”. Apelos e alertas semelhantes também podem ser identificados em artigos de outros estatísticos sêniores do ESS. Para Skaliotis (2010, p.1), “está acontecendo agora uma profunda transição tecnológica que exige mudanças em nosso paradigma profissional e em nosso modelo de negócios”. Para Struijs e Daas (2013), “a posição dos INE na sociedade da informação está se tornando menos evidente, embora o seu quadro institucional atualmente seja estável”. Já de acordo com MacFeely (2016):

após a crise financeira de 2008 e a subsequente crise econômica [...] uma nova série de pressões mais imediatas confronta as estatísticas oficiais. As inovações disruptivas, como as redes sociais e o big data, os desafios de medição decorrentes da globalização e a necessidade de compreender o progresso social, econômico e ambiental, estão forçando as estatísticas oficiais a adaptarem-se e a evoluir mais uma vez (MacFeely, 2016, p.1, tradução nossa).

Europeia (Eurostat), a Organização de Cooperação e Desenvolvimento Económico (OCDE) e a Comissão Económica e Social para a Ásia e o Pacífico (ESCAP) organizaram uma Reunião sobre a Gestão de Sistemas de Informação Estatística. A principal conclusão da reunião foi que as organizações estatísticas deveriam combinar esforços para enfrentar problemas comuns em relação ao uso de big data (UNECE, 2013a). Em agosto de 2013, o 59º. Congresso Mundial de Estatística do Instituto Internacional de Estatística (ISI) dedicou especial atenção ao tema do big data, com uma sessão de artigos convidados sobre o tema e uma sessão temática especial sobre o potencial da Internet e big data para estatísticas oficiais e um evento satélite sobre big data e computação estatística. Em Setembro de 2013, a reunião anual dos Diretores-Gerais dos institutos nacionais de estatística europeus dedicou uma sessão ao tema do big data da qual resultou o Memorando de Scheveningen sobre Big Data e estatísticas oficiais (DGINS, 2013). Como encaminhamento estratégico do memorando, os DGINS conclamaram o Sistema Estatístico Europeu a desenvolver uma estratégia e plano de ação em Big Data para estatísticas oficiais no âmbito da União Europeia, a ser integrada no programa de trabalho do Eurostat. Finalmente, em março de 2014 a Comissão de Estatística da ONU em sua 45ª. sessão apoiou a proposta de criação de um grupo de trabalho global sobre Big Data para estatísticas oficiais, resultado no projeto ONU Big Data, que analisamos na segunda parte desta tese.

Conforme é possível extrair da literatura nativa do campo estatístico (Skaliotis, 2010; Strujis e Daas, 2013; Hand, 2015; Reimsbach-Kounatze, 2015) e de depoimentos colhidos com alguns de seus agentes, o novo contexto, caracterizado por alguns como um ‘dilúvio de dados’, passou a implicar nas seguintes pressões e desafios principais: i) o aumento da demanda dos usuários de estatísticas oficiais por maior tempestividade e informações em ‘tempo real’; ii) a emergência de novos fenômenos na economia e sociedade digital difíceis de captar com fontes tradicionais de dados e iii) a concorrência dos INEs com novos atores privados produtores de novas modalidades de informações estatísticas. A seguir analisamos este alegado conjunto de ‘pressões’ sobre o campo estatístico, com o objetivo de compreender as motivações que conduzem à articulação de um discurso de modernização das estatísticas oficiais com uso de big data, neste novo contexto. Isto será importante para posteriormente analisarmos, na parte empírica da tese, o sentido de uma série de ações e estratégias desencadeadas nas estruturas de coordenação do campo estatístico internacional com efeitos nos países do Sul-Global e, em particular, no Brasil e América Latina.

Em primeiro lugar, a pressão por maior tempestividade, conforme vimos, já vinha se expressando no período neoliberal. A busca por maior tempestividade conduziu, por um lado, em uma proliferação de indicadores estatísticos, e por outro, foi incorporada na gestão de qualidade e utilizada pelos INEs como um elemento favorável para a ampliação de sua autonomia, por exemplo com a adoção de calendários de divulgação pública. Entre a primeira e segunda década do Séc. XXI, entretanto, alega-se um incremento na ‘demanda’ por tempestividade que agora passaria a incluir ‘informações em tempo real’.

Assim, de acordo com Skaliotis (2010), “Esta pressão crescente por estatísticas mais e mais rápidas é agora amplamente sentida em todo o Sistema Estatístico Europeu. Como consequência, o ESS não consegue satisfazer todas as exigências e/ou não consegue fornecer as estatísticas necessárias em tempo útil”. Ou ainda, de acordo com o primeiro relatório estratégico do Grupo de Alto Nível de Modernização das Estatísticas Oficiais da UNECE (2011, pp.2-3), “estamos vendo outros agentes criando estatísticas semelhantes às nossas, mas muito, muito mais rápidas (por exemplo, Google) e em escala quase global [...] as mudanças na nossa sociedade aumentaram a necessidade de estatísticas mais rápidas”. Já para Strujis e Daas (2013, p.2), “outros fornecedores de informações sobre a sociedade surgem por toda

parte. Frequentemente, são muito rápidos e considerados bem informados. Com isso, a sociedade fica menos dependente das informações dos INE”. Na mesma direção, Hand (2015, p.8) aponta para “várias implicações importantes e ameaças para os INEs”, representadas por “ferramentas de coleta e análise de dados em tempo real”. Finalmente de acordo com Reimsbach-Kounatze (2015):

[...] um número crescente de fornecedores de dados está construindo as suas próprias estatísticas e indicadores em muitas áreas que, até agora, eram domínio exclusivo dos INE. Muitos destes indicadores de fontes não tradicionais têm recebido atenção considerável devido à sua tempestividade, detalhe e frequência. Devido a estas qualidades, os decisores políticos começaram a utilizá-las e, de um modo mais geral, as expectativas começaram a aumentar (Reimsbach-Kounatze, 2015, p.31, tradução nossa).

Tudo leva a crer que as reações do campo estatístico ao alegado ‘aumento na demanda do público e tomadores de decisão por maior tempestividade’ vinculam-se objetivamente a oferta de uma nova modalidade de informações estatísticas que emerge a partir da primeira década do Séc. XXI. Esta nova oferta passa a configurar um mercado de *data analytics* que abrange novos serviços estatísticos baseados em big data desenvolvidos por plataformas digitais (e.g. Google Trends, Facebook Analytics, Twitter Analytics), empresas de telefonia celular (e.g. Telefónica Tech, Orange Business), empresas de tecnologia (e.g. Adobe Analytics, Price Stats) e empresas de pesquisa de mercado (e.g. Ipsos, Gartner, Nielsen).

Paralelamente aos novos serviços de data analytics, desponta uma nova geração de informações estatísticas em tempo real. Destacam-se estatísticas baseadas em buscas dos usuários, como por exemplo o Google Flu Trends (GFT), que passou a estimar a incidência de casos de influenza (Google, 2010); estatísticas de preços e inflação geradas por meio da coleta de dados na internet como o Billion Prices Project (Cavallo e Rigobon, 2016) ou com dados de scanner de grandes varejistas como as geradas pela NielsenIQ (Nielsen, 2023); estatísticas de emprego produzidas com dados de folha de pagamentos como a Automatic Data Processing (ADP) (Yildirmaz, 2016); estatísticas instantâneas de tráfego geradas por empresas de telefonia celular, utilizando dados de posicionamento dos usuários como o Floating Mobile Data (FMD) da empresa Orange (Orange, 2012), etc.

O incremento na disponibilidade de dados de usuários de serviços digitais impulsionou por sua vez novas aplicações da estatística bayesiana em modelos preditivos voltados à ‘previsão do presente’ ou ‘nowcasting’. Estes modelos passaram

a incorporar novas fontes de dados como o Google Trends, para a antecipação do resultado provável de índices oficiais tradicionais como PIB, Inflação, desemprego, etc., normalmente divulgados com menor frequência e com um período de defasagem (Choi; Varian, 2009; Banbura, M; Reichlin, L.; Giannone, D., 2010; Scott; Varian, 2013). De acordo com o economista chefe da Google, esta nova modalidade de estatísticas em tempo real 'pode ajudar a prever o presente':

Economistas, investidores e jornalistas acompanham avidamente as divulgações mensais de dados governamentais sobre as condições econômicas. No entanto, estes relatórios só estão disponíveis com algum atraso: os dados de um determinado mês são geralmente divulgados no meio do mês seguinte e são normalmente revistos vários meses mais tarde. O Google Trends fornece relatórios diários e semanais sobre o volume de consultas relacionadas a diversos setores. Nossa hipótese é que esses dados de consulta podem estar correlacionados com o nível atual de atividade econômica em determinados setores e, portanto, podem ser úteis na previsão de divulgações de dados subsequentes. Não estamos afirmando que os dados do Google Trends ajudam a prever o futuro. Em vez disso, afirmamos que o Google Trends pode ajudar a prever o presente. (Choi e Varian, 2009, p.ii)

Estas novas modalidades de informação estatística em 'tempo real' passaram a exercer não apenas pressão, mas um considerável fascínio sobre alguns segmentos do campo estatístico. Conforme é possível verificar nas observações de um estatístico sênior do Eurostat:

A principal atratividade destas abordagens é a oportunidade e o (baixo) custo. Em termos de tempestividade, estão na vanguarda daquilo que podemos esperar: estatísticas em tempo real. Esta é a principal característica do que intuitivamente chamo de Estatísticas Oficiais 2.0, ou seja, reduzir o intervalo de tempo entre o 'período de referência'/'período de ocorrência' e a 'data de publicação' para quase zero. Estou ciente de que isto pode parecer um sonho hoje, mas provavelmente será a característica predominante – rotineira – da Estatística na próxima década (Skaliotis, 2010,pág.3, tradução nossa).

A qualidade das estatísticas e predições obtidas a partir das novas modalidades de dados como o Google Trends, entretanto, não demoraram a ser objeto de críticas qualificadas. Em artigo publicado na revista Nature, Butler (2013, p.155) demonstrou que “a comparação com os dados de vigilância tradicionais mostrou que o Google Flu Trends (GFT) superestimou drasticamente os níveis máximos de gripe”. Na mesma direção, Olson et. al (2013, p.1) identificaram “falhas substantivas nos modelos original e atualizado do GFT e concluíram que “os dados do Google podem não fornecer uma vigilância confiável para a gripe sazonal ou pandêmica e devem ser interpretados com

cautela até que o algoritmo possa ser melhorado e avaliado”. As críticas publicadas em artigos revisados por pares, em periódicos de alto impacto, levaram ao encerramento do projeto pela Google em 2014 (Lazer; Kennedy, 2015).

De acordo com Lazer et al (2014, p.1203), em artigo publicado na revista *Science*, problemas em estatísticas produzidas a partir de novas fontes de dados não se limitam ao Google Trends, na medida em que pesquisas semelhantes se tornaram um lugar comum. De acordo com os autores, “embora estes estudos tenham mostrado o valor destes dados, estamos longe de um lugar onde eles possam suplantar métodos ou teorias mais tradicionais”. Lazer et al. (2014, p.1203) destacam dois problemas que conduziram aos erros do GFT e que afetam igualmente outras iniciativas semelhantes. O primeiro refere-se a ‘arrogância do big data’, ou seja, “a suposição muitas vezes implícita de que o big data é um substituto, e não um complemento, da coleta e análise de dados tradicionais”. Notadamente, os autores chamam atenção para o fato de que “a maior parte dos big data que têm recebido atenção popular não são o resultado de instrumentos concebidos para produzir dados válidos e confiáveis, passíveis de análise científica”. O segundo diz respeito à ‘dinâmica dos algoritmos’, ou seja, “mudanças feitas pelos engenheiros para melhorar o serviço comercial e pelos consumidores na utilização desse serviço” (Lazer et al. 2014, p.1204) que afetam o resultado da informação obtida a partir destas fontes. De acordo com os autores, estas dinâmicas também podem envolver a manipulação do processo de geração de dados para atingir objetivos próprios.

A existência de uma troca entre tempestividade e precisão envolvendo o big data foi observada pelos estatísticos do Eurostat (2014). De acordo com um documento intitulado “Big data, uma oportunidade ou um ameaça para as estatísticas oficiais” apresentado na Conferência dos Estatísticos Europeus, o Eurostat (2014, p.3) observou que ‘as estatísticas derivadas de big data não oferecem qualquer garantia de alta qualidade’ e que o comprometimento das estatísticas oficiais pelas novas fontes de dados poderia resultar em ‘perda de transparência com impactos na democracia’:

Existe claramente uma troca entre tempestividade e precisão até que os números finais sejam entregues. As estatísticas derivadas de Big Data podem ter uma vantagem competitiva neste aspecto. Uma das razões para uma melhor tempestividade é que a fase de coleta de dados do processo estatístico genérico é normalmente substituída nas fontes de Big Data pela disponibilidade automática e instantânea de dados [...] Os políticos e outros

utilizadores de dados poderiam decidir utilizar o primeiro indicador disponível, especialmente se os problemas de precisão forem recompensados por atrasos notavelmente mais curtos. Isto poderia tornar irrelevantes algumas das estatísticas conjunturais oficiais e prejudicar gradualmente a confiança no atual sistema estatístico. No entanto, as estatísticas derivadas de fontes de Big Data não oferecem qualquer garantia de alta qualidade, como independência profissional e disponibilidade regular. A perda de fontes de dados oficiais poderá, no final, resultar numa perda de transparência com impactos na democracia (Eurostat, 2014, p.3, tradução nossa).

Em segundo lugar, verifica-se que as novas fontes de dados abriram possibilidades para a mensuração de novos fenômenos relacionados ao universo digital por agentes de fora do campo estatístico. Assim, de acordo com a provocação de Strujis e Daas (2013, p.2), “parece que chegamos a um ponto em que a forma tradicional de fazer estatísticas não dá respostas às novas questões que surgem”. No caso da economia digital, por exemplo, a primeira iniciativa para a produção de estatísticas de comércio eletrônico utilizando métodos de extração de dados da web (webscraping) foi realizada em 2013 no Reino Unido por um instituto privado financiado pela Google (Nathan et. al., 2013). Do ponto de vista da ‘sociedade digital’, verificam-se também o desenvolvimento de iniciativas voltadas à mensurar o ‘pulso’ ou o ‘humor’ da nação, com base na análise massiva de postagens em redes sociais. Um dos primeiros trabalhos neste campo foi desenvolvido por pesquisadores das universidades de Northeastern e Harvard com dados do Twitter (Biever, 2010).

Finalmente, a alegada ‘pressão’ decorrente da concorrência com agentes privados, conforme os próprios exemplos acima ajudam a ilustrar, foi o resultado objetivo da emergência de novas fontes de dados, tecnologias e métodos que passaram a viabilizar a produção de novas modalidades de informação por agentes fora do campo estatístico. Por outro lado, a noção de que doravante os INEs passariam a atuar em um novo ambiente competitivo, em que já não deteriam sua tradicional condição monopolística, passou a constituir um importante elemento discursivo para a mobilização do campo estatístico em torno de uma estratégia de modernização que visava a incorporação de fontes de big data.

Uma das primeiras referências à potencial competição dos INEs com o setor privado, em um artigo nativo do campo estatístico, pode ser encontrada em Skaliotis (2010). É interessante observar que o estatístico sênior do Eurostat se refere a um ‘mercado de estatísticas oficiais’ no alerta que dirige a seus pares. Isto é revelador da maior permeabilidade de alguns segmentos do campo estatístico à lógica de mercado, que conforme vimos resultou de processos de subjetivação da ordem neoliberal:

Existe algum interesse dos agentes privados em entrar no mercado das estatísticas oficiais? A resposta é sim. Embora o interesse dos agentes privados neste momento - impulsionado por fins comerciais - se concentre principalmente em áreas que não são o núcleo das estatísticas oficiais, a crescente consciência do poder da mineração de dados e das análises estatísticas resultará num crescimento importante no número de analistas de dados (competentes) que competirão (por que não) com os INE numa série de domínios que hoje pertencem ao mercado algo 'fechado' das estatísticas oficiais [...]. O forte interesse das grandes empresas em desempenhar um papel de liderança no mercado de estatísticas é manifestado pelas recentes grandes aquisições de 'empresas de business intelligence' [...] a concorrência no mercado de estatísticas irá muito provavelmente favorecer aqueles que terão acesso rápido às pegadas digitais relevantes, ou seja, ao tipo de dados predominante no futuro. Com base nas evidências atuais, a minha profunda preocupação é que as enormes bases de dados emergentes que contêm estes novos tipos de dados são propriedade quase inteiramente de agentes privados, como o Google, fornecedores de serviços de Internet, sites de redes sociais, E-bay, Amazon, supermercados e muitos outros. (Skaliotis, 2010, p.2)

Um série de outras referências à competição com o setor privado se seguiram em relatórios e artigos nativos do campo. De acordo com Strujis e Daas (2013, p.8), na medida em que “a sociedade se torna menos dependente [das estatísticas oficiais] os INEs estarão sujeitos a uma maior concorrência com agentes externos às instituições estatísticas tradicionais”. Para MacFeely (2019, p.793-794), “estas tecnologias e os seus proprietários têm o potencial de ameaçar o papel dominante das estatísticas oficiais ou, de forma mais pessimista, a sua existência”.

Conforme observado por Ruppert e Scheel (2021, p.2), em resposta ao conjunto de pressões e apelos “para concretizar os potenciais e expectativas associados a estas tecnologias e dados”, os INEs europeus iniciaram uma série de experiências com fontes de big data. Para os autores, juntamente com a substituição das fontes tradicionais dos censos e surveys por registros administrativos, “estas mudanças nos métodos estatísticos são as mais fundamentais desde o início das estatísticas populacionais modernas, há mais de dois séculos”.

Compreendemos que a introdução das fontes de big data nos INEs marca o início do processo de transição para um regime estatístico-algorítmico de dataficação, caracterizado pela abertura de intersecções entre os dois campos de práticas (estatísticas e algorítmicas). Conforme veremos, este processo é permeado por disputas em torno de mudanças ontológicas dos dados — enquanto mercadorias e bens públicos — e epistemológicas nos métodos estatísticos — interpretações frequentista e bayesiana. Na próxima seção, discutiremos como as novas fontes de

dados e métodos passaram a ser abordados e introduzidos no campo estatístico, a partir da análise dos discursos e estratégias delineados pelos seus agentes.

2.9 O big data e a ‘modernização’ das estatísticas oficiais

Conforme vimos, a emergência de novas fontes de dados fora do controle e governabilidade do campo estatístico gerou uma série de reações ao que foi identificado como um novo conjunto de pressões e ameaças externas às estatísticas oficiais. Notadamente, no início da segunda década do Séc. XXI, verifica-se a articulação de um discurso de modernização das estatísticas oficiais que têm entre suas prioridades a utilização de fontes de big data pelos INEs. Neste contexto, o debate no campo estatístico passou a centrar-se na discussão sobre a usabilidade das novas fontes de dados para a compilação de estatísticas oficiais, o que envolveu uma série de considerações sobre oportunidades, limitações e riscos.

A UNECE passou a exercer considerável protagonismo na agenda de modernização do campo estatístico a partir da criação em 2010 do Grupo de Alto Nível para a Modernização das Estatísticas Oficiais (HLG) no âmbito do escritório da Conferência dos Estatísticos Europeus (CES)⁵⁶. É da lavra deste grupo o *working paper* ‘O que o Big Data significa para as estatísticas oficiais’ (UNECE, 2013). O risco de perda de relevância do campo estatístico para o setor privado e a necessidade de incorporar novas fontes de dados e novos métodos para proteger o ‘papel e importância das estatísticas oficiais’ foi o principal argumento dos estatísticos do HLG para conclamar os seus pares a ‘subirem à bordo’ das iniciativas em big data.

O Big Data tem potencial para produzir estatísticas mais relevantes e oportunas do que as fontes tradicionais de estatísticas oficiais. As estatísticas oficiais têm-se baseado quase exclusivamente na coleta de dados de inquéritos e na aquisição de dados administrativos de programas governamentais, muitas vezes uma prerrogativa legal das Instituições Nacionais de Estatística (INE). Mas este não é o caso do Big Data, onde a maioria dos dados está prontamente disponível ou com empresas privadas. Como resultado, o setor privado poderá tirar partido da era dos Big Data e produzir cada vez mais estatísticas que tentem superar as estatísticas oficiais em termos de tempestividade e relevância. É pouco provável que os INE percam a marca registada de ‘estatística oficial’, mas poderão perder

⁵⁶ A Conferência dos Estatísticos Europeus (CES) é um espaço de coordenação do campo estatístico no âmbito da Comissão Económica da Nações Unidas para a Europa (UNECE). Fundada em 1953 como corpo diretivo da produção estatística da UNECE, suas origens remontam à primeira Conferência Estatística da Liga das Nações em 1928. Ver: <https://unece.org/statistics/ces/about-conference-european-statisticians-ces>

lentamente a sua reputação e relevância, a menos que subam a bordo [...] Ao incorporar fontes relevantes de Big Data no seu processo, os INE estarão melhor posicionados para medir a sua precisão, garantir a consistência de todo o sistemas de estatísticas oficiais e fornecer interpretação, enquanto trabalham constantemente na relevância e oportunidade. O papel e a importância das estatísticas oficiais serão assim protegidos (UNECE, 2013, p.2, tradução nossa).

O Encontro de Gestão de Sistemas de Informações Estatísticas (MISIS)⁵⁷ articulado pelo HLG em 2012, reuniu chefes estatísticos e gestores de INEs, da UNECE, ESCAP e OCDE para debater a modernização da produção estatística. Neste encontro, os participantes chegaram a conclusão de que “este é o momento ideal para começar a colaborar em Big Data” (UNECE, 2013a, p.10) e concordaram em combinar esforços para enfrentar questões comuns, dentre as quais: adoção de novos métodos e infraestruturas, produção de estatísticas de ‘curto prazo’, promoção de ‘mudanças na cultura organizacional para encorajar a inovação’, criação de novas funções ‘incluindo cientistas de dados’ e cooperação com o setor privado (UNECE, 2013a). Finalmente, entre os encaminhamentos estratégicos do encontro constam a elaboração de um projeto colaborativo em big data para estatísticas oficiais e a produção de uma classificação comum sobre as novas fontes de dados (UNECE, 2013a, p.10).

Do trabalho inicial do HLG sobre o tema, resultaram uma definição de big data — “fontes de dados que podem ser geralmente descritas como de alto volume, velocidade e variedade de dados que exigem formas de processamento inovadoras e econômicas para melhorar a percepção e a tomada de decisões” — e uma classificação das principais fontes com potencial para a produção de estatísticas oficiais, conforme a tabela abaixo (UNECE 2013b).

Tabela 2 - Classificações de Fontes de Big Data (UNECE, 2013)

Categoria	Tipo	Fontes
Redes Sociais	Informações de origem humana.	Redes Sociais: Facebook, Twitter, Tumblr etc. Blogs e comentários Documentos pessoais Fotos: Instagram, Flickr, Picasa etc. Vídeos: Youtube etc. Pesquisas na Internet Conteúdo de dados móveis: mensagens de texto Mapas gerados pelo usuário E-mail
Sistemas de Negócios	Dados mediados por processos.	Dados produzidos por Órgãos Públicos Prontuários médicos Dados produzidos pelas empresas

⁵⁷ O encontro do Grupo MSIS foi promovido pela UNECE, ESCAP e OCDE entre 23 e 25 de Abril de 2013 em Paris e Bangkok (MSIS, 2013).

		Transações comerciais Registros bancários/de ações Comércio eletrônico Cartões de crédito
Internet das Coisas	Dados gerados por máquinas.	Dados de sensores Sensores fixos Automação residencial Sensores climáticos/poluição Sensores de tráfego/webcam Sensores científicos Vídeos/imagens de segurança/vigilância Sensores móveis (rastreamento) Localização do celular Automóveis Imagens de satélite Dados de sistemas de computador Registros Registros da Web

Fonte: UNECE (2013b)

O paper do HLG (UNECE, 2013) e os debates travados no encontro do MSIS (UNECE, 2013a) são relevantes para compreender as motivações dos agentes na introdução de big data e ciência de dados nas práticas estatísticas. Apesar de haver diferentes abordagens sobre o tema entre os integrantes do grupo, a percepção geral do campo estatístico neste momento é a de que “novas fontes de dados, novos atores e novas expectativas têm impacto na forma como trabalhamos e como organizamos o nosso trabalho” (UNECEa, 2013, p. 2) e apesar de não haver um consenso sobre “quão viável é o Big Data para a produção real de estatísticas oficiais” (UNECEa, 2013, p. 10), “temos que lidar com Big Data para manter a nossa relevância, e uma abordagem colaborativa pode reduzir o risco” (UNECEa, 2013, p. 11). Na mesma direção, os estatísticos do Eurostat (2014, p.3) conclamaram seus pares na Conferência dos Estatísticos Europeus em 2014: “Existe alguma maneira de evitar que aconteçam possíveis danos ao sistema de estatísticas oficiais? A resposta poderia ser explorar as possibilidades de utilização de fontes de Big Data nas estatísticas oficiais”.

Do ponto de vista das oportunidades oferecidas pelas novas fontes para as estatísticas oficiais, verifica-se em um conjunto de artigos elaborados por agentes do campo estatístico (Struijs e Daas, 2013; Struijs et al., 2014; Kitchin, 2015; MacFeely, 2019) a promoção do uso de big data para: **i)** substituir total ou parcialmente fontes estatísticas tradicionais; **ii)** prover informação complementar para estatísticas existentes sob novas perspectivas; **iii)** melhorar estimativas de fontes estatísticas tradicionais; **iv)** produzir novas informações estatísticas. De acordo com estes trabalhos, a incorporação de big data ao processo de produção de estatísticas oficiais

poderia potencialmente contribuir para: o aumento da tempestividade, a melhoria da qualidade do processo estatístico (maior precisão, granularidade, conexão e escalabilidade), a melhoria da comparabilidade internacional, o aumento do valor das informações para o público, a análise de fenômenos complexos e globais, a melhoria de análises transversais, a maior compreensão de causalidade, a redução da carga sobre os informantes e a redução de custos com a realização de surveys.

Em um dos primeiros artigos nativos do campo estatístico sobre o tema, intitulado ‘Big Data, Big Impact’, Strujis e Daas (2013, p.4-5) apresentaram à Conferência dos Estatísticos Europeus (EEC) estudos iniciais desenvolvidos pelo Netherlands Statistics (CBS) sobre as potencialidades de três novas fontes de dados: sensores de tráfego, que “poderiam resultar em estatísticas mais rápidas e robustas sobre o tráfego”; dados de posicionamento de telefones celulares, que “revelaram vários usos para estatísticas oficiais, como atividade econômica, turismo, densidade populacional para mobilidade e uso de estradas”; e mensagens de redes sociais, que alegadamente demonstraram ser “possível produzir indicadores semanais sobre a confiança dos consumidores”.⁵⁸

Com relação às limitações e desafios das fontes de big data para a compilação de estatísticas oficiais, os debates iniciais no campo estatístico apontavam para três fatores fundamentais: acesso aos dados, qualidade das fontes e capacidade computacional. Reflexões sobre estas limitações também levaram em consideração riscos subjacentes às estatísticas oficiais, como a perda de autonomia dos INEs, comprometimento da reputação e confiança pública nas estatísticas oficiais, bem como questões relacionadas à privacidade dos informantes.

Tabela 3 - Big Data: Oportunidades, Limitações e Riscos

Oportunidades	Limitações	Riscos
<p>Usabilidade potencial: Substituição total ou parcial de fontes estatísticas; Informação complementar para estatísticas existentes; Melhoria das estimativas de fontes estatísticas; Produção de novas informações estatísticas.</p> <p>Benefícios Potenciais: Aumento da tempestividade.</p>	<p>Acesso: Dificuldade de acesso do campo estatístico às fontes de big data; Em grande medida o Big Data consiste em dados guardados e controlados por empresas privadas; Altos custos para aquisição das bases de dados; Restrições de uso relacionadas à legislação de proteção de dados;</p>	<p>Autonomia: Perda da autonomia dos INEs sobre o processo estatístico Baixo ou nulo controle e mandato dos INEs sobre os big data guardados por empresas privadas. Desvio de missão: inversão da lógica dedutiva do processo de produção de estatísticas oficiais, orientada por</p>

⁵⁸ Referências a estas três experiências também constavam entre os exemplos de potencialidades das fontes de Big Data no artigo produzido pelo Grupo de Alto Nível em Modernização das Estatísticas oficiais (HLG) para a Conferência dos Estatísticos Europeus (UNECE, 2013), o que é revelador da liderança e influência dos estatísticos holandeses na agenda do big data para estatísticas oficiais.

<p>Melhoria da qualidade do processo estatístico. Melhoria da comparabilidade internacional. Aumento do valor das informações para o público Análise de fenômenos complexos e globais. Melhoria de análises transversais. Maior compreensão de causalidade. Redução da carga sobre os informantes. Redução de custos com a realização de surveys.</p>	<p>Capacidade Computacional: Desafios associados ao volume e velocidade dos Big Data. Qualidade: Usabilidade: subproduto de aplicações que não foram concebidas para produção de informações estatísticas. Perda de dados: inatividade de servidores e interrupções de rede. Volatilidade dos dados: Problemas na análise de séries temporais. Não-Representatividade: ausência de uma população-alvo definida e(ou) dificuldade de relacionar os dados à uma população conhecida.</p>	<p>questões, para uma lógica indutiva, orientada pelos dados. Confiança Pública: Comprometimento da qualidade das estatísticas oficiais pode 'minar a confiança pública em sua validade. Exposição da reputação dos INEs: associação a parceiros privados, percepção de utilização indevida de dados pessoais e vigilância. Privacidade dos informantes: Sistemas de confidencialidade e segurança dos INEs não foram projetados para lidar com big data. Ampliação das possibilidades de combinação, identificação e vazamentos de dados. Novo conjunto de requisitos associados às 'Leis Gerais de Proteção de Dados'.</p>
---	--	--

Fonte: do autor a partir de Struijs e Daas, 2013; Struijs et al., 2014; Kitchin, 2015 e MacFeely, 2019.

A abordagem do campo estatístico às limitações, desafios e riscos subjacentes à utilização de big data para a compilação de estatísticas oficiais nos possibilita identificar a emergência de controvérsias no campo estatístico relacionadas aos processos de dataficação. Nas próximas seções nos deteremos nos debates em torno das questões de acesso e qualidade, na medida em que trazem a tona dois elementos fundamentais para nossa análise: i) o obstáculo ontológico da nova economia política dos dados e; ii) o obstáculo epistemológico dos novos métodos associados às fontes de big data.

2.10 O obstáculo ontológico: a nova economia política dos dados e o duplo movimento do campo estatístico

O advento de uma nova economia de dados no Séc. XXI corresponde não apenas à emergência de novos métodos, abordagens e aplicações baseadas em dados para a melhoria de processos produtivos, comerciais e tecnológicos, mas à transformação dos dados em si em mercadorias comercializáveis e ativos capitalizáveis. Esta transformação ontológica dos dados é acompanhada, por sua vez, pela emergência de agentes que constituem um novo campo de práticas, que aqui intitulamos de campo algorítmico, diretamente envolvidos na construção social e econômica de novos mercados de dados.

Conforme vimos, as estatísticas oficiais e os dados para fins estatísticos foram historicamente constituídos como bens públicos sob o controle dos Estados Nacionais e de campos estatísticos relativamente autônomos que se diferenciaram da estrutura

estatal, na medida em que se constituíram como um corpo científico e burocrático. Neste sentido, a formatação atual dos INEs, como repositórios públicos e confiáveis de dados para fins estatísticos, é resultado de um processo histórico que envolveu a construção de uma legitimidade apoiada na autoridade do Estado e da ciência.

Os dados estatísticos sob a custódia dos INEs são bens públicos, na medida em que servem à produção de informações públicas e podem ser acessados de forma livre e gratuita pelos governos e população em geral. Restrições de acesso relativas à proteção de microdados podem ser aplicáveis, o que, entretanto, não descaracteriza o caráter público dos dados. Notadamente, os INEs detêm mandato e responsabilidades legais que os obrigam a assegurar o sigilo e a privacidade dos informantes, o que implementam por meio de técnicas de anonimização e agregação dos dados, além de sistemas de segurança contra o vazamento de informações sensíveis.

De acordo com a clássica definição de Samuelson (1954), são considerados públicos os bens que possuem características de não rivalidade e não exclusividade. A rivalidade implica que um bem só pode ser utilizado por uma parte consumidora de cada vez e que o consumo deste bem reduz a quantidade disponível do mesmo para o restante da sociedade. Já a exclusividade se aplica quando é possível impedir o consumo de um determinado bem. Dados e informações são não-rivais por natureza, na medida em que podem ser facilmente distribuídos e simultaneamente utilizados sem que isso implique em seu esgotamento para outros usuários. No que se refere à exclusividade, embora restrições de acesso possam ser interpostas pelos detentores de dados, o que caracteriza efetivamente a exclusão é a possibilidade de exercer direitos de propriedade sobre um determinado bem (Pantelis; Aija, 2013).

No que se refere ao big data, em especial aos dados gerados por usuários e consumidores de produtos e serviços em meios digitais, embora sejam efetivamente tratados como 'bens privados' pelos detentores de bases de dados, não há até o momento direitos de propriedade efetivamente regulamentados sobre esta categoria de dados (Douilhet e Karanasiou, 2016; León, 2016). Deste modo, o processo de comodificação dos dados têm se efetuado principalmente pela via da exclusão do acesso, por meio de cercamentos totais ou parciais de dados (e.g. *open data*)⁵⁹ que

⁵⁹ A partir da abordagem de Ostrom e Hess (2007) que propõe o tratamento do conhecimento e da informação como bens comuns, verifica-se a emergência de uma nova categoria de 'open data' ou 'dados abertos', que se referem a dados disponíveis publicamente mas sujeitos à determinados termos

constituem a essência do modelo de negócios das plataformas digitais (Kuerbis; Mueller, 2022). Na ausência de uma regulamentação sobre a propriedade dos dados, os cercamentos efetuados pelas plataformas possibilitam que as mesmas estabeleçam seus próprios ‘regimes de propriedade’ (Ostrom e Hess, 2007).

Partindo das três principais características do conceito de propriedade em geral — o direito de uso do bem (*usus*), o direito de transferência do bem (*abusus*) e o direito de gozar o direito sobre o bem (*fructus*) — Douilhet e Karanasiou (2016) observam que:

Na ausência de um direito formal de propriedade sobre o big data, as partes que se beneficiam desses direitos devem demonstrar estes elementos de propriedade. Considerando os grandes montantes investidos pelos grandes controladores de dados, pode parecer que os dados recolhidos e agregados pelas empresas estão sob a sua propriedade – elas os mantêm em suas bases de dados, os processam e agregam (*usus*), e extraem valor da sua análise (*fructus*) e de sua venda para terceiros (*abusus*).

O principal processo que passa a possibilitar a construção dos novos mercados de dados é o cercamento privado de grandes quantidades de dados de usuários por meio de plataformas digitais (Srnicek, 2017). Em linha com Verdegem (2022) e Dean (2016), entendemos o cercamento de dados como o processo pelo qual o livre acesso e controle sobre as informações da atividade de usuários geradas pelas interações digitais são mantidos longe dos próprios usuários e do público em benefício privado dos fornecedores (plataformas e empresas de tecnologia) onde os dados são gerados. Com isso, acordos abertos ou compartilhados de acesso e controle público sobre os dados tornam-se proprietários e exclusivos. Cercado, o big data constitui um capital informacional em disputa e uma mercadoria e(ou) ativo em potencial.

As iniciativas de ‘modernização’ das estatísticas oficiais com big data deparam-se assim, desde o primeiro momento, com os cercamentos privados das novas fontes. Apesar de existirem fontes públicas, abertas ou relativamente abertas de big data, a grande maioria das fontes relevantes para a produção estatística (e.g. telefonia celular, buscadores na internet, redes sociais, scanners, sensores) são detidas por

de acesso que limitam a quantidade, frequência e usos dos dados disponibilizados. De acordo com a taxonomia proposta por Pantelis e Aija (2013), o que diferenciaria os dados públicos dos dados abertos é o fato de que no caso dos dados abertos direitos de propriedade podem ser definidos e aplicados por meio de termos de uso. Conforme observado por diversos autores (León, 2016; Rikap, 2021; Muller e Kuerbis, 2022), a despeito das intenções de seus idealizadores, a criação de categorias híbridas como ‘commons’ e ‘open’, na medida em que estabelecem regimes de propriedade, foram invariavelmente apropriadas por agentes de mercado, constituindo-se igualmente em um meio de comodificação dos dados.

privados. Conforme veremos, a questão do acesso aos dados passou a expor uma contradição entre a lógica pública do campo estatístico e a lógica de mercado do campo algorítmico que conduzem a estratégias distintas de ação, caracterizando um duplo movimento.

A restrição de acesso do campo estatístico às fontes de big data foi destacada por Kitchin (2015) — em artigo publicado em uma coletânea especial sobre big data organizada pela International Association for Official Statistics (IAOS)⁶⁰ — como a principal limitação ao uso desta categoria de dados para a compilação de estatísticas oficiais. Neste sentido, o autor apontava para um duplo caminho estratégico, a formação de parcerias com o setor privado ou o empoderamento legal dos INEs para o acesso compulsório aos dados.

Uma primeira questão é obter acesso ao big data necessário em primeiro lugar para avaliação, experimentação, teste e adoção. Embora alguns big data sejam produzidos por agências públicas, como dados meteorológicos, alguns sites e sistemas administrativos, e alguns dados de transporte, muitos big data são atualmente gerados por empresas privadas, como telefonia móvel, mídias sociais, serviços públicos, financeiros, de crédito, seguradoras e empresas de varejo. Esses big data são mercadorias valiosas para essas empresas, fornecendo um recurso que gera vantagem competitiva ou constituindo um produto-chave, e geralmente não estão disponíveis publicamente para fins comerciais, análise oficial ou pública em formas brutas ou derivadas. Para que os INE tenham acesso a esses dados, é necessário formar parcerias estratégicas vinculativas com estas empresas ou criar/alterar instrumentos jurídicos (como leis estatística) para obrigar as empresas a fornecer tais dados. Essas negociações e reformas legislativas são demoradas e politicamente carregadas, especialmente se os INE não pagarem nem compensarem as empresas pelo fornecimento de dados para estatísticas oficiais (Kitchin, 2015, p.474-475, tradução nossa).

Para MacFeely (2019, p.31), os potenciais benefícios das novas fontes para as estatísticas oficiais só poderiam ser efetivamente considerados, ‘assumindo que os problemas de acesso possam ser ultrapassados’.

Uma das maiores barreiras ao uso de big data é a falta de acesso. Muitos big data são proprietários, ou seja, dados de propriedade comercial ou privada e que não estão disponíveis publicamente [...] O atual estatuto de propriedade de alguns dados pode mudar no futuro, à medida que as pessoas em todo o mundo percebam que os seus dados estão sendo utilizados e comercializados. Mas, no momento, muitos conjuntos de dados não estão atualmente acessíveis aos INEs, quer porque os custos são proibitivos, quer porque a legislação de proteção de dados o impede, quer porque a propriedade proprietária o torna impossível (MacFeely, 2019, p.35).

⁶⁰ A IAOS é uma organização não governamental internacional, fundada em 1985 como uma seção especializada do Instituto Internacional de Estatística (ISI).

O caminho de aproximação e parceria com o setor privado será fortemente defendido pela estratégia de modernização das estatísticas oficiais capitaneada pelo HLG da UNECE. O grupo passou a adotar pressupostos de mercado, legitimando a propriedade privada dos dados e a concepção de que constituem mercadorias sujeitas a um custo financeiro de aquisição. Apesar de reconhecer que em alguns países os INEs tem amparo legal para acessar dados fora do âmbito governamental e que em outros países a inexistência desta prerrogativa “pode resultar em limitações ao acesso a certas fontes de big data” (UNECE, 2013, p.5), o HLG passou a advogar que os encargos financeiros com a compra de dados pelos INEs seriam compensados por ‘benefícios potenciais’ que ‘superam largamente os custos’:

É provável que haja um custo para os INE adquirirem Big Data, especialmente Big Data detidos pelo setor privado e especialmente se a legislação for omissa sobre as modalidades financeiras que envolvem a aquisição de dados externos [...] Os custos podem até ser significativos para os INE, mas os benefícios potenciais superam largamente os custos, com os grandes volumes de dados fornecendo potencialmente informações que poderiam aumentar a eficiência dos programas governamentais (por exemplo, sistema de saúde) (UNECE, 2013, p.6, tradução nossa).

Partindo do reconhecimento de que, além de deter as principais fontes de dados, “é principalmente o setor privado que lidera o trabalho em ferramentas e soluções de análise de Big Data” (UNECE, 2013, p.9), o discurso de ‘modernização’ do HLG implicará em uma vigorosa defesa da necessidade de envolvimento, colaboração e sinergias dos INEs com o setor privado: “As sinergias entre os INE e o setor privado não se limitam apenas às questões tecnológicas. A colaboração dos INEs com os proprietários de fontes de dados privadas é de importância crítica” (UNECE, 2013, p.9).

A produção de uma série de artigos por agentes do campo estatístico que passaram a liderar a pauta de modernização das estatísticas oficiais com big data, indica o delineamento de uma estratégia de inovação dos INEs fortemente orientada pela lógica de mercado⁶¹. Esta orientação é patente no artigo inicial de Strujis e Daas

⁶¹ Entre estes agentes destacam-se os membros do Statistics Netherlands que exerceram considerável influência no HLG e na pauta de modernização das estatísticas oficiais com introdução de fontes de big data nos INEs europeus. Piet Daas desde 2010, coordenava o Grupo de Alto Nível de Modernização das Estatísticas Oficiais da UNECE. Peter Strujis que em 2013 ocupava o papel de coordenador de Big Data no CBS (Netherlands Statistics) em 2015 viria a ser o coordenador do projeto de Big Data do Sistema Estatístico Europeu (ESSnet) e membro do Grupo de Trabalho Global de Big Data para Estatísticas Oficiais da Comissão Estatística da ONU.

(2013, p.5) sobre big data, apresentado na Conferência dos Estatísticos Europeus. Nesta ocasião, após a apresentação de experiências iniciais com o uso de big data pelo Statistics Netherlands (CBS), os autores instaram os seus pares a refletir estrategicamente sobre 'que posição pretendem os INEs ocupar na futura sociedade da informação'. Os termos de tal reflexão, partiam do pressuposto de que 'a informação é uma mercadoria' e de que as instituições estatísticas deveriam ocupar uma nova posição no 'mercado da informação':

Em que posição pretende o INE estar no mercado da informação dentro de dez anos? Que tipos de resultados pretende produzir, que papel pretende desempenhar numa sociedade em que a informação é uma mercadoria fornecida por um grande número de empresas e instituições, muitas vezes com alta velocidade, mas de qualidade desconhecida? As respostas a tais questões serão cruciais para o efeito que o Big Data terá nas estatísticas oficiais (Strujis e Daas, 2013, p.5, tradução nossa).

Conforme observado anteriormente, a adoção de pressupostos de mercado é um indício da permeabilidade da agenda de modernização das estatísticas oficiais às concepções neoliberais. Neste sentido, ainda em seu artigo apresentado na CES, Strujis e Daas (2013, p.7) introduziram questões sensíveis que denotam o esboço de uma estratégia, que contemplava: **i)** a abdicação do controle dos INEs sobre às bases de dados: "talvez os dados possam permanecer na fonte. Talvez seja possível fazer com que as consultas sejam feitas pelo detentor dos dados na fonte"; **ii)** a mudança do papel dos INEs: "O INE quer continuar a fazer estatísticas para as quais existe uma alternativa de mercado? Seria melhor transferir o papel do INE da produção de informação estatística para a validação da informação produzida por outros?"; **iii)** o incentivo às parcerias público-privadas, "outra possibilidade é colaborar com terceiros, por exemplo, um instituto de pesquisa com instalações para processamento de Big Data"; e, **iv)** o fomento a um mercado de dados, "em alguns casos, os Big Data têm de ser pagos, embora os INE não estejam habituados a fazê-lo".

Estes elementos serão posteriormente desenvolvidos e aprofundados por Strujis et al. (2014) em um artigo publicado no periódico de alto impacto 'Big Data & Society'. No que se refere ao controle das bases de dados, os autores voltam a defender que "para a produção estatística, pode ser mais eficiente ter os dados processados no local de armazenamento", ou seja dentro da estrutura dos 'provedores de big data'. Com relação à compra de dados pelos INEs, voltam a sugerir que, "a

suposição de que os dados podem ser fornecidos gratuitamente pode não ser mais válida” (Strujis et al. 2014, p.4).

É digno de nota no discurso de ‘modernização’ apresentado por Strujis et al. (2014) a relação de subordinação que é estabelecida entre os INEs e os ‘provedores de Big Data’. Segundo esta visão, os INEs, destituídos de qualquer autoridade legal, dependeriam inteiramente de parcerias com o setor privado para o acesso ao big data: “os potenciais provedores de Big Data são parceiros essenciais, se eles não concederem acesso aos seus dados a história termina antes de começar” (Strujis et al. 2014, p.3). Neste sentido, verifica-se o delineamento de uma estratégia de mobilização do capital simbólico da confiança nas estatísticas oficiais, como meio de aproximação com o setor privado. Sua proposição é a de que instituições estatísticas e Big Techs atuem como ‘parceiros confiáveis’ em um ‘mercado competitivo’:

Um papel específico que os INE poderiam desempenhar é o de parceiro de confiança. Num mercado competitivo, os concorrentes estarão relutantes em partilhar dados sensíveis entre si. Mas poderão estar dispostos a partilhá-los com um INE que compile informação estatística que seja benéfica para todos [...] Há muitos parceiros comerciais com os quais os INEs podem colaborar. Google e Facebook são dois exemplos em que o Big Data constitui o núcleo do seu modelo de negócios. O seu conhecimento e os dados a que têm acesso podem ser muito relevantes para os INE. As empresas de TI também possuem conhecimentos relevantes sobre processamento e armazenamento de Big Data, segurança, processamento em nuvem, etc. Além da prestação de serviços pagos, a colaboração pode ser do seu interesse com vista à obtenção de conhecimentos estatísticos e para benchmarking ou validação dos seus produtos de informação (Strujis et al. 2014, p.4, tradução nossa).

Em uma direção similar, um paper apresentado por estatísticos do Eurostat na conferência de Big Data para Estatísticas Oficiais em Beijing promovida pela Comissão Estatística da ONU, apontava que “a transformação das estatísticas oficiais já começou com muitos INEs entrando no mercado do big data com convicção”. Neste trabalho, Florescu et al. (2014, p.11) defendem que “os INEs precisam adquirir competências adicionais necessárias para a utilização de big data e ser capazes de trabalhar com proprietários de fontes de dados privadas”. Além disso, para lidar com uma série de questões problemáticas como a privacidade, ética, confiança pública e vantagens competitivas, os autores apontam que: “a melhor maneira de abordar muitas destas questões é envolver-se ativamente em parcerias público-privadas (PPP), onde os INE e agentes relevantes participam em cooperação mútua” (Florescu et al. 2014, p.5).

Ainda neste trabalho, os estatísticos do Eurostat apresentaram uma proposta de certificação das novas fontes de dados desenvolvida por Petrakos et al. (2013). De acordo com este modelo de certificação, os critérios de qualidade para a avaliação de fontes de Big Data pelas autoridades estatísticas deveriam ao mesmo tempo “estar totalmente em conformidade com o padrão de qualidade e os princípios da estatística oficial”⁶² e ser “suficientemente flexíveis para considerar fontes que o julgamento inicial excluiria, mas que, após um exame aprofundado, se revelam adequadas para utilização”. Aqui a estratégia de aproximação com agentes privados difere da ideia de ‘parceiro confiável’ delineada por Struijs et al.(2014). A proposta aponta igualmente para a construção de um mercado de dados para estatísticas oficiais, mas neste caso, com maior controle das autoridades estatísticas, que operariam como certificadoras dos fornecedores de big data:

Embora o sistema estatístico oficial não tenha o monopólio dos dados nem possa assumir a responsabilidade de policiar o mundo dos dados, tem uma autoridade moral em virtude do seu histórico de longa data em termos de qualidade e, como tal, tem um papel importante a desempenhar [...] Em vez de tentar identificar se uma organização está disposta a cooperar e partilhar os seus dados, o ponto de partida seria determinar se a organização pretende ser certificada como produtora de dados nessa área específica. Adotar uma abordagem prospectiva neste tipo de situação abriria a opção de certificação como produtor de dados num determinado domínio estatístico (Florescu et al. 2014, p.10, tradução nossa).

Em um artigo apresentado na mesma conferência, o economista estadunidense Landefeld (2014, p.18) argumentou que uma proposta semelhante de certificação de fontes de big data, como a discutida pelos estatísticos do Eurostat, “foi apresentada nos Estados Unidos e encontrou forte resistência por parte dos usuários de dados e da comunidade empresarial”. Entre os motivos para a resistência, Landefeld (2014, p.18) apontava para o fato de que “os produtos de informação são uma importante fonte de receitas para as empresas e estas não desejam correr o risco de divulgação dos métodos proprietários e dos dados de origem que utilizam para produzir os dados que vendem”. De acordo com o autor há uma preferência das

⁶² A nível internacional, as estatísticas oficiais são orientadas pelos Princípios Fundamentais da ONU. A nível regional, a exemplo da Europa e América Latina, códigos de práticas estatísticas estabelecem princípios que abrangem o ambiente institucional, os processos e os resultados estatísticos. Estes princípios são geralmente adaptados por cada sistema estatístico nacional em códigos de boas práticas estatísticas.

empresas pela parceria voluntária com os governos e “tal papel [de certificação] para as agências oficiais de estatística pode produzir um ambiente adverso”.

Em seu artigo, Landenfeld (2014, p.14) defendeu o modelo de parcerias voluntárias entre os INEs e empresas privadas, destacando que, nos Estados Unidos, “alguns dos exemplos mais bem sucedidos do uso de big data vieram de projetos conjuntos entre os proprietários de dados privados e agências oficiais de estatística”. Entre os exemplos apontados, constam parcerias entre o Bureau de Análises Econômicas e a IBM, Chrysler e Google. Finalmente, o economista estimulou que os institutos de estatística sejam proativos na promoção de PPPs em seus países: “Embora pareça óbvio, a pouca frequência da colaboração público-privada em alguns países pode inibir essa colaboração e exigir que as agências estatísticas iniciem e delineiem os benefícios de tal colaboração”.

O modelo de PPPs voluntárias praticada pelos órgãos estatísticos nos EUA e proposto por Landenfeld (2014) baseia-se na adoção de um contrato comercial entre o INE e o provedor de Big Data de acordo com protocolo desenvolvido pelo *US Office of Management and business*. De acordo com o protocolo, o contrato deve abordar cinco pontos principais: objetivos, usos e qualidade dos dados, papéis e responsabilidades para a proteção dos dados e aspectos chave como duração, custos e pagamento pela aquisição ou compartilhamento dos dados. Trata-se de um modelo de mercado de dados em que a instituição estatística é um consumidor de bases de dados ‘proprietárias’. Neste modelo, a proteção da propriedade intelectual dos dados é um aspecto fundamental. De acordo com Landenfeld (2014, p.16): “Os dados como um produto de informação são propriedade intelectual da empresa, tem valor econômico e podem ser vendidos. O governo deve garantir que não haja divulgação que possa distribuí-los gratuitamente”.

Embora com modulações táticas distintas, as três proposições estratégicas de acesso à fontes de big data por meio do relacionamento dos INEs com o setor privado adotam pressupostos mercadológicos. Nos três casos os agentes do campo estatístico assumem que os big data são mercadorias sobre a propriedade de empresas privadas e buscam delinear modos de acesso aos dados coerentes com a lógica de mercado. Seja posicionando as instituições estatísticas como parceiras ou intermediárias de empresas privadas (Struijs et. al. 2014), como certificadoras de empresas produtoras de dados para estatísticas oficiais (Florescu et. al., 2014), ou finalmente como simples consumidoras em um mercado de dados (Landenfeld, 2014).

Estes três modelos ganharam proeminência nos debates sobre o reposicionamento dos INEs em um novo ‘ecossistema de dados’, expressão que passou a ser utilizada como uma espécie de eufemismo para ‘mercado de dados’.

A orientação pró-mercado do Grupo de Alto Nível de Modernização das Estatísticas oficiais prevaleceu inicialmente e concorreu para que a incorporação de big data nas estatísticas oficiais via parcerias público-privadas fosse adotada como uma prioridade pela Conferência de Estatísticos Europeus. Além disso, verifica-se a adoção em Setembro de 2013 do ‘Memorando de Sheveningen: Big Data e Estatísticas Oficiais’ pela Conferência dos Diretores Gerais dos Institutos Nacionais de Estatística (DGINS), articulada pelo Eurostat no âmbito do Sistema Estatístico Europeu (ESS)⁶³.

No memorando de Sheveningen, os DGINS “reconhecem que os Big Data representam novas oportunidades e desafios para as estatísticas oficiais e, portanto, incentivam o Sistema Estatístico Europeu e os seus parceiros a examinarem eficazmente o potencial das fontes de Big Data” (DGINS, 2013). Além disso, há um reconhecimento da necessidade de novas ‘capacidades’ e ‘habilidades’ bem como de ‘parcerias’ e ‘sinergias’ com agentes de outros domínios incluindo ‘proprietários de fontes de dados privados’. Como encaminhamento estratégico do memorando, verifica-se o desenvolvimento de um plano de ação em Big Data para estatísticas oficiais no âmbito da União Europeia que resulta em uma série de projetos experimentais e parcerias público-privadas envolvendo INEs europeus e provedores de big data.

Na medida em que a pauta de cooperação e parcerias com o setor privado ganha proeminência junto aos INEs europeus, é possível identificar a emergência de controvérsias e contramovimentos no campo estatístico. Isto se dá por dois motivos principais, por um lado a disseminação do modelo de parcerias público-privadas voluntárias em projetos experimentais de big data dá margem a preocupações sobre riscos relacionados à autonomia dos INEs e à questões éticas, reputacionais e referentes à confiança pública nas estatísticas oficiais. Por outro lado, as parcerias

⁶³ O Sistema Estatístico Europeu (ESS) é um espaço de coordenação do campo estatístico europeu que reúne a autoridade estatística da União Europeia (Eurostat), Institutos Nacionais de Estatística da Europa e outras autoridades nacionais credenciadas. A Conferência dos Diretores Gerais dos INEs (DGINS) é hospedada anualmente por um INE membro do ESS para discutir um tema importante relacionado com o programa estatístico ou com os métodos e processos utilizados para produzir estatísticas europeias. Ver: <https://ec.europa.eu/eurostat/web/european-statistical-system>

voluntárias com o setor privado apesar de possibilitarem inicialmente a experimentação de novas fontes de dados por meio de projetos-piloto, revelam pouco resultado efetivo na produção regular de estatísticas oficiais. Estas controvérsias conduzem, no contexto europeu, à ações que visam o fortalecimento do mandato legal dos INEs para o acesso compulsório à fontes de dados guardadas por privados para fins estatísticos.

Conforme vimos, a construção social da confiança pública nas estatísticas oficiais e nos INEs foi o resultado de um trabalho persistente, ao longo de mais de um século, que se valeu da autoridade e legitimidade do Estado e da ciência. Neste sentido, alguns autores passam a apontar para o risco de que a utilização de fontes de big data venha a comprometer a confiança nas estatísticas oficiais. Para Kitchin (2015, p.477), além das questões de qualidade associadas às novas fontes, que podem 'minar a confiança na validade e confiabilidade das estatísticas oficiais', há um alto risco de exposição da reputação dos INEs quando associados à parceiros privados:

Um risco crítico para os INE na implementação de um novo conjunto de meios e métodos para produzir estatísticas oficiais é a sua reputação e confiança pública serem minadas. A reputação de fornecedor justo, imparcial, objetivo e neutro de estatísticas oficiais de elevada qualidade é vista pelos INE como uma missão crítica e é normalmente a sua prioridade número em sua matriz de risco institucional. A parceria com terceiros privados e a utilização dos seus dados para compilar estatísticas oficiais expõem a reputação do INE à do parceiro. Um escândalo relacionado à segurança de dados e violações de privacidade, por exemplo, pode muito bem refletir no INE (Kitchin, 2015, p.477, tradução nossa).

Riscos relativos à independência e autonomia dos INEs na produção e distribuição das estatísticas oficiais também foram apontados por Kitchin (2015) e MacFeely (2019). Tais riscos estão relacionados principalmente ao baixo nível de governabilidade dos INEs sobre as fontes de big data controladas por empresas privadas:

A propriedade dos dados de origem é uma questão preocupante. À medida que um INE se afasta dos dados baseados em inquéritos e se torna mais dependente de dados administrativos e outros dados secundários como big data, ele abdica do controle do seu sistema de produção. O principal insumo, os dados de origem, dependem de fatores externos, expondo o INE ao risco de choques exógenos. Parcerias com fornecedores de dados terceirizados significam não apenas perder o controle da geração de dados, mas talvez também da amostragem e do processamento de dados (MacFeely, 2019, p.37).

Os INE obtêm os seus dados através de inquéritos específicos sob o seu controle e registros administrativas que acessam por meio de mandato legal. No entanto, têm pouco controle ou mandato no que diz respeito aos big data detidos por entidades privadas. Ao estabelecer parcerias com terceiros, os INE perdem o controle global da geração, amostragem e processamento de dados e têm capacidade limitada para moldar os dados produzidos (Kitchin, 2015, p.477).

A percepção de riscos associados à introdução de big data nas estatísticas oficiais levou o Conselho Consultivo de Governança das Estatísticas Europeias (ESGAB) a expressar preocupações com as implicações das parcerias entre INEs e empresas privadas (ESGAB, 2016). Em seu relatório anual de 2016, o ESGAB destacou a necessidade de ‘fortalecer a independência dos INEs’ frente aos provedores de dados e recomendou que: “a próxima revisão do Código de Práticas deverá abordar preocupações éticas associadas à utilização de ‘big data’. O Código deve incluir pelo menos um princípio e indicadores adequados para abordar a relação entre os INE e os fornecedores de dados privados” (ESGAB, 2016, p.15).

As recomendações do ESGAB (2016) foram seguidas pelo Eurostat e resultaram na contratação de “serviços sobre questões éticas, comunicacionais e de habilidades e cooperação metodológica relacionadas ao uso de big data pelas estatísticas europeias” (SOGETI, 2017). No que se refere aos aspectos éticos, o relatório destacou que os INEs envolvidos em parcerias com empresas privadas deveriam “respeitar o princípio de independência profissional, segundo o qual as estatísticas devem ser produzidas e disseminadas de forma independente. Os acordos com empresas privadas devem evitar qualquer pressão comercial para colocar os seus interesses acima do interesse público” (SOGETI, 2017, p.3).

Adicionalmente, o serviço contratado pelo Eurostat também envolveu uma revisão das legislações estatísticas no que se refere ao acesso e uso de big data pelos INEs europeus. De acordo com as conclusões do relatório, na maioria dos casos, os INE estão autorizados por legislação estatística nacional a exigir informações de entidades privadas, no entanto, a base jurídica pode não se estender explicitamente à obtenção de dados de usuários guardados por empresas privadas (SOGETI, 2017a).

Ainda de acordo com o estudo contratado pelo Eurostat, “mesmo quando os INE dispõem dos poderes necessários, as condições relevantes estabelecidas na legislação nacional podem ser complexas e levar a discussões” (SOGETI, 2017a,

p.47). Por exemplo: o tempo e esforço para fazer valer a lei, o envolvimento da Autoridade de Proteção de Dados, regras jurídicas pouco claras, necessidade de conhecimento prévio dos dados para estabelecer os requisitos de acesso, etc. Por estes motivos, o relatório concluiu que “como resultado, alguns INE e parceiros privados [...] têm testado a cooperação voluntária nos últimos anos para o acesso ao big data”. O estudo, entretanto, identificou diversos problemas no modelo de parcerias público-privadas que frustraram as ‘entusiásticas’ expectativas iniciais dos INEs europeus:

Contudo, parece agora que mesmo essa cooperação não funciona necessariamente bem na prática. Mesmo nos casos em que foram inicialmente estabelecidas parcerias entusiásticas, a sua implementação prática fica aquém, muitas vezes porque as empresas receiam partilhar dados pessoais ou confidenciais e comprometer a sua reputação. Para os INE isto é uma grande desvantagem depois de todos os investimentos na parceria. Tais desenvolvimentos nas relações bilaterais frustram a necessidade primária do INE de um acesso confiável. Além disso, o nível de cooperação difere consideravelmente. Muitas entidades privadas mostram-se relutantes em entregar conjuntos de dados aos INE porque estes dados são considerados sensíveis pela sua natureza (por exemplo, informações sobre preços, informações sobre clientes, informações sobre fontes e métodos de produção de dados, quotas de mercado ou propriedade intelectual de dados ou métodos). Outros ainda tentam cobrar quantias consideráveis pela entrega dos dados (argumentando que os dados têm valor comercial e podem ser vendidos facilmente) (SOGETI, 2017a, p.47-48, tradução nossa).

Tendo em vista as dificuldades de acesso aos dados pelos INEs, as recomendações do estudo contratado pelo Eurostat, sugeriram uma estratégia legislativa de curto e médio prazos. No curto prazo, o relatório propôs uma abordagem de base legal leve (*soft-law basis*) que “poderia ser um documento de posição a nível do ESS”, enquanto a médio prazo uma estratégia mais ‘dura’ (*hard-law basis*), “pode ser considerada e preparadas obrigações de direito vinculativo, especialmente se a abordagem de direito não vinculativo não conduzir a resultados concretos”. A adoção das recomendações pelo Eurostat foi imediata, tendo o Sistema Estatístico Europeu (ESS) publicado dois meses depois o “Documento de posicionamento sobre o acesso a dados privados de interesse público” (ESS, 2017), de acordo com o qual:

Projetos-piloto e experiências bem-sucedidas realizadas por institutos de estatística em parceria com detentores de dados privados demonstraram o enorme potencial do acesso a big data para fins estatísticos [...] No entanto, os obstáculos persistentes e a falta de clareza relativamente às condições de acesso a dados privados que são de interesse público dificultam a capacidade de explorar plenamente o potencial [...] A UE deve tomar uma

iniciativa política forte sobre este assunto como uma questão prioritária [...] O EES convida, portanto, a Comissão Europeia a incluir, numa futura proposta, disposições relevantes que estabeleçam um princípio geral de acesso a dados guardados por privados que sejam de interesse público (ESS, 2017, p.2, tradução nossa).

De fato, os dados disponíveis sobre o uso de big data em estatísticas oficiais revelam que a despeito das expectativas iniciais e das estratégias empreendidas pelo campo estatístico, as dificuldades de acesso às fontes de big data resultaram em poucos resultados efetivos na produção de estatísticas oficiais. Segundo MacFeely (2019), o Inventário de Projetos de Big Data da Divisão Estatística da ONU indicava àquela altura que o uso de big data nas estatísticas oficiais ainda era muito limitado em todo o mundo. Embora o inventário incluísse 109 iniciativas de 34 INEs, “vários projetos são especulativos ou aspiracionais, nos quais a fonte de big data ainda não foi identificada ou onde o acesso aos dados (especialmente o de telefones móveis) ainda não foi assegurado” (MacFeely, 2019, p.31). MacFeely (2019) destaca que as estatísticas de preços que utilizam dados da Web estão entre os projetos mais frequentes, uma vez que estas abordagens “normalmente têm menos problemas de acesso aos dados”. A dificuldade dos INE em acessar os dados guardados por privados é um indício do processo contemporâneo de cercamento de dados, o qual, se configura como elemento de uma disputa mais ampla entre Estados e empresas privadas pelo controle de capital informacional.

No que diz respeito à América Latina, uma pesquisa recente (Hub, 2022) com dezesseis INEs também aponta para a restrição no uso de fontes de dados guardadas por privados. Os dados revelam que há sete países que já utilizam big data em estatísticas oficiais na região, quatro países em fase de testes ou experimental e três países em planejamento de testes.⁶⁴ No que diz respeito às estatísticas oficiais, as principais fontes de big data utilizadas são imagens de satélite (33%), web scraping (27%) e medidores de energia (20%). Fontes de big data públicas ou abertas (imagens de satélite, dados web) ou de concessionárias de serviços públicos (medidores de energia e registros de saúde) estão entre as mais utilizadas (87%), enquanto fontes detidas por empresas privadas (dados de cartões de crédito e scanners) representam apenas 13%. Outros dados privados, como os telefones móveis e as redes sociais,

⁶⁴ O rótulo ‘estatísticas experimentais’ vem sendo adotado pelos INEs para diferenciar as estatísticas que estão em fase de testes, das estatísticas oficiais compiladas através de fontes de dados e métodos consolidados.

estão em fase de testes ou apenas numa fase experimental e ainda não foram utilizados em estatísticas oficiais.

Em face da dificuldade de acesso às fontes de big data, verifica-se a partir de 2017 um esforço de revisão legislativa dos sistemas estatísticos nacionais nos países europeus. No Reino Unido a nova lei de dados, *Digital Economy Act* de 2017, passou a garantir o acesso compulsório do Office for National Statistics (ONS) aos registros administrativos governamentais e a fontes de dados guardadas por privados para fins estatísticos. Na União Europeia a UNECE e o Eurostat propuseram em 2018 um novo modelo para a atualização das legislações estatísticas nacionais dos países europeus, incorporando mecanismos para o acesso compulsório aos dados guardados por privados (UNECE, 2018).

Entre 2020 e 2021, na intercorrência da pandemia da COVID-19, a evidência pública da dificuldade de acesso dos Estados a dados de telefonia móvel para a produção de estatísticas de isolamento social marcou uma mudança política considerável no debate sobre big data e estatísticas oficiais (Biancotti et al., 2021). A partir daí verificam-se duas estratégias principais no campo estatístico europeu: i) a incidência junto à Comissão Europeia para a regulamentação do acesso aos dados baseada na ideia de 'reuso' de dados de interesse público guardados por privados (Eurostat, 2022); ii) o desenvolvimento de novas modalidades de compartilhamento de dados com o uso de computação distribuída e tecnologias de melhoria de privacidade (PETs).

Em primeiro lugar, com relação à incidência legislativa, a avaliação do Eurostat (2022) concluiu que as parcerias público-privadas em big data “muitas vezes conduziram a projetos pontuais, inovação e investigação exploratória de ponta. No entanto, raramente conduziram à reutilização sustentável de dados na produção regular de estatísticas oficiais” (Eurostat, 2022, p.6). Entre as limitações do modelo de PPPs voluntárias, o relatório do Eurostat (2022) apontou para a dependência dos INEs quanto à concordância das empresas para a efetivação da parceria, às condições financeiras estabelecidas pelas empresas e à dinâmica competitiva do mercado. De acordo com o Eurostat (2022, p.20), “com base nestas limitações da abordagem voluntária, o Sistema Estatístico Europeu chegou à conclusão de que [...] é necessário incluir obrigações compulsórias no quadro legislativo, obrigando empresas a permitir a reutilização de dados privados”.

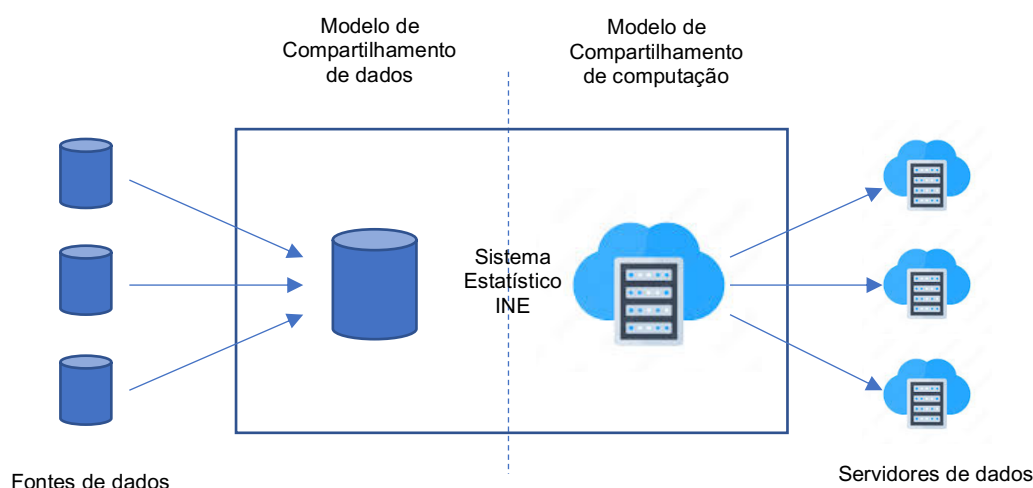
Coerente com esta estratégia, no âmbito das consultas públicas sobre a nova Lei de Dados (Data Act) da União Europeia, o Sistema Estatístico Europeu passou a defender expressamente a regulamentação do acesso compulsório a dados guardados por privados (ESS, 2021). Em contraste, empresas privadas (e.g. IBM, Microsoft, Telefônica, Vodafone, Fujitsu) e associações empresariais (e.g. Câmara de Comércio dos EUA) enviaram objeções à UE contra qualquer abordagem de acesso baseada em requisitos obrigatórios e em defesa do modelo de parcerias voluntárias. (EC, 2021). Finalmente, a nova legislação incluiu um mecanismo que concede aos governos o acesso compulsório a dados guardados por privados em emergências públicas e circunstâncias excepcionais (EC, 2022).

Na disputa pelo acesso aos dados, a defesa das estatísticas oficiais como um ‘bem público’ passou a ser expressamente mobilizada no discurso do Eurostat (2022, p.21): “As estatísticas oficiais são um bem público, o que justifica as atuais obrigações de prestação de informações estatísticas para as empresas. Pela mesma lógica, estas obrigações seriam alargadas aos dados detidos de forma privada, para os quais o benefício público poderá ser muito maior”. Além disso, o apoio dos cidadãos e outros entes públicos também passou a ser mobilizado pelo Sistema Estatístico Europeu. Assim, de acordo com o Eurostat (2022, p.20): “uma larga maioria das partes interessadas (em especial cidadãos e administrações públicas) expressou a opinião de que a partilha de dados deveria ser obrigatória – embora com salvaguardas claras – para casos de utilização específicos que tenham um claro interesse público, incluindo o campo das estatísticas oficiais”.

Em segundo lugar, com relação às novas modalidades de acesso aos dados, segmentos do campo estatístico vinculados à agenda de modernização das estatísticas oficiais, vêm apoiando um novo modelo de computação distribuída ‘para fora’ em oposição ao modelo de concentração dos dados ‘para dentro’. Neste novo modelo as bases de dados brutas não seriam transferidas para o INE, mas ficariam nos servidores dos ‘proprietários’ dos dados ou de um intermediário, e seriam acessados pelos INEs por meio de protocolos seguros com uso de tecnologias de melhoria de privacidade (e.g. anonimização, criptografia) apenas para a realização de processamentos. De acordo com Ricciato, Wirthmann e Hahn (2020, p.6) “Esta tendência para a transferência de computação em vez de dados é reforçada pelo fato de novos dados digitais serem frequentemente produzidos pelo setor privado (dados guardados por privados) e, portanto, podem ser sensíveis do ponto de vista

comercial”. Conforme veremos na próxima seção, esta nova modalidade de compartilhamento de computação com uso de PETs tem sido priorizada no âmbito do Projeto ONU Big Data o que tem envolvido a participação direta da Microsoft e outras empresas privadas.

Tabela 4 - Modelos de Acesso: Compartilhamento de Dados/Computação



Fonte: do autor com base em Ricciato, Wirthmann e Hahn (2020)

Concluimos que as lógicas conflitantes da mercantilização e do interesse público conduziram a um duplo movimento no campo estatístico (Polanyi, 2001)⁶⁵. Por um lado, segmentos buscam a cooperação com empresas privadas interessadas na criação de mercados de dados, com vistas à obter acesso às novas fontes de dados e tecnologias, capitais em disputa no campo. Por outro lado, verifica-se um contramovimento protetivo que opera na defesa do caráter público dos dados, com vistas a garantir o acesso dos Estados Nacionais e seus INEs para a produção de estatísticas oficiais. Conforme veremos na parte empírica deste tese, este duplo movimento também pôde ser identificado nos contextos do campo estatístico

⁶⁵ Conforme vimos anteriormente, o conceito polanyiano de duplo movimento que instrumentalizamos na análise afirma que a sociedade moderna é governada por dois princípios opostos. O movimento para expandir a extensão do mercado autorregulado – ou para separar a economia da sociedade, é frequentemente enfrentado por um contramovimento protetivo de resistência social “que visa a conservação do homem e da natureza, bem como da organização produtiva” (Polanyi, 2001:138). O Estado – também um participante e arquiteto do mercado – toma partido em ambos os polos deste duplo movimento. É esta ‘posição contraditória’ que sustenta a relativa autonomia do Estado face ao interesse econômico imediato.

transnacional e no caso brasileiro, na medida em que as parcerias público-privadas voluntárias e um novo modelo de plataforma de estatísticas oficiais é induzido na região e no país com o financiamento de Big Techs.

Conforme veremos na próxima seção, o debate sobre a usabilidade do big data em estatísticas oficiais também passou a envolver uma série de questões referentes à qualidade das novas fontes. Tais discussões ‘técnicas’ revelam a emergência de um obstáculo epistemológico no campo estatístico, na medida em que a estatística frequentista e a teoria da amostragem se revelam insuficientes para lidar com as novas fontes. Conforme veremos, o discurso de ‘modernização’ das estatísticas oficiais também passa a abranger a defesa de uma ‘mudança cultural’ no campo estatístico por meio da introdução da ‘ciência de dados’ e da gestação de uma nova geração de ‘cientistas de dados’ nos INEs.

2.11 O obstáculo epistemológico: não representatividade do big data, modelagem e estatística bayesiana

A incorporação de big data nas estatísticas oficiais tem suscitado debates no campo estatístico que dizem respeito à qualidade das novas fontes. Na medida em que estas fontes são, em geral, subprodutos de transações digitais, sensores e aplicações que não foram concebidas originalmente com a finalidade de produzir informações estatísticas, o conhecimento e a compreensão do conteúdo e estrutura das fontes de big data passaram a constituir um pré-requisito essencial para a análise de seu potencial de usabilidade estatística. Conforme vimos na seção anterior, a dificuldade de acesso às fontes de dados, em geral cercadas por organizações privadas, impôs uma série de limitações neste sentido. Entretanto, na medida em que o acesso experimental do campo estatístico ao big data foi viabilizado, outras limitações de ordem técnica e metodológica passaram a emergir.

As análises iniciais de algumas fontes de big data por agentes do campo estatístico logo demonstraram uma série de limitações para a compilação de estatísticas oficiais. Daas et al. (2015) apontaram para problemas de perda de dados em sensores de tráfego e dados de posicionamento de telefones celulares decorrente da inatividade de servidores e interrupções de rede. Braaksma e Zeelenberg (2015, p.193) apontaram que “os big data podem ser altamente voláteis e seletivos: a

cobertura da população a que se referem pode mudar de dia para dia, conduzindo a saltos inexplicáveis nas séries temporais”.

Entre as questões de qualidade do big data, os problemas de cobertura e representatividade logo se apresentaram como um obstáculo à sua utilização na produção de estatísticas oficiais dentro do quadro teórico da estatística frequentista. De acordo com um *working paper* do Eurostat (2014, p.4), “a utilização da amostragem probabilística nas estatísticas tradicionais fornece um quadro teórico que garante a confiança nos números com base em erros de amostragem. A maior parte do Big Data disponível não pode ser adaptado a este quadro e outros procedimentos devem ser desenvolvidos”.

A incongruência do big data à teoria frequentista tradicional decorre de alguns motivos, conforme explicam Braaksma e Zeelenberg (2015, p.196), “muitas fontes de Big Data são compostas por dados observacionais orientados por eventos que não são projetados para análise de dados. Faltam-lhes populações-alvo bem definidas, estruturas de dados e garantias de qualidade. Isto dificulta a aplicação de métodos estatísticos tradicionais, baseados na teoria da amostragem”.

A dificuldade, ou mesmo impossibilidade, de relacionar os eventos observáveis a uma população-alvo específica, bem definida e com características conhecidas, caracteriza o big data como uma fonte de dados de natureza não representativa. Conforme ilustram Strujis e Daas (2013; 2014), “os celulares podem ser transportados por outras pessoas que não o proprietário, os veículos que passam por um sensor podem ser veículos particulares ou de empresa, e o que sabemos sobre a população que utiliza as redes sociais? Como essas populações mudam com o tempo? Quão estáveis são?” (Strujis e Daas, 2013, p.6). Ainda de acordo com Kitchin (2015), “os grandes volumes de dados, embora exaustivos, geralmente não são representativos de uma população inteira, uma vez que se referem apenas a quem utiliza um serviço”.

A questão da representatividade das fontes de big data emerge por um lado, como uma importante limitação ao uso destas fontes para a compilação de estatísticas oficiais de acordo com as práticas estatísticas tradicionais apoiadas na teoria da amostragem. Por outro lado, a defesa da inovação dos métodos e práticas estatísticas dos INEs para lidar com o ‘desafio’ das fontes de big data torna-se um importante elemento do discurso de modernização das estatísticas oficiais. Isso pode ser observado, por exemplo, na provocação dos estatísticos do HLG (UNECE, 2013), para quem, com big data é preciso ‘pensar fora da caixa’:

[...] a representatividade é a questão fundamental do Big data. A dificuldade em definir a população-alvo, a população do inquérito e o enquadramento do inquérito põe em risco a forma tradicional como os estatísticos oficiais pensam e fazem inferências estatísticas sobre a população-alvo. Com um inquérito tradicional, os estatísticos identificam uma população-alvo/inquérito, constroem uma estrutura de inquérito para atingir essa população, extraem uma amostra, recolhem os dados, etc. Com o Big data, os dados vêm em primeiro lugar e o reflexo dos estatísticos oficiais seria construir uma caixa! Isto levanta a questão: será esta a única forma de produzir um sistema nacional coerente e integrado de estatísticas oficiais? É hora de pensar fora da caixa? (UNECE, 2013, p.5)

Para os promotores do uso de big data na produção de estatísticas oficiais, 'pensar fora da caixa' passa a significar a utilização de novas abordagens baseadas em modelos orientados por algoritmos desenvolvidas em outros campos de práticas. De acordo com Daas et al. (2015, pp. 258-259), a complexidade do big data "faz com que muitas das abordagens padrão utilizadas nas estatísticas oficiais sejam limitadas em utilidade e desempenho. Os modelos orientados por algoritmos desenvolvidos em campos fora da estatística podem ser mais aplicáveis aqui". Na mesma direção, Braaksma e Zeelenberg (2015) defendem uma 'mudança no paradigma' das estatísticas oficiais para a utilização de big data:

Em primeiro lugar, podemos aceitar os big data como eles são: um indicador imperfeito, mas oportuno, dos fenômenos da sociedade. Esses dados existem e por isso são interessantes. Em segundo lugar, podemos estender esta abordagem por meio de uma modelagem explícita. Novos métodos, como técnicas de aprendizado de máquina, podem ser considerados juntamente com técnicas bayesianas. Os institutos nacionais de estatística sempre se mostraram relutantes em utilizar modelos, exceto em casos específicos, como estimativas de pequenas áreas. Com base na experiência do Statistics Netherlands, argumentamos que os INE não devem ter medo de utilizar modelos, desde que a sua utilização seja documentada e tornada transparente para os utilizadores (Braaksma e Zeelenberg, 2015, p.193, tradução nossa).

Para melhor compreendermos a mudança de paradigma a que se referem Braaksma e Zeelenberg (2015), retornaremos às construções cognitivas da 'urna de Bernoulli' e da 'curva normal', que, conforme vimos anteriormente, subsidiaram a abordagem frequentista aplicada à produção de estatísticas oficiais ao longo de dois séculos.

Em linha com Desrosières (2008), traçamos uma analogia das modernas operações censitárias que emergiram no Séc. XIX, como um esforço de construção de urnas probabilísticas na escala da nação, que garantiam o conhecimento integral

de uma população por meio de sua enumeração exaustiva. O domínio sobre os parâmetros populacionais de uma população conhecida foi o que posteriormente possibilitou aos estatísticos do Séc. XX a aplicação da teoria da amostragem, apoiada na noção de intervalo de confiança que deriva da distribuição normal, para a obtenção de informações confiáveis e mais ágeis por meio de inquéritos amostrais. Deste modo, 'retirando apenas algumas bolas da urna', tornou-se possível inferir com segurança outras características da população. Esta abordagem à inferência estatística se baseia em probabilidades objetivas que derivam de frequências observadas em uma população conhecida. Deste modo, a produção de estatísticas oficiais se consolidou no Séc. XX com uma 'abordagem orientada para o desenho' de inquéritos censitários e amostrais, sob o marco da teoria da amostragem e de acordo com uma epistemologia estatística frequentista e objetivista.

Utilizando novamente a analogia da 'Urna de Bernoulli', agora para o caso do big data, podemos pensar que os estatísticos do Séc. XXI passam a lidar com 'urnas' cuja construção e composição lhes é desconhecida. O desconhecimento da população e de seus parâmetros, ou seja, do espaço amostral contido nestas 'urnas', constitui um obstáculo epistemológico à teoria da amostragem. Notadamente, as novas fontes de dados fazem emergir no campo estatístico oficial um problema similar àquele que se dedicara Bayes na construção de seu teorema, ou seja, o de como inferir a probabilidade de um evento do qual as causas são desconhecidas.

Conforme vimos, a solução da estatística bayesiana para o problema das 'causas desconhecidas' envolve a utilização de dados parciais e crenças subjetivas para estimar uma probabilidade à priori que é utilizada para calcular uma probabilidade à posteriori, com base na verossimilhança das evidências utilizadas. Na medida em que novos dados são obtidos, as crenças iniciais podem ser atualizadas, otimizando a probabilidade estimada. Para a resolução de problemas complexos, tal procedimento se apoia em geral na construção de modelos que tem por fundamento a construção cognitiva da árvore ou rede bayesiana. A árvore possibilita aproximações sucessivas e a atualização dos parâmetros na medida em que novas informações são obtidas. É justamente nesta direção que passam a apontar as propostas de alguns agentes do campo estatístico envolvidos na pauta do big data:

Um caminho a seguir é utilizar o big data junto com informações adicionais e ver se podemos modelar o fenômeno que queremos descrever. Nos últimos anos, tem surgido uma onda nas estatísticas matemáticas com o

desenvolvimento de novos métodos avançados para big data. Eles vêm em vários sabores, como regressão de alta dimensão, técnicas de aprendizado de máquina, modelagem gráfica, ciência de dados e redes bayesianas [...] De maneira geral, pode-se dizer que o big data é um caso em que temos informações insuficientes sobre as relações da fonte de dados com os fenômenos estatísticos que queremos descrever. Muitas vezes, isso é causado pela falta de informações sobre o próprio processo de geração de dados. Os modelos são então úteis para formular suposições explícitas sobre estas relações e para estimar questões de seletividade ou cobertura. Por exemplo, uma forma de reduzir a possível seletividade numa fonte de mídia social poderia ser traçar perfis de contas individuais, a fim de descobrir mais sobre características de fundo. Se conseguirmos determinar se uma conta pertence a um homem ou a uma mulher, seremos capazes de lidar melhor com o preconceito de gênero no sentimento. Técnicas para fazer isso já foram desenvolvidas e estão se tornando cada vez mais sofisticadas. O mesmo se aplica à distribuição etária, ao nível de escolaridade e à localização geográfica. Os problemas de cobertura com fontes individuais de redes sociais podem ser reduzidos através da combinação de múltiplas fontes; e a forma sensata de o fazer é através da utilização de um modelo, por exemplo, o uso de um filtro Bayesiano para reduzir a volatilidade (Braaksma e Zeelenberg, 2015, p.198, tradução nossa).

A construção de modelos estatísticos, conforme defendida por Braaksma e Zeelenberg (2015), apresenta-se, portanto, como um recurso para que os estatísticos possam inferir sobre a construção e composição dessas ‘urnas’ desconhecidas, ou seja, sobre o próprio processo que gerou o big data, e a qual população os dados se referem. Tais modelos estatísticos visam emular as características populacionais desconhecidas do big data utilizando outros dados disponíveis, bem como as crenças subjetivas dos estatísticos para a formulação de ‘suposições’ que permitam relacionar os dados aos fenômenos que se pretende descrever. Neste sentido, a introdução de big data nos INEs passa a exigir não apenas novos conhecimentos e habilidades, mas implica em uma mudança substantiva na abordagem teórica aplicada à produção de estatísticas oficiais. Conforme os próprios autores reconhecem, “os institutos nacionais de estatística sempre foram relutantes em utilizar métodos baseados em modelos nas estatísticas oficiais” (Braaksma e Zeelenberg, 2015, p.198). Não à toa, a sua defesa da modelagem estatística é cercada por salvaguardas:

[...] o uso de modelos deve ser explicitado. Deve ser documentado e tornado transparente para nossos usuários. Além disso, os modelos não devem ser usados indiscriminadamente: não devemos esquecer que o objetivo principal de um INE é descrever e não prescrever ou julgar. Portanto, deveríamos abster-nos de fazer previsões e de utilizar modelos puramente comportamentais. Além disso, devemos ter cuidado para evitar a falha do modelo quando as suposições subjacentes a ele falharem. Portanto, qualquer modelo deverá basear-se em dados efetivamente observados para o período em consideração, que se relacionem com os fenômenos econômicos e sociais que tentamos descrever através de estimativas estatísticas; e a

construção de modelos deve ser acompanhada por extensos testes de especificação (Braaksma e Zeelenberg, 2015, p.201, tradução nossa).

Tudo parece funcionar como se a introdução de big data no processo de produção de estatísticas oficiais conduzisse a uma ruptura epistemológica na teoria e prática estatística incorporada pelos INEs. Esta ruptura pode ser caracterizada por dois obstáculos fundamentais. Em primeiro lugar, passa a ocorrer nos INEs uma confrontação entre os quadros teóricos da estatística frequentista e da estatística bayesiana. Em segundo lugar, verifica-se uma oposição entre a lógica dedutiva da ‘abordagem de desenho’ das estatísticas oficiais e a lógica indutiva da ‘abordagem de modelo’ do big data.

Tabela 5 - Corte Epistemológico: Abordagens Frequentista e Bayesiana.

Característica	Abordagem Frequentista	Abordagem Bayesiana
Interpretação de probabilidade	Frequência relativa	Grau de crença
Conhecimento prévio	Não considerado	Incorporado através de distribuições a priori
Inferência	Baseada em significância estatística (p-valor)	Baseada em probabilidade a posteriori (atualização de crenças)
Amostragem Aleatória	Essencial para garantir a representatividade da população.	Pode ser flexibilizada
Epistemologia	Objetivista	Subjetivista
Foco	Descritivo: Estimar parâmetros populacionais "verdadeiros"	Prescritivo: atualizar crenças sobre os parâmetros à medida que novas informações são coletadas

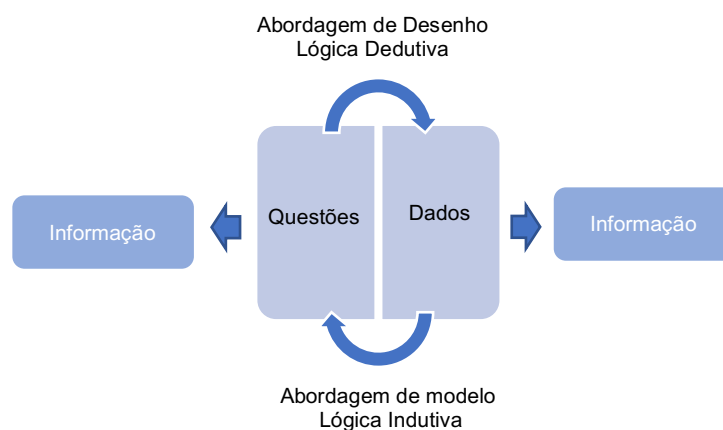
Fonte: do autor.

Conforme vimos, as principais diferenças entre as abordagens frequentista e bayesiana residem na interpretação da probabilidade, na incorporação de conhecimento prévio à análise e na forma de fazer a inferência estatística. De acordo com a abordagem frequentista, a probabilidade é compreendida como a frequência relativa de um evento e tomada como uma observação ‘objetiva’ de um fenômeno ‘real’. Já de acordo com a abordagem bayesiana, a probabilidade de um evento é interpretada como um grau de crença sobre a sua ocorrência, e portanto uma medida que incorpora suposições ‘subjetivas’. Uma diferença fundamental desta abordagem é a possibilidade de incorporação de conhecimentos prévios (dados parciais) que possibilitam estimar uma probabilidade à priori e atualizar esta probabilidade à medida

que novos conhecimentos são obtidos. Isto implica em diferenças importantes na abordagem à inferência estatística. No caso frequentista, a inferência baseia-se na noção de significância estatística (valor p), ou seja, a probabilidade de que os resultados observados correspondam efetivamente ao fenômeno investigado. No caso bayesiano, a inferência ampara-se em probabilidades a posteriori que refletem uma crença atualizada por novas informações. Assim, enquanto para a abordagem frequentista a amostragem aleatória representativa é um aspecto essencial para a inferência estatística, para a abordagem bayesiana estes elementos (representatividade e aleatoriedade) podem ser flexibilizados.

Em segundo lugar, no que se refere às diferenças na abordagem da investigação, a 'abordagem de desenho' tradicionalmente orienta à produção de estatísticas oficiais pode ser considerada, em grande medida, uma abordagem dedutiva, na medida em que parte de uma definição prévia de questões que se deseja responder com os dados coletados. Essa definição guia todo o processo de coleta e análise, direcionando o foco da pesquisa. Tais questões geralmente se baseiam em hipóteses e teorias pré-existentes sobre o tema de interesse que orientam o desenho da coleta de dados junto a uma população-alvo. Os dados obtidos portanto constituem um meio para testar as hipóteses e verificar sua sustentação. Neste sentido, a análise dos dados coletados segue uma lógica dedutiva, partindo de premissas gerais (hipóteses) para verificar se elas se aplicam aos casos específicos (dados). Finalmente, o objetivo final da abordagem de desenho é generalizar os resultados obtidos para uma população de interesse, por meio de probabilidades objetivas, assumindo que a amostra coletada é representativa.

Figura 12 - Indução e Dedução: Abordagens de Modelo e de Desenho



Fonte: do autor.

A 'abordagem de modelo', por outro lado, caracteriza um processo indutivo, enquanto os modelos estatísticos são construídos com base nos dados observados. Ao invés de partir de problemas e hipóteses pré-definidos, a análise explora os dados para identificar padrões e relações entre as variáveis que os modelos buscam explicar, relacionando-os a fenômenos que pretendem descrever. O objetivo desta abordagem também é a generalização dos resultados para uma população de interesse. Entretanto, conforme vimos, há dificuldades adicionais para a inferência estatística no caso do big data, na medida em que, em geral, não é possível associar os dados à uma população específica. Neste caso, a própria relação dos resultados com uma população de interesse é igualmente objeto de modelagem, o que pode incluir, no caso dos modelos bayesianos o recurso à probabilidades subjetivas.

De acordo com Kitchin (2015), uma potencial inversão do processo tradicional de produção de estatísticas oficiais (do desenho de inquéritos para a modelagem dos dados) poderia representar riscos políticos e institucionais consideráveis, na medida em que os dados passassem a conduzir a agenda e missão dos INEs:

Tradicionalmente, o motor da produção de estatísticas tem sido uma preocupação ou questão fundamental; dados foram gerados para responder a um conjunto específico de consultas. Na era dos big data, há potencial para que esta situação seja revertida, com a abundância e o custo-benefício dos big data definindo a agenda do que é medido. Em outras palavras, as estatísticas oficiais podem desviar-se para seguir os dados, em vez de os dados serem produzidos para a compilação de estatísticas oficiais. Para além de ter implicações no trabalho institucional dos INE, tal medida representa uma clara ameaça à integridade e à qualidade das estatísticas oficiais. É absolutamente crítico, portanto, que os INE permaneçam centrados nas questões para as quais os dados são utilizados, avaliando a adequação dos big data à sua atividade principal, em vez de deixar que os big data conduzam a sua missão. (Kitchin, 2015, p.477)

Enquanto modifica as práticas estatísticas, a ruptura epistemológica implicada na introdução de big data nos INEs também produz efeitos na estruturação do campo estatístico. Estes efeitos estão vinculados ao estabelecimento de interseções com o campo de práticas algorítmicas que resultam na internalização de novas práticas e de um novo perfil profissional que lhe é próprio, o cientista de dados. Na próxima seção analisamos como este processo ganha materialidade na medida em que o discurso de modernização das estatísticas oficiais passa a difundir a ideia de que é necessária uma 'mudança cultural' nos INEs para 'encorajar o empreendedorismo e a inovação' (UNECE, 2013a, p.8).

2.12 Habitus algorítmico e mudança cultural: a introdução da ciência de dados nos INEs

A ideia de que o novo contexto de dataficação requer uma ‘mudança cultural’ nos INEs passou a ser um lugar comum em uma série de artigos e eventos do campo estatístico que abordam o tema do big data (Grommé; Cakici; Takala, 2021, p.238). É interessante observar que tais argumentos partem dos mesmos agentes do campo estatístico que, conforme analisamos nas seções anteriores, apoiaram uma estratégia de modernização das estatísticas oficiais via parcerias público-privadas e aproximação com os provedores de big data. A estratégia defendida por estes agentes visa o ‘cultivo’ de uma nova geração de cientistas de dados, oriunda do setor privado, nos INEs. Esta estratégia pode ser observada, entre outros, no artigo introdutório do HLG sobre big data para estatísticas oficiais:

A ciência de dados associada ao Big data que está emergindo no setor privado não parece estar ainda ligada à comunidade estatística oficial. Os INEs terão de realizar rastreamentos internos e nacionais para identificar onde estão os cientistas de dados e conectá-los à área de estatísticas oficiais. (UNECE, 2013, p.5)

A ideia de que “para usar Big Data, são necessários estatísticos com uma mentalidade diferente e novas habilidades” (UNECE, 2013), e de que tais atributos desejáveis — e.g. afinidade com TI, conhecimento de programação, uso de técnicas de modelagem, extração de teorias a partir dos dados — corresponderiam aos emergentes ‘cientistas de dados’, em geral provenientes da iniciativa privada, é um mantra que passou a ser propagado pela agenda de modernização das estatísticas oficiais. Conforme ilustrado nas citações a seguir:

Os modernos cientistas de dados podem estar mais bem equipados do que os estatísticos tradicionalmente treinados. Provavelmente mais importante é a necessidade de uma mentalidade diferente, uma vez que a utilização de Big Data pode implicar em uma mudança de paradigma, incluindo um uso aumentado e modificado de técnicas de modelação. (Strujis et al. 2014, p.3)

Talvez mais importante seja a atitude das pessoas envolvidas. Trabalhar com Big Data requer uma mentalidade aberta e a capacidade de não ver todos os problemas a priori em termos de teoria de amostragem, uma vez que Big Data é mais semelhante a grandes conjuntos de dados observacionais. O termo “cientista de dados” foi cunhado para designar investigadores com as competências identificadas acima [...] já que eles são mais orientados para a prática e estão mais acostumados a derivar teorias a partir de dados. (Daas et al. 2015, p.259)

A ciência de dados se tornará cada vez mais importante para extrair informações significativas dos dados. Os INEs precisarão contratar cientistas e engenheiros de dados e precisarão treinar seu pessoal nessas áreas em rápido desenvolvimento [...] a colaboração das estatísticas oficiais com o setor privado, a academia e a sociedade civil será a nova forma de trabalhar. (Schweinfest e Jansen, 2021, p.2)

Em sua defesa do ‘cultivo’ de uma nova geração de cientistas de dados nos INEs, Strujis e Daas (2013) foram os primeiros a se referir especificamente à necessidade de uma determinada ‘orientação mental’ e ‘habilidade comportamental’ para lidar com big data. Em seu artigo direcionado à conferência dos Estatísticos Europeus, os autores revelaram ainda receios quanto a uma provável resistência interna nos INEs à introdução destes novos profissionais, o que intitulam de ‘uma questão cultural’. Notadamente, propõem como estratégia de ação a criação de núcleos de cientistas de dados no interior das instituições estatísticas nacionais para superar tal resistência:

Para trabalhar com Big Data são necessárias competências técnicas específicas, tais como competências informáticas avançadas, um domínio justo de matemática e estatística, competências de modelação e competências de engenharia de dados. Mas igualmente importantes são a orientação mental e as competências comportamentais [...] Para este tipo de profissional foi cunhado o termo cientista de dados. Contudo, não é evidente que a cultura dos INE consiga absorver suavemente este tipo de profissional. Uma forma de lidar com essa questão cultural é criar um ou mais núcleos de cientistas de dados trabalhando com Big Data, e deixar esses núcleos crescerem, o que será um processo natural caso tenham sucesso. (Strujis e Daas, 2013, p.8)

A estratégia proposta por Strujis e Daas (2013) foi inicialmente implementada em 2016 nos Países Baixos com a criação do ‘Center for Big Data Statistics’ vinculado ao Netherlands Statistics (CBS)⁶⁶. Iniciativa semelhante foi posteriormente adotada pelo Office of National Statistics (ONS)⁶⁷ do Reino Unido em 2017 com a criação do ‘Data Science Campus’. Grommé, Cakici e Takala (2021), em um trabalho recente sobre o processo de dataficação nos INEs europeus, realizaram um estudo etnográfico que investigou estas duas iniciativas. A pesquisa revelou a existência de um conjunto de lutas em curso no interior do campo estatístico pela “valorização relativa do capital cultural e do habitus necessários para trabalhar com big data”

⁶⁶ <https://www.cbs.nl/en-gb/about-us/innovation/nieuwsberichten/big-data/cbs-launching-center-for-big-data-statistics>

⁶⁷ <https://datasciencecampus.ons.gov.uk/>

(Grommé; Cakici; Takala, 2021, p.239), processo caracterizado como uma competição entre os estatísticos nacionais e a facção emergente de cientistas de dados nos INEs:

Embora os estatísticos ocupem há muito tempo uma posição dominante, os cientistas de dados são uma facção emergente que desafia esse domínio neste campo. Para os cientistas de dados, o que está em jogo é o reconhecimento do big data e dos métodos analíticos relacionados como legítimos e autorizados e, por sua vez, o capital cultural e simbólico que isso conferirá. Para os estatísticos, o que está em jogo é proteger e promover a sua autoridade e posição em relação uns aos outros e a esta facção. (Grommé; Cakici; Takala, 2021, p.241)

De acordo com Grommé, Cakici e Takala (2021, p. 242), estatísticos nacionais e cientistas de dados distinguem-se pela “valorização e apropriação de certas formas de capital cultural em detrimento de outras”, o que envolve tanto novos conjuntos de habilidades e competências profissionais, como o reconhecimento “do que valorizar e qual ética profissional apoiar”. Em sua pesquisa, os autores verificaram que “para alguns estatísticos, a resposta adequada ao que consideram ser o desafio do big data é a de que os INE precisam se tornar mais semelhantes aos seus concorrentes do setor privado”. Para Grommé, Cakici e Takala (2021), a “modernização das estatísticas frequentemente refere-se a adoção de uma mentalidade do setor privado e de um habitus empreendedor”. Por outro lado, de acordo com os autores, “quase com a mesma regularidade, o futuro da profissão também foi definido em contraste com os valores defendidos no setor privado, através do reforço e da defesa de valores de longa data do serviço público na produção de estatísticas oficiais [...] como a confiança, a responsabilidade e função públicas e a legitimidade democrática”. Finalmente, os resultados da pesquisa apontaram para a emergência de tensões explícitas nos INEs entre a disposição empreendedora dos cientistas de dados e o habitus do serviço público incorporado pelos estatísticos de Estado.

Coerente com nosso conceito de campo de práticas algorítmicas, compreendemos a disposição ‘empreendedora’ dos cientistas de dados como um elemento do ‘habitus algorítmico’ incorporado por estes agentes. De acordo com os achados de Grommé, Cakici e Takala (2021), apoiados em outros estudos sobre os agentes do setor de tecnologia (Irani, 2019) e ciência de dados (Mackenzie, 2013), esta disposição inclui uma forte crença na inovação tecnológica como o principal motor da mudança social e um sentimento de otimismo e urgência para alcançar e

realizar esta inovação. A prática da colaboração e experimentação são amplamente valorizadas por estes sujeitos na medida em que possibilitam, mais do que um resultado concreto e imediato, a possibilidade de geração de valores futuros, invariavelmente relacionados ao mercado, sejam valor financeiro ou valor social.

Grommé, Cakici e Takala (2021) apontam que a incorporação da disposição empreendedora dos cientistas de dados no campo estatístico oficial é acompanhada por estímulos a um 'ethos filantrópico' por meio da aplicação das competências de otimização dos dados em iniciativas para o 'bem comum'. Estas estratégias de justificação e encobrimento de motivações profissionais e mercadológicas são reforçadas por meio de competições e *hackathons*, que envolvem a retórica de abordar os cientistas de dados "como 'pessoas maravilhosas': um grupo altamente desejável, equipado com uma combinação única de competências para enfrentar os desafios sociais dos nossos tempos" (Grommé; Cakici; Takala, 2021, p.246).

Conforme veremos na segunda parte deste trabalho o discurso de modernização das estatísticas oficiais com big data que emergiu no contexto europeu ganhou uma escala global a partir de 2015 com a sua incorporação na pauta da Comissão Estatística da ONU. Veremos como iniciativas articuladas pela estratégia da 'revolução de dados para o desenvolvimento' com apoio de Big Techs passaram a incentivar a introdução das fontes de big data e a formação de uma nova geração de cientistas de dados nos INEs do Sul-Global. Por meio de grupos de trabalho com especialistas da iniciativa privada, festivais de dados, hackathons, etc. verificam-se esforços para moldar um novo perfil profissional nos INEs do Sul-Global, o que conforme veremos, conduz a controvérsias e disputas, na medida em que este novo conjunto de disposições entra em choque com o habitus dos estatísticos de Estado.

Conclusão: Obstáculos ontológicos (dados) e epistemológicos (métodos) à transição para o regime de dataficação conduzem a um duplo movimento no campo estatístico

Neste capítulo, vimos como a crise do regime de acumulação fordista-keynesiano deu margem a uma nova crise de governamentalidade que abriu caminho para o neoliberalismo sob um novo regime de acumulação financeirizada. Neste contexto, apontamos um conjunto de transformações no campo estatístico que conduziram à configuração de um regime estatístico de comensuração: i) a influência

da lógica de financeirização sobre as práticas estatísticas com a disseminação de uma ‘política de indicadores’; **ii**) a ‘virada empresarial’ dos INEs com a introdução de uma nova linguagem e práticas corporativas no processo produtivo das estatísticas oficiais; **iii**) a emergência de novos agentes quantificadores internacionais e privados que passam a desafiar a lógica tradicional das estatísticas oficiais, limitadas a um espaço nacional e impulsionadas por instituições estatais; e **iv**) o impulsionamento das Tecnologias de Informação e da Comunicação que implicaram em maior descentralização e agilidade no processo estatístico com o uso de registros administrativos.

Desta análise, concluímos que as transformações no campo estatístico ao longo do período neoliberal, por um lado, abriram flancos que tornaram os sistemas estatísticos nacionais mais permeáveis ao mercado, mais responsivos às inovações tecnológicas e mais susceptíveis à influência de agentes internacionais e privados. Por outro lado, a despeito destas mudanças, vimos como os INEs se mostraram resilientes e capazes de se adaptar e se reposicionar no novo contexto, preservando a sua autonomia e o caráter público e nacional das estatísticas oficiais por meio de contramovimentos protetivos. Vimos ainda como no caso brasileiro, a resiliência às reformas neoliberais vinculou-se à sua atenuação pelos segmentos desenvolvimentistas e por um novo posicionamento do IBGE no contexto da redemocratização, reforçado pela organização e resistência de seus quadros.

A crise do regime de acumulação financeirizada, evidenciada a partir de 2007/2008, aponta para uma crise contemporânea da governamentalidade neoliberal e para uma reconfiguração do regime estatístico. Tal reconfiguração se dá no contexto de uma nova economia orientada aos dados e de um novo regime de acumulação do capital cada vez mais pautado pelas plataformas e finanças digitalizadas. Este novo contexto implica em duas transformações sem precedentes para o campo estatístico. Em primeiro lugar, a nova economia política dos dados, orientada pela comoditização e mercantilização, desafia a concepção tradicional dos dados como bens públicos. Em segundo lugar, a epistemologia bayesiana e a modelagem indutiva associada às novas fontes de dados não representativas, desafia as práticas estatísticas tradicionais, fundamentadas no frequentismo, na teoria da amostragem e na abordagem dedutiva do desenho de inquéritos.

Na medida em que o novo campo algorítmico, que congrega Big Techs, empresas de tecnologia e um conjunto de produtores e intermediários privados de

dados e informações, passa a ameaçar a hegemonia informacional dos Estados e a relevância das estatísticas oficiais, o campo estatístico é pressionado a se ‘modernizar’ e a se engajar com as fontes de big data. O cercamento privado dos dados, condição para a constituição de novos mercados, veda o acesso dos Estados aos novos capitais informacionais e as estatísticas oficiais passam a constituir um potencial nicho de mercado para o campo algorítmico. Configura-se assim um novo espaço de lutas em que se verifica a emergência de um duplo movimento no campo estatístico.

Por um lado, a articulação de um discurso de ‘modernização das estatísticas oficiais’ caracteriza um movimento pró-mercado, que conduz ao estabelecimento de intersecções entre os campos estatístico e algorítmico, por meio de uma agenda de parcerias público-privadas e incentivos à gestação de uma nova geração de cientistas de dados nos INEs, mais receptivos à inovação e à incidência do mercado. Por outro lado, verifica-se um contramovimento protetivo da autonomia dos INEs e dos métodos estatísticos tradicionais, por meio da regulamentação do acesso às fontes de dados como bens públicos, visando a sua incorporação controlada sob a governança dos estatísticos de Estado e dos sistemas estatísticos nacionais.

Concluimos que o processo contemporâneo de transição para um novo regime estatístico-algorítmico de dataficação é caracterizado pelo duplo movimento do campo estatístico e por uma série de lutas que envolvem: i) disputas entre estratégias de mercado e estratégias legislativas para o acesso aos novos capitais informacionais; ii) confronto entre os métodos e modelos tradicionais de produção de estatísticas oficiais e os novos métodos e modelos aplicados à análise de big data; e iii) conflitos entre o habitus empreendedor e algorítmico dos cientistas de dados e o habitus burocrático e científico dos estatísticos de Estado.

Conforme veremos na segunda parte da tese, o discurso de modernização das estatísticas oficiais com big data que emergiu no contexto europeu, ganhou uma escala global a partir de 2015 com a sua incorporação na pauta da Comissão Estatística da ONU, no contexto de implementação da nova agenda global de 2030 e dos Objetivos do Desenvolvimento Sustentável (ODS). Na medida em que a União Europeia gradualmente se afasta da agenda de parcerias público-privadas em big data e busca reforçar seus sistemas estatísticos nacionais pela via da regulação, verificamos a disseminação da agenda pró-mercado para o Sul-Global, por meio de um projeto de plataformização das estatísticas oficiais com apoio de Big Techs.

PARTE II

3. O PROJETO ONU BIG DATA E A PLATAFORMIZAÇÃO DAS ESTATÍSTICAS OFICIAIS

Todos concordam que educação, saúde e segurança são 'resultados desejados', porém como alcançá-los? No passado, quando deparamos com a dura escolha política de consegui-los ou por meio do mercado ou por meio do Estado, as linhas do debate ideológico eram claras. Hoje, quando se supõe que a escolha seja entre o digital e o analógico, ou entre a retroalimentação dinâmica e a lei estática, não há mais essa clareza ideológica — como se a própria escolha de como alcançar esses 'resultados desejados' fosse apolítica e não nos obrigasse a optar entre concepções de vida comunitária distintas e muitas vezes incompatíveis. Ao assumir que o mundo utópico dos ciclos infinitos de retroalimentação é tão eficiente que transcende a política, os proponentes da regulação algorítmica caem na mesma armadilha dos tecnocratas do passado. Sim, esses sistemas são extremamente eficientes [...] isso não significa que seu regime não possa ser avaliado fora do pântano linguístico da eficiência e da inovação, isto é, a partir de referências políticas, e não econômicas (Morozov, 2020, p. 87-88).

Neste capítulo, a hipótese de que a introdução de novas fontes de dados e métodos no processo de produção das estatísticas oficiais conduz a uma interseção entre os campos de práticas estatísticas e algorítmicas e a uma transição para um novo regime estatístico-algorítmico de dataficação, é testada por meio da investigação da Plataforma Global das Nações Unidas (UNGP). A Plataforma Global é um projeto voltado à produção de estatísticas oficiais com uso de big data e métodos algorítmicos. A iniciativa foi desenvolvida sob os auspícios da Comissão Estatística das Nações Unidas (UN-StatCom) e é coordenada pela sua Divisão Estatística (UNSD), no escopo do projeto ONU Big Data.

Analisamos como os Objetivos de Desenvolvimento Sustentável (ODS) foram instrumentalizados pela estratégia da 'revolução dos dados para o desenvolvimento' com vistas a promover o uso de big data em estatísticas oficiais. A estratégia é caracterizada pela adoção de princípios de mercado e incentivos ao envolvimento do setor privado no setor estatístico estatal por meio de Parcerias Público-Privadas (PPPs). Nossa investigação segue o percurso de ações de incidência promovidas por Big Techs e outras corporações privadas na Comissão e Divisão Estatística da ONU. Analisamos discursos e articulações de alto nível entre agentes transnacionais dos campos estatístico e algorítmico e desdobramos controvérsias que derivam da tentativa de implantação de um mercados de dados para estatísticas oficiais, inspirado no modelo de plataforma das Big Tech.

A adoção de princípios de mercado e das lógicas de comodificação dos dados e de plataforma das estatísticas oficiais é analisada criticamente do ponto de

vista dos riscos de subsunção dos sistemas estatísticos nacionais dos países do Sul-Global a um mercado de dados, métodos e tecnologias controlado por grandes corporações privadas e governos do Norte-Global. A análise se vale do conceito de ‘colonialismo de dados’ (Thatcher et al. 2016; Couldry; Mejias, 2019) para investigar “assimetrias de poder inerentes às formas contemporâneas de mercantilização de dados” (Thatcher et al. 2016, p.992) e as condições de possibilidade para o surgimento de políticas neocoloniais de dados no campo das estatísticas oficiais, que podem potencialmente comprometer a soberania nacional dos países do Sul-Global.

A primeira sessão discute a afinidade da agenda de desenvolvimento global pós-2015 com o ‘neoliberalismo pragmático’ (Sandbrook, 2000) e apresenta o contexto em que a promoção de fontes de Big Data e novos métodos para estatísticas oficiais foram introduzidos na ONU. A segunda e terceira sessões investigam como o discurso da revolução dos dados instrumentalizou os ODS para promover uma agenda de mercado no campo estatístico global com a incidência e financiamento de Big Techs. Os percursos e resultados das estratégias que deram origem à Plataforma Global e as controvérsias e desdobramentos que derivam de sua implantação são analisados na terceira e quarta sessões.

3.1 A abordagem neoliberal na agenda de desenvolvimento global pós-2015

Conforme analisamos anteriormente no capítulo 2, a intensificação dos processos de globalização no final do Séc. XX foi acompanhada pelo fortalecimento de espaços internacionais e transnacionais de quantificação. Do ponto de vista das Organizações Internacionais, indicamos como a partir de meados dos anos 1990, verificou-se a emergência de um novo modelo de ‘governança global’ que tem como fundamento o estabelecimento de metas nacionais calculáveis e comparáveis por meio de indicadores estatísticos e rankings.

Este modelo de ‘governança global’ foi embasado em uma nova concepção de ‘desenvolvimento humano’ (Sen, 1999) que se apresentou como uma crítica ao economicismo e como uma defesa de uma nova abordagem multideterminada do desenvolvimento. Do ponto de vista político e econômico, o modelo de ‘desenvolvimento humano’ subordina as políticas nacionais e globais de desenvolvimento à lógica do mercado autorregulado e eleva o livre mercado ao lócus central de realização das capacidades humanas. Seguimos Sandbrook (2000) na

caracterização deste modelo como uma nova roupagem, mais sofisticada, do projeto neoliberal, a qual intitula de ‘neoliberalismo pragmático’.

Do ponto de vista do campo estatístico transnacional, conforme analisamos anteriormente, a ‘virada estatística’ na ‘governança global’ implicou em uma participação mais pronunciada dos órgãos técnicos da Comissão e Divisão Estatística da ONU⁶⁸ nos processos de negociação das políticas de desenvolvimento e na coordenação estatística global, por meio da agenda dos Objetivos de Desenvolvimento do Milênio (ODM). Apontamos que, enquanto quadro normativo para as políticas nacionais, os ODM implicaram em uma série de efeitos políticos e cognitivos problemáticos.

Na primeira década do Séc. XXI, verifica-se uma aprofundamento das teses do neoliberalismo pragmático e da ‘governança global’ por indicadores, na medida em que a concepção de desenvolvimento humano passa a englobar as questões ambientais sob o paradigma da sustentabilidade. Em 2015, no âmbito da ONU, a promoção de um acordo global em torno da Agenda 2030 propõe 17 Objetivos de Desenvolvimento Sustentável em substituição aos 8 Objetivos de Desenvolvimento do Milênio (ODM). Para a “erradicação da pobreza em todas as suas formas e dimensões” (UN, 2015, p.2), os ODS se desdobram em 169 metas e em algumas centenas de indicadores.

Com os ODS, a comensuração generalizada característica da governamentalidade neoliberal encontra o seu clímax. Não deixa de ser curioso observar como o sonho alimentado por Quetelet e pelos primeiros estatistas do Séc. XIX é aqui revivido em sua máxima plenitude. Conforme vimos no capítulo 1, a ilusão de uma governança técnica, apolítica e planetária pelos números faz parte das origens genéticas do internacionalismo estatístico e tudo parece funcionar como se ainda reverberasse em concepções contemporâneas como os ODS e o big data.

Analistas têm apontado diversos aspectos críticos nas premissas e no conteúdo dos ODS que comprometem os próprios objetivos que a agenda de desenvolvimento pretende alcançar: há retrocessos no escopo em relação a agendas

⁶⁸ A Divisão Estatística das Nações Unidas (UNSD) é subordinada ao Departamento de Assuntos Econômicos e Sociais das Nações Unidas (DESA) e funciona como mecanismo central do Secretariado da Organização das Nações Unidas para suprir as necessidades estatísticas e coordenar as atividades do sistema estatístico mundial. A Divisão é supervisionada pela Comissão Estatística das Nações Unidas, criada em 1947, mais alto órgão decisório para coordenar as atividades internacionais de estatística, que reúne os chefes dos Sistemas Estatísticos Nacionais dos Estados-membros. <https://unstats.un.org/home/>

anteriores (Koehler, 2017), verifica-se a predominância de uma episteme de mercado nos fundamentos da agenda (Weber, 2017), há uma influência assimétrica de grandes corporações privadas no conteúdo e implementação da agenda (Scheyvens et. al, 2016) e o aprofundamento de uma ortodoxia do desenvolvimento (Gabay; Ilcan, 2017).

Em uma análise retrospectiva das agendas de desenvolvimento anteriores da ONU (1960 – 2000), Khoeler (2017) enfatiza que muitos pontos importantes foram perdidos em sua tradução para os ODM e posteriormente para os ODS. Esses pontos incluem: i) o trabalho decente deixado de lado nas agendas; ii) as recomendações para aumentar os gastos sociais e a proteção contra cortes no orçamento público foram retiradas; iii) o silêncio da agenda sobre o tema da regulamentação das corporações transnacionais, e iv) a não abordagem da assimetria das relações de poder nos contextos nacional e transnacional.

Para Weber (2017), a agenda pós-2015 consolida uma lógica formulada desde os anos 1980 que assume o mercado como princípio organizador do desenvolvimento (Weber 2017, p.408). Essa abordagem está implícita no conteúdo dos ODS. A ausência de compromissos para assegurar garantias universais permite que as necessidades vitais sejam tratadas como mercadorias, priorizando os interesses comerciais sobre os direitos coletivos. O compromisso de “não deixar ninguém para trás” é, portanto, apresentado como “um discurso estrategicamente utilizado para justificar a implementação de um projeto político altamente problemático como marco para o desenvolvimento global” (Weber, p.399).

Shceyvens et. al. (2016) destacam que uma das mudanças mais significativas introduzidas pelos ODS foi ter dado destaque ao papel do setor privado na definição da agenda global de desenvolvimento. Isso se deveu em parte à crise financeira global de 2007/08 e à subsequente restrição dos orçamentos públicos para o desenvolvimento. O setor empresarial assumiu um papel de destaque na agenda 2030 neste contexto. Isso resultou em uma visão de desenvolvimento na qual o papel dos governos seria o de criar ambientes propícios para o desempenho do setor privado.

Uma característica discursiva comum das agendas globais é a de que o processo de desenvolvimento deve atuar para animar as capacidades dos países em desenvolvimento tornando-os sujeitos ativos de acordo com uma determinada visão de mundo que privilegia o livre mercado, o que Gabay e Ilcan (2017) denominaram de

‘ortodoxia do desenvolvimento’. Os autores destacam três temas relevantes para a compreensão da ortodoxia dos ODS: parcerias, capacitação e big data.

Na agenda pós-2015, o discurso das parcerias atua para ocultar e reproduzir relações de poder desiguais, sancionando uma forma particular de ação baseada na racionalidade liberal. Iniciativas de capacitação “muitas vezes funcionam para governar grupos e populações, tornando-os atores autorregulados e responsáveis por meio de sua participação nas promessas de relações baseadas no mercado” (Gabay e Ilcan 2017, p.479). O big data, por sua vez, complementa a promessa de “não deixar ninguém para trás”, introduzindo a ideia de que “ninguém deve ser invisível”. Nesses termos, a parceria e a capacitação para o uso de Big Data no contexto dos ODS abrem espaços para a abertura de novos mercados de dados e informações em torno dos objetivos de desenvolvimento nos países do Sul-Global. Finalmente, o apelo à inovação e ao empoderamento dos cidadãos para tomar melhores decisões por meio dos dados “destinam-se a induzir o movimento em direção a certos tipos de subjetividade pró mercado” (Gabay e Ilcan, p.482).

Parcerias, capacitação e big data são elementos centrais da ‘revolução de dados para o desenvolvimento’, discurso que foi articulado no âmbito dos ODS. A próxima sessão investiga como a revolução dos dados passou a configurar uma agenda dentro da Comissão Estatística das Nações Unidas (StatCom) para intensificar o uso de big data em estatísticas oficiais e o envolvimento do setor privado no campo estatístico por meio de Parcerias Público-Privadas (PPPs).

3.2 O discurso da revolução de dados

A primeira menção ao termo ‘revolução de dados’ dentro da ONU foi articulada em 2013 pelo *High-Level Panel of Eminent Persons (HLP)*, um grupo constituído para oferecer recomendações ao secretário-geral da ONU sobre a agenda de desenvolvimento pós-2015. Seu relatório enfatiza: “Também pedimos uma revolução de dados para o desenvolvimento sustentável, com uma nova iniciativa internacional para melhorar a qualidade das estatísticas e informações disponíveis para pessoas e governos” (UN-HLP 2013, p.21).

As recomendações do HLP foram seguidas em 2014 pelo *Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG)*, grupo constituído pela ONU para identificar medidas destinadas a fechar lacunas de dados

e fortalecer as capacidades estatísticas nacionais para monitorar os ODS (UN-IEAG, 2014). O relatório '*A world that counts*' do IEAG é um manifesto político que consolida o discurso e a estratégia da revolução dos dados e apresenta diretrizes para sua implementação. O diagnóstico efetuado neste relatório destaca limitações das estatísticas dos governos nacionais para prover os dados necessários para o monitoramento dos indicadores dos ODS. Os problemas apontados incluem defasagem de tempo, desagregação insuficiente, invisibilidade de certos grupos populacionais, viés de gênero e riscos de interferência política dos governos. Finalmente, o relatório aponta para a necessidade de ampliação da participação do setor privado no manejo dos dados e informações.

As agências com mandato para coletar informações públicas nem sempre estão bem adequadas para garantir que suas informações sejam usadas pelas partes interessadas, enquanto a sociedade civil e o setor privado podem desempenhar um papel crítico na tradução de dados em uma forma que sejam mais facilmente utilizáveis. (UN-IEAG, 2014, p.15, tradução nossa.)

De acordo com os especialistas do IEAG, “muitos países ainda têm dados ruins, os dados chegam tarde demais e muitos problemas ainda não são cobertos adequadamente pelos dados existentes” (UN-IEAG, 2014, p.11). Neste sentido, o relatório conclui que para alcançar os ODS, os países, especialmente do Sul-Global, devem ‘abraçar a revolução dos dados’ e modernizar seus Institutos Nacionais de Estatística por meio de uma forte colaboração com o setor privado.

São necessárias novas instituições, novos atores, novas ideias e novas parcerias, e todos têm algo a oferecer à revolução de dados. Os institutos nacionais de estatística, os guardiões tradicionais dos dados públicos para o bem público, continuarão a ser fundamentais para todos os esforços governamentais para aproveitar a revolução de dados para o desenvolvimento sustentável. Para desempenhar este papel, contudo, terão de mudar, e mais rapidamente do que no passado, e continuar a adaptar-se, abandonando processos de produção dispendiosos e complicados, incorporando novas fontes de dados [...] Em muitos casos, serão necessários investimentos técnicos e financeiros para permitir que essas mudanças aconteçam, e uma forte colaboração entre as instituições públicas e o setor privado pode ajudar as agências oficiais a saltar diretamente para novas tecnologias e formas de fazer as coisas (UN-IEAG, 2014, p.9, tradução nossa).

É interessante observar como a estratégia da revolução de dados instrumentalizou o discurso de modernização das estatísticas oficiais — que conforme vimos anteriormente, já vinha sendo articulado no contexto europeu a partir de 2010

— e os ODS para a promoção de novos arranjos institucionais que passam a envolver a participação de corporações privadas no campo estatístico transnacional. Notadamente, o relatório do IEAG fez uma série de recomendações para que a ONU passasse a exercer a liderança na criação de um ‘consenso global de dados para o desenvolvimento’. As propostas incluíam, entre outras, “alavancar os recursos e a criatividade do setor privado, incluindo uma análise de sugestões para a criação de incentivos para investimentos do setor privado, considerando as expectativas das empresas em termos do horizonte temporal e retornos” (UN-IEAG, 2014, p.25). O principal instrumento proposto pelo IEAG para a implementação da revolução de dados é a criação de uma Parceria Global de Dados de Desenvolvimento Sustentável (GPSDD).

De acordo com as recomendações do relatório, a nova organização, GPSDD, seria responsável pela mobilização e coordenação das iniciativas e instituições para alcançar os objetivos da revolução de dados, incluindo, a intermediação de parcerias público-privadas e a realização de um Fórum Mundial de Dados e eventos associados como festivais de dados e hackathons para “promover o envolvimento regular entre produtores e usuários de dados privados, públicos e comunitários” (IEAG, 2014, p.26).

A GPSDD poderia promover diversas iniciativas, tais como [...] intermediar parcerias público-privadas globais importantes com empresas privadas e organizações da sociedade civil para o compartilhamento de dados. Com base nos esforços existentes já em curso, as parcerias forneceriam modelos de melhores práticas, úteis para organismos nacionais e regionais que tentam negociar acordos semelhantes, identificariam incentivos e restrições específicos para várias indústrias, permitiriam economias de escala e permitiriam demonstrar o valor e a possibilidade do compartilhamento de dados e colaboração entre os setores público e privado. (UN-IEAG, 2014, p.26-27)

A estratégia da revolução de dados, consolidada no relatório do IEAG, exerceu forte influência na Comissão Estatística da ONU e serviu de base para a mobilização de recursos e promoção de mudanças na arquitetura institucional da Divisão de Estatísticas da ONU, com vistas a criação de um programa de big data e ciência de dados para estatísticas oficiais, informalmente intitulado de Projeto ONU Big Data. O discurso subjacente à esta estratégia defendia que, diante dos desafios representados pelo monitoramento dos ODS, seria preciso ‘modernizar’ os INEs para a incorporação de novas fontes de dados às estatísticas oficiais. Tal ‘modernização’ seria alcançada por meio de novos arranjos institucionais no campo estatístico, com o envolvimento do setor privado e por meio da adesão dos INEs a um ‘ecossistema global de dados’,

mediado por espaços supranacionais com governança compartilhada entre atores públicos e privados.

Pelo menos dois desdobramentos importantes no ambiente da ONU referem-se explicitamente às recomendações do IEAG: i) a criação da Parceria Global de Dados para o Desenvolvimento Sustentável (GPSDD) e ii) a constituição de um grupo de trabalho global em big data para estatísticas oficiais (GWG) posteriormente institucionalizado como o Comitê de Experts em Big Data (CEBD). O envolvimento de grandes corporações privadas globais, em particular da Microsoft e Google, no financiamento das iniciativas da GPSDD e o lobby que estas organizações passam a desempenhar no sistema estatístico global é um aspecto relevante da implementação da revolução de dados e do projeto ONU Big Data, conforme passamos a analisar na próxima seção.

3.3 Incidência de corporações privadas no campo estatístico transnacional

Constituída a partir das recomendações do relatório do IEAG (2014), a Parceria Global de Dados para o Desenvolvimento Sustentável (GPSDD) tem operado como um poderoso instrumento para a incidência política de corporações privadas na Comissão Estatística da ONU e, especialmente, nos sistemas estatísticos dos países do Sul-Global. Constituída como uma organização sem fins lucrativos, a GPSDD é hospedada pela Fundação das Nações Unidas, organização filantrópica privada estadunidense que articula uma rede de empresas multinacionais e fundações corporativas que financiam e incidem em ações de diversos segmentos da ONU⁶⁹.

A GPSDD opera como um *think tank* e secretaria global de uma rede que congrega membros de governos, organizações internacionais e empresas privadas empenhadas em “disseminar a revolução de dados como uma força para o bem” (GPSDD 2020a, p.2). Os parceiros privados no setor de dados e tecnologia incluem Big Techs como Facebook, Google, Microsoft, IBM, além de outras empresas de

⁶⁹ A Fundação das Nações Unidas é uma fundação privada sem fins lucrativos de direito privado sediada nos EUA constituída e presidida pelo magnata das telecomunicações Ted Turner, fundador do canal CNN e um dos maiores sócios do grupo Time Warner. <https://unfoundation.org/>. Constam entre as parceiras da UN Foundation a Fundação Rockefeller, Ford Foundation, IBM, Microsoft, Amazon, Google, Coca Cola, Nestlé, Unilever, Pfizer, etc. Para uma lista completa: <https://unfoundation.org/who-we-are/our-partners/more-of-our-partners/>

grande e médio porte⁷⁰. Entre os principais financiadores da iniciativa constam a Fundação Bill & Melinda Gates e a Google.org. A GPSDD está atualmente dedicada a alcançar três objetivos principais: i) expandir novas tecnologias e fontes de dados; ii) mobilizar um movimento global para promover o uso responsável de dados; e iii) incorporar padrões de interoperabilidade de dados em estruturas globais de dados e estatísticas (GPSDD 2020b, p.2).

De 2015 a 2019, a GPSDD dedicou-se à estruturação de sua rede global e à implementação de projetos-piloto em colaboração com governos de países em desenvolvimento (GPSDD 2020c). Atualmente a organização opera em 30 países do Sul Global e os seus planos para 2023 incluíam aumentar o número de governos que utilizam big data de fontes privadas por meio da intermediação de pelo menos 10 Parcerias Público-Privadas e aconselhamento técnico para pelo menos 20 INEs (GPSDD, 2020a)

Em parceria com o Banco Mundial, a GPSDD lançou a iniciativa ‘*Data for Now*’. De acordo com o projeto, “para alcançar os ODS, e não deixar ninguém para trás, todos no mundo devem estar representados em dados atualizados e oportunos” (GPSDD, 2019, p.1). A iniciativa tem entre seus objetivos avançar no uso de dados gerados pelos cidadãos (*citizen-data*) por meio de PPPs com uma estratégia explicitamente orientada para os INEs do Sul-Global:

A iniciativa envolverá o trabalho em estreita colaboração com os Institutos Nacionais de Estatística e todas as agências governamentais relevantes, para que o uso de novas fontes de dados, como dados gerados pelos cidadãos, complemente as estatísticas oficiais [...] O objetivo será integrar novos métodos e fontes nos sistemas estatísticos nacionais existentes, e explicitamente não criar uma infraestrutura paralela de dados no nível do país. Embora os sistemas governamentais estejam no centro desta iniciativa, o objetivo será melhorar as oportunidades de colaboração e compartilhamento de informações [...] e catalisar novas parcerias de dados entre entidades públicas, privadas e acadêmicas (GPSDD, 2019, p.3, tradução nossa).

Iniciativas como o ‘*data for now*’, evidenciam como os ODS são instrumentalizados para promover parcerias público-privadas e um ambiente favorável

⁷⁰ Constam como parceiros corporativos da rede o Facebook, Google, Microsoft, IBM entre outros e entre os financiadores a Fundação Bill & Melinda Gates e a Google.Org. Para uma lista completa dos parceiros:

<https://docs.google.com/spreadsheets/d/1OizuPo28fBbFbjswKyFF2aD7YUeklGVBSrD2eEULNSM/edit#gid=0>

ao desenvolvimento de mercados de dados envolvendo os sistemas estatísticos nacionais de países do sul global. Essa estratégia ocorre, entre outros, por meio da influência política de alto nível da GPSDD na Comissão Estatística da ONU, que detém poder decisório sobre o sistema estatístico global. Uma das instâncias de incidência da GPSDD é o Grupo de Alto Nível para Parcerias, Coordenação e Capacitação para estatísticas da Agenda de 2030 (HLG-PCCB). A UN-StatCom designou ao grupo, entre outras tarefas, o desenvolvimento de um plano de ação global para o estabelecimento de uma parceria global de dados para o desenvolvimento sustentável (UN-StatCom 2016). O HLG-PCCB é composto por representantes de INEs e parceiros externos; entre eles estão a GPSDD e o Grupo Paris 21.

Criado em 1999 como uma rede para o aperfeiçoamento das estatísticas no âmbito dos ODM, o Grupo Paris 21, sediado na OCDE, também atua na promoção do uso de Big Data para estatísticas oficiais em países em desenvolvimento. Por meio do projeto *Informing Data Revolution*, financiado pela Fundação Bill e Melinda Gates, o grupo promoveu PPPs e a proposição de modelos de financiamento para o setor estatístico, intitulados de *Data Compacts* (Paris21, 2015).

Os modelos de parcerias público privadas no setor estatístico foram objeto de avaliação por um documento estratégico da OCDE/Paris 21 em que são analisadas lições aprendidas e possíveis cenários futuros (Robin; Klein; Jutting, 2016). Um conjunto de projetos envolvendo big data para estatísticas oficiais é analisado de acordo com quatro modelos: (i) produção interna de estatísticas pelo provedor privado de dados; (ii) transferência de dados privados para INEs; (iii) transferência de dados privados para intermediários (*third-parties*); (iv) terceirização das funções dos INEs. O modelo de parcerias público-privadas mediado por empresas intermediárias foi avaliado como aquele com maior potencial de sucesso, considerando que “o setor privado está disposto a cooperar com intermediários” (Robin; Klein; Jutting, 2016, p.25). Neste sentido, o relatório sugere que instâncias estatísticas regionais e supranacionais poderiam operar como potenciais intermediários para as PPPs:

os escritórios regionais de estatística podem possuir a infraestrutura mais apropriada para lidar com os desafios técnicos da transferência de grandes conjuntos de dados [...] Eles também poderiam atuar como intermediários confiáveis e ajudar a economizar em custos de coordenação a) entre detentores de dados privados e b) entre INEs. Isso, no entanto, implicaria uma harmonização dos quadros jurídicos nacionais, o que provavelmente só ocorrerá a longo prazo (Robin; Klein; Jutting, 2016, p.23, tradução nossa).

Lançado no primeiro Fórum de Dados das Nações Unidas, realizado na África do Sul em 2017, o Plano de Ação Global da Cidade do Cabo, como resultado do trabalho do HLG-PCCB, foi ratificado pela Comissão Estatística da ONU no mesmo ano em sua 48ª. sessão. O plano acomodou objetivos estratégicos da GPSDD, dos quais destacamos: **(i)** identificar e remover barreiras ao uso de novas fontes de dados; **(ii)** desenvolver arranjos institucionais necessários para a cooperação público-privada, incluindo o uso de dados de fontes não oficiais; **(iii)** criar oportunidades para atores não estatais participarem do financiamento de atividades estatísticas por meio de mecanismos inovadores de financiamento (UN-StatCom, 2017). Estas iniciativas contribuíram para pavimentar um caminho institucional para a ampliação da participação do setor privado nos sistemas estatísticos global e nacionais. Paralelamente, conforme analisamos a seguir, o programa de Big Data da ONU dedicou-se à concepção e desenvolvimento de uma estrutura técnica, política e mercadológica para a indução das parcerias público-privadas no campo estatístico.

3.4 A Plataforma Global da ONU.

Em 2014 a Comissão Estatística da ONU criou o Grupo de Trabalho Global (GWG) posteriormente intitulado de Comitê de Especialistas em Big Data e Ciência de Dados para Estatísticas Oficiais (UN-CEBD). As seguintes tarefas foram delegadas ao grupo: **(i)** fornecer uma visão estratégica e coordenação de um programa global de big data para estatísticas oficiais; **(ii)** promover o uso prático de fontes de big data; **(iii)** promover a comunicação e a defesa do uso de big data em aplicações políticas; **(iv)** aumentar a confiança do público no uso de big data para estatísticas oficiais (UN-StatCom, 2015).

O UN-CEBD é composto por representantes de 31 INEs de países membros da ONU, 16 Organizações Internacionais e pela GPSDD. A estrutura de governança do comitê conta com um conselho consultivo de caráter estratégico que se reúne quatro vezes por ano, um bureau de caráter operacional que se reúne quinzenalmente, um conselho técnico que revisa e aprova os projetos em andamento e um comitê científico que coordena as atividades dos grupos de trabalho dedicados ao desenvolvimento de pesquisas e projetos específicos voltados à utilização de big data e ciência de dados para estatísticas oficiais. Os grupos de trabalho atuam nos seguintes temas: (i) observação da terra, (ii) dados de telefonia móvel, (iii) dados de

scanner, (iv) dados de rastreamento de embarcações (AIS), (v) big data para os ODS, (vi) tecnologias de melhoria de privacidade (PET) e (vii) capacitação e treinamento.

Entre as atribuições do UN-CEBD também constam a realização anual de Conferências Internacionais de Big Data para Estatísticas Oficiais, oportunidade em que também ocorrem as reuniões anuais da plenária do UN-CEBD. Nestas reuniões o comitê realiza a avaliação anual de suas ações e encaminha deliberações que são sistematizadas em um relatório apresentado para a ratificação da Comissão Estatística da ONU. As atas de reuniões e relatórios do comitê compõe uma importante fonte de dados secundários para esta pesquisa.

As Conferência Internacionais do UN-CEBD têm se constituído como espaços de articulação entre representantes seniores de INEs e OIs com empresas privadas do setor de dados e tecnologia, incluindo Big Techs como Microsoft, Google, Amazon, Facebook, e IBM e empresas de grande e médio porte como Nielsen, Samsung, Likedin, HP, Galois, etc. O release da 3ª. Conferência Internacional em Big Data para Estatísticas Oficiais da ONU, realizada em 2016 em Dublin, era bastante representativo das motivações do evento:

Este jogo entre os setores público e privado no domínio das estatísticas oficiais levanta a questão de saber o que os institutos nacionais de estatística têm de fazer para permanecerem relevantes. Evidentemente, o setor privado está avançando muito mais rapidamente do que o sector público na utilização de fontes de Big Data disponíveis. As parcerias com o setor privado na utilização de Big Data para estatísticas oficiais parecem ser o único caminho a seguir. Se não pode vencê-los, junte-se a eles” (UN-DESA, 2016, tradução nossa).

Um dos principais desdobramentos da conferência de 2016 foi o esboço de uma proposta de parceria pública privada global inspirada no modelo de plataforma das Big Techs. Um dos painéis do evento intitulado ‘Big Data para indicadores ODS’ contou apresentações da Microsoft, IBM e Google. Consta no relatório do evento que “todos os apresentadores mostraram vontade de construir parcerias público-privadas em um esforço para utilizar novas fontes de dados e tecnologias modernas para uma melhor tomada de decisões” (UN-Statcom, 2017a, p.14). Interessante observar que a apresentação da Microsoft versou especificamente sobre um modelo de plataforma de dados que

trabalha com o conceito de cofre de dados, que é um repositório hiper escalável de dados potencialmente confidenciais de múltiplas partes que

facilita o compartilhamento seguro de dados e fornece mecanismos para consumo de dados, marcação de políticas, criptografia, armazenamento e acesso aos dados, mantendo a conformidade com um conjunto diversificado de políticas eficazes (UN-Statcom, 2017a, p.14).

O painel de conclusão da conferência de 2016, moderado por Ronald Jansen, coordenador do Programa ONU Big Data⁷¹, apontou para algumas recomendações para os INEs, com destaque para “considerar a entrada no ambiente em nuvem, ao mesmo tempo que se afastam dos equipamentos físicos de TI e adotam aplicações comuns para armazenamento, processamento e disseminação num ambiente tecnológico confiável e seguro” (UN-StatCom, 2017a, p.15). Ainda neste painel de encerramento, os moderadores destacaram que a introdução de um novo ambiente tecnológico na produção de estatísticas oficiais e nas funções dos INEs requer ‘novas parcerias colaborativas e estratégicas’ com o setor privado no nível regional e global e destacaram a necessidade de “investimentos em plataformas regionais e internacionais comuns para acesso a dados, serviços e aplicações através de parcerias estratégicas” (UN-Statcom, 2017a, p.15). Da articulação de alto nível entre agentes transnacionais dos campos estatístico e algorítmico, nasce neste evento a proposta de uma Plataforma Global de big data para estatísticas oficiais.

3.4.1 O modelo de mercado e extrativismo de dados da Plataforma Global

A articulação realizada pelo UN-CEBD com os representantes de Big Techs na Conferência de 2016 resultou em um plano para o desenvolvimento de uma Plataforma Global de dados, serviços e aplicações para apoiar a comunidade estatística internacional. De acordo com a proposta apresentada pelo UN-CEBD em 2017 à Comissão Estatística da ONU: “o objetivo é desenvolver um caso de negócios para explorar especificamente a possibilidade de construir uma parceria público-privada que deverá ampliar as iniciativas atuais e futuras para fazer melhor uso de fontes de dados inovadoras nos níveis nacional e regional” (UN-StatCom, 2017, p.1).

É interessante observar que em sua proposta original, o projeto da Plataforma Global defendia a adoção de uma política “para fornecer dados, serviços e aplicações open source e APIs associadas como um bem público”. (UN-StatCom, 2017, p.5). Três

⁷¹ Ronald Jansen é Diretor de inovação de dados e capacitação da Divisão Estatística da ONU e coordenador global do Projeto ONU Big Data. Conforme analisamos mais à frente, Jansen consta entre os entrevistados em nossa pesquisa de campo.

anos após o primeiro plano apresentado na Comissão Estatística da ONU, entretanto, a Plataforma Global passa a ser apresentada como uma infraestrutura de ‘negócios’ que utiliza ‘princípios de mercado’ e que tem um ‘marketplace’ como interface com os usuários. Notadamente, a ‘confiança’ é um aspecto bastante reforçado na apresentação da iniciativa, que alegadamente se baseia em ‘parceiros confiáveis, dados confiáveis, métodos confiáveis e aprendizado confiável’.

A Plataforma Global da ONU é uma infraestrutura política, técnica e de negócios que apoia um ecossistema internacional de estatísticos, cientistas de dados e outros parceiros para trabalharem juntos na resolução de problemas e publicação aberta de seus resultados. Ela oferece um ambiente confiável para análise de dados colaborativa com base em quatro pilares: parceiros confiáveis, dados confiáveis, métodos confiáveis (algoritmos) e aprendizado confiável. Usando princípios de rede e de mercado, a plataforma facilita a troca, desenvolvimento e compartilhamento de dados, métodos, ferramentas e expertise com o objetivo de acelerar a inovação de dados. A plataforma também será usada para coordenar atividades de capacitação relacionadas ao desenvolvimento de habilidades em ciência de dados e ao uso de big data e técnicas de aprendizado de máquina. O *marketplace* é atualmente o “front-end” da plataforma (UN-StatCom, 2019, p.10, tradução nossa).

A reorientação da Plataforma Global, de um projeto pautado pelo ‘bem público’ para uma iniciativa orientada por princípios de mercado, se efetiva no processo de discussão de seu ‘modelo de negócios’ em 2019. A esta altura, uma versão de testes da Plataforma já estava concluída e os membros do UN-CEBD passaram a discutir planos para o funcionamento da iniciativa. Notadamente, no documento de base do modelo de negócios da Plataforma Global apresentado e ratificado pela Comissão Estatística da ONU, o objetivo de apoiar a realização de lucros para parceiros privados é explícito:

O modelo de negócio da UNGP [Plataforma Global da ONU] garantirá que dados, métodos, aprendizagem, serviços e aplicações de software confiáveis sejam trocados de uma forma que contribuam para uma melhor compreensão da economia e da sociedade globais, a concretização dos ODS e a modernização do sistema estatístico, ao mesmo tempo que apoia lucros sustentáveis para parceiros comerciais (UN-StatCom, 2019a, p.6, tradução nossa).

De acordo com o modelo de negócios da Plataforma Global haveria cinco principais partes interessadas no projeto, compreendidos como ‘parceiros de confiança’: Institutos Nacionais de Estatística, organizações da sociedade civil, organizações acadêmicas, setor privado e financiadores. De acordo com a proposta

“espera-se que todos os parceiros contribuam com ativos para a plataforma para, em troca, obterem benefícios dela” (UN-StatCom, 2019a, p.26). Constam potenciais benefícios institucionais, financeiros e reputacionais de acordo com o tipo de parte interessada.

Dos INEs espera-se que contribuam com uma série de ‘ativos’ para a Plataforma, desde bases de dados, métodos, algoritmos, materiais de treinamento, bem como aporte de recursos humanos e financeiros. Em troca, são previstos benefícios como a ampliação de capacidades, redução de custos e a ‘comoditização’ dos produtos da plataforma.

Os INEs se beneficiarão do desenvolvimento de capacidades através do acesso a novas fontes de dados, novas tecnologias e novos conhecimentos especializados através da rede de parceiros confiáveis. Se beneficiarão da redução do custo de desenvolvimento de métodos e conjuntos de dados. Sob a égide da ONU, os INEs podem colaborar para a comoditização e padronização de produtos na plataforma, o que melhorará a sua capacidade, ao mesmo tempo que permitirá que os recursos sejam concentrados em estatísticas inovadoras e personalizadas (UN-StatCom, 2019a, p.34, tradução nossa).

A participação de empresas multinacionais na iniciativa também é prevista no modelo, seja no provimento de serviços de nuvem, dados e tecnologias, bem como no desenvolvimento de APIs com incorporação de métodos, algoritmos e dados da Plataforma. A radicalidade da orientação mercadológica da proposta é explícita na descrição de potenciais benefícios para as empresas parceiras em um ‘mercado global’ de estatísticas oficiais.

A colaboração na plataforma dará às empresas multinacionais oportunidades de testar os seus produtos e serviços numa comunidade global e obter acesso a potenciais clientes governamentais e outros. A redução dos custos da inovação através da colaboração levará a um aumento na taxa e na qualidade do desenvolvimento de produtos. A comoditização de produtos na plataforma ganhará com a qualidade e a confiança da plataforma e com o acesso a parceiros da plataforma neste mercado global. O estatuto de parceiro de confiança terá valor reputacional para as organizações comerciais, o que se traduzirá em mensagens positivas para a promoção da empresa e dos produtos. Os provedores de dados comerciais poderiam reduzir o custo de integração de conjuntos de dados proprietários e confidenciais com outros conjuntos de dados. Compartilhar dados em uma plataforma confiável significa que uma base expandida de usuários confiáveis pode crescer rapidamente a um custo reduzido. Por exemplo, dados sensíveis móveis, de varejo ou geoespaciais podem ser integrados com estatísticas oficiais e incorporados em novos produtos que oferecem utilidade adicional (UN-StatCom, 2019a, p.34, tradução nossa).

O modelo estabelece ainda um ‘serviço de faturamento global’ (global billing) na Plataforma direcionado para os financiadores e parceiros privados que por este meio obteriam ‘lucro comercial sustentável’:

Um sistema de faturamento global garante que os doadores da Plataforma tenham visibilidade de onde o financiamento está sendo alocado, garantindo que ‘ninguém seja deixado para trás’. O faturamento global permite lucro comercial sustentável para organizações privadas que têm a oportunidade de se beneficiar da UNGP (Plataforma), ao mesmo tempo que torna a participação gratuita para todos. (UN-StatCom, 2019a, p.31)

O modelo de negócios da Plataforma Global evidencia a concepção de uma arquitetura assimétrica e problemática para os países do Sul-Global, na medida em que a relação entre ‘ativos’ e ‘benefícios’ das distintas partes interessadas efetivamente reforça as desigualdades tecnológicas entre os países participantes. A intenção de extração de dados de países do sul-global por meio da iniciativa fica explícita na passagem a seguir do modelo de negócios:

Espera-se que todos os parceiros contribuam com ativos para a Plataforma para, em retorno, receber benefícios. O modelo misto de negócios deve possibilitar que cada parceiro contribua com aquilo que pode contribuir. Por exemplo, um país em desenvolvimento pode ter poucos métodos de big data, mas a Plataforma espera que ele contribua com algo de interesse público mais amplo – talvez um conjunto de dados abertos. Espera-se que o parceiro do setor privado forneça dados e/ou serviços com uma boa relação de custo-benefício, ao mesmo tempo que ainda obteria lucros (UN-StatCom, 2019a, p.26).

Neste modelo, tudo parece funcionar como se a extração de dados dos INEs dos países em desenvolvimento ‘com poucos métodos de big data’, operasse como uma moeda de troca para a redução de custos de dados e serviços do setor privado, em benefício dos INEs do Norte-Global e das próprias empresas, que ‘ainda obteriam lucros’ com a venda de produtos e serviços. Finalmente, a arquitetura assimétrica da Plataforma Global é enfatizada em seu aspecto geopolítico na medida em que propõe a implantação de hubs regionais em países do Sul-Global que funcionariam como instâncias subordinadas a uma organização central responsável pela gestão do projeto a partir do Norte-Global. Notadamente, entre as opções apresentadas no plano de negócios para a personalidade jurídica desta organização — organização privada ou pública vinculada à ONU — a opção adotada pelo UN-CEBD para a gestão do projeto foi a de criação de uma ‘organização privada sem fins lucrativos’.

3.4.2 Controvérsias na discussão do modelo da Plataforma Global

Observamos a sessão de apresentação do modelo de negócios da Plataforma em um evento paralelo na 50ª. Sessão da Comissão Estatística da ONU em 2019. Isto nos ofereceu uma oportunidade de analisar os discursos dos agentes promotores da iniciativa e as percepções de outros agentes do campo estatístico presentes na reunião. O evento que foi mediado por Sylvie Mitchel do *Statistics Canada*, contou com apresentações de Ronald Jansen da UNSD, Mark Craddock do ONS, Sumair Sayane vice-presidente da empresa Nielsen, e finalmente, com uma sessão de perguntas e respostas com representantes de INEs e outras organizações presentes na sessão.

Após uma breve apresentação do modelo de negócios da Plataforma Global por Mitchel, que não tocou em vários dos pontos sensíveis que discutimos anteriormente, a discussão foi pautada pela escolha da personalidade jurídica da organização gestora da Plataforma. Ronald Jansen defendeu a criação de uma organização filantrópica com sede no Reino Unido.

Eu acho que só será sustentável se for um instituto, qualquer que seja o nome, pode ser Global Platform Institute, onde teremos pessoal totalmente dedicado, que estará mantendo o sistema, atraindo financiamentos e mantendo os serviços [...] o que vai garantir que existam muitas pesquisas acontecendo neste marketplace, então pensem sobre isso, penso que este é um momento único na história em que a comunidade de estatísticas oficiais pode realmente ter um instituto para a inovação de dados (UN-StatCom, 2019b, tradução nossa).

Mitchel também defendeu o modelo de gestão por uma organização filantrópica, o que, segundo ela, facilitaria a alavancagem de financiamentos de empresas privadas. Seu depoimento é importante porque revela em maiores detalhes os tipos de acordos público-privados existentes no projeto ONU Big Data, que envolvem, por exemplo, créditos de serviços de computação em nuvem oferecidos pelas Big Techs parceiras. Adicionalmente, Mitchel revelou que determinados acordos de financiamento com empresas privadas determinam a escolha de temas dos grupos de trabalho da ONU Big Data, o que revela um alto grau de ingerência dos interesses privados na condução do projeto.

Atualmente algumas filantrópicas são capazes de alavancar financiamentos do setor privado e nós temos alguns financiamentos do setor privado agora

em termos de créditos de nuvem para podermos desenvolver a plataforma, então as vezes o financiamento vem como dinheiro direto, algumas vezes há uma organização financiadora que tem provido recursos tanto para projetos específicos que eles acreditam que podem aprender em um aspecto específico e eles dão financiamento para criar um grupo de trabalho para trabalhar nisso, algumas vezes tem sido créditos, aplicações, ferramentas, pesquisas e métodos que eles desenvolveram e estão dispostos a oferecer para a comunidade [estatística], para todos nós usarmos essa infraestrutura, então por causa disso as filantrópicas tem a oportunidade de dar deduções (UN-StatCom, 2019b, tradução nossa).

Convidado a participar do evento para apresentar a parceria da empresa Nielsen⁷² com o projeto ONU Big Data, o vice-presidente Sayane, abordou a participação da empresa no grupo de trabalho de modernização de índice de preços. Seu depoimento é ilustrativo de como empresas multinacionais identificam um potencial de mercado no setor de estatísticas oficiais. A incidência no campo estatístico global pode beneficiá-las na medida em que influenciam globalmente os métodos de trabalhos dos INEs, convertendo-os em potencial consumidores de sua metodologia e bases de dados. O discurso de Sayane também oferece uma ilustração do que intitulamos 'neoliberalismo pragmático', na medida em que os evidentes objetivos de mercado da empresa são encapsulados em slogans como 'comunidade', 'sustentabilidade', 'colaboração', 'modernização', 'inovação' e 'ganho mútuo'.

Basicamente, estamos doando dados e algumas capacidades para criar metodologias confiáveis que podem estar na Plataforma e ser aplicadas por diferentes INEs para que estejam aptos a modernizar o cálculo dos índices de preços e estamos trabalhando com o grupo para ver como podemos apoiar a construção de capacidades entre os INEs para serem capazes de utilizar essa metodologia, para usar essa base de dados que a Nielsen tem. Por que estamos fazendo isso? Há uma razão interna e uma razão externa. A razão interna é porque [...] a menos que as comunidades melhorem suas condições, a atividade econômica não vai acontecer e ninguém vai se beneficiar então é do interesse de todos que apoiemos as comunidades para se desenvolver sustentavelmente e uma das razões mais importantes de nossa missão é tornar os mercados mais eficientes, o que significa justamente isso e a segunda questão, é que o time que estamos trabalhando [ONU Big Data] tem sido fenomenal no sentido de que as expertises, capacidades, a visão é incrível e muito importante há um grande reconhecimento de encontrar novas formas de construir essas colaborações entre o setor público e privado, porque eu concordo totalmente com o que esse time acredita, que temos que encontrar esses novos modos de colaborar, e que uma relação de ganho mútuo para trabalhar juntos é o único modo que podemos avançar (UN-StatCom, 2019b, transcrição nossa).

⁷² A Nielsen é uma empresa multinacional presente em mais de 100 países e uma líder global no mercado de informações, análise de dados e pesquisas de mercado.

A sessão de perguntas e respostas, que seguiu às apresentações, trouxe à tona interessantes controvérsias. O estatístico Inga Fletcher do *Statistics Austria* endereçou uma pergunta para o vice-presidentes da Nielsen que abordou diretamente o conflito de interesses na participação da empresa no projeto. Fletcher apontou para a necessidade de acesso aos dados brutos pelos INEs e para os riscos à sua autonomia na medida em que se tornem dependentes dos dados da Nielsen.

A Nielsen também vende dados de scanner para institutos de estatística, então da forma como vejo há um certo conflito de interesses em sua participação e eu me pergunto como você lida com isso, por que o dado de scanner que a Nielsen tem é um produto, não é o dado bruto. Vocês reprocessam, não é big data, é realmente estruturado, de muito boa qualidade, então qual a sua estratégia? você realmente quer oferecer esses dados, por que realmente o que ajudaria seria prover os dados brutos e os INEs realmente aprenderem a processar os dados que vem dos varejistas e não exatamente processar o dado que vem da Nielsen, que é um bom produto, mas é um produto que tem um custo e que normalmente você tem que pagar por isso. Então eu vejo um conflito de interesses para a Nielsen ajudar a comunidade estatística, de certo modo nos fazendo dependentes da Nielsen, então a pergunta é: qual a sua estratégia de longo prazo para apoiar a comunidade estatística? (UN-StatCom, 2019b, transcrição nossa).

Em uma resposta jocosa à Fletcher, Sayane evidenciou que de fato existe o interesse de comercialização dos produtos da Nielsen para os INEs, apesar de defender a independência do grupo de trabalho da ONU e a existência de outras opções.

Sim, é verdade que isto é o 'pão com manteiga' da Nielsen nos últimos 90 anos, que paga o meu salário, que paga aquele terno que meu filho está usando, muito bonito por sinal [risos] então sim, não vamos ser capazes de nos afastar disso [...] sabendo disso, como avançamos? como podemos achar áreas de convergência em que se torna um benefício mútuo para todas as partes? acho que a forma como este grupo de trabalho esta funcionando é bastante independente, acho que estão buscando pelos melhores meios [...] olhando para múltiplas fontes de dados [...] então de nossa parte eu acho que estamos trabalhando com o time baseado nas requisições que temos e se chegarmos a uma situação em que este é o dado que seria o mais demandado pelos INEs então podemos trabalhar nessas discussões e em como viabilizamos para isso avançar, como essa colaboração poderia se dar, e há países que foram para um outro lado como a Holanda que faz o trabalho por eles mesmos indo na fonte, há várias opções, não será só um sentido e uma única solução. Ronald, você quer falar algo sobre a colaboração com o setor privado? (UN-StatCom, 2019b, transcrição nossa).

Instado por Sayane a se manifestar sobre a questão das parcerias público-privadas, Jansen fez uma fala muito breve e que de certo modo confirmou os receios

do estatístico austríaco, evidenciando o objetivo de padronizar a metodologia da Nielsen para um grande número de INEs, aproveitando a presença global da empresa.

A razão que gostamos de trabalhar com a Nielsen é que eles operam em muitos países, então podemos padronizar esse tipo de metodologia que estamos desenvolvendo e ter um modo de trabalhar com isso, para nós também foi uma experimentação de como podemos trabalhar com o setor privado, é uma via de mão dupla, nós damos pequenos passos mas estamos avançando (UN-StatCom, 2019b, transcrição nossa).

Este debate é relevante enquanto evidencia controvérsias no interior do campo estatístico sobre os potenciais conflitos de interesse envolvendo a participação do setor privado. Além disso, a discussão demonstra como a estratégia de parcerias globais adotada pela ONU Big Data reforça assimetrias tecnológicas entre os países participantes. Por exemplo, o fato de que condições institucionais, legais e tecnológicas, que possibilitam a alguns INE no Norte-Global, como o CBS da Holanda, ir direto às fontes de dados e desenvolver seus próprios métodos, não estão dadas na maioria dos casos, especialmente entre os INEs do Sul-Global. A ‘padronização’ da metodologia da Nielsen, neste caso, conforme defendida por Jansen, concorre, portanto, para reduzir a autonomia destas instituições, convertendo-os em potenciais consumidores dos produtos da Nielsen.

As desigualdades regionais e os diferentes níveis de maturidade e capacidade dos sistemas estatísticos nacionais foram abordados por Oliver Chinganya, Diretor do Centro Africano de Estatísticas da Comissão Econômica das Nações Unidas para a África (UNECA). Em seu discurso, Chinganya apontou preocupações em como o países ‘com sistemas mais fracos’ vão poder realmente se beneficiar de ‘todas essas coisas bonitas que promovemos’.

Nossa região é seriamente limitada e desafiada em termos de capacidades e eu penso que não apenas a África mas também outras regiões globais que são seriamente desafiadas e o que eu me pergunto é se considerados esses desafios que eles tem e que mesmo os sistemas estatísticos nacionais são muito fracos e nós estamos chegando, com todas essas coisas bonitas que promovemos, que são muito úteis, incluindo a Plataforma Global, como eles vão acompanhar, como eles vão lidar com essas coisas, com estes sistemas mais fracos, e essa é uma área que nós temos que abordar [...] isso não quer dizer que o trabalho com big data não está acontecendo, alguns países já fazem algum trabalho com isso mas como eles realmente podem explorar o seu potencial e como assegurar que eles realmente se beneficiam de todas essas tecnologias (UN-StatCom, 2019b, transcrição nossa).

Finalmente, Rachel Baven do Departamento para o Desenvolvimento Internacional do Reino Unido (DFID), abordou a questão da confiança, elemento que conforme vimos, é repetidamente reforçado nos slogans da Plataforma Global. Afinal, perguntou Baven, quem decide o que é confiável e com qual propósito: “você fala bastante sobre ter dados confiáveis, parceiros confiáveis, metodologias confiáveis, mas quem decide quais desses são realmente confiáveis? e também um outro aspecto, confiável com qual propósito?”. A questão foi respondida por Mitchel, representante do Statistics Canadá.

Estamos tentando descobrir como isso funcionaria, é esperado que exista algum tipo de validação, não espero que todos tenham acesso a tudo e estamos tentando descobrir qual a logística por detrás disso. O objetivo comum são as estatísticas oficiais e todos temos os princípios fundamentais das estatísticas oficiais. Sabendo que temos um objetivo comum, isso nos dá uma comunidade para a qual nós queremos usar os dados e a plataforma se destina às estatísticas oficiais, não para a comercialização. Nós esperamos que dentro da plataforma tudo que for desenvolvido seja open source, acessível a todos, não haveria um aspecto de propriedade intelectual nisso, então estamos tentando estabelecer a partir dos princípios em que operamos, nem tudo está finalizado ainda, mas esse é o tipo de pensamento quando pensamos nos parceiros confiáveis (UN-StatCom, 2019b, transcrição nossa).

Em sua resposta, Mitchel evidenciou dois pontos importantes. Em primeiro lugar, não havia naquele momento uma conceituação ética, política e operacional do que o projeto entendia por ‘confiança’, nem sequer um processo de validação do que seria propriamente ‘confiável’. Assim, tudo parece funcionar como se a Plataforma buscasse operar uma conversão do capital simbólico da confiança nas estatísticas oficiais, em capital cultural institucionalizado na forma do título ‘parceiro confiável’ para convertê-lo em outras espécies de capitais (dados, métodos, lucros) nas relações com os parceiros privados. Finalmente, há uma evidente contradição entre o discurso de Mitchel — uma Plataforma da ‘comunidade estatística’ em que tudo é ‘open source’ — e o modelo de negócios efetivamente proposto, orientado por princípios de mercado e visando a geração de lucros para parceiros comerciais.

3.4.3 The hacker, the hipster and the hustler: a construção de um mercado global de estatísticas oficiais.

Entre 2017 e 2019, a Plataforma Global foi desenvolvida pelo Office of National Statistics (ONS) do Reino Unido. A partir de 2020, a iniciativa entrou em produção

com o oferecimento de acesso a serviços de computação em nuvem da Amazon Web Services (AWS), Google Cloud e Microsoft Azure e uma variedade de outros serviços para colaboração de código, publicação de métodos, observação da Terra e análise de dados de localização. Aos ‘parceiros confiáveis’, o acesso à plataforma passou a possibilitar a pesquisa, construção, implantação e consumo de dados, algoritmos e métodos estatísticos, utilizando linguagens de programação como R, Python, Java e Scala, bem como a hospedagem de modelos de aprendizado de máquina e sua publicação em APIs (UN-StatCom, 2019c).

De acordo com a documentação do projeto, a coordenação técnica da Plataforma ficou a cargo de Mark Craddock, consultor freelance e engenheiro de TI contratado pelo ONS⁷³. Profissional ativo em comunidades de desenvolvedores de TI, startups e plataformas de dados, verificamos que Craddock apresentou o projeto da Plataforma Global da ONU em um programa de entrevistas da comunidade de desenvolvedores em operações de aprendizado de máquina (MLOPs) e no encontro do grupo de usuários da Amazon Web Services (AWS). De acordo com as apresentações de Craddock nestes eventos, pudemos verificar que a abordagem utilizada no desenvolvimento dos modelos tecnológico e de negócios da Plataforma Global baseou-se amplamente nos modelos de Plataformas digitais comerciais. De acordo com Craddock:

O que é a Uber, o que é o LinkedIn, o que é o Spotify, o que faz deles um business digital e uma plataforma digital? O e-bay, a Amazon [...] então realmente o modelo de negócios de uma plataforma funciona como um site de namoro, como fazer com que consumidores e fornecedores se encontrem e como você constrói uma plataforma digital para apoiar esse modelo de negócios, então basicamente todas as plataformas são sites de namoro, então, no uber você pega o motorista e marca um encontro com uma pessoa que quer ir para algum lugar [...] então o que é chave em uma plataforma é que é o único modelo de negócios capaz de orquestrar uma ampla variedade de produtos e serviços em um ecossistema, você só pode construir uma plataforma em um ecossistema existente, você não pode realmente criar um ecossistema, então na ONU há uma série de pessoas usando dados, usando big data, criando modelos algoritmos e a plataforma é a melhor forma de agregá-los (MLOPs, 2020, transcrição e tradução nossa).

⁷³ Apesar de apresentado institucionalmente nas atividades do UN-CEBD e em vários eventos da ONU como representante do ONS ou Diretor Técnico da Plataforma Global, não identificamos a existência de um vínculo formal entre o engenheiro e o ONS ou a ONU. De acordo com o seu perfil na Amazon, onde consta a publicação de um livro sobre a Plataforma Global, que não conseguimos ter acesso, consta ainda que “Mark é cofundador de uma organização sem fins lucrativos que desenvolve treinamentos e certificações nas tecnologias mais recentes”. Em um evento da AWS, verificamos que Craddock se apresentou como funcionário da ONU: “Eu sou Mark Craddock, eu trabalho para as Nações Unidas e sou o Diretor Técnico da Plataforma Global da ONU” (BIG, 2019, transcrição nossa).

O discurso de Craddock nos permite compreender que o desenvolvimento da Plataforma Global envolveu a adoção de uma lógica de mercado que é própria do campo de práticas algorítmicas e que se manifesta nas disposições empreendedoras dos sujeitos que atuam em comunidades de desenvolvedores, startups, empresas de tecnologia e plataformas. De acordo com Craddock, por exemplo, o recrutamento e composição da equipe da Plataforma Global, baseou-se na estratégia intitulada ‘the hacker, the hipster and the hustler’, configuração que é utilizada pelo campo para descrever os perfis desejáveis de profissionais para um projeto bem sucedido. Assim, o hacker corresponderia a um perfil profissional técnico normalmente envolvido no ‘back-end’, ou seja, na estrutura tecnológica do projeto. O hipster, seria o sujeito inovador, descolado e criativo que em geral atua no ‘front-end’, no design e interação com o usuário. Já o hustler, ou traficante, é o empreendedor, o responsável pela divulgação e marketing do projeto. De acordo com Craddock, “o hustler é sobre sair e conversar com as pessoas e promover a Plataforma Global, então nós estamos fazendo isso por todo o mundo nos últimos dois anos” (BIG, 2019).

Verificamos ainda que o desenvolvimento tecnológico da Plataforma Global adotou fundamentos disruptivos das Plataformas comerciais, que contrariam os princípios fundamentais das estatísticas oficiais, tradicionalmente baseados em métodos consolidados e controlados. Isto pode ser evidenciado, por exemplo, no conceito de ‘fricção livre’ utilizado no modelo da Plataforma Global, ou seja, na opção deliberada de possibilitar que os usuários criem novos procedimentos que não foram inicialmente previstos. Trata-se, portanto, de criar um ambiente que deliberadamente estimula a ‘desobediência’ para a criação de potenciais ‘novos mercados’. De acordo com Craddock:

Esse modelo permite que o ecossistema seja capaz de prover uma combinação improvável de atributos. Então entre o ecossistema e a plataforma, ele permite que outras pessoas façam coisas na Plataforma que você provavelmente não pensou antes, por exemplo o Uber Eats sobre o Uber, isso provavelmente não tinha sido pensado no começo [...] e realmente você deve construir sua plataforma para a desobediência, então construir de um modo que as pessoas façam coisas que você não espera que elas façam, porque então potencialmente ela pode construir novos mercados e novos conhecimentos (MLOPs, 2020, transcrição nossa).

Entre 2017-2019, a direção geral do projeto da Plataforma Global ficou sob a responsabilidade de Heather Savory, *Deputy National Statistician* do Reino Unido, Diretora de Capacidade de Dados no ONS e co-chair do UN-CEBD. Verificamos que

em junho de 2019 Savory se desligou de suas atribuições públicas à frente do ONS e do sistema estatístico do Reino Unido, para “liderar o desenvolvimento da Plataforma Global da ONU em Big Data” (Deputy, 2019, p.1). Em suas palavras de despedida ao ONS, Savory afirmou que:

A nossa parceria com a ONU na Plataforma Global nos permitiu testar algumas das tecnologias inovadoras que incorporamos em nossos próprios e novos sistemas do ONS. Quero ajudar a desenvolver o que foi alcançado até agora aqui, para alcançar massa crítica e dar à Plataforma Global uma vida própria para o futuro. Através deste trabalho, apoiaremos os tomadores de decisões em todo o mundo com bons dados para ajudar a informar as suas soluções para os desafios do nosso mundo de hoje (Deputy, 2019, p.1).

Notadamente, cerca de dois meses após se desligar do ONS, em agosto de 2019, Savory efetuou em seu nome o registro do Global Platform Institute no Reino Unido, empresa de sociedade limitada constituída por ela, como sócia principal, e por mais dois outros sócios⁷⁴. Cabe destacar que o cargo de Deputy National Statistician do Reino Unido é mandatário e envolve um processo rigoroso de seleção, neste sentido, além da atuação profissional de alto nível no ONS, o cargo é investido de um valor honorífico, nacionalmente reconhecido. A abdicação de Savory para assumir o controle da Plataforma Global por meio de uma organização privada, sem mencionar o conflito de interesses desse movimento, é uma evidencia do valor conferido à iniciativa e do que estava em jogo neste processo.

Em março de 2020, de acordo com o relatório apresentado pelo UN-CEBD na 51ª. Sessão da Comissão Estatística da ONU, a Plataforma Global já era apresentada como uma ‘realidade’:

A Plataforma Global evoluiu de um conceito para uma realidade como um ambiente colaborativo para a entrega de dados, métodos e aprendizagem. O Grupo de Trabalho Global está agora trabalhando nos detalhes para o controle e acesso aos vários conjuntos de dados na Plataforma, o grau em que os dados e algoritmos devem ser tornados abertos e como o software, serviços e ferramentas disponíveis serão independentes do fornecedor. Estas questões têm uma relação direta com o modelo de negócio da Plataforma, descrito no presente relatório (UN-StatCom, 2019c, p.2, tradução nossa).

⁷⁴ Consta no Registrar of Companies do Reino Unido a incorporação do Global Platform Institute sob o número 12161673 em 19 de agosto de 2019 em nome dos sócios Heather Jane Savory, Graham John Tyson e Carol Anne Tullo. <https://find-and-update.company-information.service.gov.uk/company/12161673/filing-history>

Apesar do UN-CEBD mencionar o modelo de negócios, não há efetivamente uma descrição ou discussão do mesmo no relatório apresentado para a Comissão Estatística da ONU. O que há é a indicação de algumas decisões tomadas pelo grupo e outras ainda em discussão para colocar a Plataforma Global em produção. De todo modo, os elementos mercadológicos que passaram a caracterizar a iniciativa são fartamente evidenciados ao longo do documento.

Em primeiro lugar, há um desvio da proposta originalmente apresentada à UN-StatCom em 2017, que propunha a criação de uma Plataforma baseada em uma política de dados e tecnologias totalmente *open source* que permitiria o compartilhamento de dados, serviços e aplicações como um bem público (UN-StatCom, 2017). De acordo com o relatório, o grupo estaria trabalhando nos detalhes de controle e acesso à Plataforma o que incluía uma definição sobre “o grau em que os dados e algoritmos devem ser tornados abertos” (UN-StatCom, 2019c, p.2).

Em segundo lugar, é mencionado o desenvolvimento de “uma funcionalidade de mercado para a Plataforma Global” sob a extensão ‘marketplace.officialstatistics.org’⁷⁵ (UN-StatCom, 2019c, p.6). De acordo com a concepção do projeto, a interface com os usuários se daria por meio do marketplace, possibilitando o acesso a recursos abertos/gratuitos e proprietários/pagos.

Em terceiro lugar, é anunciado que “o Global Platform Institute foi estabelecido como uma organização sem fins lucrativos no Reino Unido” para a gestão e coordenação dos serviços da plataforma (UN-StatCom, 2019c, p.9). De acordo com o relatório do UN-CEBD, caberia ao recém criado instituto, assumir a gestão da Plataforma Global a partir do segundo trimestre de 2020. O instituto também assumiria a coordenação do projeto junto a quatro hubs regionais, sediados em INEs de países da Ásia (Indonésia), África (Ruanda), Oriente Médio (Emirados Árabes Unidos) e América Latina (Brasil). Entre as atribuições do instituto constam ainda, a criação de um estrutura de monetização do consumo de dados e serviços, bem como dos custos e receitas financeiras, por meio de um ‘sistema de faturamento global’ ou ‘*global billing*’:

O Instituto criará uma estrutura global de apoio à Plataforma Global, para garantir que os projetos colaborativos decorram sem problemas, com contabilidade transparente dos serviços consumidos e dos dados utilizados

⁷⁵ Com uso do wayback machine do Internet Archives, verificamos que o domínio marketplace.officialstatistics.org esteve em operação de agosto de 2020 a janeiro de 2023.

em cada projeto. Os custos operacionais da prestação de serviços serão recuperados através do sistema de faturamento global, que deverá indicar para cada projeto ou atividade de formação quais os serviços consumidos e a que custo (UN-StatCom, 2019c, p.9, tradução nossa).

O relatório versa ainda sobre a constituição da ‘rede de parceiros da Plataforma Global’, que englobaria todos os membros da ‘comunidade estatística’. Os INEs e órgãos estatísticos vinculados às OIs, são intitulados ‘parceiros principais’. Adicionalmente, a rede também seria constituída por parceiros externos, como empresas privadas, organizações da sociedade civil e órgãos de pesquisa, “que se tornaram parceiros confiáveis da Plataforma Global por indicação de parceiros principais” (UN-StatCom, 2020, p.9). A lista de ‘parceiros confiáveis’ do setor privado incluía: Microsoft Research, Positium, Flowminder, Telenor Group, Nielsen, Maritime Research, Galois and Cybernetica (UN-StatCom, 2020). De acordo com o relatório, o mecanismo de adesão de um parceiro externo consistiria na simples indicação por pelo menos um dos parceiros principais, o que lhe garantiria o título de ‘parceiro confiável’ e o acesso à Plataforma Global. Finalmente, o relatório versa sobre a possibilidade de elaboração de acordos de parceria estratégica entre a UNSD com ‘certas organizações’ em benefício da rede da Plataforma Global:

Tais acordos beneficiariam a rede, no sentido de que a parceria proporcionaria acesso a grandes fontes de dados detidas por proprietários de dados específicos, a plataformas adicionais ou a redes adicionais de especialistas do setor público ou privado. Algumas dessas parcerias estratégicas estão atualmente em consideração (UN-StatCom, 2019c, p.10).

A construção da Plataforma Global como uma espécie de mercado global de dados, serviços e métodos para estatísticas oficiais alcançava, portanto, no início de 2020 um status de ‘realidade’. As evidências reunidas sobre esta experiência revelam evidentes conflitos de interesses, contradições entre os discursos e práticas dos agentes e aspectos problemáticos para os sistemas estatísticos nacionais à medida em que são estabelecidas interfaces entre os setores público e privado e entre os campos estatístico e algorítmico.

O que nos parece relevante observar neste caso é como o discurso e a prática do neoliberalismo pragmático foram capazes de produzir um alto nível de permissividade à incidência e aos interesses de agentes privados, mesmo em um sistema burocrático solidamente estruturado, como é o caso da Coordenação Estatística da ONU. Notadamente, a estratégia da ‘revolução de dados’ e o

encapsulamento dos interesses mercadológicos sob o discurso da ‘sustentabilidade’, foi, em última instância, o que possibilitou este ensaio para a construção de um mercado global de estatísticas oficiais, com uma inspiração nitidamente neocolonial, que compreende os países do Sul-Global como repositórios de dados a serem extraídos em benefício de corporações e governos do Norte-Global.

A passagem do ensaio à realidade, entretanto, não se deu ‘livre de fricção’ como gostariam os desenvolvedores da Plataforma. O mundo social é cheio de atritos, controvérsias e contramovimentos, que retardam, alteram e reconfiguram as estratégias dos agentes. Conforme veremos, o movimento arriscado da ex-diretora do ONS e a declaração aberta dos objetivos mercadológicos da Plataforma não atenderam adequadamente o conjunto de interesses envolvidos, especialmente o necessário encobrimento simbólico destes interesses.

3.4.4 Entre a Comunidade e as Big Tech: controvérsias na implantação da Plataforma Global.

Entre 2020 e 2021, verifica-se uma mudança radical na execução dos planos de implantação da Plataforma Global. Em julho de 2020, Heather Savory deu entrada em um pedido de dissolução do Global Platform Institute⁷⁶. Conforme informações obtidas em off junto a uma fonte que atua no projeto da ONU Big Data, a criação do instituto pela ex-diretora do ONS gerou controvérsias e disputas internas no UN-CEBD. Paralelamente à dissolução do instituto, verifica-se a transferência da gestão da Plataforma Global para a GPSDD nos Estados Unidos e a adoção de um novo modelo de negócios, intitulado de ‘modelo comunitário’.

Em novembro de 2020, sob o contexto de isolamento social durante a pandemia de COVID-19, a reunião do UN-CEBD foi realizada em modo virtual, o que nos possibilitou a oportunidade de observá-la. Na reunião, foram apresentados o novo ‘modelo comunitário’ da plataforma e a organização doravante responsável pela gestão do projeto, a GPSDD. A tarefa de introduzir o ponto de pauta sobre a Plataforma Global coube a Tom Smith, diretor do Data Science Campus do ONS anteriormente sob a liderança de Savory. Em sua fala, bem como ao longo da reunião,

⁷⁶ Consta no Registrar of Companies do Reino Unido o pedido de dissolução voluntária do Global Platform Institute por Heather Savory em Julho de 2020. O processo de dissolução foi concluído em Outubro de 2020.

a dissolução do Instituto não foi mencionada, nem tampouco o motivo pelo qual o projeto teria sido ‘afastado’ ou ‘forçado a deixar’ a equipe de Savory⁷⁷.

O escritório de estatísticas oficiais do Reino Unido ficou muito feliz e orgulhoso em ajudar a desenvolver o programa e os projetos pilotos [...] então no começo deste ano este piloto foi meio que afastado do trabalho no Reino Unido e em particular do time coordenado por Heather Savory e do campus de ciência de dados, mas aquilo foi só um piloto, foi a fase um, talvez até a fase zero, digamos assim, o verdadeiro valor aqui é em escalar isso para a comunidade global e é por isso que a intenção sempre foi que isso fosse entregue para a Divisão Estatística da ONU liderada por Ronald, Stefan e outros, para que isso realmente alcance integralmente suas atribuições para apoiar o trabalho em todo o globo para os INEs e para os outros. Então eu vejo que há três princípios, o primeiro é em torno das fontes de dados [...] o segundo é em torno da infraestrutura e processamento [...] mas o terceiro é talvez o mais importante e isso é a comunidade. Então a Plataforma Global é realmente construída em torno de uma comunidade de usuários, uma comunidade que está enfrentando os mesmo tipos de problemas e desafios em diferentes partes do mundo [...] então vou resumir isto assim: não é só técnico, não são só os dados, é também a comunidade (UN-Big Data, 2020, tradução nossa).

É curioso observar no discurso de Smith e de outros agentes que lideraram a reunião como a palavra ‘comunidade’ passou a ser fortemente instrumentalizada para defender o ‘valor’ da Plataforma Global. A ideia de comunidade parece ser mobilizada aqui, tanto para encapsular as mudanças efetuadas no projeto, sem necessariamente explicá-las, quanto para lançar um véu sobre os distintos interesses e assimetrias entre os agentes envolvidos.

Após a fala introdutória de Smith, a diretoria executiva da GPSDD, Claire Melamed, apresentou-se como a nova gestora da Plataforma. É importante destacar que Melamed foi uma das autoras do relatório do IEAG que cunhou o termo ‘revolução de dados’ em 2014. Em sua apresentação, a CEO destacou o papel da GPSDD em ser a ponte entre as estatísticas oficiais e o setor privado, sugerindo que a Plataforma foi desenhada justamente para ser a estrutura técnica desta ‘ponte’. Interessante observar como Melamed mobiliza em seu discurso os ‘ODS’, a ‘mágica’ do mundo tecnológico e, especialmente, a ‘comunidade’.

Eu quero focar aqui na importância da comunidade. Tem sido fantástico nos últimos anos vendo do lado de fora a evolução da Plataforma e olhando com grande interesse, então é realmente um privilégio desde julho estar mais

⁷⁷ Na reunião, Tom Smith do ONS usou a expressão ‘driven out’ que a depender do contexto pode significar, expulso, extirpado, coagido, afastado ou forçado a deixar. De todo modo, a fala sugere a ideia de que o projeto não saiu da responsabilidade da equipe de Savory por vontade própria, mas por alguma pressão ou circunstância externa.

envolvida e ter sido convidada para assumir este papel de trabalhar próxima à Divisão Estatística da ONU na gestão da Plataforma. Para aqueles que não nos conhecem, a Parceria Global para Dados do Desenvolvimento Sustentável [GPSDD] foi criada em 2015 para aproveitar o grande momento em torno dos dados, especialmente focando em alcançar progresso nos ODS, mas obviamente essa é uma agenda bastante larga [...] Nós fomos criados principalmente para ser a ponte entre a comunidade estatística e os outros mundos dos dados que existem lá fora e pela mágica desse novo mundo tecnológico em que vivemos [...] fomos realmente criados para fazer a ponte entre as estatísticas oficiais, a sociedade civil e o setor privado e espero que seja nesta capacidade que possamos ser um parceiro útil na Plataforma que é a infraestrutura técnica que foi desenhada para alcançar esse tipo de parcerias e colaborações. Em termos de nosso papel, estamos gerenciando os contratos, aspectos logísticos da Plataforma e toda essa conversa sobre financiamento [...] mas o que quero destacar e onde eu queria a contribuição de vocês é realmente sobre o valor agregado da Plataforma, estamos falando em construir uma comunidade, então temos que entender as suas necessidades, quem é a comunidade, o que eles querem e como a Plataforma pode ajudá-los (UN-Big Data, 2020, tradução nossa).

No relatório apresentado pelo UN-CEBD para a Comissão Estatística da ONU em 2021, sem fazer qualquer menção à decisão anterior em que foi criado o Global Platform Institute, a GPSDD é indiretamente apresentada como a gestora da plataforma:

O Grupo de Trabalho Global [UN-CEBD] e a Parceria Global [GPSDD] estão muito bem alinhados nos seus objetivos e a colaboração na Plataforma Global enquadra-se em todas as suas aspirações. A Divisão de Estatística e a Parceria Global tomaram medidas concretas para promover a sua colaboração na gestão da Plataforma Global. A Parceria Global concordou em gerir os contratos com fornecedores de dados e serviços e avaliar as necessidades dos utilizadores da Plataforma Global (UN-StatCom, 2020, p.14, tradução nossa).

A menção à ‘comunidade’ é enfatizada no relatório do UN-CEBD de 2021. O objetivo da Plataforma agora seria o de “possibilitar à comunidade estatística global trabalhar colaborativamente em projetos utilizando as suas bases de dados e infraestrutura tecnológica” (UN-Statcom, 2020b, p.6). O modelo de financiamento da Plataforma é agora apresentado como um ‘modelo comunitário’ e já não há menções à monetização de produtos e serviços em um marketplace, tampouco ao global billing no relatório apresentado à Comissão Estatística. Doravante o ‘modelo comunitário’ de financiamento da Plataforma seria baseado em três eixos principais: contribuições em espécie de especialistas atuando nos grupos de trabalho, financiamento de base para manter os recursos da plataforma e financiamentos especiais por meio de projetos com parceiros da Plataforma (UN-StatCom, 2020).

O relatório do UN-CEBD apresenta ainda sete projetos em andamento na Plataforma Global: i) projeto do grupo de trabalho em observação da Terra para construir um pipeline de dados de satélite para estimar estatísticas de culturas agrícolas no Senegal; ii) um segundo projeto da equipe de trabalho em observação da Terra em modelagem de Inteligência Artificial para medição de serviços ambientais; iii) projeto do grupo de trabalho em sistemas de dados de rastreamento de navios (AIS) para medir a atividade comercial internacional; iv) projeto do ONS no Reino Unido sobre o processamento de dados de AIS para indicadores económicos rápidos; este projeto está em produção; v) projeto para hospedar a base de dados nacional do Camboja na Plataforma Global utilizando a tecnologia .STAT; vi) projeto da equipe de trabalho em dados de scanner para construção de pipeline de dados na Plataforma Global para produção de índices de preços; vii) Projeto para manter uma base de dados para um registo global de empresas abrangendo as 500 maiores empresas multinacionais. É importante observar que nenhum dos projetos em curso na Plataforma Global em 2021 abordava diretamente o monitoramento dos ODS. De fato, o relatório do grupo de trabalho responsável pelo tema era desalentador:

No momento não existe nenhuma aplicação concreta de Big Data que possa ser usada por países individuais no acompanhamento estatístico dos ODS. No entanto, existem iniciativas que podem ajudar a monitorar os indicadores dos ODS em nível global. A equipe de trabalho em Big Data para o SDG está tentando sintetizar essas atividades (UN-StatCom, 2020a, p.1, tradução nossa).

Na medida em que o projeto da Plataforma Global entrava em fase de produção, verifica-se uma menor ênfase aos ODS, o que evidencia a instrumentalização da agenda de 2030 como uma estratégia discursiva que foi utilizada pelo projeto com o objetivo de mobilizar recursos e efetuar as mudanças institucionais necessárias à sua implementação na UNSD. Neste sentido, o relatório aponta agora para uma articulação do UN-CEBD com a recém criada Rede de Estatísticos Econômicos das Nações Unidas, visando a constituição de um novo grupo de trabalho voltado à aquisição de dados de empresas multinacionais por meio de acordos globais de cofinanciamento:

Esta equipa abordará empresas globais para negociar o acesso às suas fontes de dados globais por meio de acordos globais e trabalhará em estreita colaboração com a modalidade de investimento na aquisição, troca e partilha de dados recomendado pelo grupo de Amigos da Presidência sobre

estatísticas económicas [FOC] e Rede de Estatísticos Econômicos das Nações Unidas (UN-StatCom, 2020, p.3)

Constituído pela Comissão Estatística da ONU em 2020, ao Grupo de Amigos da Presidência em Estatísticas Econômicas (FOC) foi atribuída a missão de fornecer recomendações para o sistema global de estatísticas econômicas. A aquisição de dados de empresas globais é um dos pontos estratégicos abordados pelo Grupo. De acordo com o FOC: “muitos institutos nacionais de estatística estão fazendo pedidos individuais de acesso para usar dados de empresas multinacionais [...] Um coletivo de Institutos Nacionais de Estatística e Organizações Internacionais podem ser capazes de oferecer mais em retorno a essas empresas” (UN-StatCom, 2020b, p.11). Notadamente, o FOC recomendou à Comissão Estatística da ONU a adoção de uma nova agenda de investimento e codesenvolvimento da infraestrutura estatística global voltada à aquisição de dados de empresas multinacionais utilizando a ‘marca das Nações Unidas’ por meio da Plataforma Global:

Devem ser ampliados os mecanismos de governança e coordenação em uma plataforma tecnológica global para permitir o investimento e o codesenvolvimento para a aquisição e acesso global de dados; compartilhamento de dados; infraestrutura comum de dados e recursos comuns; e as soluções tecnológicas comuns. Uma abordagem global à aquisição de dados melhorará a comparabilidade internacional das estatísticas económicas produzidas pelos países. A este respeito, as organizações internacionais têm um papel a desempenhar ajudando os países a obter acesso aos dados recolhidos pelo setor privado. No entanto, deve ser aplicada uma abordagem pragmática às soluções globais de dados tecnológicos, com práticas de desenvolvimento, acesso e utilização de métodos acordados. Neste sentido, a utilização de uma plataforma tecnológica global comum para acesso e utilização de dados, métodos e formação exigirá um compromisso significativo em investimento e coprodução antes de se tornar uma realidade. Não obstante, deve ser adotada uma abordagem global sistemática para a aquisição, acesso e partilha de dados, utilizando a Plataforma Global das Nações Unidas e a marca das Nações Unidas, para promover o investimento e a coprodução em parcerias com proprietários de dados do setor privado que operam a nível mundial e instituições académicas internacionalmente reconhecidas (UN-StatCom, 2020b, p.17).

Em 2021 o resultado das articulações entre a Divisão Estatística da ONU, GPSDD, UN-CEBD e FOC evidenciam a construção da Plataforma Global como uma estratégia e arquitetura para viabilizar a aquisição de dados de empresas multinacionais utilizando a ‘marca das Nações Unidas’. A estrutura técnica possibilitaria a formalização de acordos para a aquisição de dados de multinacionais, compartilhamento de dados públicos dos países membros e o desenvolvimento e

padronização de métodos algorítmicos para a produção de estatísticas oficiais com influência direta do setor privado, por meio de ‘coinvestimentos’ e ‘coprodução’ com os ‘proprietários de dados do setor privado que operam a nível mundial’. O desenvolvimento desta complexa arquitetura no campo estatístico transnacional evidencia a ação eficiente do lobby das Big Tech e de outras empresas globais do setor de dados e tecnologia em sua incidência na Comissão Estatística e Divisão Estatística da ONU.

Concebida para ter alcance global por meio de hubs regionais em países do Sul-Global, a implantação desta arquitetura se dará igualmente por meio da instrumentalização dos ODS, de modo a mobilizar recursos e engajamento político nas estruturas supranacionais implantadas em INEs da Ásia, Oriente Médio, África e América Latina. De acordo com o relatório do UN-CEBD, “Estes centros regionais reunirão a comunidade de estatísticos oficiais a nível regional para trabalhar em projetos que utilizam big data e ciência de dados para a produção de estatísticas e indicadores dos Objetivos de Desenvolvimento Sustentável e para realizar atividades correspondentes de formação e capacitação” (UN-StatCom, 2020, p.8). Para o UN-CEBD, de modo a ‘não deixar ninguém para trás’, a Plataforma Global ofereceria “oportunidades para os institutos nacionais de estatística complementarem e melhorarem as suas atividades de inovação, que podem ficar ‘presas’ a provisões e limitações internas” (UN-StatCom, 2021, p.7).

No próximo capítulo, analisaremos a implantação do hub regional da Plataforma Global para a América Latina no Brasil. Antes disso, entretanto, na próxima seção finalizamos a investigação sobre a Plataforma Global com dados primários obtidos por meio de duas entrevistas com agentes de alto nível do campo estatístico transnacional.

3.5 O elefante na sala: controvérsias de alto nível no campo estatístico transnacional

Nesta seção, cotejamos nossos achados empíricos sobre a Plataforma Global da ONU com dados primários coletados em duas entrevistas realizadas com agentes de alto nível do campo estatístico transnacional: Ronald Jansen (Diretor de inovação de dados e capacitação da Divisão Estatística da ONU e coordenador global do Projeto ONU Big Data) e Steve MacFeely (Estatístico Chefe da Organização Mundial

de Saúde e coordenador do Comitê de Estatísticos Chefe da ONU). O confronto entre as duas entrevistas é relevante para evidenciar diferentes pontos de vista e a existência de controvérsias entre agentes graduados do campo estatístico transnacional no que se refere à agenda de big data para estatísticas oficiais. Estas controvérsias reforçam nossa tese de que a transição para um regime estatístico-algorítmico de dataficação é caracterizada por um duplo movimento no campo estatístico.

Ronald Jansen é um dos principais promotores da estratégia da ‘revolução de dados’ e da agenda de big data para estatísticas oficiais na ONU, conforme ele mesmo afirmou em nossa entrevista: “eu fui um dos impulsionadores dos relatórios da revolução de dados, que saíram em 2014, nossa ideia era que o big data poderia ser usado para ajudar a obter dados mais oportunos e mais granulares para os indicadores dos ODS” (Ronald Jansen, entrevista em 12 de setembro de 2022). É interessante observar como Jansen mobiliza o discurso da competição com o setor privado como uma justificativa para a necessidade de introdução de fontes de big data na ‘comunidade estatística’.

A Comissão [Estatística da ONU] estava particularmente interessada, por volta de 2014, em como a comunidade de estatísticas oficiais poderia fazer bom uso de big data, novas fontes de dados, ciência de dados e novas tecnologias para que os escritórios de estatística, permaneçam na vanguarda do que é tecnologicamente possível na sociedade digital. Também tendo em conta o fato de o setor privado estar começando a utilizar e publicar muitos dados em dashboards, por isso a comunidade estatística também tem de se manter atualizada e mostrar o que é possível e ver os benefícios e os desafios da utilização de big data (Ronald Jansen, entrevista em 12 de Setembro de 2022, tradução nossa).

Steve MacFeely é uma das vozes críticas do campo estatístico transnacional sobre o uso de big data em estatísticas oficiais, o que pode ser observado em alguns artigos que escreveu sobre o tema (MacFeely, 2016, 2019; MacFeely; Barnat, 2017). Conforme o Estatístico Chefe mencionou em nossa entrevista, “há opiniões divergentes, algumas pessoas tendem a se concentrar nos lados positivos do big data e outras tendem a se concentrar nos lados mais fracos. Mas, no geral, há um reconhecimento de que o big data é apenas outro tipo de dado, ele tem viés, tem pontos fracos, têm pontos fortes, assim como todas as fontes de dados”. Para MacFeely, a ‘revolução de dados’ levantou uma série de esperanças e promessas sobre as fontes de big data, mas “anos depois, não vimos muitos produtos

provenientes de big data”. Do ponto de vista da agenda de big data na ONU, um dos aspectos problemáticos apontados por MacFeely refere-se a introdução de possíveis vieses decorrentes da exclusão digital, especialmente nos países do Sul-Global.

Uma das questões que certamente preocupa muitas pessoas é porque muitas fontes de big data são, na verdade, subprodutos de algum tipo de produto digital. A exclusão digital cria potencialmente uma divisão de dados, então muitas comunidades, muitos países que talvez não sejam tão fortes digitalmente podem ser excluídos de muitas fontes de big data, então se você começar a usar esses dados, especialmente na perspectiva da ONU, talvez isso introduza vieses muito importantes. (Steve MacFeely, entrevista em 19 de Maio de 2022, tradução nossa).

Ainda no tema da ‘revolução de dados’, abordamos os dois agentes sobre o uso de big data para o monitoramento dos ODS. Notadamente, nossa entrevista com Ronald Jansen confirmou o que já havíamos verificado nos relatórios do grupo de trabalho da ONU Big Data sobre os ODS, ou seja, que os resultados são bastante limitados. De acordo com Jansen, os melhores resultados foram alcançados com uso de imagens de satélite, já os resultados com outras fontes de big data são ‘menos evidentes’.

Para alguns indicadores usando dados de satélite é onde realmente acho que funcionaram, como para os serviços de água, para verificar se os lagos estão ficando menores ou maiores ou os rios e assim por diante [...] Agora em outras fontes de dados é mais difícil e, como dizer isso? são menos evidentes os resultados. (Ronald Jansen, entrevista em 12 de Setembro de 2022, tradução nossa).

É curioso constatar que o uso de big data para o monitoramento dos ODS, objetivo que em tese motivou a agenda da revolução de dados e uma série de investimentos no ambiente da ONU, tenha alcançado resultados tão limitados. Neste sentido, as informações de Jansen corroboram nosso argumento de que os ODS foram instrumentalizados com vistas à mobilizar os sistemas estatísticos global e nacionais para uma agenda que passou a gravitar de fato em torno das parcerias público-privadas e mercados de dados, sob a incidência de grandes empresas transnacionais.

Para Steve MacFeely, o monitoramento dos ODS implicou em enormes pressões para o trabalho dos INEs em todo o mundo, na medida em que as novas demandas de informação em nível nacional não foram acompanhadas pela ampliação de recursos e capacidades necessárias para atendê-las. Foi neste contexto, que, de

acordo com MacFeely, ‘com muita esperança e talvez em desespero’, o big data surgiu como uma alternativa:

Não existe solução mágica para nenhum indicador. O escopo dos ODS é enorme [...] não é surpreendente que no nível nacional o instituto de estatística começou a procurar e a se mexer, ‘como vamos produzir esses indicadores? e mesmo que sejam produzidos pela organização internacional usando big data, ainda temos que avaliar e validar a qualidade. Como vamos fazer isso?’ Quer dizer, foi como se uma bomba tivesse explodido na comunidade de estatísticas oficiais, quase pânico porque como deveríamos fazer isso se simplesmente não tínhamos os recursos nem a experiência. Então não é nenhuma surpresa que as pessoas começaram a olhar para o big data, com muita esperança, talvez em desespero (Steve MacFeely, 19 de Maio de 2022, tradução nossa).

MacFeely também confirmou que os resultados concretos do monitoramento dos ODS com big data foram restritos à poucos indicadores e ao uso de imagens de satélite. Do ponto de vista da utilização destas fontes pela ONU, o Estatístico Chefe apontou preocupações com a soberania nacional dos países membros. Para MacFeely, a potencial inversão do modelo tradicional de compilação e harmonização de estatísticas nacionais, para um modelo onde informações estatísticas seriam produzidas em um nível global, é uma importante dicotomia do big data, que pode implicar em problemas diplomáticos. Essa discussão, entretanto, segundo nos afirmou, não tem sido devidamente realizada na ONU:

O modelo de produção, tradicionalmente na ONU, tem sido: os dados são recolhidos a nível nacional e depois são fornecidos a nível internacional, mas agora, se eu os extrair de um satélite, potencialmente estamos invertendo o modelo. A organização internacional produz os dados e depois fornece as estimativas ao país [...] O modelo de produção seria invertido, então essa é uma outra questão super sensível, porque você está realmente desafiando a soberania nacional de dados e esta é uma das grandes dicotomias do big data. Isso não é realmente discutido, você pode ir a uma reunião da ONU e ouvirá pessoas falando sobre oportunidades de big data e em muitos casos elas realmente não pensaram nas implicações do que estão dizendo, porque se você apresentar isso aos países, em muitos casos eles apenas dirão, o que está acontecendo aqui? esses dados não vieram do meu ministério, não tenho como aceitar isso (Steve MacFeely, entrevista em 19 de Maio de 2022, tradução nossa).

Perguntamos a MacFeely qual a sua avaliação sobre o projeto da Plataforma Global da ONU e dos hubs regionais de Big Data, considerando que o modelo aqui proposto também poderia potencialmente alterar o processo tradicional de produção de estatísticas oficiais pelos Estados nacionais. Sua resposta foi ao mesmo tempo crítica e cética. Para o Estatístico Chefe, há muito ‘entusiasmo’ em torno do big data,

mas há desafios técnicos e políticos relevantes na implementação do projeto que precisam ser avaliados com o tempo. Notadamente, MacFeely se refere ao receio dos países em perder o controle de sua produção estatística.

Acho que teremos que esperar para ver. Como eu disse, há muito entusiasmo, muito drama em torno do big data, mas são apenas fontes de dados, no que me diz respeito. Quer dizer, cada fonte de dados tem seus desafios únicos. Então, digamos que não há nada inerentemente especial no big data. É apenas um tipo de dado. Muitas das questões serão mais a implementação técnica prática e depois a diplomática, porque os hubs regionais penso que ainda vão enfrentar o desafio, talvez porque estejam na região, talvez seja mais fácil para os países da região aceitarem, por que não estão tão longe como as sedes internacionais, mas penso que muitos países têm realmente medo de perder o controle da sua própria narrativa nacional. Então você sabe, o tempo dirá. Quer dizer, espero que isso ofereça alguma coisa, acho que os países precisam ser tranquilizados. (Steve MacFeely, entrevista em 19 de Maio de 2022, tradução nossa)

Em nossa entrevista com Jansen, tivemos a oportunidade de abordar o tema da Plataforma Global em maiores detalhes. O Diretor da UNSD nos apresentou o projeto como um ‘plataforma comunitária’ de dados e tecnologias que se distingue fundamentalmente do ambiente interno de trabalho dos INEs, na medida em que oferece possibilidades de colaboração internacional com ‘especialistas’. De acordo com Jansen, a grande vantagem da Plataforma é a sua maior flexibilidade e abertura, em oposição a estrutura rígida e fechada dos INEs que ‘tem dados muito sensíveis’ e ‘não vão deixar ninguém de fora trabalhar com isso’.

A plataforma global é uma plataforma comunitária, e a comunidade é uma comunidade estatística. Portanto, temos um ambiente onde podemos realmente trabalhar juntos [...] O verdadeiro benefício desta plataforma é que dá acesso a dados, tecnologias e especialistas para que qualquer instituto de estatística, pequeno ou grande, possa iniciar um projeto muito rapidamente [...] Portanto, você não precisa passar por todos os tipos de procedimentos difíceis para obter infraestrutura e ter computadores suficientes e, em seguida, ter armazenamento suficiente disponível. Temos isso disponível para esses projetos. E o fato de você poder colaborar internacionalmente é importante porque muitos dos grandes escritórios de estatística têm instalações de TI muito boas, mas eles não podem deixar ninguém entrar, porque eles têm dados muito sensíveis. Então, eles não vão deixar ninguém de fora trabalhar com isso. É por isso que a plataforma, a plataforma global da ONU, é algo que pode ser acessado de qualquer lugar, e podemos trabalhar juntos nas coisas (Ronald Jansen, entrevista em 12 de setembro de 2022, tradução nossa).

Perguntamos a Jansen como o projeto lidava com as diferenças de capacidades tecnológicas entre os países envolvidos, considerando a sua dimensão global. De acordo com o Diretor da UNSD, a iniciativa atende tanto aos países

desenvolvidos, oferecendo a possibilidade de cooperar em ‘coisas de alta tecnologia’, quanto aos países ‘que não tem muita infraestrutura’, podendo ‘ajudá-los’ hospedando suas bases de dados.

Então, para os países desenvolvidos, eles podem trabalhar em pares, em coisas de alta tecnologia como as tecnologias de melhoria de privacidade, que às vezes são muito complicadas, então podem trabalhar juntos nisso, mas para os países que não têm muita infraestrutura, nós podemos ajudar com nossa plataforma e já fazemos isso no Camboja há alguns anos. Neste caso, hospedamos o seu banco de dados na plataforma e utilizamos o que há de mais moderno, a tecnologia .Stat que foi desenvolvida por um consórcio de escritórios de estatística de países desenvolvidos. Eles desenvolveram este software e nós o instalamos para o Camboja e eles podem simplesmente manter seu banco de dados na Plataforma. Eles apenas mantêm e gerenciam os dados, mas a tecnologia que usam é realmente tecnologia de ponta. Portanto, é possível que escritórios menores tenham acesso à alta tecnologia (Ronald Jansen, entrevista em 12 de setembro de 2022, tradução nossa).

A abordagem de Jansen ao problema da assimetria tecnológica entre os países participantes da Plataforma evidencia a lógica de colonialismo e extrativismo de dados que destacamos anteriormente na análise do modelo de negócios da iniciativa. É interessante observar o contraste entre a concepção de Jansen, e as preocupações apontadas anteriormente por MacFeely acerca da soberania nacional dos países membros da ONU. Neste sentido, destaca-se principalmente a implantação dos hubs nos países do Sul-Global, que apontam para a criação de uma estrutura estatística supra regional paralela aos Estados Nacionais e coordenada globalmente por uma organização privada controlada por Big Techs. De acordo com Jansen:

Os hubs regionais são mesmo para aproximar esse tipo de modernização dos países da região. Assim, no caso do Brasil, os países latino-americanos poderiam mais facilmente desenvolver capacidades na sua região. Então é assim que vemos. Isso pode ser em seminários, mas também pode ser em projetos. No caso dos projetos eles podem usar a plataforma global para fazer a colaboração, então, se eles fizerem seminários ou quiserem também iniciar projetos para a região, eles não precisam fazer isso por conta própria com sua infraestrutura. Eles podem usar a plataforma para fazer isso. Então, é assim que vemos que isso funciona (Ronald Jansen, entrevista em 12 de setembro de 2022, tradução nossa).

Em nossa entrevista, MacFeely se referiu criticamente a uma certa ‘pressão’ para a introdução de big data nos sistemas estatísticos do Sul-Global. De acordo com o Estatístico Chefe, o uso de big data por países que ainda não tem uma forte governança estatística é um grande risco, não apenas para estes países individualmente, mas para todo o sistema estatístico global.

INEs utilizando big data onde não há uma governança forte, eu diria que é um acidente à espera de acontecer. Para os INEs, a única coisa que não se pode acelerar é a governança. Você deve implementar uma boa governança, independentemente do tipo de dados que estiver usando, sejam dados primários, secundários, administrativos, big data, não importa. É preciso ter governança para antecipar os desafios. Portanto, você precisa de lei estatística, de governança, de um código de práticas, de uma estrutura de garantia de qualidade. Não há uma maneira fácil de contornar isso e é um grande risco quando há essa pressão para enfiar big data goela abaixo das pessoas, se elas não tiverem isso em vigor, porque com certeza é aí que o desastre vai acontecer, minha preocupação é que alguém faça algo estúpido com dados de telefonia celular em um país, por exemplo, isso não afetará apenas aquele país, mas afetará todo mundo. Portanto, o efeito de repercussão pode ser enorme (Steve MacFeely, 19 de Maio de 2022).

No que se refere à ‘governança’ estatística sobre as fontes de big data, abordamos o tema do acesso dos INEs aos dados guardados por privados, em particular, perguntamos a opinião dos dois agentes sobre a via da legislação e a via das parcerias público-privadas como estratégias de acesso aos dados. Ronald Jansen defendeu as parcerias com o setor privado, alegando que apenas a legislação é insuficiente para a garantia de acesso aos dados:

Em alguns casos, já existe uma lei que determina que todas as empresas privadas teriam de fornecer dados ao Instituto de Estatística e isso ainda não acontece porque dificultam o processo por qualquer outra razão. Portanto, tem que haver uma espécie de parceria e um benefício para a empresa privada trabalhar com a comunidade estatística. E acho que é assim que gostaríamos de trabalhar com eles. É bom ter a lei atrás de você. Mas a lei por si só não é suficiente [...] Então, penso que deve haver um modelo de parceria onde também existam benefícios claros para o fornecedor de dados trabalhar com o Instituto de Estatística e isso poderia ser, não só para a marca da empresa, mas também em termos de obtenção de melhores insights sobre seus mercados. Portanto, o Instituto de Estatística não dará um benefício específico a uma empresa para trabalhar com eles, mas penso que apenas trabalhando com eles e desenvolvendo indicadores, as empresas terão melhores informações de mercado do que se tivessem de fazer tudo sozinhas (Ronald Jansen, entrevista em 12 de setembro de 2022, tradução nossa).

O modelo de parcerias público-privadas defendido por Jansen seria estruturado tecnicamente por meio do modelo de computação compartilhada (vide capítulo 2) com utilização de tecnologias de melhoria de privacidade (PET). Este modelo vem sendo desenvolvidas na Plataforma Global e no Grupo de Trabalho de melhoria de privacidade da ONU Big Data que conta com o financiamento e participação direta da Microsoft. De acordo com Jansen, o uso de PETs viabilizaria o acesso seguro dos INEs às bases de dados mantidas sob o controle de seus ‘proprietários’ privados sem ‘saída de informações confidenciais’.

Estamos trabalhando no Comitê de especialistas das Nações Unidas [UN-CEBD] em tecnologias que melhoram a privacidade. Então, como você pode obter acesso a dados confidenciais sem violar a privacidade? Há um grande número de tecnologias por aí e elas estão sendo testadas e acho que mostrarão cada vez mais resultados. Por meio dessas tecnologias, você interrogaria o conjunto de dados e seria capaz de obter resultados e ao mesmo tempo garantir que não há vazamento ou saída de informações confidenciais. Na medida em que o governo e as empresas se familiarizarem mais com esses tipos de tecnologias e houver evidências de que isso é realmente verdade, que você pode acessar os dados sem ter nada saindo, então acho que esse poderia ser um bom modelo daqui para frente. (Ronald Jansen, entrevista em 12 de setembro de 2022, tradução nossa).

MacFeely, por sua vez, defendeu a via da regulação como um meio de acesso aos dados e apontou para mudanças em curso na legislação estatística europeia que visam garantir o acesso dos INEs aos dados guardados por privados para fins estatísticos: “No contexto europeu, eu sei que o sistema estatístico está buscando ativamente alterar a lei estatística para lidar com esta questão de como obter acesso aos dados [guardados por privados]”. O Estatístico Chefe defendeu o caráter dos dados estatísticos como bens públicos e demonstrou preocupações com assimetrias no ‘mundo dos dados’, entre o processo de abertura dos dados governamentais e os cercamentos de dados por empresas privadas.

Em geral, eu apoio os dados abertos. Apoio, certamente, a ideia de que os dados estatísticos devem ser bens públicos. Minha preocupação, é que estamos vendo dois movimentos diferentes no mundo dos dados. Existe uma enorme pressão sobre os INE e outros departamentos governamentais para que abram os seus dados, mas do lado do setor privado, vemos uma redução nos dados abertos, porque agora eles estão começando a perceber o valor comercial dos dados. Eu acho que isso está criando uma assimetria ou desequilíbrio que pode ser problemático no futuro. (Steve MacFeely, 19 de Maio de 2022, tradução nossa).

MacFeely nos informou ainda que no âmbito do Conselho dos Chefes Estatísticos da ONU está sendo estudada uma nova iniciativa que visa a proteção dos dados como bens públicos:

No nível da ONU, existe uma preocupação em torno da fragmentação do mundo dos dados, então, estamos analisando a ideia de um Pacto Global de Dados e uma das questões com as quais estamos lidando é: nem todos os dados deveriam ser bens públicos, mas alguns dados claramente deveriam ser. Então, como proteger a parte pública dos dados para garantir que não sejam privatizados ou sugados e se tornem privados [...] É muito importante que os dados necessários a nível global para avaliar e tomar decisões permaneçam públicos, e não se torne um ativo comercial, porque então não

há como lidar com questões públicas globais. (Steve MacFeely, 19 de Maio de 2022, tradução nossa).

Apesar de reconhecer oportunidades para os INEs nas parcerias público-privadas com os detentores das bases de dados, afirmando que, “quem quer que esteja compilando esses dados, geralmente terá uma noção muito melhor de quais são os pontos fortes e fracos”, MacFeely também identificou riscos nas parcerias, relacionados à reputação dos INEs e à confiança pública nas estatísticas oficiais. De acordo com o Estatístico Chefe, “para os institutos nacionais de estatística que coletam muitas informações muito sensíveis, a confiança é absolutamente fundamental, então devem ter muito cuidado para não justificar ou tentar não violar a confiança pública” (Steve MacFeely, 19 de Maio de 2022).

Seguindo no tema das parcerias público-privadas, questionamos Jansen sobre a natureza da participação de empresas privadas na Plataforma Global. Sua resposta incorreu em uma contradição relevante para compreendermos a estratégia discursiva em torno da utilização do conceito de ‘comunidade’.

Eu não vejo muita independência do setor privado na Plataforma. Quer dizer, usamos a nuvem AWS, mas o gerenciamento atualmente é feito pelo Centro de Computação da ONU, que também tem Microsoft Azure e Google. Então, poderíamos usar qualquer um desses três. Não estamos completamente vendidos para a AWS, pois na verdade a ONU está usando muito mais Microsoft do que Amazon. Então, acho que há alguma concorrência e não somos vendidos para a AWS de forma alguma. Temos um conjunto de dados do setor privado, os dados AIS, e há um pouco de concorrência aí [...] O único outro link comercial com um provedor de serviços é um aplicativo de rede que ajuda a otimizar o uso da nuvem. E é isso. Portanto, temos muito pouca relação direta com o setor privado. Todo o resto que fazemos na plataforma é feito pela Comunidade e a Comunidade também contém especialistas do setor privado, mas eles trabalham conosco sem quaisquer incentivos financeiros (Ronald Jansen, entrevista em 12 de setembro de 2022, tradução nossa).

A alegada concorrência entre os provedores de nuvens e dados parece ser utilizada por Jansen para conferir legitimidade a estas relações comerciais. Ao afirmar, entretanto, haver ‘pouca relação direta com o setor privado’, para logo em seguida apontar que tudo é ‘feito pela comunidade’ e que a ‘comunidade também contém especialistas do setor privado’, Jansen incorre em uma contradição que revela o sentido implícito no uso do termo ‘comunidade’. Trata-se de uma estratégia discursiva que visa justamente encobrir a relação direta da Plataforma Global com o setor privado. Ao incluir a iniciativa privada na ‘comunidade’, a ‘comunidade estatística’ é

convertida em um agrupamento público-privado em que tais relações seriam legítimas. Perguntamos então a Jansen se identificava possíveis conflitos de interesse na participação dos especialistas de empresas privadas na Plataforma:

Então, somos críticos no sentido de que se pensarmos que os especialistas do setor privado têm em mente certos tipos de benefícios comerciais, então verificaremos isso, é possível porque na área das tecnologias que melhoram a privacidade, estamos com um pouco de cuidado com o Facebook ou o Google trabalhando conosco, mas trabalhamos com empresas menores do setor privado, temos a Samsung trabalhando conosco. Então, todos trabalham conosco e se trabalham conosco o fazem sem nenhum benefício comercial. Quer dizer se você contar que desenvolvemos algum tipo de conhecimento. Em geral, eles levam isso para casa, para o trabalho. Esse é o benefício deles, mas não desenvolvemos nenhum produto que eles possam explorar comercialmente (Ronald Jansen, entrevista em 12 de setembro de 2022, tradução nossa).

Nesta resposta ficam evidenciadas mais algumas contradições e também algumas omissões relevantes. Ao afirmar que é crítico à existência de interesses comerciais na participação dos especialistas das empresas, para logo em seguida reconhecer estes interesses, Jansen incorre em uma segunda contradição, que revela justamente a existência de tais conflitos. É notável ainda a afirmação de que trabalham apenas com ‘empresas menores do setor privado’ como a Samsung. Para além do fato da Samsung ser uma empresa multinacional, Jansen omitiu deliberadamente as parcerias com empresas de porte global como a Microsoft e a Nielsen. Finalmente, a afirmação de que não desenvolvem nenhum produto que as empresas possam explorar comercialmente é no mínimo ingênua. Conforme vimos, o modelo de negócios da Plataforma defendia explicitamente a obtenção de lucros para as empresas. Além disso, verificamos que há financiamento das empresas para a criação de grupos de trabalho específicos de acordo com os seus interesses comerciais, a exemplo da Microsoft na área de PETs e da Nielsen na área de estatísticas de preços.

Finalmente, questionamos Jansen sobre a contradição fundamental da Plataforma Global, ou seja, como um projeto que atua no campo público e estatal das estatísticas oficiais pôde ser concebido de acordo com princípios de mercado. Apesar de reconhecer este fato, cuja responsabilidade atribuiu aos ‘consultores’ do ONS, Jansen alegou que a orientação de mercado, é ‘do passado’ e que atualmente a Plataforma se basearia totalmente em um ‘princípio comunitário’. Desnecessário dizer que ao englobar a iniciativa privada, a ‘comunidade’ constitui um artefato para encobrir a prevalência dos princípios de mercado.

Sim, isso é do passado, Oscar, isso é do passado. Qualquer coisa envolvendo o mercado não está mais na plataforma. Então sim, foi assim, a plataforma foi desenvolvida por consultores do Instituto de Estatística do Reino Unido e eles tinham em mente torná-la um mercado para que pudesse desenvolver-se com receitas. Mas nos afastamos completamente disso, pelo que você diz, por conflitos de interesse comercial. Portanto, não existe modelo de receita. O modelo que temos agora é um modelo comunitário é realmente um princípio comunitário agora. (Ronald Jansen, entrevista em 12 de setembro de 2022).

Para concluir, destacamos uma autocrítica relevante de MacFeely, que aponta para a prevalência na ONU de discussões técnicas quando há de fato a necessidade de uma ‘conversa honesta’ sobre uma série de questões políticas e econômicas que envolvem a agenda global do big data para estatísticas oficiais. Para MacFeely, ‘o elefante está na sala nos encarando, mas ninguém quer tocar’:

Talvez parte dos desafios seja o fato de não termos tido uma conversa honesta sobre estas questões. Todo o material sobre big data tende a se concentrar na parte técnica. Acho que realmente precisa haver uma discussão muito maior sobre a necessidade dos dados como um bem público, quais dados deveriam ser bens públicos? A necessidade de dados abertos [...] Então, sim, de certa forma, acho que parte do problema é que talvez estejamos dançando em torno do grande problema, então estamos fazendo muito trabalho técnico e o elefante está lá na sala nos encarando, mas ninguém quer tocar.

Conclusão: Plataformização das estatísticas oficiais e disseminação do regime de dataficação alinhado à estratégia econômica e política das Big Tech

As evidências que analisamos neste capítulo demonstram como o regime estatístico de comensuração, característico da governamentalidade neoliberal alcançou sua expressão máxima na construção do esquema global de indicadores dos Objetivos do Desenvolvimento Sustentável (ODS). Ao estabelecer demandas irrealizáveis de quantificação, os ODS impulsionaram o uso de big data e novos métodos algorítmicos pelos sistemas estatísticos nacionais e global, contribuindo assim para mudanças nas estruturas, normas e práticas do campo estatístico e para a disseminação de um novo regime estatístico-algorítmico de dataficação.

Do ponto de vista da ONU, este processo se dá em um contexto de fragilização institucional da organização, após a crise financeira de 2007/2008, que se reflete em uma maior dependência de financiamentos de corporações privadas globais. Neste contexto, o setor privado assumiu um maior protagonismo na agenda de

desenvolvimento pós-2015, que passou a reproduzir a lógica do projeto neoliberal com um maior nível de sofisticação e pragmatismo. Notadamente, elementos discursivos como ‘sustentabilidade’, ‘comunidade’, ‘cooperação’, ‘bem comum’, etc., são instrumentalizados para encapsular simbolicamente objetivos estritamente mercadológicos dos agentes.

No contexto da agenda 2030, o campo algorítmico, formado por Big Techs e corporações privadas do setor de dados e tecnologia, passou a incidir diretamente sobre as estruturas de coordenação do sistema estatístico global da Comissão Estatística e Divisão Estatística da ONU, por meio de um lobby que se materializa na Parceria Global de Dados para o Desenvolvimento Sustentável (GPSDD). A articulação de alto nível entre agentes do campo estatístico e algorítmico transnacional se expressa na estratégia da ‘revolução de dados para o desenvolvimento sustentável’, que passou a instrumentalizar os ODS para promover o uso de big data em estatísticas oficiais por meio de parcerias público privadas, novos modelos de financiamento da infraestrutura estatística global e mudanças na arquitetura institucional da Divisão Estatística da ONU. Notadamente, a ‘revolução de dados’ abriu um espaço político para a ingerência direta das corporações privadas nos sistemas estatísticos nacionais e global, se revelando ao mesmo tempo uma estratégia econômica e política das Big Tech.

O Projeto ONU Big Data, que resulta desta articulação, constitui um núcleo global de experimentação e disseminação de novas práticas de dataficação e algoritmização das estatísticas oficiais, com o protagonismo de INEs do Norte-Global e com o financiamento e envolvimento direto de Big Techs e outras corporações privadas. Neste contexto, verifica-se o desenvolvimento de um projeto de plataforma das estatísticas oficiais, de acordo com princípios mercadológicos e tecnológicos das plataformas comerciais de dados.

As evidências reunidas demonstram como a Plataforma Global da ONU adota uma concepção neocolonialista de dados, na medida em que intensifica assimetrias tecnológicas, conferindo aos países do Sul-Global um status de provedores de dados e receptores passivos de métodos e tecnologias desenvolvidas por países e corporações do Norte-Global. Destaca-se a implantação de hubs regionais em países do Sul-Global que funcionariam como estruturas supranacionais intermediárias e facilitadoras da inserção subordinada dos INEs à Plataforma, coordenada a partir do Norte-Global por uma organização privada financiada por Big Techs.

Desdobrando controvérsias do Projeto ONU Big Data, demonstramos como as parcerias público-privadas, a adesão à premissas de mercado e a incorporação de um novo habitus algorítmico empreendedor no campo estatístico conduzem a uma série de conflitos de interesses. Verificamos fragilidades nos processos de controle da Comissão Estatística da ONU, na medida em que a organização se torna mais permeável à influência e aos interesses das corporações privadas globais. Evidenciamos ainda como o capital simbólico da ONU e das estatísticas oficiais é instrumentalizado para a promoção de mercados de dados. Estes fatores representam riscos aos sistemas estatísticos nacionais e global, em vista da potencial exposição da reputação destes sistemas e do comprometimento da confiança pública nas estatísticas oficiais.

Conforme vimos no capítulo 1, a construção de um sistema estatístico global composto por um conjunto de sistemas nacionais data de 1947 e constitui um dos alicerces do sistema ONU. Identificamos controvérsias entre agentes de alto nível do campo estatístico transnacional e verificamos que ainda estão em debate as estratégias de apropriação de novas tecnologias e fontes de dados na produção de estatísticas oficiais. As entrevistas com agentes do campo estatístico transnacional evidenciam um duplo movimento no campo estatístico. Por um lado, o movimento pró-mercado dos promotores das parcerias público-privadas, orientado pela comoditização dos dados, e, por outro, o contramovimento da regulamentação e fortalecimento da governança dos sistemas estatísticos nacionais e global, protetivo dos dados como bens públicos. Em qualquer caso, estes modelos deveriam estar sujeitos aos princípios fundamentais das estatísticas oficiais, historicamente adotados pela ONU. A geração de lucros, a privatização e a transnacionalização dos sistemas estatísticos nacionais, não encontram respaldo nestes princípios.

Os dados mobilizados e armazenados pelos sistemas estatísticos nacionais e global, o potencial de consumo dos governos nacionais e do sistema ONU de novas fontes de dados, métodos e tecnologias, e a confiança e legitimidade das estatísticas oficiais são elementos-chave para compreendermos o interesse das Big Tech e do campo algorítmico em geral, em seu movimento de incidência e interação com o campo estatístico. Trata-se de construir novos meios de geração de valor a partir dos dados em novos mercados globais que passam a envolver o setor tradicionalmente público e estatal das estatísticas oficiais. Neste sentido, a Plataforma Global é um experimento de vanguarda, na medida em que poderia potencialmente inverter o

processo tradicional de compilação de estatísticas oficiais, transferindo o lócus da produção nacional para os níveis supranacional e global.

O alcance global da Plataforma por meio dos hubs regionais pode potencialmente habilitar novas cadeias de valor e mercados de dados, envolvendo, entre outros: i) a padronização global de métodos para a produção de estatísticas oficiais cada vez mais dependentes de dados extraídos e processados por corporações globais; ii) a comercialização de dados guardados por privados para a produção de estatísticas oficiais; iii) a criação de serviços de computação distribuída para a monetização do acesso aos dados; iv) o uso dos dados para treinamento de modelos de aprendizado de IA. Enfim, tratam-se de modelos que se encontram em experimentação e cuja implementação encontra resistências vinculadas principalmente ao caráter estatal, público e nacional das estatísticas oficiais.

Os hubs regionais no Sul-Global para a promoção de big data em estatísticas oficiais podem ser compreendidos, portanto, como uma estratégia para vencer as resistências nacionais. Assim, se valem das mesmas táticas de instrumentalização dos ODS, por meio de ações de sensibilização como capacitações, festivais de dados e datathons, para fomentar a gestação de uma nova geração de cientistas de dados nos INEs, mais susceptíveis à lógica de mercado, dataficação e algoritmização. No próximo capítulo analisamos como se tem efetivado no Brasil, a implantação do Hub Regional de Big Data para a América Latina no IBGE.

4. O IBGE E O HUB REGIONAL DE BIG DATA NO BRASIL

O 'modelo brasileiro' é algo que deve ser pesquisado em nossa história, e não foi decretado por ninguém, nem pode ser mudado por decreto ou por ato institucional. Ora, o primeiro atributo do nosso 'modelo' é o fato de corresponder a uma formação periférica, subdesenvolvida, gravitando em torno da área mais desenvolvida do mundo. Essa formação se desenvolve pela absorção da cultura de vanguarda do mundo, e a importação de tecnologia é o cerne desse processo. Que esse estado de coisas não esteja fadado a eternizar-se, prova-o o fato de estarmos reduzindo a distância que nos separa da vanguarda, mas seria pura tolice fazer de conta que não somos mais periféricos e dependentes. Ao contrário, o que importa é que aprendamos a conhecer bem as particularidades, tanto de nossa perifericidade, como de nossa dependência. Só assim poderemos acelerar o processo e, afinal, superá-las a ambas [...] Somos assim, periféricos e dependentes, dado que os movimentos mais profundos de nossa economia são resposta a movimentos engendrados no centro dinâmico da economia mundial. Não obstante, essa perifericidade reveste caráter ativo e não passivo. Noutros termos, não nos podemos furtar aos efeitos dos impulsos gerados no centro dinâmico, mas podemos escolher o modo como organizamos nossa resposta aos ditos impulsos (Rangel, 1982, p. 100-102).

Neste capítulo abordamos a implantação do 'Hub regional de Big Data no Brasil em apoio à Plataforma Global das Nações Unidas' e suas repercussões no IBGE. Conforme vimos no capítulo 3, a implantação de hubs regionais em países do Sul-Global está inserida no contexto do Projeto ONU Big Data e na iniciativa da Plataforma Global da ONU. Estas ações consistem em desdobramentos da estratégia da 'revolução de dados para o desenvolvimento sustentável' que prevê a ampliação da participação do setor privado nos sistemas estatísticos nacionais, o incentivo à parcerias público-privadas na utilização de fontes de big data e a formação de uma nova geração de cientistas de dados nos INEs.

A análise que efetuamos aqui se vale principalmente de dados primários obtidos por meio de entrevistas com servidores e ex-servidores do IBGE — o que inclui ex-presidentes, servidores graduados e profissionais mais novos do quadro — bem como pela observação não participante de atividades promovidas pelo Hub regional com o envolvimento de outros INEs da América Latina, OIs e empresas privadas.

Na primeira seção do capítulo, apresentamos uma contextualização do Brasil e do IBGE no período posterior às crises financeiras de 2007/08. A crise de governamentalidade que caracteriza este período têm sua culminância com a eleição de Jair Bolsonaro. Na segunda seção, analisamos como as gestões do IBGE sob o governo de extrema-direita contribuíram para desestabilizar a instituição, ao mesmo

tempo em que atuaram na defesa de uma agenda de ‘modernização’ das estatísticas oficiais com big data e ciência de dados pautadas pela cooperação internacional. Na terceira seção, investigamos a implantação do Hub regional de Big Data no Brasil e analisamos as percepções dos informantes sobre o projeto da Plataforma Global e sobre a incidência de Big Techs no campo estatístico. Na quarta seção, a partir da análise de webinários promovidos pelo hub regional, discutimos como estas atividades ao mesmo tempo em que oportunizam a integração dos INEs da região, vêm caracterizando-se como um espaço de articulação público-privado com participação de empresas parceiras da Plataforma Global. Ainda nesta seção, analisamos ações de incidência das empresas e modelos de acesso aos dados pelos INEs que vêm sendo propostos e discutidos neste espaço. Na quinta seção, investigamos três casos de uso de big data no IBGE e discutimos as percepções dos agentes sobre a viabilidade das experiências e sobre os modelos de acesso aos dados. Finalmente, na sexta seção discutimos como a introdução do big data e da ciência de dados é percebida pelos agentes do IBGE.

4.1 Antecedentes: O Brasil e o IBGE no contexto da crise de governamentalidade neoliberal

A partir da crise financeira mundial de 2007/2008, verifica-se uma desaceleração no ritmo de crescimento da economia brasileira que conduz a uma gradual deterioração das condições de sustentação do pacto capital-trabalho que conferiu estabilidade política e institucional ao governo Lula (Singer, 2012). A conjuntura favorável representada pela alta demanda internacional por commodities, que possibilitou a manutenção do crescimento econômico sustentado no setor agroexportador, chega ao final com a nova crise financeira de 2011 no início do governo de Dilma Rousseff (Singer, 2016).

A despeito de representar um governo de continuidade sob o Partido dos Trabalhadores, a presidência de Rousseff marcou uma importante inflexão política. Tal inflexão se caracterizou por uma postura mais combativa, em contraste com o estilo conciliador de Lula e pela determinação de implantar uma política macroeconômica nacional-desenvolvimentista. A reorientação da política econômica — com a redução da taxa de juros e aumento da intervenção do Estado na economia,

por meio de políticas de investimento e proteção da indústria nacional — passou a confrontar diretamente o setor financeiro (Singer, 2016).

A sustentação da política desenvolvimentista de Rousseff dependia fundamentalmente do apoio da coalização produtivista formada durante o Governo Lula, que reunia as centrais sindicais e o setor industrial. Na medida em que a burguesia industrial abandonou esta coalização e se associou ao capital rentista, o governo de Rousseff ficou cada vez mais vulnerável às pressões dos campos financeiro e político. Após a eclosão de grandes protestos de rua em 2013, o caráter errático das políticas de governo que tentavam atender tanto as pressões dos setores rentistas como a sua base de apoio popular, apesar de lograr a recondução para um segundo mandato, foi incapaz de garantir uma base consistente de apoio político, culminando na deposição de Rousseff pela via do golpe institucional de 2016 (Singer, 2020).

No IBGE, Rousseff nomeou a ibgeana Wasmália Bivar (2011-2016) para a presidência do órgão, priorizando a indicação de quadros técnicos da casa para o cargo, a exemplo do que fizera anteriormente Lula. A gestão de Bivar notabilizou-se por assumir a presidência da Comissão Estatística da ONU, denotando uma ampliação da articulação internacional do IBGE. Do ponto de vista interno, entretanto, a despeito da implantação de importantes projetos — como a PNAD-Contínua, que passou a fornecer indicadores socioeconômicos trimestrais — o período caracterizou-se por um elevado grau de instabilidade, com conflitos no nível da direção e intercorrência de greves. Destaca-se a entrega de cargos por duas diretoras descontentes com a suspensão da divulgação de indicadores econômicos, o que reputaram à motivações políticas que teriam se sobreposto a autonomia do órgão.

Após o golpe de 2016, o governo Temer reorientou a agenda do país para uma liberalização econômica pautada pela regressão de direitos sociais e trabalhistas. Verifica-se a implantação de um radical ajuste fiscal com a aprovação da lei de limitação dos gastos públicos, reforma da legislação trabalhista e tentativas de reforma da previdência. Quanto ao IBGE, após uma breve gestão de 11 meses do economista Paulo Rabelo de Castro, o ibgeano Roberto Olinto assumiu a presidência do órgão entre 2017 e 2019. No conturbado período político do golpe e transição de governo, verificou-se a desestruturação de importantes espaços de coordenação do sistema estatístico e geocientífico nacional, em que o IBGE vinha assumindo papel protagonista, como o Comitê de Estatísticas Sociais. Neste sentido, verificam-se

esforços da gestão de Olinto em pautar um processo de discussão sobre a reestruturação do sistema estatístico e geocientífico nacional e atualização da legislação estatística.

O Planejamento Estratégico do IBGE (2017-2027) realizado sob a gestão de Olinto incluiu entre as ações estratégicas a atualização da regulamentação do sistema estatístico nacional com o objetivo de fortalecer o papel de coordenação do IBGE e garantir o acesso sistemático a registros administrativos e novas fontes de dados (IBGE, 2017). É neste plano que pela primeira vez o termo *big data* é mencionado em um documento institucional do IBGE. A referência ao termo, encontra-se no objetivo estratégico referente à implementação de novas estratégias de aquisição de dados que previu a criação de um Comitê de Inteligência de Dados (*Big Data*). A compilação de estatísticas a partir de *big data* também constava entre os objetivos de convênios firmados em 2018 pelo IBGE com o *Eurostat* e *INEGI*, respectivamente os órgãos estatísticos da União Europeia e México.

Conforme previsto no Plano Estratégico 2017-2027, um comitê gestor de Big Data (CGBigData), diretamente subordinado ao Conselho Diretor, foi formalmente constituído, juntamente com grupos de trabalho sobre o tema (IBGE, 2017a). O comitê tinha por objetivo “avaliar a relevância dos dados abertos governamentais e dos dados de outras instituições públicas ou privadas que possam servir de insumos para projetos institucionais” (IBGE, 2017a, p.1). A coordenação do colegiado ficou a cargo do próprio presidente do IBGE, com assessoria do Diretor de Informática, responsável pela secretaria executiva. O peso institucional conferido ao comitê é revelador da importância estratégica que o IBGE já vinha atribuindo ao tema dos dados.

A partir de 2019, a descontinuidade de diversas ações institucionais no IBGE e a desarticulação do sistema estatístico e geocientífico nacional estão relacionadas às ações do governo de extrema direita de Jair Bolsonaro, que decretou a dissolução de todas as instâncias colegiadas na administração pública federal⁷⁸, o que resultou na extinção do CGBigData, do Comitê de Estatísticas Sociais, da CONCLA, CONCAR e de outros espaços de articulação interinstitucional.

⁷⁸ O decreto 9.759 de 11 de abril de 2019 extinguiu e restringiu instâncias colegiadas na administração pública federal durante o Governo Bolsonaro.

4.2 O big data e a ‘modernização’ do IBGE: controvérsias no Governo Bolsonaro

A emergência de governos de extrema direita e profascistas em todo o mundo é um importante sintoma da crise da governamentalidade neoliberal (Saidel, 2023) e do capitalismo democrático (Streeck, 2013). Estes governos podem ser compreendidos como o resultado da desestruturação do tecido político que decorre do desenraizamento radical da economia de sua base social, processo promovido pela ortodoxia neoliberal e enfatizado no contexto da financeirização digitalizada. Na medida em que a própria democracia representativa liberal e os poderes constitucionais dos Estados Nacionais apresentam-se como fatores limitante ao novo regime de acumulação financeirizada do capital, o autoritarismo ultraliberal emerge como uma nova fórmula disruptiva que encontra apoio e ressonância entre os promotores do capitalismo de plataforma.

No Brasil, a emergência do bolsonarismo está diretamente relacionada à crise de hegemonia que se seguiu ao golpe de 2016, e, em particular, à incapacidade das forças políticas da direita tradicional, que promoveram a ruptura, em constituírem-se como força dirigente. De acordo com Boito (2020, p.10), o golpe mudou profundamente o processo político brasileiro, na medida em que “colocou em movimento forças conservadoras e antidemocráticas até então adormecidas, desacreditou o voto popular, ensejou o retorno dos militares como atores políticos de primeiro plano, judicializou de modo extremado os conflitos políticos, estimulou conflitos entre os ramos e instituições do Estado e desacreditou a democracia”. Ao minar as bases de sua própria legitimidade, a direita liberal repetiu o golpismo trágico de 1964, mas desta vez como farsa, abrindo mais uma vez o caminho para o autoritarismo na política brasileira⁷⁹.

No Governo Bolsonaro (2019-2022), o IBGE foi dirigido por dois presidentes sem experiência anterior na coordenação de estatísticas nacionais e foi exposto a intensas controvérsias. A primeira nomeada para o cargo pelo ministro ultraliberal Paulo Guedes, foi a jovem cientista social, egressa do Banco Mundial, Susana

⁷⁹ Nos referimos aqui à famosa frase de Marx (2011a, p.25) no 18 Brumário de Luís Bonaparte: “Em alguma passagem de suas obras, Hegel comenta que todos os grandes fatos e todos os grandes personagens da história mundial são encenados, por assim dizer, duas vezes. Ele se esqueceu de acrescentar: a primeira vez como tragédia, a segunda como farsa.”

Cordeiro Guerra⁸⁰. Guerra assumiu a presidência do IBGE com um forte discurso de ‘modernização’ das estatísticas oficiais e de defesa do big data. Em sua mensagem de abertura no relatório de gestão do órgão em 2019 destacou que: “O IBGE, alinhado aos movimentos internacionais, caminha em direção ao Big Data, ao uso de novas ferramentas de coleta, como o webscraping, e rumo à adoção de diferentes formas de processamento e de armazenamento de dados” (IBGE, 2020, p. 6).

Sob a gestão de Guerra, o Plano Diretor de Tecnologia da Informação e Comunicação 2019/2020 previu entre as ações de ‘maior vulto’ do IBGE no período, a recepção e uso de registros administrativos e bases de dados externas para as quais: “o uso de tecnologias de Business Intelligence, Big Data, Data Lake e outros deverá ser intensificado.” (IBGE, 2019, p.47). Na mesma direção, a estratégia geral de Tecnologia de Informação e Comunicação 2019/2020 apresentava entre seus objetivos a implementação de novas estratégias de aquisição de dados e a formação de “uma equipe de big data habilitada a explorar esses acervos na busca de dados para uso do IBGE” (IBGE, 2019a, p.10). Além dessas medidas, a presidente anunciou a negociação de projetos de cooperação com o Banco Mundial (BIRD) e o Banco Interamericano de Desenvolvimento (BID), voltadas à “identificar métodos mais eficientes de coleta” e à implantação de “um sólido programa de capacitação no uso de tecnologias de dados (Data Science)” (IBGE, 2020, p.6).

4.4.1 A controvérsia do Censo Demográfico

Logo no início da nova gestão, o discurso de ‘modernização’ e a defesa de ‘novos métodos’ avançou em uma área bastante sensível para o corpo técnico do IBGE, o censo demográfico⁸¹. Neste caso, o uso de registros administrativos em

⁸⁰ Susana Guerra é graduada em Ciências Sociais pela Harvard University, concluiu o mestrado em Administração Pública e Desenvolvimento Internacional pela Harvard Kennedy School e doutorado em Ciências Políticas pelo Massachusetts Institute of Technology. Trabalhou no Banco Mundial nas áreas de reforma do Estado, descentralização fiscal e administrativa e melhoria da qualidade dos serviços públicos.

⁸¹ O primeiro censo demográfico realizado pelo IBGE ocorreu em 1940. Realizado desde então com periodicidade decenal, à exceção do Censo de 1990 que ocorreu em 1991, a nona operação censitária do IBGE estava inicialmente prevista para o ano de 2020. A realização do censo, entretanto, foi afetada pela intercorrência da pandemia de Covid-19, cortes orçamentários e controvérsias que culminaram em processos judiciais. Estes fatores contribuíram para o adiamento da operação, iniciada em agosto de 2022. Normalmente prevista para ser realizada em um período de três meses, a coleta do Censo 2022 se estendeu por dez meses, sendo finalizada em maio de 2023.

substituição à coleta de um conjunto de dados primários na operação censitária foi a justificativa utilizada pela presidência do IBGE para implementar, de forma unilateral e à revelia do Conselho Técnico, uma redução significativa do número de questões nos questionários do censo 2020⁸². A redução do questionário do censo havia sido precedida pelo anúncio de um corte significativo no orçamento da operação censitária⁸³ (Canuto, 2019) e por comentários jocosos do ministro da economia já na cerimônia de posse da nova presidente:

Falta dinheiro para o Censo, mas o presidente fica de frente para o Pão de Açúcar, a diretoria fica no Centro e a turma da ralação fica aqui. Devia todo mundo estar junto em um prédio só. Ou quem sabe a gente vende os prédios e bota dinheiro para complementar, para fazer o Censo bem feito. Esse é um desafio. [...] Tem que acabar com o privilégio e acabar com a vista para o mar do presidente [...] O Censo é importante. Vamos tentar simplificar. O Censo de países ricos tem dez perguntas. O brasileiro tem 150 perguntas. Sejamos espartanos e façamos o essencial (Rosa, 2019).

A redução do questionário do censo às vésperas de sua realização, sob um contexto de cortes orçamentários e realizada de modo unilateral, provocou uma forte controvérsia dentro e fora da instituição. A reação do corpo técnico veio a público em Junho de 2019 com o lançamento da campanha “Todos em Defesa do Censo” pelo Sindicato dos Trabalhadores do IBGE – ASSIBGE. A mobilização do sindicato foi acompanhada pelo pedido coletivo de exoneração de cargos de chefia de membros do Comitê Técnico do Censo Demográfico (Pacheco, 2019)⁸⁴ e pela formalização de

⁸² A alteração efetuada pelo Conselho Diretor reduziu o ‘questionário básico’ do censo de 34 para 25 perguntas, e o ‘questionário da amostra’ censitária de 102 para 76 perguntas. As principais modificações realizadas no questionário básico referem-se à mudança do quesito de renda total, não mais voltado à coleta da informação junto a todos os membros do domicílio, mas apenas à pessoa responsável; além da exclusão dos quesitos sobre a migração interna, migração internacional e emigração. Já em relação ao questionário amostra, foram excluídas questões sobre a posse de bens duráveis (geladeira, automóvel, telefonia celular etc.), sendo mantidas apenas em relação à existência de máquina de lavar e acesso à internet no domicílio, além do valor do aluguel pago pelos domicílios alugados e perguntas relativas ao deslocamento para estudo.

⁸³ Em 2019 houve uma redução de cerca de 26% entre o orçamento inicialmente previsto de R\$ 3,1 bilhões para cerca R\$ 2,2 bilhões (Canuto, 2019).

⁸⁴ Entregaram seus cargos: Andrea Bastos, assessora da Diretoria de Pesquisas; Marcos Paulo Soares, coordenador da Coordenação de Metodologia e Qualidade (Comeq); Barbara Cobo, coordenadora da Coordenação de População e Indicadores Sociais (Copis); Leila Ervatti, gerente de população; José Guedes, gerente de metadados (Pacheco, 2019).

denúncia encaminhada ao Ministério Público Federal, que resultou em uma representação junto ao Tribunal de Contas da União⁸⁵.

O posicionamento da ASSIBGE, que por ocasião da campanha contou com subsídios técnicos de servidores graduados e ex-presidentes do IBGE, reforçava não haver oposição ao uso de registros administrativos. A crítica dos trabalhadores do IBGE dizia respeito sobretudo ao caráter autoritário da decisão, não validada pelo Conselho Técnico e em descumprimento a procedimentos técnicos básicos, como a realização de testes de validação dos questionários e dos métodos propostos para a substituição das informações retiradas dos instrumentos. De acordo com o posicionamento da ASSIBGE:

A redução do questionário atinge temas como o valor do aluguel, renda domiciliar, posse de bens duráveis, imigração/migração, registro sobre frequência à rede pública ou privada de educação. Desses pontos derivam políticas públicas fundamentais. A direção do IBGE insiste que as perdas do questionário do censo serão compensadas por Registros Administrativos. Por parte dos trabalhadores não há nenhuma oposição à incorporação sistemática e crescente dos registros, como parte do Sistema Estatístico Nacional, à produção do IBGE [...] Entretanto, não é razoável que os registros sejam oferecidos como alternativa para cobrir as lacunas deixadas pelos cortes no questionário do Censo 2020 sem que haja testes que garantam a viabilidade dessa substituição (ASSIBGE, 2019).

Em direção semelhante, o ex-presidente Roberto Olinto⁸⁶ expôs alguns elementos da controvérsia do Censo ao aposentar-se da instituição. Em sua carta aos 'ibgeanos', Olinto se referiu a riscos para o IBGE relacionados a um 'papel subserviente' nos acordos com órgãos internacionais:

⁸⁵ A redução do número de questões do Censo foi objeto de representação do Ministério Público Federal (MPF) ao Tribunal de Contas da União (TCU) como resultado de uma denúncia apresentada pelo sindicato dos trabalhadores do IBGE (ASSIBGE). Em junho de 2019 o parecer do ministério público de contas apontou irregularidades nas ações do Conselho Diretor do IBGE, em face das alterações do questionários não terem sido aprovadas pelo Conselho Técnico. A recomendação do MP ao TCU solicitava que fossem apuradas as responsabilidades dos agentes envolvidos tendo em vista a prática de ato com grave infração à norma regulamentar. Finalmente, em março de 2021, embora reconhecendo a desconformidade do ato do Conselho Diretor, o TCU decidiu acatar a redução dos questionários do censo por entender que o IBGE demonstrou razões metodológicas suficientemente embasadas. O tribunal, entretanto, recomendou ao IBGE a elaboração de normas para disciplinar o uso de registros administrativos integrado à coleta direta de informações junto à população.

⁸⁶ Roberto Luís Olinto Ramos, presidiu o IBGE entre Junho de 2017 e Fevereiro de 2019. Olinto é engenheiro de sistemas pela PUC/RJ, com mestrado em Engenharia de Sistemas e doutorado em Engenharia de Produção na COPPE/UFRJ. Servidor de carreira do IBGE, ingressou no órgão em 1980 no Departamento de Estatísticas Derivadas e Estudos Econômicos. Foi coordenador de Contas Nacionais de 2005 a 2014 e diretor de Pesquisas de 2014 a 2017

O acesso aos registros administrativos é realizado pelo IBGE há muito anos, temos convênios, há anos [...] Experiência existe, não é novidade. Mas é necessário ir além. É necessário ter um plano integrador das nossas pesquisas com os registros, é necessário ter acesso aos microdados dos registros para que possam ser usados para produção da estatística oficial, é necessário ter acordos que garantam o recebimento tempestivo. Para que se avance, todo esse conhecimento deve ser respeitado. Nada é inédito. Não se reduz o escopo de uma operação estatística dizendo que as lacunas serão complementadas por registros. O que se faz, e no mundo todo, é fazer uma pesquisa completa de base e durante os anos seguintes ir discutindo, analisando, institucionalizando um plano de integração. Os países, excessivamente citados atualmente, que eliminaram os censos ou vem reduzindo seus questionários não o fizeram abruptamente foram processos de anos. Propusemos a discussão ampliada dos registros para uso da estatística oficial. Enviamos ao Ministério do Planejamento um projeto de decreto-lei que permitisse ao IBGE o acesso às bases de registros necessárias ao aperfeiçoamento de nossa produção, seja para complementar, substituir ou mesmo basear a nossa análise. Essa proposta era o primeiro passo para iniciar a discussão da lei mais geral [...] Me entristece que todo esse ativo de conhecimento e experiência esteja sendo ignorado, lendo a última entrevista sobre os acordos com o BID fico pasmo com o desconhecimento do que fizemos. Toda ajuda é bem-vinda, mas um papel subserviente é destrutivo (Olinto, 2019).

O estreitamento de relações do IBGE com o BID e Banco Mundial, bem como a influência de quadros destas instituições nas decisões da presidência do órgão constituíram mais um elemento controverso da nova gestão. Um dos episódios polêmicos desta relação com atores internacionais se deu por ocasião de missão da diretora do Banco Mundial, Arianna Legovini⁸⁷ no IBGE, em maio de 2021. Legovini fez duras críticas ao projeto técnico do Censo e defendeu a redução do número de questões e do plano amostral da operação. A postura autoritária da diretora motivaram uma queixa formal do sindicato dos trabalhadores do IBGE ao Banco Mundial.

Causou imenso desconforto ao corpo técnico da instituição a abordagem utilizada pela pesquisadora, que ignorava absolutamente todo processo de trabalho desenvolvido ao longo do atual ciclo censitário brasileiro, iniciado em 2016, e marcado, como de costume, por um sólido programa de trabalho que envolveu grande parte do seu corpo técnico [...] Legovini foi questionada sobre algumas das suas recomendações, que incluíam a redução pouco fundamentada do tamanho do questionário [...] Da mesma forma, causou estranheza a todos a possibilidade de alteração do plano amostral da próxima operação censitária brasileira [...] Questionou os responsáveis técnicos de forma desrespeitosa, lançando dúvidas sobre as decisões que nortearam um projeto estratégico para a sociedade brasileira, e se disse convencida de que a proposta não possuía consistência técnica [...] a intervenção da Sra. Legovini se desdobrou em um desgastante processo de questionamentos ao projeto técnico original, amparados parcialmente na intervenção da Sra. Legovini, e que podem significar um prejuízo incalculável ao censo

⁸⁷ Arianna Legovini ingressou no Banco Mundial em 2004 e atualmente é Diretora do Departamento de Avaliação de Impacto de Desenvolvimento (DIME).

demográfico brasileiro, à imagem do IBGE, e à imagem do Banco Mundial (ASSIBGE, 2019a).

A controvérsia em torno dos cortes orçamentários e redução do questionário do censo gerou ainda uma inusitada divisão de posicionamentos entre ex-presidentes da instituição, favoráveis e contrários às medidas adotadas pela nova gestão. Por um lado, o grupo formado pelos ex-presidentes mais recentes e oriundos ‘da casa’ — que se revezaram na presidência após 2003 — aderiu à campanha promovida pelo sindicato, com uma carta de repúdio em que acusaram a nova gestão do IBGE de “improvisação técnica” (Ex-Presidentes, 2019)⁸⁸. Por outro lado, um segundo grupo formado exclusivamente pelos ex-presidentes mais antigos e ‘de fora da casa’ — que presidiram o IBGE entre 1986 e 2003 — apresentou uma carta de apoio à iniciativa de redução do questionário do censo (Correio Braziliense, 2019)⁸⁹.

É interessante observar como o segundo grupo de ex-presidentes, caracterizado pela orientação política liberal de seus membros, recorreu repetidamente à ideia de ‘modernização’ para apoiar as iniciativas de redução do questionário do censo, ao mesmo tempo justificadas por um ‘quadro de restrições financeiras’. Notadamente, a ‘modernização’ é apresentada pelo grupo como uma ‘tendência que se observa internacionalmente’ e que inclui, além da intensificação no uso de registros administrativos, a adoção das ‘modernas metodologias de big data’:

Acreditamos que o momento é oportuno para pensar qual é o papel do Censo Demográfico, e como ele pode ser modernizado para se tornar cada vez mais eficiente, confiável e factível no quadro de restrições financeiras em que vivemos [...] a tendência que se observa internacionalmente é a de reduzir o tamanho dos questionários dos Censos Decenais, ao mesmo tempo em que se intensifica o uso de pesquisas amostrais, de registros administrativos e de análises de grandes agregados de informações produzidas por diversas fontes em tempo real, com uso das modernas metodologias de *big data*. É nesse sentido que nos parece que o IBGE precisa evoluir, de uma agência organizada para a coleta detalhada de informações para uma instituição modernizada capaz de produzir, compatibilizar e fazer uso de dados das mais

⁸⁸ Além de Roberto Olinto, dois outros ex-presidentes ‘da casa’ endossaram a carta de repúdio: Eduardo Pereira Nunes (presidente do IBGE de 6 de fevereiro de 2003 a 14 de setembro de 2011) e Wasmália Socorro Barata Bivar (de 14 de Setembro de 2011 a 04 de Julho de 2016). Também assinou o documento o ex-presidentes sem vinculação institucional com o IBGE, Paulo Rabello de Castro (de 04 de julho de 2016 a 01 de junho de 2017).

⁸⁹ Compuseram este grupo Edmar Lisboa Bacha (presidente do IBGE de 10 de maio de 1985 a 27 de novembro de 1986), Edson de Oliveira Nunes (de 28 de novembro de 1986 a 13 de abril de 1988), Charles Curte Mueller (de 03 de maio de 1988 a 18 de abril de 1990), Eduardo Augusto Guimarães (de 18 de abril de 1990 a 26 de março de 1992), Silvio Augusto Minccioti (de 15 de junho de 1993 a 30 de março de 1994), Simon Schwartzman (de 02 de maio de 1994 a 31 de dezembro de 1998) e Sérgio Besserman Viana (de 18 de janeiro de 1999 a 31 de janeiro de 2003).

diversas fontes. (Garcia, 2019, p.2)

De acordo com o sociólogo e ex-presidente do IBGE, Simon Schwartzman⁹⁰, seu endosso à carta de ex-presidentes que apoiou a redução do questionário do censo justificou-se pela defesa dos recursos necessários para a realização da operação e por sua concordância com a intensificação do uso de registros administrativos em um contexto de ‘constrangimento financeiro’. Conforme vimos no capítulo 2, quando presidente do IBGE em 1998, Schwartzman propôs uma reestruturação radical do IBGE, nos moldes de uma agência regida sob as regras da iniciativa privada. Em nossa entrevista, o ex-presidente afirmou que o IBGE ‘custa muito caro como instituição’ e tem uma estrutura ‘obsoleta de coleta de informações’. Finalmente, o sociólogo defendeu que um papel ampliado na coordenação nacional, visando a produção de registros administrativos com qualidade estatística, poderia ser ‘uma saída’ para o órgão:

A qualidade do dado administrativos hoje em dia aumentou e é muito mais possível essa função do IBGE em se envolver efetivamente em negociar com diferentes áreas do governo pra produção de dados administrativos que tenham qualidade estatística, eu acho que é uma função que vai ser cada vez mais importante. E eu acho que eles [direção do IBGE], eu diria que por necessidade também estavam tentando fazer isso. É claro que se fizer isso é mais barato. É mais barato você efetuar o acordo com diferentes órgãos pra coletar os dados e coordenar do que você ir lá e fazer a coleta primária. Então há um constrangimento financeiro, o IBGE não conseguiu resolver o problema de que ele custa muito caro como instituição, porque tem uma estrutura muito grande e obsoleta de coleta de informações [...] então digamos assim, a saída pelo lado de você concentrar mais competência no trabalho de coordenação, de negociação com os outros órgãos de governo eu acho que faz sentido (Simon Schwartzman, entrevista em 01 de Setembro de 2022).

Importante observar que não havia oposição entre os dois grupos de ex-presidentes quanto à necessidade de intensificação do uso de registros administrativos pelo IBGE ou quanto ao exercício de um papel ampliado do órgão na coordenação estatística nacional. A crítica sustentada pelo grupo contrário à redução do orçamento e questionário do censo dizia respeito sobretudo ao caráter de improvisação das medidas tomadas, à falta de discussão com o corpo técnico da

⁹⁰ Simon Schwartzman (1939 -) foi presidente do IBGE entre 1994-1998. Schwartzman é graduado em Sociologia Política e Administração Pública pela Universidade Federal de Minas Gerais, mestre em Sociologia pela Facultad Latinoamericana de Ciencias Sociales, em Santiago do Chile e doutor em Ciências Políticas pela University of California, em Berkeley (1973)

instituição e ao abandono de ações que visavam justamente o fortalecimento institucional do papel de coordenação do IBGE. Neste sentido, de acordo com Roberto Olinto, iniciativas que estavam em andamento, como a revisão da legislação do sistema estatístico nacional para viabilizar o acesso a registros administrativos e outras fontes de dados, foram completamente interrompidas a partir de 2019. Interessante observar como o ex-presidente associa a mudança de orientação na política institucional, ao discurso de ‘modernização’ do IBGE com big data introduzido pela nova gestão:

Nós traduzimos a lei [marco legal europeu] e começamos uma discussão dentro do IBGE, do sindicato aos coordenadores. A ideia era que todo mundo discutisse. Então começamos a discutir com o Ministério do Planejamento, com algumas pessoas dentro do ministério que tinham interesse nisso, como o Banco Central. Dentro do Planejamento, criamos um comitê pra discutir com a Receita Federal e com o Ministério da Fazenda. Isso estava avançando no Congresso também, junto com a frente parlamentar de geografia, estatística e meio ambiente [...] Como o meu mandato estava acabando e a gente não ia conseguir fazer a lei toda, decidimos fazer uma lei complementar só pra ter acesso aos dados da Receita Federal, aos dados fiscais. Isso chegou a ser um pré-projeto encaminhado e que ia ser aprovado. Aí mudou tudo. Aí entrou a Suzana [Guerra] e ela abandonou completamente essa história. E por quê? Na cabeça dela ela tinha que modernizar o IBGE. A inovação era trazer o Big Data. Ela não entende que a questão tecnológica é coberta por uma questão institucional [...] O Big Data é mais um exemplo dessas coisas genéricas que todo mundo fala. Então, por exemplo, usar a nota fiscal que é uma base de dados gigantesca operada como um registro administrativo, para a estatística oficial é uma questão muito mais pé no chão, do que ficar discutindo teoria. O instituto tem que ter diretriz, temos que trabalhar em uma linha e essa linha, desta forma, não é big data, é preciso estabelecer a diretriz e conceituar claramente. (Roberto Olinto, entrevista virtual em 9 de Junho de 2022).

O renomado estatístico ibgeano Pedro Nascimento⁹¹, enfatizou igualmente em nossa entrevista a necessidade de reforma do marco legal para reforçar a função de coordenação do IBGE junto ao sistema estatístico nacional. Seu depoimento também confirma as informações que obtivemos, quanto ao andamento deste debate na gestão de Olinto e sua interrupção no governo Bolsonaro.

O Brasil precisa com urgência de uma reforma do seu arcabouço legal pra produção de estatística. Isso vinha sendo objeto de atenção da direção do

⁹¹ Pedro Luís do Nascimento Silva (1959 -) ingressou no IBGE em 1981 e se aposentou em 2019. Nascimento é bacharel em estatística pela ENCE, mestre em matemática aplicada e estatística pela Associação Instituto Nacional de Matemática Pura e Aplicada e doutor em estatística pela University of Southampton. Foi diretor da ENCE, presidente do Instituto Interamericano de Estatística (IASI), presidente da International Association of Survey Statisticians (IASS) e presidente do Instituto Internacional de Estatística (ISI).

IBGE até à saída do Roberto Olinto. Ele vinha promovendo uma tentativa de promover um debate e ter um projeto de lei estatística com alguns atores no parlamento que pudessem levar adiante a iniciativa, mas esse debate desapareceu com o governo Bolsonaro que na verdade decidiu diminuir, apequenar, ou se tivesse tido condições, até destruir o IBGE, porque não faz outra coisa. A única promessa de campanha bem cumprida é que vai acabar com tudo que está aí. Então, a gente precisa desse arcabouço renovado para reforçar uma função de coordenação do IBGE (Pedro Nascimento, entrevista virtual em 21 de Junho de 2022).

4.4.3 Modernização subalterna, desestabilização institucional e legado disruptivo

As controvérsias analisadas nesta seção nos ajudam a compreender como o discurso de modernização das estatísticas oficiais foi utilizado para justificar ações disruptivas em um momento de instabilidade institucional do IBGE. Conforme vimos, o órgão vinha construindo uma arquitetura político-institucional para lidar com os desafios da dataficação. Neste sentido, as ações reativas dos quadros do IBGE que se seguiram à desarticulação destas iniciativas institucionais não se tratam de uma oposição conservadora ao uso de novas tecnologias, como a querela entre os ex-presidentes pode levar a crer, mas caracterizam contramovimentos protetivos ao caráter desestruturador de ações que buscaram justificação no discurso de modernização.

A agenda de 'modernização' perseguida por Guerra e continuada por seu sucessor Eduardo Rios Neto⁹² foi pautada principalmente pela cooperação internacional. Destacam-se os projetos de cooperação com o Banco Interamericano de Desenvolvimento (BID) e Banco Mundial para a intensificação do uso de registros administrativos (IBGE, 2021) e o acordo internacional com a Divisão Estatística da ONU para a implantação do Hub Regional para Big Data no Brasil em apoio à Plataforma Global da ONU (IBGE, 2021a).

Curiosamente, a assinatura do memorando de entendimento para a implantação do Hub regional em 08 de Abril de 2021 consta como último ato administrativo de Guerra à frente do órgão (IBGE, 2021a). O seu pedido de exoneração foi encaminhado logo após um corte expressivo dos recursos

⁹² Eduardo Luiz Gonçalves Rios Neto ingressou no IBGE em 2019 no cargo de Diretor de Pesquisas e assumiu a presidência do IBGE entre abril de 2021 e janeiro de 2023. Eduardo é doutor em Demografia pela Universidade da Califórnia/Berkeley e pós-doutor pela Universidade do Texas/Austin. Foi Chefe do Departamento de Demografia da Universidade Federal de Minas Gerais, coordenou o curso de pós-graduação em Demografia da UFMG e presidiu a Associação Brasileira de Estudos Populacionais ABEP.

orçamentários para a realização do censo demográfico, o que inviabilizou a realização da operação em 2021 (Caram e Resende, 2021)⁹³. Alguns meses após sua saída do IBGE, Guerra foi indicada pelo Governo Bolsonaro para ocupar um cargo de gerência na sede do BID em Washington⁹⁴. Em seu perfil profissional publicado na página da instituição, o banco destaca que a nova gerente “foi Presidente do Instituto Brasileiro de Geografia e Estatística, de 2019 a 2021, organização com cerca de 12.000 funcionários, onde liderou reformas para modernizar a produção de estatísticas oficiais e o censo nacional” (BID, 2021). Ainda de acordo com apuração de Stefano (2021):

[...] pesaram a favor do nome da economista a experiência que ela acumulou à frente do IBGE durante a pandemia. Com os desafios impostos pela pandemia, incluindo restrição de locomoção de recenseadores, Guerra impulsionou o uso de tecnologia para que o instituto mantivesse a produção de dados estatísticos, incluindo ferramentas como big data, inteligência artificial e pesquisas assistidas por computador. O avanço na digitalização é justamente uma das prioridades do BID na América Latina (Stefano, 2021, p.1).

Em sua carta de despedida aos ‘ibgeanos’, a ex-presidente destacou entre os feitos de sua gestão: “a realização de esforços disruptivos, com o objetivo de acelerar o processo de modernização do IBGE, em especial no que tange à coleta de dados” (Guerra, 2021, p.1). Na próxima seção analisamos um dos legados ‘disruptivos’ de sua gestão, a implantação do hub regional de big data no Brasil.

4.5 O Hub Regional de Big Data no Brasil

Conforme vimos no capítulo três, o projeto da Plataforma Global previa a implantação de quatro hubs regionais em Institutos Nacionais de Estatística do Sul-Global, mais especificamente na Ásia (Indonésia)⁹⁵, no oriente médio (Emirados

⁹³ O corte introduzido no relatório final do orçamento federal de 2021 pelo Senador Marcio Bittar (MDB-AC), aprovado pelo Congresso Nacional em 25 de março de 2021 foi superior à 90% e reduziu o orçamento do Censo de um total de R\$ 2 bilhões para R\$ 191 milhões. Posteriormente um novo corte foi efetuado pelo Governo Federal na sanção do orçamento federal em 22 de abril de 2021, reduzindo os recursos destinados ao Censo à cerca de R\$ 50 milhões (Caram e Resende, 2021).

⁹⁴ Susana Guerra assumiu a gerência de Instituições para o Desenvolvimento no BID em Setembro de 2021.

⁹⁵ Inicialmente o projeto previa a implantação de um hub regional na China, entretanto, a partir de 2023 um novo hub regional na Ásia foi implantado na Indonésia, enquanto a categoria do hub da China foi alterada de ‘hub regional’ para ‘hub setorial’.

Árabes Unidos), na África (Ruanda) e América Latina (Brasil). Ainda de acordo com nossa discussão na seção anterior, vimos como o discurso de modernização foi instrumentalizado pela nova gestão do IBGE durante o governo Bolsonaro, para promover ações disruptivas que comprometeram a realização do censo demográfico, interromperam iniciativas institucionais relevantes, motivaram pedidos de aposentadoria e exoneração de servidores graduados e afetaram a imagem pública do órgão.

Foi em meio a este contexto controverso de desestabilização institucional que o IBGE firmou o memorando de entendimento com a ONU para a implementação do 'Hub Regional de Big Data no Brasil em apoio à Plataforma Global das Nações Unidas' (IBGE, 2021a). Os objetivos e atividades previstas para o hub no Brasil, segundo o acordo firmado, incluíam:

Os principais objetivos do Hub Regional no Brasil são a facilitação de projetos no uso de Big Data e ciência de dados para estatísticas oficiais e indicadores dos ODS, o compartilhamento de conhecimento recém-desenvolvido sobre métodos, algoritmos e ferramentas e a oferta de treinamentos no uso de Big Data e ciência de dados para a comunidade de estatísticas oficiais nas regiões das Américas Central e do Sul [...] O Hub Regional no Brasil abrangerá as seguintes atividades: pesquisa sobre o uso de Big Data e ciência de dados para o desenvolvimento de novos métodos e técnicas estatísticas; capacitação e treinamento para estatísticos e cientistas de dados de institutos de estatística nacionais, regionais e globais; seminários e workshops internacionais sobre ciência de dados e modernização das estatísticas oficiais; desenvolvimentos e manutenções adicionais da Plataforma Global da ONU - em colaboração com outros hubs regionais - para o compartilhamento de dados confiáveis, métodos confiáveis e aprendizado confiável entre a comunidade de estatísticas oficiais e seus stakeholders (IBGE, 2021a, p.2).

Fisicamente, o hub regional está sediado na Escola Nacional de Ciência Estatística (ENCE) do IBGE⁹⁶. De acordo com o estatístico Pedro Nascimento: “essa

⁹⁶ Fundada em 06 de março de 1953, a Escola Nacional de Ciências Estatísticas - ENCE é uma instituição Federal de Ensino Superior que faz parte do Instituto Brasileiro de Geografia e Estatística - IBGE. Além das atividades de ensino e pesquisa, a escola é responsável pelas atividades de capacitação e treinamento de servidores do IBGE, e demais servidores públicos, oferecendo também cursos avulsos de atualização e extensão. Atualmente a escola oferece os seguintes cursos: Bacharelado em Estatística (desde 1953), Pós-graduação Lato Sensu: Especialização em Análise Ambiental e Gestão do Território (desde 1997), Pós-graduação stricto-sensu: Mestrado e Doutorado em População, Território e Estatísticas Públicas (desde 1998); Atividades de Capacitação e treinamento para servidores do IBGE. Desde sua fundação a ENCE passou por pelo menos duas importantes reformulações que resultaram de missões de avaliação externa da escola. Em 1975 na gestão de Isaac Kerstenetzky a missão francesa de avaliação da ENCE foi chefiada pela École Nationale de la Statistique et de l'Administration Économique (ENSAE) associada ao Institut Nationale de la Statistique et des Études Économiques (INSEE) e resultou em uma reestruturação do curso de bacharelado. Em 1986 no âmbito da reforma administrativa do IBGE realizada na gestão de Edmar

negociação começou em dois mil e dezenove. Eu ainda estava ativo na ENCE. Eu participei dos trâmites iniciais pra que essa conexão fosse feita e o processo levou um tempo grande e só veio se materializar no ano passado. Eu achei uma oportunidade incrível para a ENCE e o IBGE estar nesse ambiente” (Pedro Nascimento, entrevista em 21 de Junho de 2022).

Conforme verificado nas entrevistas com membros do hub, o projeto começou a operar formalmente a partir de agosto de 2021 com a elaboração de um primeiro plano de trabalho e divulgação de sua página na internet. O plano de trabalho do hub foi estruturado em quatro linhas de ação: i) fortalecer laços e promover a cooperação entre produtores de estatísticas oficiais na região; ii) capacitar e fomentar o interesse de jovens estatísticos sobre a utilização de big data nas estatísticas oficiais; iii) apoiar pesquisas sobre o uso de big data e ciência de dados; iv) organizar e realizar seminários e conferências.

De acordo com as informações disponíveis na página web do hub, a gestão da iniciativa é realizada por um grupo consultivo formado por representantes de organizações internacionais (ONU, BID, CEPAL), INEs latino-americanos (INEGI do México, DANE da Colômbia), pesquisadores nacionais (Nic.Br e UFPE), além de uma representante do ONS, instituto estatístico do Reino Unido. O grupo consultivo é coordenado por Andrea Diniz (ENCE/IBGE) e pelo representante da Divisão Estatística da ONU, Ronald Jansen. Há ainda um grupo técnico formado por oito servidores do IBGE, que além da ENCE, representam a Diretoria de Pesquisas (DPE), Diretoria de Geociências (DGC), Diretoria de Informática (DI) e Centro de Documentação e Disseminação de Informações (CDDI).

De acordo com informações do estatístico Delta⁹⁷, os recursos para o funcionamento da iniciativa, provém integralmente do IBGE:

O IBGE financia nove técnicos para se dedicar ao hub, não em tempo integral, mas a gente tem nove servidores trabalhando no hub. O IBGE financia outros servidores que trabalham pontualmente por exemplo preparando o material pra divulgação, a comunicação do hub, desenvolvimento de website e

Bacha, uma missão externa de avaliação coordenada por Simon Schwartzman foi contratada para avaliar as opções de manutenção da ENCE no IBGE ou sua transferência para alguma universidade. A missão concluiu pela importância de manutenção da vinculação institucional da escola ao IBGE, fortalecendo seu perfil acadêmico e as atividades de formação dos quadros da instituição (Senra, 2017).

⁹⁷ Delta é um estatístico com nível de doutorado. É servidor público em atividade e analista sênior no IBGE com mais de 20 anos de atuação. Delta participa da equipe do hub.

atualização de website. O IBGE coloca um monte de recursos humanos. O IBGE cede a estrutura pro Hub funcionar, toda a estrutura computacional, a página está hospedada no IBGE. Toda a estrutura é o IBGE que está provendo. Por enquanto a ONU provê apoio, mas não apoio financeiro nesse momento (Delta, entrevista em 30 de Junho de 2022).

Apesar de o IBGE ser o provedor de recursos humanos e materiais para o funcionamento da iniciativa, Delta destacou em mais de uma oportunidade durante nossa entrevista, que o caráter regional e internacional do hub se sobrepõe à sua vinculação nacional e institucional. Por exemplo, ao se referir ao modo correto de citar o hub enfatizou que: “em Inglês é Regional Hub for big data, você pode em português, chamar de Hub regional pra big data no Brasil, mas sempre com o cuidado de dizer que serve à América Latina e Caribe, pra ninguém ficar pensando que esse hub é brasileiro. Não é brasileiro, é parte da iniciativa de Big Data das Nações Unidas” (Delta, entrevista em 30 de Junho de 2022). Em outro momento da entrevista, ao ser perguntado sobre as iniciativas com big data desenvolvidas internamente pelo IBGE, Delta destacou que ‘o hub regional não é um hub pra servir ao IBGE, ao Brasil, é um hub que serve a América Latina e o Caribe’ e que, portanto, ‘não olha pra dentro, olha pra fora’:

Dentro do IBGE a gente tem algumas iniciativas. A gente sabe porque a gente conversa com os colegas e eu sei por dois outros motivos, Primeiro porque o hub regional pra big data ele tem um grupo técnico com representantes de todas as diretorias do IBGE mais o CDDI e a ENCE. Então a gente sabe mais ou menos o que está acontecendo em todas as áreas do IBGE em relação a isso. É um hub regional pra uso de big data, as pessoas são do grupo técnico e a gente só fala sobre uso de big data. É verdade que o hub regional não é um hub pra servir ao IBGE, ao Brasil, é um hub que serve a América Latina e Caribe. Então a gente não olha pra dentro, a gente olha pra fora, olhar pra dentro é uma das coisas que a gente faz primeiro porque a gente troca com os colegas, os colegas estão dando horas do seu trabalho para o hub, e, segundo, porque o hub regional conduziu uma consulta [sobre uso de big data na América Latina que incluiu o IBGE] (Delta, entrevista em 30 de Junho de 2022).

A avaliação de que há uma desvinculação da iniciativa do hub com os objetivos institucionais do IBGE, bem como de que há uma apreciação ‘ingênua’ dos interesses envolvidos no projeto, foi objeto de críticas do ex-presidente Roberto Olinto em nossa entrevista:

É interessante essa história, chamar o IBGE pra fazer o UN Big Data Hub. O que o IBGE vai ganhar com isso? pessoas correndo pra participar, pra fazer viagem internacional ou pra dizer que é importante, colocar no currículo que fez uma coisa internacional e no fundo ninguém pensa o seguinte, o que o

IBGE vai ganhar com isso? [...] Vamos ver, tem duas coisas, como gestor do IBGE eu mandei muita gente viajar, por que eu sabia que o aprendizado daquele técnico era a viagem em si e não o seminário. Era ir pra um outro mundo, conhecer um seminário internacional, isso é aprendizado e retorna pro IBGE como um treinamento. A outra coisa é a ingenuidade, porque existem outros interesses envolvidos ali (Roberto Olinto, entrevista em 09 de junho de 2022).

4.5.1 Hub regional e Plataforma Global: posicionamentos e controvérsias.

A vinculação do hub com o projeto da Plataforma Global da ONU foi um dos aspectos que exploramos nas entrevistas realizadas com membros do hub regional de Big Data no Brasil. Notadamente, identificamos distintos entendimentos sobre o projeto entre os agentes entrevistados.

Para o estatístico Pedro Nascimento, a Plataforma Global é uma iniciativa voltada ‘principalmente para ajudar países que não tem capacidade’ tecnológica e estatística, como os países africanos. De acordo com sua visão, INEs de países mais bem estruturados tecnologicamente, como é o caso do IBGE, não necessitariam da infraestrutura da plataforma:

É experimentação, uma facilidade de computação na nuvem em que você pode depositar dados volumosos e operar sobre eles com ferramentas computacionais adequadas, principalmente pra ajudar países que não tem capacidade mesmo de ter isso. Eu, por exemplo, já ministrei cursos em vários países africanos, fui a Moçambique, é uma situação que o IBGE, não que a gente seja o topo do mundo em termos de ferramentas, mas quando comparados com Moçambique nós somos extremamente ricos e afortunados [...] a gente aqui tem muitas dificuldades, mas ao mesmo tempo a gente precisa reconhecer que a gente tem coisas fenomenais no IBGE. O IBGE é uma instituição muito bem posicionada em tecnologia e tem muitos países que não chegam nem perto da gente. E pra apoiar esse tipo de país que não tem tamanho, não tem infraestrutura, não vai ter as competências pra ter isso, eles criaram esse ambiente. É compartilhamento inclusive de capacidade computacional pra você explorar e operar com essas fontes novas, compartilhamento também de processamento. (Pedro Nascimento, entrevista em 21 de Junho de 2022).

O estatístico Beta⁹⁸, por outro lado, apresentou uma interpretação bastante diferente da iniciativa da ONU. Ao contrário da visão ativa de Nascimento quanto a capacidade tecnológica e estatística do IBGE, o discurso do jovem estatístico revela uma percepção de inferioridade frente aos INEs do Norte-Global. Para Beta, a Plataforma Global foi concebida justamente para estimular a ‘cooperação’ de INEs de

⁹⁸ Beta é um estatístico com mestrado em ciência da computação. É servidor público ativo e analista pleno no IBGE com menos de 10 anos de atuação. Beta é membro da equipe do hub.

países em desenvolvimento como o Brasil, com instituições mais avançadas, ‘inclusive das esferas privadas’. É interessante observar a expressão de um sentimento de frustração de Beta, pelo fato de o IBGE ainda não estar plenamente alinhado com a iniciativa:

Eu entendo que a Plataforma Global é um centralizador de bases de dados de metodologias implementadas em um servidor de processamento seguro que você pode acessar tanto pra desenvolver quanto até eventualmente pra colocar em produção algum projeto. O fato de ser uma Plataforma Global centralizada, permite a interação de pessoas de diferentes instituições, inclusive das esferas privadas. Entendi que a proposta é nessa linha. Que eu acho superinteressante mas a gente, infelizmente, ainda não está muito nessa pegada (Beta, entrevista em 15 de julho de 2022).

Perguntado sobre a relação entre o hub do Brasil e a Plataforma Global, Beta inferiu uma relação de superioridade hierárquica da Plataforma para com os hubs regionais, ‘nestas localizações estratégicas’, ou seja, no Sul-Global. O jovem estatístico, entretanto, não faz uma apreciação crítica desta provável relação de subordinação geopolítica. Pelo contrário, a posição de ‘inferioridade’ dos INEs de ‘países em desenvolvimento’ em relação aos seus pares nos países do Norte-Global é justificada pela ausência de cooperação com estas instituições. O discurso de Beta revela uma crença no caráter eminentemente positivo da inovação e nas dádivas que ‘eles’, agentes hierarquicamente superiores do Norte-Global, proporcionam para ‘a gente’, INEs do Sul-Global.

A Plataforma Global é um esforço maior que o hub, na verdade os hubs eles estão ali inseridos dentro da Plataforma Global de um ponto de vista talvez hierárquico, não sei se de fato é assim, mas os hubs fazem parte da Plataforma Global e essa estrutura é uma espécie de meio pra viabilizar todo tipo de interação que for necessário. Essa infraestrutura que eles estão oferecendo, todos os eventos que eles patrocinam, vem nessa essência de incentivar a cooperação, nessa linha, eu acho que os hubs regionais, nessas localizações estratégicas, vêm realmente pra trazer esse corpo, porque se você falar em Estados Unidos e Europa, eles já se falam há muito tempo, em termos de estatística oficial. Eles já têm uma interação muito maior [...] A participação de países como Estados Unidos, Canadá, Nova Zelândia, Austrália, ela é bem presente também, são os INEs que estão na vanguarda do trabalho que a gente faz, realmente esses países em desenvolvimento, vamos colocar assim, ficam totalmente à margem. A gente muitas vezes sequer consegue acompanhar a discussão, porque eles estão num nível de desenvolvimento muito distante (Beta, entrevista em 15 de julho de 2022).

Ao exemplificar a assimetria ‘técnica’ existente entre ‘eles’ (INEs do Norte-Global) e ‘a gente’ (INEs do Sul-Global), e os ganhos com as iniciativas de

cooperação, Beta revelou uma forte aderência ao discurso de modernização das estatísticas oficiais. A ideia de 'modernização' aqui se revela atrelada a adoção de 'boas práticas' desenvolvidas pelos INEs do Norte-Global.

A gente teve um documento muito interessante que foi publicado, um Guidelines, e a gente pegou esse documento quando a gente já estava nessa discussão da modernização, e falou assim: 'vamos modernizar, vamos usar isso como referência pra modernizar até onde a gente puder, claro, com as nossas limitações, dentro das boas práticas que eles adotam'. Só que eles davam ali como ponto pacífico coisas que a gente ainda estava tentando discutir e muitas vezes a gente sequer conseguia participar da discussão. É muito interessante que a Divisão Estatística da ONU em todos os braços que estão envolvidos, eles sempre tiveram essa preocupação, que está muito evidente agora. Foi interessante que eles pediram pros países membros comentarem sobre esse guidelines, e colocamos lá algumas coisas que a gente aproveitou e nesse comentário, eu falei 'aqui e ali tem umas coisas que ninguém aqui fala sobre isso' e eles fizeram uma nova versão e contemplaram nosso pedido. Então é essa interação acontecendo, essa cooperação de fato acontecendo, eles que já desenvolveram, já avançaram, compartilhando também com a gente (Beta, entrevista em 15 de julho de 2022).

Evidentemente existem ganhos potenciais com as iniciativas de cooperação técnica, na medida em que novos métodos e práticas que já foram implantados e testados, podem ser mais facilmente adaptados em outros contextos, na medida em que os técnicos tenham acesso à documentação destas experiências e mesmo aos seus pares em outras organizações, que enfrentaram problemas similares. Por outro lado, o desenvolvimento de ferramentas estatísticas não é um problema meramente técnico, mas engloba questões sensíveis de ordem econômica, política e geopolítica. Neste sentido, a visão de Beta sobre a cooperação com os INEs do Norte-Global, mediada pela Divisão Estatística da ONU revela uma postura de subordinação, que busca recompensas simbólicas por meio da participação em instâncias internacionais, onde o próprio reconhecimento da alegada posição de 'inferioridade' é valorizado e recompensado.

Algumas críticas do ex-presidente Olinto à iniciativa do hub vão nesta direção, ou seja, do tratamento de 'boas práticas' como questões meramente técnicas e da participação em espaços internacionais sem a devida avaliação de suas implicações políticas. Utilizando a OCDE como um exemplo, o ex-presidente, que coordenou por muitos anos a área de contas nacionais do IBGE, chamou atenção para a relação intrínseca entre as questões de economia política e as ferramentas estatísticas, bem

como para as implicações políticas e econômicas da adesão a determinados acordos internacionais:

Vou te dar um exemplo análogo [ao hub da ONU Big Data]. Durante muito tempo tive o pé atrás com o IBGE entrar na OCDE. Continuo com o pé atrás. Mas por quê? Na minha experiência pessoal, quando nós fomos convidados pra ser observadores de algumas reuniões, junto com os BRICS, eles mandavam na época pra área de contas nacionais pedidos de dados e você tinha que obedecer. Era uma coisa meio compulsória, ‘a gente quer que vocês produzam as contas nacionais com uma classificação x, assim e assim, com essas tabelas’ e mandavam as tabelas pra gente preencher, como se a gente fosse um miquinho adestrado. Eles são extremamente autoritários. Então quando você assina com a OCDE, politicamente você está se declarando uma economia liberal e estatisticamente você vai ter que seguir o que a OCDE determinar. Não quer dizer que você tenha que fazer só aquilo. Esses acordos internacionais são muito complicados, e em algum momento você vai ter alguns compromissos que você vai ter que seguir, e colocar pessoas pra cuidar disso, o que é pior, por que não está sobrando, ao contrário, falta (Roberto Olinto, entrevista em 9 de junho de 2022).

Retomando a questão da Plataforma Global, conforme vimos anteriormente no capítulo três, o projeto nasceu como uma parceria público-privada com a participação de Big Techs e orientada por objetivos de mercado que visavam intensificar a participação do setor privado no campo estatístico. Funcionários de alto nível do ONS criaram uma organização sem fins lucrativos, o Global Platform Institute para gerenciar o projeto, o que não chegou a se concretizar. A posterior transferência da gestão do projeto para a GPSDD, que tem entre seus financiadores fundações filantrópicas vinculadas à Google e Microsoft, configura potenciais riscos à soberania nacional dos países da América Latina e em particular do Brasil, que hospeda a iniciativa.

Notadamente, evidências colhidas nas entrevistas com membros do hub regional apontaram para uma tentativa de desvincular o hub do projeto da Plataforma Global a partir de 2022. Durante nossa entrevista com Delta, por exemplo, ao nos referirmos ao ‘hub da Plataforma Global’, denominação que consta no acordo formal assinado pelo IBGE, fomos corrigidos e informados que, “o hub é uma coisa e a plataforma é outra”:

Não é só Plataforma Global. O Nome da iniciativa é UN Big Data, é mais do que a plataforma global. Na UN Big Data você tem a plataforma global, tem os task teams e os hubs, que são três coisas diferentes. A gente no começo chamava Hub da Plataforma Global mas não é da Plataforma Global. Plataforma global é uma coisa. O hub é outra coisa e os task teams também. São três frentes diferentes (Delta, entrevista em 30 de junho de 2022).

A retificação feita por Delta, entretanto, divergiu da apresentação do hub, realizada por sua coordenadora, a servidora do IBGE Andrea Diniz, na primeira atividade pública do projeto no Brasil. Nesta ocasião, a iniciativa havia sido assim apresentada por Diniz:

A missão do hub regional da América Latina é fazer essa ponte, é garantir que a Plataforma Global consiga trabalhar junto com os institutos de estatística e outras instituições de informação de um modo geral e com isso fomentar e apoiar o avanço do uso de big data para a produção de estatística oficial e também para monitorar os Objetivos de Desenvolvimento Sustentável (Hub, 2021).

De acordo com informações obtidas em off com um dos membros do hub do Brasil⁹⁹, em reunião ocorrida em janeiro de 2022 representantes dos hubs regionais fizeram questionamentos à coordenação global do projeto da ONU big data sobre a natureza de sua vinculação ao projeto da Plataforma Global. De acordo com as informações obtidas, havia muitas dúvidas sobre o funcionamento, gestão e objetivos da Plataforma. Nossa fonte tinha conhecimento sobre as controvérsias envolvendo os membros do ONS. Não identificamos, entretanto, maiores conhecimentos desta fonte, nem tampouco de outros membros do hub do Brasil, sobre a GPSDD. Nesta reunião, portanto, ainda segundo esta fonte, foi tomada a decisão de desvincular os hubs da Plataforma Global e de tratá-los como iniciativas independentes.

Finalmente, compreendemos os esforços de autonomização dos hubs em relação à Plataforma Global como indícios de lutas que emergem na medida em que ações globais do projeto são traduzidas por suas contrapartes regionais. Os hubs regionais são formados por estatísticos de Estado, sujeitos a um conjunto de regras, normas e disposições sancionadas por seus campos estatísticos nacionais e instituições. Tais lutas, evidenciam contradições entre a lógica pública dos INEs e a lógica de mercado da Plataforma.

4.5.2 Percepções sobre a incidência de Big Techs no campo estatístico

Vimos no capítulo dois como a lógica de mercado e de comodificação dos dados passou a constituir um dos fundamentos implícitos no discurso de

⁹⁹ Durante a entrevista, ao tratarmos especificamente sobre os motivos que levaram a autonomização do hub com relação ao projeto da Plataforma Global, esta fonte solicitou que interrompêssemos a gravação e que as informações fossem utilizadas em off.

modernização das estatísticas oficiais que emergiu na Europa. Ainda conforme analisamos no capítulo três, o projeto da ONU Big Data tem sido pautado pelo incentivo às parcerias público-privadas e experimentação de novos modelos de negócios no campo estatístico com participação do setor privado, o que inclui a presença de Big Techs.

Verificamos que a participação de grandes empresas de tecnologia no nível global do projeto é de conhecimento de alguns servidores do IBGE envolvidos na iniciativa do hub. Beta, por exemplo, confirmou a participação da Microsoft e outras empresas de grande porte em um dos grupos de trabalho. A princípio, a participação destas empresas foi apresentada pelo estatístico como pautadas por um caráter não comercial:

O papel de empresas como a Microsoft, por exemplo, que eu tenho visto, do trabalho que eu tenho acompanhado, tem sido no desenvolvimento da tecnologia mesmo, você tem ali um profissional que é um grande especialista em algoritmos, em criptografia, em coisas desse tipo, e eles estão dentro desse grupo. Além da Microsoft tem duas ou três empresas grandes e mais uma empresa menor que também entrou agora nessa área, enfim, eles dão contribuições bem pertinentes, mas eu acho que envolve mais essa questão da contribuição, em termos de conhecimento, de experimentação, com acesso a toda infraestrutura que eles tem lá na realidade deles (Beta, entrevista em 15 de julho de 2022).

Ao ser perguntado sobre possíveis motivações de mercado na participação destas empresas, Beta apontou para o potencial de negócios representado pela área de computação distribuída e tecnologias de ampliação de privacidade. De acordo com Beta, prováveis interesses comerciais da Microsoft podem estar relacionados ao desenvolvimento de serviços e aplicações para novas modalidades de consumo e compartilhamento de dados.

Nessa questão especificamente da privacidade eu diria que sim, porque uma das aplicações desse tipo de tecnologia é realmente você conseguir consumir dados [...] para uma Microsoft estar envolvida num caso desses é talvez pra de fato oferecer esse serviço, um Privacy Cloud da Microsoft em que esse tipo de processamento pode ser realizado. Eu acho que é isso que a Microsoft está pensando ali, em ser o provedor desse serviço. Ela talvez vá ser a primeira gigante nisso (Beta, entrevista em 15 de julho de 2022).

Beta é um entusiasta das tecnologias de ampliação de privacidade e nos explicou em maiores detalhes como vem sendo concebida uma nova estrutura tecnológica para o compartilhamento 'seguro' de dados entre órgãos públicos,

universidades e empresas privadas. De acordo com suas informações, podemos compreender melhor as potenciais vantagens comerciais vislumbradas pela Microsoft:

Qual é a lógica, qual é a estrutura que está sendo criada, você vai ter um servidor local com tudo que você precisar de proteção e você vai ter um nó de rede, um servidor de processamento que vai ter acesso, por exemplo, ao dado do IBGE, da Receita Federal, do Ministério do Trabalho, um nó do setor público. Aí você vai ter um outro servidor de processamento que ele vai ter acesso ao dado do servidor das universidade e você tem aqui um nó acadêmico, aí você vai ter um nó do setor privado, etc. Esses acordos vão sendo feitos e aquele nó está lá, nada muda, o que acontece é que agora o IBGE fez um acordo e ele passou a conseguir acessar o nó acadêmico, que antes ele não conseguia. Então além de você administrativamente fazer esse acordo, de ter acesso e de dar acesso ao seu dado para aquele servidor de processamento, o teu dado tem que estar preparado pra interagir com um protocolo de processamento seguro que está implementado naquele servidor. Então você imagina, a primeira gigante de tecnologia que oferecer esse serviço em nuvem e estabelecer o protocolo, vai ser quem vai tocar isso pela próxima década (Beta, entrevista 15 de julho de 2022).

Para Pedro Nascimento, o setor de estatísticas oficiais, quando considerado de um ponto de vista global, é um mercado importante para empresas de tecnologia, o que constituiria uma motivação legítima para a participação destas empresas em projetos como a ONU Big Data: “Eu acho que a participação da Microsoft nesse ambiente é para se posicionar como fornecedor de tecnologia, produtos, serviços, para serem consumidos por esses agentes [INEs] que, no mundo inteiro, quando somados, são um mercado importante pra produtos e serviços como os da Microsoft”. Nascimento, descarta interesses ‘explícitos’ de privatização do setor de estatísticas oficiais, mas especula sobre a possibilidade de que algumas empresas, estejam investindo na criação de um mercado futuro:

Eu não vejo o mundo privado ainda se interessando por este negócio do Estado, ainda é um negócio do Estado no mundo inteiro. Eu não vi iniciativa nos países, mesmo nos países mais capitalistas, mais avançados, dessas empresas dando sinais de querer entrar nesse mercado, nesse espaço, [mas] o mundo privado não é homogêneo, tem corporações que enxergam valor e investem. Algumas organizações, como a Google, podem estar investindo em criar, em potencialmente ajudar a criar, um mercado que ainda não é viável hoje mas que daqui a algum tempo ela possa tirar proveito (Pedro Nascimento, 21 de junho de 2022).

Já Roberto Olinto apontou para riscos de que a participação de grandes empresas privadas na iniciativa, em particular da Microsoft, possa implicar em acesso privilegiado aos dados dos países envolvidos.

Eu não confio em nada dessa coisa de ‘bom samaritano’, porque empresas como a Microsoft começam a ter um acesso num determinado nível de detalhe até a dados que teoricamente seriam sigilosos [...] Imagina você sendo um hub com dados de um monte de países. Qual é o qual o poder que a Microsoft vai ter neste projeto, de fazer análises e inferências sobre aquele monte de países, é muito poder (Roberto Olinto, entrevista em 9 de junho de 2022).

Para o demógrafo Ômega¹⁰⁰, os INEs vivem um cenário de concorrência. Seu discurso, em certo sentido, reproduz argumentos que vimos ser instrumentalizados no campo estatístico europeu e na ONU para incentivar a ‘modernização’ dos INEs. É interessante observar, entretanto, como o demógrafo do IBGE destaca a emergência de novos agentes e interesses relacionados à comodificação das informações. Para Ômega, estes agentes, ‘interessados no simples valor de mercado’, seriam distintos dos agentes do campo estatístico, movidos pelo interesse público.

Eu acho que é uma coisa que o Instituto de estatística tem que tomar um certo cuidado, é uma competição, não deixa de ser uma competição. Se a Google resolver competir com o IBGE, a gente não tem chance nenhuma. Eles têm dinheiro, eles têm tecnologia, eles vão ter tudo. Agora a gente tem que zelar pelo que a gente tem de melhor, que é a confiança e a sustentabilidade, porque a Google pode estar interessada nisso hoje e amanhã não está mais. Então a gente tem que segurar no que a gente tem de melhor. A gente é bom em fazer estatísticas. A gente garante a sustentabilidade disso. A gente tem séries históricas imensas e a gente está fazendo isso porque a estatística é um bem público, a gente não está interessado em ganhar dinheiro, nem em mercado. Agora a gente tem que reconhecer que cada vez mais informação é uma commodity. A gente não pode negar isso, informação tem valor de mercado e cada vez mais vai ter gente interessada nisso pelo simples valor de mercado e não pela produção do dado em si, pelo bem que ela pode gerar (Ômega, entrevista em 25 de maio de 2022).

Simon Schwartzman, por sua vez, identifica uma mudança no cenário internacional com o fortalecimento do setor privado em relação às instituições de coordenação internacional. Apesar de reconhecer a importância de vínculos com o setor produtivo, que conferem ‘vitalidade’ em áreas como a pesquisa científica, o ex-presidente do IBGE aponta para eventuais problemas de ‘conflitos de interesse’ na relação público-privada que precisam ser ‘negociados’ e regulados.

Eu acho que o cenário mudou e talvez no sentido de um enfraquecimento das instituições de coordenação internacional e fortalecimento do setor privado. E aí evidentemente você tem problemas de eventuais conflitos de interesse,

¹⁰⁰ Ômega é doutor em demografia. É um analista sênior e servidor público ativo no IBGE com mais de 20 anos de atuação e faz parte da equipe do hub.

quando você coloca uma grande empresa com todos os seus interesses comerciais em um negócio como esses. Eu acho que isso é uma questão geral, por exemplo, na pesquisa científica hoje em dia as universidades não podem trabalhar em isolado, a pesquisa tenta se ligar ao setor produtivo, às empresas, mas daí quando você começa a fazer isso, começam a surgir problemas de conflito de interesse. Então você tem que regular, você tem que estar atento ao que é o interesse público e o que é o interesse privado, estabelecer regras. É um processo de negociação que tem que ser feito o tempo todo. Mas evidentemente que tem que ter cuidado, ou seja, você não pode simplesmente deixar que o setor privado se aproprie dos recursos públicos pra seus fins privados de lucro (Simon Schwartzman, entrevista em 01 de Setembro de 2022).

Finalmente, vimos nesta seção que a implantação do Hub regional do Brasil, em um contexto de fragilidade institucional do IBGE, tem suscitado posicionamentos distintos sobre as prioridades e o papel da instituição no debate nacional e internacional da dataficação das estatísticas oficiais. Notadamente, a desarticulação de iniciativas internas que visavam o fortalecimento institucional do IBGE e do sistema estatístico nacional foi acompanhada pela adesão a um projeto global pautado por interesses internacionais e comerciais que envolve a participação de empresas privadas de tecnologia e Big Techs.

Os riscos de conflitos de interesse e as implicações políticas e econômicas da incidência de interesses internacionais e comerciais no campo estatístico nacional, embora identificados por agentes mais graduados do IBGE, não parecem ser devidamente equacionados entre todos os membros do hub no Brasil, em particular entre os quadros mais novos da instituição e os quadros mais militantes em favor do big data que atuam na coordenação do hub. Verificamos um contraste entre a posição ativa e crítica de quadros mais antigos da instituição, e a disposição de subordinação e encantamento técnico com as iniciativas internacionais de ‘cooperação’ e ‘modernização’ dos quadros mais novos. De todo modo, a tentativa de dissociação entre o Hub e a Plataforma indica tensões na tradução dos interesses da iniciativa global em suas instâncias regionais, coordenadas por estatísticos de Estado que estão sujeitos às regras, normas e disposições sancionadas pelos campos estatísticos nacionais e suas instituições.

4.6 A interface público-privada do Hub Regional

A realização de webinários tem consistido em uma das principais atividades realizada pelo Hub da ONU big data no Brasil. Participei como ouvinte de alguns

webinários em tempo real e acessei o conteúdo dos demais vídeos dos eventos no Youtube. Nesta seção procedemos a uma discussão do conteúdo destas atividades, que nos oportunizaram analisar a participação, posicionamentos, discursos e ações dos agentes públicos e privados, globais e nacionais envolvidos na iniciativa.

Entre dezembro de 2021 e outubro de 2023, foram realizados treze webinários sobre diversas temáticas relacionadas ao uso de big data em estatísticas oficiais, dentre os quais destaca-se o uso de dados de telefonia móvel, tema que contou com oito webinários. As inscrições são, em geral, realizadas por formulários virtuais do Hub do Brasil ou da Rede de Transferência de Conhecimentos da CEPAL. A maioria dos webinários foi mediada por representantes do IBGE que atuam no conselho técnico do Hub, mas também se verificou a mediação de representantes da CEPAL, da GPSDD e da Divisão Estatística da ONU. A audiência dos eventos é formada principalmente por profissionais de INEs da América Latina e a programação consiste em apresentações realizadas por representantes de INEs, agências da ONU e empresas privadas¹⁰¹.

As apresentações nos webinários tem envolvido principalmente casos de sucesso de INEs latino-americanos e de outras regiões no uso de novas fontes de dados para a compilação de estatísticas oficiais. Notadamente, todos os casos apresentados pelos INEs latino-americanos trataram de projetos que tinham como característica comum o desenvolvimento autônomo de métodos pelos próprios INEs, sem intermediação de empresas privadas, para a incorporação de novas fontes de dados abertos na produção de estatísticas oficiais. São exemplos: os casos do Brasil, Chile e Uruguai que utilizam dados da internet coletados por web scraping para a produção de índices de preço; os casos de uso de dados abertos de sensores orbitais por Colômbia, México, Paraguai e República Dominicana; o caso de uso de dados

¹⁰¹ Verificamos a realização dos seguintes webinários pelo Hub de big data do Brasil: i) Ferramentas de Capacitação da UN Big Data em 21 de setembro de 2022, ii) Using Mobile Phone for Official Statistics: Opportunities and Challenges em 06 de outubro de 2022, iii) Big Data for Official Statistics em 17 de novembro de 2022, iv) Using Mobile Phone Data for Official Statistics on Migration em 15 de dezembro de 2022, v) Using Mobile Phone Data for Official Statistics on Tourism em 09 de fevereiro de 2023, vi) Using Mobile Phone Data for Official Statistics on Population em 09 de março de 2023, Using Mobile Phone Data on Displacement and Disaster statistics em 06 de abril de 2024, vii) Using Mobile Phone Data for Official Statistics on Information Society em 04 de maio de 2023, viii) Using Mobile Phone Data for Official Statistics on Transportation and Commuting em 01 de junho de 2023, ix) Mobile Phone Data Access for Official Statistics em 07 de julho de 2023, x) Using Mobile Phone Data on Trending Topics em 03 de agosto de 2023, xi) Use of Earth Observation for Official Statistics em 09 de agosto de 2023, xii) Webscraping for prices em 06 de setembro de 2023, xiii) Data Science in Official Statistics em 11 de outubro de 2023.

abertos do X (ex-twitter) pelo México para a produção de indicadores de sentimento da população; e o caso do Uruguai que utiliza dados de telefonia celular de sua operadora pública para estatísticas de mobilidade, deslocamentos e desastres naturais.

Conforme pudemos verificar em nossa observação dos webinários, a apresentação de experiências pelos INEs latino-americanos tem possibilitado um intercâmbio relevante de informações sobre as novas fontes de dados, métodos e tecnologias utilizadas nos projetos. Foi ainda possível observar a interação com o público participante na forma de perguntas e respostas ao final das apresentações, que viabilizaram trocas de informações mais detalhadas sobre as experiências e debates sobre temas específicos. Notadamente, a iniciativa promovida pelo hub tem contribuído para a articulação regional dos INEs latino-americanos e apresenta um potencial relevante para desencadear iniciativas de cooperação entre os agentes e instituições do campo estatístico na região.

No que se refere à presença de empresas parceiras da Plataforma Global nos webinários, verificamos que suas participações consistiram na apresentação de ‘casos de sucesso’, ‘boas práticas’, ‘soluções’ e ‘modelos de negócios’ desenvolvidos para o uso de big data na compilação de estatísticas oficiais. Constam, neste sentido, a participação das empresas Positium, Anonos, Flowminder e Data Pop Alliance. Estas empresas podem ser caracterizadas como agentes intermediários, que buscam construir e ocupar um espaço em um novo mercado de dados para estatísticas oficiais, potencialmente constituído na relação entre os governos nacionais, os Institutos Nacionais de Estatísticas, os grandes detentores de dados (plataformas de dados e operadoras de telefonia celular) e os grandes provedores de infraestrutura tecnológica (Big Techs).

As atividades nos webinários também se constituíram em momentos de articulação e preparação para a participação de representantes de INEs latino-americanos no ‘Festival de Dados’ de 2023. Organizado pela GPSDD e pelo INE do Uruguai, que sediou o evento em Punta Del Leste, o festival contou com financiamento da Google.Org, Microsoft e Bill and Melinda Gates Foundation. A edição na América Latina deu seguimento ao ‘Festival de Dados para o Desenvolvimento’, realizado no Reino Unido em 2018. Os festivais consistem em uma espécie de feira global para estimular parcerias e negócios entre os setores público e privado na área de dados. De acordo com a GPSDD, os festivais são:

Um fórum para celebração, troca e descoberta de soluções práticas de dados para os maiores desafios do mundo [...] para profissionais de todos os setores com foco em tomada de decisões para melhorar vidas e proteger o planeta [...] que estão se esforçando para resolver desafios de desenvolvimento, oferecendo novas soluções de dados e explorando métodos para colocar as pessoas no centro dos dados. (GPSDD, 2023)

Conforme vimos anteriormente, a estratégia discursiva da ‘revolução de dados’, instrumentalizada pela GPSDD, atua para encapsular os objetivos mercadológicos da rede em um conjunto de valores positivos e humanitários que constituem um importante mecanismo de encobrimento, sensibilização e encantamento dos agentes envolvidos. Neste sentido, a ideia de que é necessária uma ‘mudança cultural’ no campo estatístico, que, conforme vimos, foi incorporada ao discurso de ‘modernização das estatísticas oficiais’, se revela igualmente um elemento estratégico da iniciativa da ONU Big Data para sensibilizar e envolver os agentes públicos do campo estatístico, de modo a torná-los mais receptivos à ‘inovação’, à ‘modernização’ e à participação do setor privado no setor estatístico.

A seguir analisamos algumas discussões e debates efetuados nas sessões dos webinários, destacando evidências da orientação estratégica global do projeto, com vistas à ampliação da participação da iniciativa privada no setor estatístico latino americano.

4.6.1 Viés mercadológico e incidência da iniciativa privada

O webinário que precedeu a realização do Festival de Dados, em novembro de 2023, foi mediado pelo coordenador Global do Projeto da ONU Big Data, Ronald Jansen. A sessão abordou o uso de dados de sensores orbitais e contou com apresentações dos INEs do Paraguai e República Dominicana. Analisaremos a seguir, o debate travado entre o mediador e a representante da República Dominicana, na medida em que nos ajuda a ilustrar o enquadramento enviesado da coordenação global do projeto, que caracteriza o que denominamos de movimento pró-mercado no campo estatístico.

Após a apresentação da experiência do ONE da República Dominicana, por sua representante Crismairy Mena, que versou sobre o uso de imagens de satélite

para a produção de estatísticas ambientais, a sessão de perguntas e respostas teve início com uma pergunta formulada pelo próprio Jansen:

Eu estou interessado em saber sobre os dados de satélite que o ONE está utilizando, quais são as suas fontes, são comerciais ou abertas? E em adição a isso, você disse que usa alguns drones. São drones do ONE ou você contrata de empresas privadas? E, em geral quem são seus parceiros, não apenas na aquisição de dados de satélite mas no processamento dos dados e no conhecimento especializado adicional (Hub, 2023).

Em sua resposta, Mena explicou à Jansen que toda a iniciativa da República Dominicana, neste projeto, é de caráter público e governamental, sendo desenvolvida de forma autônoma por técnicos e especialistas do próprio ONE:

No que diz respeito às imagens de satélite, nós trabalhamos com as imagens da NASA que são públicas, tem um site na internet que você pode acessar e os dados são gratuitos. Quanto à segunda pergunta dos drones, nós temos drones no nosso escritório, mas nesse exercício nós precisávamos de um tipo específico de drone, então estamos usando drones de uma instituição que é uma subsidiária do ONE que está nos emprestando. Com relação aos peritos, nós temos um departamento de geoestatística que analisa as imagens e este exercício é coordenado pelo nosso departamento de estatísticas ambientais (Hub, 2023)

Aparentemente não satisfeito com a resposta de Mena, Jansen insistiu em uma segunda pergunta sobre o tema das parcerias com o setor privado. É interessante observar como Jansen associou o que intitula de 'nova era de trabalho com dados não tradicionais' com as parcerias com o setor privado:

Obrigado pela resposta, mas nessa nova era de trabalho com dados não tradicionais, o seu instituto também faz acordos com o setor privado ou institutos de pesquisa? por que pelo que entendi do que você acaba de explicar tudo está num nível governamental, correto? Ou em um nível interno da República Dominicana. Vocês também tem acordos de parceria com o setor privado ou institutos de pesquisa para desenvolver mais esse tipo de trabalho? você mencionou em sua apresentação 'termos de licença de uso', então deve haver alguma tecnologia do setor privado que vocês utilizam, talvez você possa explicar um pouco mais isso? (Hub, 2023).

A resposta da representante do ONE à segunda pergunta foi taxativa em afirmar que o instituto não possui acordos com a iniciativa privada, que envolvam a participação de empresas no processo de produção estatística. Mena, entretanto, destacou a existência da interação com universidades e acordos com algumas

associações do setor privado, enquanto instituições usuárias do sistema estatístico nacional:

Com o setor privado nós não temos nenhum acordo, mas estamos trabalhando para a criação do ecossistema de dados das universidades e temos interação com as universidades porque elas participam em atividades que convidamos e o uso externo das estatísticas faz parte do plano estatístico nacional, então atualmente estamos trabalhando em um acordo entre as principais instituições da República Dominicana para criar o ecossistema de dados da academia. Temos acordos com algumas associações do setor privado mas não exploramos com eles dados de fontes não tradicionais, os acordos são mais no intuito de responder às suas necessidades de informação (Hub, 2023).

O debate revela dois elementos significativos da experiência em curso junto ao campo estatístico latino-americano. Em primeiro lugar, há que se destacar a característica pública e governamental dos projetos com uso de novas fontes de dados que vêm sendo realizados de forma autônoma pelos INEs da região. Em segundo lugar, verifica-se o foco da coordenação global do projeto da ONU no envolvimento do setor privado. Neste sentido, tudo parece funcionar como se a lógica do projeto global concebesse a participação do setor privado no setor de estatísticas oficiais como um fim em si mesmo. Notadamente, não observamos o mesmo enquadramento entre os mediadores do hub do Brasil e da CEPAL.

Em que pese o fato de que os webinários tem funcionado como um espaço relevante de interação entre os INEs da região e de que as experiências das empresas podem contribuir para a apropriação de soluções técnicas adaptadas aos contextos e prioridades nacionais dos países latino-americanos, verifica-se, por outro lado, a construção de uma espécie de mercado regional para a promoção de produtos e serviços de empresas parceiras da Plataforma.

O viés mercadológico do projeto ONU Big Data, enfatizado em eventos como o 'festival de dados' pode vir a produzir efeitos nos INEs latino-americanos, cujas abordagens têm se pautado principalmente por iniciativas públicas e governamentais. Neste sentido, determinadas relações público-privadas com agentes de empresas internacionais podem revelar-se eventualmente problemáticas do ponto de vista da ética profissional, da normatização pública e da soberania nacional dos países envolvidos. Notadamente, para além do viés discursivo, colhemos evidências de incidência comercial direta de uma das empresas participantes na iniciativa junto aos INEs latino-americanos.

Muito provavelmente em face de minha vinculação institucional com o IBGE, após minha inscrição nos webinários passei a receber e-mails comerciais da empresa Positium que é uma ‘parceira confiável’ da Plataforma Global¹⁰². Os e-mails, dirigidos pessoalmente em meu nome, visavam o agendamento de reuniões ou indicação de contatos em minha organização para a apresentação de produtos da empresa na área de big data para estatísticas oficiais. No primeiro e-mail da empresa, a ONU foi citada como referência na abertura da comunicação:

Prezado Sr. Dalva, conversei com algumas pessoas da ONU e dos institutos nacionais de estatística (INE) em todo o mundo sobre o uso de MPD [dados de posicionamento móvel] para o bem da sociedade, e gostaria de entrar em contato com você sobre o mesmo assunto. Tenho o prazer de compartilhar algumas ideias sobre a utilização de fontes de dados inovadoras, como MPD e registros oficiais, para produzir estimativas populacionais e estatísticas oficiais para censos. Por meio do MPD, os INEs podem coletar dados populacionais mais granulares, precisos e oportunos com uma taxa de eficiência mais alta em comparação com métodos mais tradicionais. Você pode acessar mais informações aqui [link]. Estarei me reunindo com alguns outros representantes de INEs na próxima semana. Qual seria o melhor momento para você conversar? Você pode encontrar minha disponibilidade que corresponda à sua programação através deste link. Alternativamente, eu também apreciaria se você me encaminhasse para alguém em sua organização com quem eu possa discutir mais detalhadamente este tópico. Obrigado. Atenciosamente, Evest Defiana - Gerente de Parceria Estratégica e Desenvolvimento de Negócios - Positium (e-mail recebido em 29 de junho de 2023).

As mensagens da empresa Positium evidenciam a utilização da articulação dos hubs regionais como forma de acessar os INEs, enquanto potenciais consumidores de produtos e serviços desenvolvidos por parceiros da Plataforma. O compartilhamento dos dados da mala direta das ações formativas do Projeto ONU Big Data com as empresas parceiras, pode potencialmente configurar uma situação de conflito de interesses¹⁰³. Conforme vimos, a participação destas empresas tem se

¹⁰² A Positium é uma empresa de tecnologia de pequeno porte para os padrões europeus (faturamento anual de 1,06 milhão de Euros e 18 funcionários em 2022) sediada em Tartum na Estônia (Storybook, 2024). A empresa é especializada em análise de dados de telefonia celular e parceira da Plataforma Global da ONU.

¹⁰³ De acordo com a legislação brasileira, a situação de ‘conflito de interesses’ é caracterizada quando o confronto entre público e privado implica prejuízo para o interesse coletivo ou para o desempenho da função pública. Neste sentido, pode haver conflito mesmo que não haja dano ao patrimônio público ou ganho financeiro decorrente da atividade privada. A lei 12.813, de 16 de março de 2023, que dispõe sobre o conflito de interesse na administração pública federal estabelece (no art. 5º. Inciso I) entre as situações que configuram esta situação: “divulgar ou fazer uso de informação privilegiada, em proveito próprio ou de terceiro, obtida em razão das atividades exercidas”.

dado por meio do estabelecimento de relações a um nível global com a Divisão Estatística da ONU e GPSDD. Estas relações são 'oficializadas' com a atribuição de um título de 'parceiro confiável' da Plataforma Global que lhes possibilita a participação e acesso privilegiado às atividades realizadas com os INEs, seja por meio de interações na própria Plataforma, nos grupos de trabalho ou nos eventos realizados pelos hubs regionais.

A participação da Positium e outras empresas especializadas no grupo de trabalho de dados de telefonia móvel da ONU Big Data foi apontada pelo analista Ômega do IBGE em uma entrevista. Para Ômega, estas empresas identificaram um 'nicho de mercado' nas estatísticas oficiais e estão no projeto para 'vender consultorias':

No task team de telefonia móvel tem algumas empresas que trabalham com a produção de dados, com o processamento desses dados, algumas empresas, são poucas empresas no mundo que trabalham com dados de telefonia móvel pra gerar estatísticas. São pouquíssimas, e tem algumas que descobriram esse nicho de mercado, pegar esses dados de telefonia e investir em produção de estatísticas. Elas estão lá na ONU porque elas ensinam os outros a usar, qual o interesse delas? elas vão vender consultoria (Ômega, entrevista em 25 de maio de 2022).

De fato, pudemos confirmar por meio de dados públicos sobre o desempenho financeiro da empresa Positium, que a venda de serviços para INEs e órgãos de governo de países da Ásia (Indonésia) e Oriente Médio (Omã e Emirados Árabes Unidos) iniciados a partir de 2017, dobraram o faturamento da empresa a partir de 2018 e alcançaram 58% das receitas da empresa em 2021 (Storybook, 2024). Estes dados confirmam que o setor estatístico de países do Sul-Global constituem o principal 'nicho de mercado' da empresa e são coerentes com o seu interesse de mercado nos países da América Latina.

De acordo com Ômega, os interesses comerciais das empresas que estão atuando dentro do grupo de trabalho são legítimos, 'é trabalho e eles querem ser pagos por isso'. O analista aponta para os benefícios concretos da colaboração, na medida em que as empresas possuem conhecimentos especializados, que os institutos de estatística necessitam e ainda não detêm. Para Ômega é mais eficiente utilizar ferramentas já desenvolvidas das empresas especialistas, do que 'inventar a roda' dentro do IBGE. O analista, entretanto, projeta uma autossuficiência futura dos INEs neste campo:

Essas empresas encontraram um nicho de mercado onde elas podem atuar. Elas têm um conhecimento profundo dos dados, em como processar esses dados, elas tem experiência, a gente não tem, a gente quer usar esses dados e eles tem como nos ensinar a usar esses dados, eles tem a experiência, eles tem o software já prontos, eles já tem a roda feita, a gente não vai precisa inventar a roda, eles já tem um script pronto pra que eu vou pegar pessoal de informática pra quebrar a cabeça, fazer uma query, um script para processar aqueles dados, se eles já tem? Eles já conhecem a fundo aqueles dados. Então assim eu acho que um ajuda o outro, e é claro, é trabalho, eles querem ser pagos por isso com a expertise que eles têm [...] Claro que eu acho que vai chegar uma hora que a gente vai ficar auto suficiente nisso, mas a gente está longe disso. A gente está engatinhando. A gente mal conhece os dados (Ômega, entrevista em 25 de maio de 2022).

Para o estatístico Gama¹⁰⁴, o envolvimento de empresas intermediárias como a Positium é uma solução adotada em alguns países para mediar o acesso aos dados entre o setor público e os grandes detentores de bases de dados. Gama, não vê ‘com maus olhos’ este ‘modelo de negócios’, desde que a relação seja ‘bem estabelecida e combinada entre todos os participantes’:

Por enquanto o Brasil não tem capacitação em uso de dados de telefonia, então essas empresas surgem. Em outros países é comum essa coexistência. Em outros países essas empresas fazem a mediação entre o setor privado e o público no acesso ao dado. Então a empresa privada ela não manda direto uma base pro setor público, pro produtor de estatística pública, mas ela tem um terceiro fazendo intermediação dessa fonte de Big Data e através desse terceiro que é feito todo o modelo de negócio e todo o modelo pra geração de indicadores e de estatísticas públicas e oficiais com os dados [...] Eu não vejo com maus olhos isso não. Se for uma relação bem estabelecida e combinada entre todos os participantes. Eu não vejo problema (Estatístico Gama, entrevista em 15 de setembro de 2022).

A ‘reinvenção da roda’, como sinônimo de atividade contraproducente, e a ‘coexistência comum em outros países’, são estratégias discursivas utilizadas por Ômega e Gama, para justificar o novo contexto de parcerias público-privadas no campo estatístico. Estas justificativas, entretanto, tendem a ignorar ou encobrir objetivos políticos e econômicos que orientam as ‘soluções técnicas’ e os ‘modelos de negócios’ desenvolvidos por agentes externos. Reinventar a roda, neste caso, pode significar levar em conta as premissas que embasam as soluções técnicas e os modelos de negócios de modo a deter maior controle e orientá-las de acordo com o interesse público, institucional e nacional. A seguir analisamos em maiores detalhes estas premissas e modelos.

¹⁰⁴ Gama é mestre em estatística pela ENCE e faz parte da equipe do hub regional.

4.7.2 A empresa Positium e os modelos de acesso aos dados guardados por privados

A empresa Positium participou de três webinários promovidos pelo hub. Nas três oportunidades verificamos que o debate sobre o acesso aos dados de telefonia foi abordado nas sessões de perguntas e respostas. A recorrência do tema é uma evidência de sua relevância para os INEs. Passamos a analisar estes debates a seguir, na medida em que nos fornecem maiores elementos sobre as premissas que norteiam os modelos de acesso aos dados, bem como sobre o posicionamento de agentes do públicos dos INEs, e em particular do IBGE acerca destas premissas e modelos.

O primeiro webinário sobre uso de dados de telefonia celular para estatísticas de Turismo foi mediado pela analista do IBGE, Maria do Carmo Bueno e contou com uma apresentação do *Statistics Indonesia* sobre um caso de uso de dados de telefonia celular realizado em parceria com a Positium. Em seguida, o representante da empresa, Siim Esko, apresentou outros projetos, produtos e modalidades de uso dos dados celulares para estatísticas oficiais desenvolvidas pela Positium. Ao final das apresentações, uma participante do webinário encaminhou uma pergunta para Esko em que questionava se o acesso aos dados de telefonia móvel pelos INEs envolvia algum tipo de pagamento ou incentivo para as empresas de telecomunicações.

Inicialmente, o representante da Positium respondeu que há uma expectativa de pagamento às operadoras de telefonia para obter acesso aos dados, uma vez que as operações de armazenamento e processamento implicam em custos para estas empresas. Esko também fez observações sobre as discussões em andamento na Europa para o acesso aos dados via acordos que envolvem o pagamento dos custos operacionais das empresas. Neste sentido, defendeu o desenvolvimento de projetos pelos INEs, que sejam ‘comercialmente viáveis’ para as operadoras. Finalmente, entretanto, apontou para a possibilidade do acesso compulsório aos dados pelos INEs, pela via da regulação.

Sim, há principalmente pagamentos pois é uma fonte de big data, então as operadoras de telefonia celular também tem custos para armazenar esses dados e também o processamento, na maioria dos casos, não em todos os casos, ocorre na própria operadora, então eles também tem o custo de ter um bom sistema de processamento onde esses dados podem ser processados. A agregação de bases de dados com um bom nível também é custosa, então há incentivos para pelo menos arcar com os custos mínimos das operadoras. No Eurostat há discussões em andamento sobre a adoção

de algum tipo de padronização pan-europeia para garantir o acesso aos dados de telefonia celular pelos institutos de estatística e arcar com o custo mínimo das operadoras, mas eu penso que o atual caminho a seguir é começar com um produto comercialmente viável para as operadoras de telefonia a não ser que você tenha uma base legal muito boa para trazer todos os dados para o instituto de estatística, algumas vezes isso é possível, [por exemplo] no caso do Omã os dados de telefonia são coletados na nuvem do governo, que é um ambiente seguro para armazenar estes dados sensíveis, nesse caso o incentivo às empresas é obedecer a legislação (Hub, 2023a).

Após a fala de Esko, a analista do IBGE, que fazia a mediação do evento, teceu comentários sobre as dificuldades de acesso enfrentadas pelos INEs. Maria Bueno, entretanto, foi interrompida pelo representante da Positium.

Essa questão é uma discussão em andamento em todo lugar, eu penso que o problema na utilização de dados de telefonia é exatamente o acesso aos dados porque as operadoras de telefonia celular, é claro que elas não querem dar os dados de graça, elas querem ser pagas ou querem alguma vantagem, mas algumas vezes não é possível...[interrupção por Esko] (Hub, 2023a).

Maria, algumas vezes pode não ser possível mas eu tenho que adicionar que há casos em que as operadoras de telefonia estão interessadas na monetização dos dados e na maioria dos casos os seus clientes não são instituição do governo, são instituições privadas mas os institutos de estatística podem ensinar as operadoras a analisar os dados de um modo em que verifiquem a qualidade, validade e precisão dos dados de um modo que eles não pensaram antes, então é precisamente sobre como devolver por meio destes projetos algumas habilidades, algumas capacidades para pensar sobre os dados de um modo estatístico e isso é algo que os institutos de estatística são excepcionalmente capazes de fazer (Hub, 2023a).

A resposta de Bueno à intervenção de Esko, por sua vez, revelou um tensionamento no que se refere à adoção de modelos de parcerias público-privadas no Brasil. Bueno, apontou para dificuldades na interação com as empresas privadas e se ressentiu da maior facilidade de acesso aos dados nos países que contam com operadoras de telefonia públicas. Finalmente, a analista do IBGE defendeu a necessidade de uma mudança na legislação brasileira para garantir o acesso dos dados para fins estatísticos.

Sim, eu entendo que há muitos modelos para interagir com as operadoras de telefonia celular mas é uma discussão difícil com eles, não é tão fácil obter os dados. No caso do Brasil, nós tentamos muitas vezes e isso não é tão fácil, aqui nós temos quatro operadoras principais e algumas são mais fáceis de conversar e outras simplesmente não querem conversar, eles não são receptivos. Em alguns países as operadoras são do governo, então é mais fácil [...] outro ponto é a mudança da legislação, no Brasil nós temos que mudar a legislação por que as operadoras de telefonia móvel dizem que não podem nos dar os dados por causa da privacidade, da confidencialidade, mas é óbvio que somos uma agência estatística e mesmo assim eles dizem que

há questões relacionadas à privacidade dos consumidores, então é uma questão legal que temos que resolver (Hub, 2023a).

Este primeiro debate nos permite compreender que as soluções tecnológicas para o uso de big data em estatísticas oficiais por empresas intermediárias como a Positium, podem operar de acordo com diferentes ‘modelos de negócios’ e premissas. Por um lado, há modelos de Parcerias Público Privadas que legitimam os cercamentos de dados, sua monetização e comercialização para os INEs. Por outro lado, há modelos de acesso público e regulado, orientados pelas premissas de que os dados para fins estatísticos são bens públicos aos quais deve ser garantido o acesso do Estado para a produção de estatísticas oficiais. Vimos que a representante do IBGE nesta ocasião fez a defesa do segundo modelo.

Em um segundo webinar sobre o uso de dados de telefonia celular para estatísticas populacionais, desta vez mediado pela coordenadora do Hub de Big Data no Brasil, Andrea Diniz, a questão do acesso aos dados voltou a figurar nas sessão de perguntas e respostas. Nesta oportunidade pudemos identificar mais algumas nuances dos modelos defendidos pela Positium, e o papel relevante que, de acordo com a empresa, os INEs poderiam desempenhar neste contexto.

Margus Tiru, representante da Positium, defendeu a transição do atual modelo de acesso fragmentado aos dados, em que uma série de agentes fazem demandas diretas às operadoras, para um outro modelo de acesso centralizado, que intitula de ‘Quadro Nacional de MPD’. De acordo com este segundo modelo, uma autoridade confiável exerceria um papel de centralização e coordenação dos dados. Para a Positium este papel poderia ser exercido pelos INEs.

Eu penso que chegou o momento para o que normalmente chamamos de quadro nacional de MPD [...] Os institutos de estatística são a fonte central de estatísticas para o governo, eles podem ter uma lei estatística que lhes permite obter acesso aos dados dos operadores de redes móveis. O processamento dos dados pode acontecer tanto do lado da operadora móvel ou pode acontecer centralmente no escritório de estatísticas usando todas as tecnologias de proteção ou privilégios administrativos que protegem os dados (Hub, 2023b).

É interessante observar que neste debate, enquanto o representante da Positium fez uma defesa do ‘modelo de regulação’, Andrea Diniz do IBGE, por outro lado, assumiu premissas do ‘modelo proprietário’. A fala de Diniz concebe os dados de telefonia como um tipo de ‘big data’ sob a propriedade das empresas, às quais

caberia a decisão de ‘oferecê-los’ ou não ao Estado para uso em finalidades estatísticas. Adotando argumentos típicos das operadoras, Diniz destacou, por exemplo, os produtos pagos e os ‘custos’ das empresas com os dados, além das questões de ‘privacidade’. Finalmente, a coordenadora do hub defendeu o modelo de processamento de dados dentro da estrutura das operadoras com utilização de tecnologias de ampliação de privacidade que vem sendo desenvolvido no âmbito do Projeto ONU Big Data.

Mesmo que as operadoras de telefonia móvel queiram oferecer os dados sem cobrar pelos seus produtos eles tem custos com os dados e em adição existem questões de privacidade, neste sentido um grupo de experts da ONU está trabalhando com tecnologias de ampliação de privacidade, em que trabalham com metodologias de proteção de ambos os lados, então vamos olhar para uns dez anos a frente no horizonte até que estejamos mais organizados globalmente e como as operadoras se organizarão para oferecer este tipo de big data (Hub, 2023b).

As diferenças de posicionamento entre as representantes do IBGE ficam evidentes em um terceiro webinar que teve como tema especificamente a questão do acesso aos dados de telefonia celular. O evento contou com a participação de Erki Saluveer da Positium, Federico Segui do INE do Uruguai e mediação de Maria do Carmo Bueno do IBGE. Saluveer detalhou diferentes modelos de acesso aos dados — em bases voluntárias, comerciais e reguladas — e apresentou diferentes esquemas de participação de empresas intermediárias como a Positium, seja na intermediação e processamento dos dados entre as operadoras e os INEs, ou na prestação direta de serviços para os INEs, quando estes realizam o processamento centralizado dos dados em sua estrutura. Segui, por sua vez, abordou a experiência do INE do Uruguai na produção de estatísticas de deslocamentos e isolamento social durante a pandemia de Covid-19, com uso de dados da operadora pública de telefonia celular Antel.

Após as apresentações, o debate teve início com uma provocação de Bueno do IBGE, sobre a questão da privacidade: “você acreditam que as questões de privacidade são apenas uma desculpa das operadoras para negar o acesso aos dados, ou são de fato um problema real?” (HubBrasil, 2024b. Transcrição nossa). Notadamente, Saluveer concordou que as questões de privacidade dos dados são frequentemente utilizadas como um subterfúgio das operadoras para dificultar o acesso aos dados, isto por que as fontes de big data são um problema relativamente

novo para os reguladores. Em resposta, Bueno fez um comentário sobre como a privacidade vem sendo utilizada pelas operadoras no Brasil para negar o acesso aos dados e apontou para o fato de que as empresas estão interessadas em vender serviços:

aqui as operadoras de telefonia celular sempre dizem que é uma questão de privacidade, é duro ouvir isso porque somos um instituto de estatística e diariamente lidamos com questões de confidencialidade e temos regras para lidar com isso [...] as empresas querem vender serviços, mas isso é muito complicado para nós, instituto de estatística, porque não sabemos quais métodos eles estão utilizando, porque precisamos de uma metodologia transparente e não aceitar qualquer coisa de um terceiro (Hub, 2023c).

A seguir o representante do Uruguai relatou a experiência do INE com as operadoras privadas. Para Segui, a alegação de confidencialidade é uma desculpa das empresas, por que efetivamente identificou nas negociações que há um real interesse das operadoras em utilizar os dados para vender informações. Ainda de acordo com Segui, para estas empresas “nós estamos tentando destruir o seu negócio”:

Nós solicitamos dados das três operadoras do Uruguai, foi fácil conseguir os dados da Antel porque é uma empresa pública e nacional e são parte do Sistema Estatístico Nacional mas com as outras empresas privadas [Claro e Movistar] [...] eles queriam compartilhar informações em um nível muito agregado não no nível detalhado e desagregado que queríamos. Depois disso fomos à autoridade de telecomunicações, a reguladora, e firmamos um memorando de entendimentos para ter acesso aos dados das três empresas, e eles estão tentando estabelecer uma regulação para forçá-los a compartilhar os dados. Estamos neste processo. Durante essas negociações, e por isso eu concordo com Saluveer, eu também acho que a confidencialidade é uma desculpa, eles disseram que nós estamos tentando destruir o seu negócio porque eles querem vender os dados, eles querem produzir as informações, vender o movimento da população e seus dashboards, eles querem vender, então se produzirmos esses indicadores e dashboards de graça eles pensam que estamos destruindo o seu potencial negócio (Hub, 2023c).

Para concluirmos, vimos como a empresa Positium identificou um nicho de mercado no setor de estatísticas oficiais e vem utilizando sua participação privilegiada no projeto da ONU big data para fins comerciais. Verificamos que os modelos de negócios desenvolvidos pela empresa podem tanto se ajustar à lógica dos modelos proprietários, atendendo aos interesses comerciais das empresas, quanto à lógica do interesse público, mediada pela regulação nacional do acesso aos dados para fins estatísticos, conferindo maior autoridade e autonomia aos INEs. Neste sentido, o

posicionamento ativo dos INEs latino-americanos neste debate e sua capacidade institucional e política são elementos relevantes para a adoção de modelos mais adequados à soberania nacional dos dados para fins estatísticos.

Finalmente, o acesso aos dados guardados por privados revela-se como uma questão emergente no campo estatístico latino-americano e traz a tona duas diferentes abordagens entre os agentes. Por um lado, a via da regulação do acesso aos dados e da ampliação da autoridade dos INEs como repositórios públicos, seguros e centralizados de dados para fins estatísticos. Por outro lado, a via das parcerias público-privadas pautadas por modelos de negócios economicamente viáveis para as empresas, que asseguram a manutenção dos dados em seus repositórios privados. Na próxima sessão analisamos o caso do Projeto OPAL, desenvolvido por outras duas empresas parceiras da Plataforma Global, a Data Pop Alliance e a Flowminder, que exploram o modelo de PPP com uso de algoritmos de ampliação de privacidade.

4.7.4 Flowminder e Data Pop Alliance: o modelo de parcerias público-privadas e as tecnologias de ampliação de privacidade

Flowminder e Data Pop Alliance (DPA) são duas ‘organizações sem fins lucrativos’ parceiras da GPSDD e Plataforma Global que tem promovido o uso de big data em estatísticas oficiais por meio do modelo de parcerias público-privadas. As duas organizações tem promovido o uso da tecnologia OPAL (open algorithms). O OPAL consiste em uma solução tecnológica de compartilhamento de dados, com uso de tecnologias de ampliação de privacidade e computação distribuída. O projeto piloto do OPAL envolveu uma parceria entre a DPA e o Departamento Administrativo Nacional de Estatísticas da Colômbia (DANE).

Em sua apresentação no webinar sobre uso de dados celulares do Hub regional, Emanuel Letouzé, coordenador da DPA, apresentou o OPAL como um exemplo de ‘inovação sócio-política para utilizar novos dados e melhorar as estatísticas oficiais, que reflete e serve a valores humanitários’ (HubBrasil, 2023c). Ainda de acordo com a GPSDD, o OPAL é uma “inovação sócio tecnológica sem fins lucrativos e um novo paradigma para a utilização de dados privados para o bem social” (GPSDD, 2018:1). Em contraste com o encapsulamento humanitário da iniciativa, em um relatório da DPA para o governo da Colômbia, o OPAL é apresentado como um

exemplo de ação “essencial para impulsionar o mercado de dados no país” (DPA, 2019:22).

Em termos técnicos, o OPAL consiste em uma tecnologia de ampliação de privacidade (PET) e método de compartilhamento de dados por computação distribuída em que algoritmos certificados operam sobre dados brutos que permanecem nos servidores do proprietário da base de dados, de modo a gerar estatísticas agregadas para usuários selecionados. No modelo OPAL (Hardjono e Pentland, 2019), o INE é concebido como um usuário com acesso a determinadas estatísticas agregadas, por algoritmos desenvolvidos por um intermediário e implantados no servidor de uma empresa privada. Enquanto consumidor de dados agregados, o grau de autonomia do INE na produção estatística é reduzido. Sem acesso à base de dados bruta mas apenas aos dados agregados, existem limitações ao processo estatístico. Tais limitações, entretanto, podem ser minoradas, caso o INE esteja envolvido na construção do algoritmo.

A apresentação de Letouzé no webinar evidencia como a questão da privacidade passou a ser instrumentalizada para legitimar os cercamentos privados de dados e para construção de um modelo de negócios que reduz substancialmente a autonomia dos INEs na produção estatística:

OPAL é um mecanismo de perguntas e respostas, o INE faz perguntas aos dados do setor privado e então ele vê apenas os indicadores — podem ser estimativa de pobreza, mobilidade, uma matriz — apenas isso é compartilhado, de forma a preservar a privacidade. O dado bruto nunca sai dos servidores da empresa privada, mas é anonimizado e agregado e você ainda tem um sistema de governança. Então estamos testando diferentes tipos de OPAL para ação humanitária e monitoramento dos ODS e nos nossos planos estamos tentando expandir não apenas para dados de operadoras de telefonia celular, mas para dados de observação da terra, transações financeiras e microdados dos INEs que são muito ricos e sensíveis em uma série de países como o Brasil (Hub, 2023d).

Em sua estratégia de Big Data para a América Latina, a DPA defende expressamente que os INEs da região devem “envolver-se com o setor privado, avaliar os modelos atuais de compartilhamento de dados corporativos e estabelecer acordos para parcerias público-privadas” (DPA, 2016:39). Para a DPA, os governos da América Latina têm potencial para desempenhar um papel de liderança nos ecossistemas nacionais de dados, 'tornando-se, por um lado, um facilitador e, por outro, um consumidor de big data' (DPA, 2021:3). As recomendações da DPA ao governo Colombiano apoiaram a remoção de barreiras administrativas e legais para

viabilizar mercados de dados no país: 'Em termos de investimento público, as melhores práticas para incentivar a cadeia de valor do big data centram-se na desregulamentação do ecossistema' (DPA, 2019:22).

Na Colômbia a implementação do Projeto OPAL, consistiu em uma PPP entre o DANE e a Telefónica Colômbia para identificar casos de uso de dados de *Call Detail Record* (CDR). Para a implementação do projeto piloto, o OPAL recebeu financiamento da GPSDD e do Banco Mundial (GPSDD, 2018). Posteriormente, o OPAL angariou 1,5 milhões de euros da Agência Francesa de Desenvolvimento (AFD, 2018). A partir de 2020, o projeto entrou em fase de teste de mercado em parceria com a Flowminder e atualmente concentra-se “nos países de baixo e médio rendimento, começando pelo Haiti em colaboração com a Digicel” (OPAL, 2022:1).

Em sua apresentação no webinar, Emanuel Letouzé destacou o fato de que a maioria dos projetos em big data para estatísticas oficiais na América Latina, a exemplo do OPAL, ainda são limitados a projetos piloto e provas de conceito. Para Letouzé o principal desafio enfrentado pelos projetos na região para ampliar a sua escala de atuação, é a necessidade de uma ‘mudança cultural’:

Este processo tem que ser parte de uma mudança cultural, é algo muito sensível, todo mundo sabe disso, e envolve não apenas os INE e as empresas privadas, mas as universidades e organizações da sociedade civil também tem que ser parte desse mix, tem que ter muita sensibilização, discussões, então é realmente uma mudança cultural mais do que qualquer coisa (Hub, 2023d).

A necessidade de uma ‘mudança cultural’ foi igualmente apontada como o principal desafio para ampliação do uso de big data pelos governos nacionais da América Latina em um estudo realizado pela DPA (2021, p.39). O estudo destacou, por exemplo, a resistência dos servidores públicos e a dificuldade em estabelecer alianças público-privadas com os governos nacionais na região, em face de ‘equivocos sobre o significado e o potencial do Big Data’:

Para dar um salto qualitativo rumo à adoção de tecnologias disruptivas como o Big Data [...] é essencial uma mudança na percepção cultural em torno dos dados e da sua utilização. Por um lado, como evidenciado nas entrevistas deste estudo exploratório, há uma baixa demanda por dados digitais resultante da preferência dos servidores públicos por dados analógicos. A crença ainda arraigada de que estes dados são mais confiáveis, constitui uma barreira à promoção da utilização de informação digitalizada como uma fonte valiosa na criação de estatísticas oficiais ou políticas públicas [...] É também inegável que, embora exista uma vontade de digitalização da informação no setor público, em alguns destes países existe uma desconfiança

relativamente ao compartilhamento de dados governamentais que dificulta a compreensão do valor que o Big Data pode trazer [...] Neste sentido, os equívocos sobre o significado e o potencial do Big Data são um grande desafio para a consolidação de um potencial ecossistema de Big Data nestes países (DPA, 2021, p.38-39).

Em resumo, as ‘organizações sem fins lucrativos’ Data Pop Alliance e Flowminder, parceiras da Plataforma Global, tem atuado na América Latina na proposição de um novo modelo de negócios em que atuam como intermediárias de mercados de dados para estatísticas oficiais. Neste caso, um modelo de compartilhamento de dados por computação distribuída com uso de tecnologias de ampliação de privacidade, que assegura a manutenção dos dados em seus repositórios privados e viabiliza a a consulta de informações agregadas pelos INEs. O modelo atende a interesses de grandes detentores de bases de dados mas em contrapartida reduz a autonomia estatística dos INEs. É interessante ainda observar como as empresas ‘filantrópicas’ instrumentalizam conceitos como “bem social”, “valor humanitário”, etc., que asseguram o acesso a parceiros e recursos para o desenvolvimento de um produto de natureza comercial.

Finalmente, são evidenciadas resistências ao big data no setor público da América Latina e dificuldades de escalar os projetos em nível nacional. Enquanto as empresas traduzem estas resistências em termos ‘culturais’, nós às interpretamos como indícios de contramovimentos protetivos dos INEs e governos nacionais à incidência de agentes do mercado em seus campos estatísticos. Na próxima seção, passamos a analisar em maior profundidade às percepções do corpo técnico do IBGE sobre a questão do acesso aos dados na medida em que analisamos experiências em curso na instituição com uso de big data.

4.8 Os casos de uso de big data no IBGE e o debate do acesso aos dados.

Identificamos pelo menos três experiências em andamento no IBGE com novas fontes de dados: *web scraping* para estatísticas de preços, uso experimental de dados de telefonia móvel e *web crawling* para estatísticas de comércio eletrônico. Nesta sessão apresentamos estas experiências à luz da avaliação dos agentes envolvidos e de suas percepções acerca dos potenciais e limitações destes projetos. Evidenciamos especialmente os seus posicionamentos sobre o debate do acesso aos dados para fins estatísticos.

4.8.1 O uso de Web scraping para estatísticas de preço e a dificuldade de acesso aos dados fiscais.

No IBGE, a iniciativa mais bem-sucedida com uso de novas fontes de dados, que efetivamente evoluiu da fase experimental para a produção de estatísticas oficiais, foi a coleta automatizada de dados na internet (*web scraping*) para o índice de preços de passagens aéreas. De acordo com o estatístico Épsilon¹⁰⁵, o projeto foi desenvolvido de forma autônoma pelo IBGE com envolvimento da Coordenação de Índices de Preço (COINP), Coordenação de Métodos e Qualidade (COMEQ) e Diretoria de Informática (DI):

O projeto iniciou no final de dois mil e dezessete em uma parceria com a COMEQ, eles desenvolveram um piloto pra coleta de passagem aérea de algumas companhias e a gente começou a acompanhar pra ver se os resultados eram coerentes com o resultado da coleta manual. Isso ficou em análise durante algum tempo e foi implantado em produção em dois mil e vinte. Então hoje a coleta de passagem aérea é feita de forma automática usando o webscraping implantado em produção [...] Na fase de estudos a gente também analisou fatores de estabilidade do site e dos robôs, se os sites teriam alguma mudança ou algum bloqueio, esse é um ponto importante, em geral do estudo até você colocar algo em produção é um processo delicado. Você precisa avaliar bem antes de fazer isso [...] Então hoje esses robôs estão implementados na TI da gerência da DI que dá suporte ao SNPC e eles fazem todo o processo de acompanhamento (Épsilon, entrevista em 20 de maio de 2022).

As iniciativas que utilizam dados da web no IBGE tendem a avançar na medida em que há maior facilidade de acesso aos dados na internet. A utilização de robôs na coleta de preços, entretanto, pode colocar o INE em situações delicadas, caso configure uma espécie de hackeamento dos sites coletados. De acordo com Épsilon, a coleta na internet pelo IBGE tenta seguir protocolos e etiquetas, embora ainda não exista uma política de coleta via web no órgão. O estatístico também mencionou o uso de procedimentos técnicos de ‘calibração dos robôs’ para evitar que a coleta seja bloqueada pelos sites:

Idealmente o que a gente tenta fazer é entrar em contato com as empresas, pelo menos no caso das companhias aéreas a gente fez isso, pra conseguir uma permissão de acesso, mas hoje a gente não tem no IBGE uma política de acesso web [...] Como órgão de estatística oficial isso é mais delicado e a

¹⁰⁵ Épsilon é estatístico com nível de doutorado. É analista sênior do IBGE com mais de dez anos de atuação. Épsilon não faz parte da equipe do hub, mas acompanha um dos grupos de trabalho da ONU Big Data.

gente quer sempre o consentimento do informante e a gente corre um risco também de ter algum bloqueio, então se botar em produção e se não tiver essa negociação isso pode ser bloqueado [...] A gente já teve bloqueio principalmente na fase de testes, em que estávamos ajustando ainda a maneira como o robô ia atuar (Épsilon, entrevista em 20 de maio de 2022).

De acordo com o estatístico Delta, apesar de ‘mais democrática’, a coleta de dados na internet traz desafios associados aos termos de uso dos sites e legislações nacionais que precisam ser respeitados pelos INEs: “A internet é uma fonte de big data democrática, as pessoas podem ir lá e acessar, mas é democrática com algumas limitações. A gente não pode pegar aquilo cujo o termo de uso diz que não pode pegar. E não pode fazer com aquilo o que o termo de uso diz que não pode fazer.” (Delta, entrevista em 30 de junho de 2022).

Ainda na área de estatísticas de preço, o IBGE vem desenvolvendo um projeto piloto para coleta automatizada na internet de preços de hospedagem em plataformas de reserva on-line. De acordo o estatístico Zeta¹⁰⁶, “especificamente pra esse projeto, substituir os dados não vai acontecer, se for pra usar os dados vai ser um uso pra apoiar a coleta, mas provavelmente substituir não vai acontecer porque existem riscos” (Zeta, entrevista em 26 de Maio de 2022). Os riscos pontuados por Zeta dão uma melhor dimensão de que apesar de apresentarem menor dificuldade de acesso, os dados na internet podem ser muito voláteis:

Nós começamos com três sites, agora a gente está pesquisando um só, porque um não valia a pena e o outro colocou um captcha, um verificador e a gente furar esse captha [...] já começa a ficar um pouco mais obscuro do que só fazer a coleta. O IBGE fazer alguma coisa nesse sentido pode ferir a credibilidade. É um passo a mais do que fazer coleta, e aí a gente preferiu não usar mais esse site. Então, por exemplo, é um risco nesse sentido. Uma outra coisa que a gente também viveu são as mudanças no layout do site. Eles não te avisam. Até porque você está ali meio de, não é intruso, mas um pouco intruso. Um dos sites, por exemplo, teve uma grande mudança de layout então o algoritmo simplesmente parou de funcionar [...] isso é um risco também pra coleta. (Zeta, entrevista em 26 de Maio de 2022)

Zeta apontou que uma das plataformas de reserva on-line selecionadas para o projeto possui uma API que possibilitaria um acesso mais estável e seguro, porém pago aos dados. A opção de uso da API, entretanto, foi descartada em razão dos custos. Perguntamos, se ele tinha conhecimento da discussão sobre parcerias público-privadas no campo estatístico, e se no caso do projeto de preços de

¹⁰⁶ Zeta é estatístico com nível de mestrado e analista júnior no IBGE com menos de 10 anos de atuação.

hospedagem havia sido cogitada alguma forma de parceria com as plataformas de reserva on-line. Sua resposta indicou preocupações de que as PPPs possam ‘ferir de alguma forma a missão do IBGE’:

O que eu fico muito em dúvida em relação a essas parcerias é a contrapartida para a empresa. Porque qual é a contrapartida que o IBGE vai ter que dar pra empresa se não o dinheiro? Porque também dinheiro é difícil, pra comprar de fato, tá comprando um serviço, aí não é uma parceria, é uma compra, mas se você faz uma parceria, eu entendo que ambas as partes precisam se beneficiar de alguma forma. E aí eu fico em dúvida de quais são os produtos que o IBGE poderia produzir dentro dessas parcerias, ou quais informações o IBGE poderia fornecer dentro dessas parcerias que não feririam de alguma forma a missão do IBGE, que não feririam a confidencialidade dos informantes, eu fico em dúvida mesmo, eu não sei qual é a resposta. (Zeta, entrevista em 26 de Maio de 2022)

De acordo com Beta, a discussão de parcerias público-privadas e de acesso aos dados guardados por privados ainda está distante do horizonte brasileiro. Isto porque o IBGE ainda não consegue sequer acesso a determinados dados governamentais do sistema estatístico nacional, como os dados fiscais.

No Brasil eu ainda vejo isso muito distante. Sinceramente, a gente ainda não consegue fazer essa discussão. A gente ainda não consegue fazer essa discussão com o dado público. O IBGE é o coordenador do Sistema Estatístico Nacional e ele não tem acesso aos dados do sistema estatístico nacional! Então a gente tem dado do governo federal, de instituições públicas que o coordenador do sistema estatístico nacional não pode ter acesso, como a receita, por exemplo (Beta, entrevista em 15 de julho de 2022).

De acordo com Épsilon, as notas fiscais são uma fonte muito mais robusta para a produção de índices de preço do que os dados coletados na web. Conforme vimos, entre as iniciativas do IBGE interrompidas pelo Governo Bolsonaro constava justamente um projeto de lei específico para facilitar o acesso aos dados fiscais para a produção de estatísticas oficiais. No caso da Plataforma Global, vimos também o envolvimento de empresas como a Nielsen no grupo de estatísticas de preço e *scanner data* da ONU Big Data. Épsilon destacou o grande potencial dos dados fiscais no caso brasileiro, na medida em que constituem registros administrativos ‘já em poder do Estado’, sob a tutela da Receita Federal.

Na área de índice de preço, um outro tipo de dado que é mais poderoso do que a web que os países do hemisfério norte tem avançado e tem estudado mais é o scanner data, que é, enfim, o dado que é similar ao que a gente tem na nossa nota fiscal, associado às informações de venda do produto efetivamente, no código de barras, então você tem informação da descrição, do momento da venda, da quantidade vendida. Em geral esse tipo de acesso

envolve uma negociação direta com as empresas, então os países que foram bem sucedidos nisso eles fazem esse tipo de negociação e alguns deles tem uma legislação estatística que garante o acesso a esse tipo de informação. No nosso caso, a gente tem uma base de dados muito rica já em poder do Estado, da receita federal. Então a gente tem esse grande potencial nas mãos do Estado, mas tem empecilhos legais, por parte dos órgãos fiscais que alegam questões de sigilo fiscal pra disponibilizar certas informações. Então a gente precisaria avançar nesse sentido (Épsilon entrevista em 20 de maio de 2022).

Em resumo, o caso de uso de web scraping para estatísticas de preços, revela que há potencial na coleta automatizada de dados na internet e que experiências viáveis vem sendo desenvolvidas de forma autônoma pelos quadros do IBGE. Em alguns casos específicos o novo método pode vir a substituir a coleta tradicional e em outros casos pode ser utilizado como um apoio à coleta tradicional.

Verificamos ainda que apesar da internet ser potencialmente uma fonte de dados mais aberta, há questões técnicas, éticas e legais que podem restringir o acesso dos INEs. Também pudemos verificar entre alguns agentes entrevistados que os modelos de parcerias público-privadas são vistos com uma certa desconfiança, na medida em que podem comprometer a missão institucional do IBGE. Por outro lado, há uma demanda de fortalecimento legal e institucional do IBGE no âmbito do sistema estatístico nacional, para ter acesso a fontes de dados fiscais bem mais robustas que a internet no caso das estatísticas de preço. Finalmente, este caso é especialmente relevante por que ilustra uma iniciativa interna e autônoma de inovação com o envolvimento de técnicos de diferentes áreas do IBGE.

4.8.2 O uso experimental de dados de telefonia celular e o debate do acesso aos dados guardados por privados.

O IBGE também vem realizando testes de uso de dados de telefonia celular, neste caso, entretanto, a iniciativa não se originou de uma demanda interna mas veio 'do exterior'. De acordo com Ômega, que atua no grupo de trabalho de dados de telefonia móvel da ONU big data, há muito interesse no IBGE em utilizar os dados de telefonia, principalmente pela possibilidade de ter dados atualizados com maior frequência para captar fenômenos não cobertos pelas fontes tradicionais e com possibilidade de redução de custos de coleta. Ômega apontou ainda exemplos do potencial de uso dos dados de telefonia para a produção de estatísticas de deslocamento e turismo: "Vieram algumas pessoas do exterior que tinham algumas

experiências oferecendo as possibilidades pra ver se o IBGE se interessava. É claro que pra gente é muito interessante. Tem muita coisa que a gente não consegue em pesquisas tradicionais, que consegue derivar de dados de telefonia móvel” (Ômega, entrevista em 25 de maio de 2022).

A experiência do IBGE com uso de dados de telefonia celular consistiu em um projeto experimental para medir o acesso à internet usando dados de uma operadora local de telefonia móvel em parceria com a empresa Positium (ITU, 2021). De acordo com Ômega, foram utilizados dados de apenas uma operadora para a cidade do Rio de Janeiro, e os resultados foram animadores: “comparamos os resultados de acesso à internet com os dados da PNAD e verificamos uma aderência maravilhosa”. Apesar de demonstrar potencial, a experiência não foi além de testes. Isto decorre da dificuldade de acesso aos dados das operadoras de telefonia:

A gente não consegue os dados. O presidente do IBGE [Eduardo Rios Neto] está super interessado, mas o que eu falei pra ele é que não adianta a gente ficar fazendo teste de novo [...] a gente tem que conseguir esses dados de uma maneira consistente, sentar com essas empresas, sentar com o governo eu acho que é uma coisa bem política, não é o presidente do IBGE que tem que falar, eu acho que tem que ser o ministro, tem que ser alguma coisa acima do IBGE, eu acho que outros ministérios podem se aproveitar desses dados [...] não tem que negociar só pro IBGE (Ômega, entrevista em 25 de maio de 2022).

Ômega descreveu ainda a dificuldade de negociação com as operadoras privadas. Para o analista, a alegada questão de privacidade dos dados é um subterfúgio das empresas para não conceder o acesso ao IBGE em face de seu interesse na venda de serviços. O analista também se posicionou contrário à compra de serviços das operadoras, o que segundo Ômega caracterizaria uma terceirização da produção estatística:

O grande problema é o acesso. Desde o ano dois mil isso vem sendo estudado na Europa e em outros países e até hoje não entrou na rotina de produção dos institutos de estatística. Por quê? Acesso aos dados, as empresas de telefonia são muito herméticas, elas não te dão acesso a esses dados, você não consegue entrar numa negociação com elas. Aqui no Brasil eu tentei várias vezes negociar, colocando Anatel na conversa e o que eles alegam? Ah esses dados são sigilosos e aí depois da LGPD que não dão mesmo. Claro que a gente sabe que os dados são sigilosos, é claro que a gente vai ter cuidado com isso. Não tem sentido querer alegar sigilo pra mandar o dado pro IBGE. Então é uma questão de mercado, é um mercado muito fechado. Elas não querem dar o dado. Elas querem vender o serviço. Isso pros institutos de estatística não interessa. Porque pra estatística pública é complicado, assim a gente estaria terceirizando. Eu nunca vi nenhum instituto de estatística terceirizar sua estatística, é uma coisa bem estranha.

Elas não têm que prestar serviço pra gente, elas têm um contrato, elas são concessionárias de um serviço público (Ômega, entrevista em 25 de maio de 2022).

De acordo com Beta, a compra de dados de telefonia na escala em que seriam necessários para colocar uma estatística oficial ‘em produção’ é inviável para os INEs. Não apenas pelos custos, mas pelo comprometimento da autonomia do órgão que ficaria mais vulnerável à instabilidades tanto no lado do governo quanto das empresas:

O privado essencialmente fala assim 'olha eu cedo o meu dado mas você paga por ele' e isso é inviável pra qualquer instituto nacional de estatística. Não é só pro IBGE, comprar o dado no volume que você precisaria, comprar um dado de telefonia móvel, você consegue comprar pra fazer um estudo, mas não pra colocar isso em produção. Inclusive porque muda o governo, diminui o orçamento do Instituto Nacional de Estatística, ou muda a administração da empresa e a empresa resolve que aquele valor negociado não cabe mais, vai dobrar o valor e aí você interrompe a tua série histórica, como é que fica? vai descontinuar um processo de produção estatística oficial. A gente não pode considerar essa hipótese, isso é um risco que não dá pra assumir. Então eu acho que é inviável não só porque o custo é muito alto, como ele é inviável porque traz uma instabilidade, uma fragilidade muito grande para o processo da estatística oficial (Beta, entrevista em 15 de julho de 2022).

Para Ômega, a solução para o acesso aos dados passa pela atualização da legislação. O analista apontou para mudanças na Europa e destacou a defasagem do marco legal estatístico no Brasil: “Vários países na Europa estão mudando sua legislação exatamente por causa dessas fontes novas. Principalmente com relação à internet e telefonia móvel e por causa da LGPD também. Então vários países estão modernizando as suas leis estatísticas. A legislação do Brasil já era defasada, com isso fica mais ainda” (Ômega, entrevista em 25 de maio de 2022).

Na mesma direção, o estatístico Delta também fez uma defesa enfática da necessidade de atualização da legislação estatística no Brasil para viabilizar o acesso aos dados guardados por privados:

A quantidade de dados privados que está na classe dos big data é enorme. Então um desafio é esse, é ter acesso aos dados. Sem acesso a gente não faz nada [...] Então a gente precisa ter uma legislação que faça o que a lei tradicional do IBGE já faz em relação a pesquisa de campo. A lei diz que todos os residentes tem que prestar informação ao IBGE, pra realização das pesquisas. A gente precisa de uma legislação que diga que todos os proprietários de dado de qualquer natureza tem a obrigação de fornecê-los ao IBGE, sob demanda, e na mesma lei estabelecer que o IBGE está comprometido com a confidencialidade e é responsável por manter o sigilo da informação (Delta, entrevista em 30 de junho de 2022).

Já para Beta, a discussão de acesso aos dados é mais complexa e não se limita à regulamentação. Entusiasta do compartilhamento de dados com uso de tecnologias de ampliação de privacidade e computação distribuída, para o estatístico este modelo de acesso descentralizado — em que o IBGE não receberia o dado bruto mas apenas faria o processamento no servidor do ‘proprietário’ do dado ou em um servidor ‘seguro’ de um terceiro — é muito mais ‘factível’ do que ‘essa discussão de acesso ao dado’.

Isso é uma discussão super ampla, não é trivial, não é assim: 'dá acesso pro IBGE', é uma discussão muito mais complexa. Eu acho que a gente no Brasil está muito longe de usar dado privado, mas uma coisa que eu acho interessante é a alternativa de você não ter acesso ao dado, não pegar na tua máquina o dado, o estatístico do IBGE não olhar o dado mas ter acesso para executar algum tipo de processamento. Isso me parece muito mais factível, ainda não no momento porque a gente não está com essa tecnologia consolidada mas eu acho isso muito mais próximo do que essa discussão de acesso aos dados que vem sendo travada com muita dificuldade já há alguns anos e eu não vejo isso sendo resolvido num futuro muito breve (Beta, entrevista em 15 de julho de 2022).

Conforme vimos, os modelos de acesso descentralizado aos dados com uso de tecnologias de ampliação de privacidade vêm sendo estimulados pela ONU Big Data. Identificamos a participação de empresas intermediárias na construção de modelos de negócios ‘economicamente viáveis’ para os ‘proprietários’ dos dados e como estes modelos podem potencialmente vir a constituir um novo mercado para grandes empresas de tecnologia como a Microsoft. Beta, entretanto, destacou que há um grande potencial para a utilização desta modalidade de compartilhamento e processamento de dados entre órgãos públicos. O IBGE já utiliza técnicas de ampliação de privacidade e segurança no armazenamento dos microdados estatísticos e na divulgação de informações para a garantia da anonimização dos dados. Neste sentido, haveria capacidade técnica no órgão para o desenvolvimento autônomo e seguro de métodos semelhantes de compartilhamento de dados por computação distribuída utilizando tecnologias de anonimização. Esta modalidade poderia ser utilizada, por exemplo, nos casos em que existam restrições de segurança, a exemplo dos dados fiscais.

Perguntado sobre o modelo de parcerias público-privadas, Ômega nos informou que na área de telefonia essa discussão ainda está muito inicial, e que teve início justamente porque os INEs em todo o mundo não conseguiram ter acesso aos

dados. O analista expressou dúvidas sobre o que os INEs poderiam oferecer em troca para as empresas e descartou a compra dos dados:

O que está assim no menu do dia é isso, o acesso aos dados, trazendo as empresas pra conversar com os institutos de estatística, pra sentar juntos e discutir. E partindo dessa ideia de parcerias, como trocar ganhos. A gente sabe o que a gente quer. Eles teriam que dar, mas assim, o que que a gente pode oferecer pra eles em troca? Isso está muito inicial (Ômega, entrevista em 25 de maio de 2022).

Ainda no tema das parcerias público-privadas, Ômega citou um caso de PPP na África com a intermediação da empresa Flowminder, que é parceira da Plataforma Global da ONU. É interessante observar como o analista diz achar a experiência africana 'interessante' do ponto de vista do acesso aos dados ao governo, mas como ao mesmo tempo descarta, no caso do IBGE, a intermediação de um parceiro privado como a Flowminder:

Na África, tem um caso com a intermediação da Flowminder, que é um instituto filantrópico, digamos assim. O Flowminder é financiado pela Fundação Bill e Melinda Gates. Ele fez a mediação entre o governo e a Orange. Os dados iam pra uso de todo o governo e a Flowminder desenvolveu um kit pra acesso aos dados. Eu achei uma coisa bem interessante essa de fazer um acordo, não que precise ter uma instituição mediando e não que precise ser uma coisa filantrópica, não necessariamente isso, mas uma coisa bem interessante é essa de fazer um acordo pros dados ficarem disponíveis pra todo o governo (Maria do Carmo Bueno, entrevista em 25 de maio de 2022).

Em resumo, este caso revela que para os agentes do IBGE, o uso de dados de telefonia móvel poderia contribuir para complementar a produção de estatísticas oficiais em áreas não cobertas pelas pesquisas tradicionais. Entretanto, o órgão não tem conseguido acesso aos dados guardados pelas operadoras de telefonia. Conforme já havíamos observado nos webinários que debateram o tema do acesso, há nuances importantes no debate sobre o acesso a dados guardados por privados no IBGE. Por um lado, há agentes que verbalizam a defesa de um modelo centralizado de acesso aos dados por meio da atualização da legislação estatística. Por outro lado, há agentes que defendem um modelo descentralizado de compartilhamento de dados, mediado por tecnologias de ampliação de privacidade e computação distribuída.

Verificamos ainda que há resistências às parcerias público-privadas, ao pagamento pelos dados e à participação de intermediários. Finalmente, o caso é relevante porque apresenta uma experiência que foi motivada por agentes externos e

que apesar dos resultados limitados do ponto de vista da produção de estatísticas oficiais, contribuiu para um maior conhecimento sobre os potenciais e limitações no uso destes dados pelo IBGE.

4.8.3 O uso de web crawling para estatísticas de comércio eletrônico e o debate da compra de dados de intermediários

Um terceiro caso de uso de fontes de big data no IBGE envolve a utilização de web crawling para estatísticas de comércio eletrônico. A ação é um desdobramento da participação do IBGE no projeto ‘Big Data para medir e fomentar a economia digital’, desenvolvido pela CEPAL entre 2016 e 2021, que envolveu além do Brasil os INEs do México (INEGI), Chile (INE) e Colômbia (DANE). Entre outras ações, o projeto tentou replicar a metodologia de medição de comércio eletrônico adotada pelo Statistics Netherlands (CBS).

No caso dos Países Baixos, o projeto teve início por meio de uma PPP patrocinada pela Google. De acordo com o CBS: “A Google abordou o Statistics Netherlands para realizar um estudo para aprofundar a compreensão da economia da Internet utilizando uma abordagem inovadora” (CBS, 2016, p.3). O projeto consistiu em combinar dados do registro de empresas do CBS com dados sobre a presença de negócios na internet, recolhidos pela empresa Dataprovider (CBS, 2016). O projeto piloto evoluiu para uma medição contínua de comércio eletrônico, com a aquisição regular de dados da Dataprovider pelo CBS (CBS, 2020). Em 2019, foi assinado um acordo de colaboração e a Dataprovider recebeu um ‘selo de qualidade’ do CBS. Segundo um dos diretores do CBS, “os atores do mercado podem expandir e enriquecer a gama dos seus produtos e serviços trabalhando conosco. Nossa rede contribui para isso” (CBS, 2019:1). Ainda de acordo com o CBS (2019, p.1), a certificação e rede de contatos do CBS possibilitou à Dataprovider expandir suas operações para Coreia do Sul (CBS, 2019:1). De acordo com um dos diretores da empresa:

O selo de qualidade concedido pelo CBS significa muito para nós [...] fomos convidados pelas Nações Unidas para fazer uma palestra em uma assembleia junto com o CBS, diante de mais de 100 diretores de estatística de todo o mundo [...] Isso se enquadra bem nos planos futuros da nossa empresa: continuar crescendo e colaborando com novos atores para expandir nossa gama de dados (CBS, 2019, p.1, tradução nossa).

Verificamos ainda que a estratégia de aquisição de dados de empresas pelo CBS foi avaliada em um relatório encomendado pelo Ministério do Interior holandês. O relatório destacou a existência de alternativas legais para o INE acessar os dados:

É importante perceber que o CBS tem alternativas. Pode comprar dados, como da Coostousto, onde obtém dados de mídia social ou do Dataprovider, que coleta dados sobre sites mas também é possível chegar a acordos sobre entregas de dados regulares ou incidentais em consulta com detentores de dados terceirizados. Além disso, está aberta a via da regulamentação, desde que a aquisição dos dados se enquadre naturalmente na missão estatutária. Outra linha que pode levar a obrigações de provisão de terceiros passa por Bruxelas, onde o compartilhamento de dados B2G [Empresas para Governos] está sendo considerado no Data Act, um novo regulamento legal no contexto da estratégia de digitalização da UE que está sendo preparado (Geonovum 2021, p.16, tradução nossa).

Passando para a América Latina, de acordo com Beta, o objetivo do projeto da CEPAL visava o monitoramento e produção de indicadores do comércio eletrônico nos países envolvidos. O estatístico destacou as dificuldades de captação da atividade econômica na internet pelas pesquisas tradicionais: “A ideia era tentar de alguma forma monitorar ou criar indicadores pra comércio eletrônico. Porque na estatística oficial tradicional isso simplesmente passa batido. O quanto a empresa está vendendo através da internet [...] na verdade a gente não sabe que parcela é essa” (Beta, entrevista em 15 de julho de 2022).

De acordo com o estatístico Sigma¹⁰⁷, o projeto seguiu a mesma metodologia utilizada pelo CBS. Inicialmente, foi realizada uma caracterização da economia da internet e uma categorização das empresas de acordo com o nível de presença na Web. Em um segundo momento, foi realizado um cruzamento dos cadastros de empresas dos INES com a base de dados da empresa Dataprovider.

A CEPAL pagou por essa base de dados da empresa de webscraping. No caso, Brasil, Chile, México e Colômbia fizeram esse exercício. Dentro da CEPAL no Chile, os consultores faziam o trabalho com a base de dados e a gente entregava as palavras-chaves classificadas por categoria, é como se a gente tivesse quase traduzido o que a Holanda fez [...] o uso dessas palavras-chaves, é uma técnica de machine learning para categorização, isso também já veio pronto, o nosso trabalho foi fazer o link com o cadastro (Sigma, entrevista em 30 de maio de 2022).

¹⁰⁷ Sigma é um estatístico com nível de doutorado e analista sênior no IBGE com mais de 20 anos de atuação na instituição.

Conforme explicado por Beta, a metodologia utilizada pela Dataprovider difere do webscraping, que visa a extração de dados específicos de páginas da web. Neste caso, o método utilizado foi o webcrawling, uma varredura mais abrangente que gera uma indexação do conteúdo de uma grande quantidade de sites, processo utilizado por mecanismos de buscas como o da Google. Ainda de acordo com Beta, os INEs latino-americanos encontraram muitas dificuldades em relacionar as bases de dados e a aderência entre os dados da Dataprovider e os cadastros de empresas se revelou muito baixa. No caso do Brasil, foi possível identificar apenas 9,2% das empresas na base de dados do Dataprovider (IBGE, 2020). Segundo as entrevistas, nos demais países os percentuais foram ainda menores. Para Beta o projeto serviu para conhecer melhor as dificuldades deste tipo de levantamento e análise de dados:

Em princípio [a base da Dataprovider] deveria conter só empresas brasileiras. Essa teoria a gente não verificou porque na verdade é muito difícil definir isso. Se é uma empresa brasileira, é muito complicado. O Brasil foi até mais ou menos bem sucedido, talvez por ser um país maior, com uma infraestrutura de TI maior, mas México e Colômbia tiveram muita dificuldade com o dado que foi retornado pela Dataprovider [...] Uma coisa que ficou claro pra gente nesse projeto é que essa correspondência entre a empresa e website tem muitas fragilidades. No desenvolvimento do escopo desse projeto sequer alguém tinha antecipado essa realidade. Então, uma coisa que esse estudo serviu foi pra trazer um pouco mais essa realidade (Estatístico Beta, entrevista em 15 de julho de 2022).

Conforme pudemos verificar em um relatório do CBS, também foram identificadas limitações nos resultados alcançados nos Países Baixos. Uma avaliação realizada por estatísticos do CBS sobre os indicadores gerados entre 2016 e 2018, revelou problemas na consistência dos dados:

Os problemas mais importantes são decorrentes da combinação da variabilidade no big data (dados da Internet) e do método utilizado para delinear a economia da Internet e as suas categorias. As análises dos nossos resultados relativamente à consistência a nível micro mostram que existe uma instabilidade substancial na forma como as empresas são categorizadas ao longo do tempo. Isto significa que é necessário ter cuidado ao interpretar os resultados, especialmente as tendências ao longo do tempo. (CBS, 2020:5).

Sigma apontou igualmente para as dificuldades enfrentadas no projeto e para a 'ilusão' em torno do uso generalizado de big data. Para o estatístico, há casos de usos pontuais bem sucedidos, mas em áreas mais complexas, apesar das

experiências gerarem ‘indicativos’ ainda são insuficientes para a efetiva produção de estatísticas oficiais.

Utilizar técnicas de big data e pensar que o dado está aí pronto é pura ilusão. Pura ilusão. Eu acho que em alguns casos dá pra fazer realmente, como o pessoal dos preços fez, mas quando você vai pra uma coisa um pouco mais profunda, é difícil, não é fácil, você não sabe o que vai encontrar. Sempre tem algum problema, mas de alguma forma eu acho que é um indicativo. Agora você não pode dizer que aquilo ali é estatística oficial, ainda está muito longe da gente chegar em alguma coisa desse tipo. Não tem mágica, porque vem muita sujeira, muita coisa que a gente tem que limpar, que tem que estudar (Sigma, entrevista em 30 de maio de 2022).

De acordo com Beta, o aprendizado gerado pelo projeto resultou em uma nova iniciativa que vem sendo desenvolvida de forma autônoma pelo IBGE. É interessante observar como Beta destaca que nesse caso o IBGE está fazendo a coleta ‘por conta própria’, sem intermediação de empresas como a Dataprovider. Outro ponto interessante é a adaptação da metodologia holandesa ao contexto e necessidades do IBGE, o que se reflete na definição de uma amostra específica e na construção de um *scrap* que identifique especificamente o CNPJ das empresas. A iniciativa, entretanto, ainda está em curso e já revelou novas dificuldades:

Existe por conta desse desenvolvimento um projeto hoje no IBGE pra tratar esse tema, mas a gente está coletando as informações por conta própria a gente não está usando Dataprovider nem nada do tipo. A gente está tentando fazer uma coisa com um foco diferente porque o objetivo essencial é tentar produzir um indicador [...] O escopo fechado desse projeto no momento é partir de uma amostra de empresas da PAC [Pesquisa Anual do Comércio], em torno de cinco mil empresas. Primeiro porque a gente já tem uma experiência com comércio e segundo porque você espera que o comércio quer se fazer presente na internet hoje em dia. Um terceiro ponto é que pelo Código de Defesa do Consumidor todo mundo que atua no comércio tem que ter no rodapé da página principal do website o CNPJ. Então a nossa expectativa era entrar lá, olhar o rodapé, que é um *scrap* super trivial de fazer, pegar o CNPJ e relacionar ao cadastro de empresa do IBGE. Só que apenas dez por cento dos websites de comércio tem o CNPJ.

Em resumo, vimos como um modelo de parceria público-privada na Holanda com apoio da Google foi replicado na América Latina com intermediação da CEPAL. O caso oferece mais elementos de como Big Techs e OIs tem apoiado e divulgado modelos de negócio que fomentam a criação de mercado de dados para estatísticas oficiais. O caso também apresentou um outro exemplo de como o capital simbólico da ‘confiança’ foi mobilizado para a parceria entre o CBS e a Dataprovider por meio de um ‘selo de qualidade’. As dificuldades no pareamento das bases de dados da

empresa Dataprovider, se por um lado inviabilizaram o objetivo inicial de produção de indicadores de comércio eletrônico, por outro, ampliaram os conhecimentos dos INEs da América Latina sobre as limitações e dificuldades no uso de big data. Por fim, o caso evidenciou contramovimentos na Holanda, na forma de recomendações ao CBS sobre alternativas legais para o acesso aos dados, quanto no Brasil, onde a replicação da experiência com base nos aprendizados do projeto passou a ser feita de forma autônoma sem intermediação de empresas privadas.

4.9 Percepções sobre a introdução de big data e ciência de dados no IBGE.

O uso de novas fontes de dados para produção de estatísticas oficiais no IBGE, tem sido acompanhado pela introdução dos termos ‘ciência de dados’ e ‘cientista de dados’. Vimos anteriormente no capítulo 3 que a ciência de dados se apresenta como um novo campo epistêmico híbrido entre a estatística e a ciência da computação e que o cientista de dados corresponde a um novo perfil profissional associado a um novo conjunto de competências e habilidades relacionadas ao processamento e análise de big data.

Conforme vimos, o Plano Estratégico do IBGE (2017-2027) previa a criação de um comitê de Big Data e ciência de dados, que chegou a ser constituído na gestão de Roberto Olinto e posteriormente extinto no Governo Bolsonaro. Vimos ainda que a ex-presidente Susana Guerra promoveu acordos com o BID e Banco Mundial que previam programas de formação em ciência de dados no órgão. Além disso, o Hub regional no Brasil tem entre seus objetivos e linhas de ação a implantação de projetos, pesquisas e capacitações em ciência de dados.

Verificamos ainda a criação de um grupo de trabalho em ciência de dados por iniciativa interna de técnicos da Diretoria de Pesquisas em 2019. As atividades do grupo de ciência de dados da DPE, entretanto, tiveram curta duração e foram interrompidas no contexto da pandemia de COVID-19. De acordo com Beta, o grupo estaria atualmente se rearticulando sob a sua coordenação.

De acordo com o estatístico Delta, que está ativamente envolvido em processos de capacitação na área de ciência de dados no hub regional, o perfil profissional associado ao cientista de dados se diferenciaria do estatístico tradicional, na medida em que passa a englobar maior autonomia computacional e domínio de novos métodos.

O que muda é a autonomia computacional dos profissionais. Hoje não basta ser um bom analista que sabe olhar uma tabela e escrever um bom texto. Hoje precisa de um analista que sabe acessar uma nuvem, fazer processamento de alto desempenho, mexer com algoritmos, usar linguagens que deem conta de processar big data. Não basta um bom teste de hipótese, não basta saber bem métodos estatísticos tradicionais, considerando que tem um punhado de métodos que ainda não foram desenvolvidos ou consolidados e eles vão precisar aprender também. Então eu diria que autonomia computacional pra começar e pra continuar precisa adquirir habilidades no uso desses métodos que estão por vir. Então é isso que muda no perfil profissional do meu ponto de vista (Estatística Delta, entrevista virtual em 30 de Junho de 2022).

Em nossas entrevistas realizadas com os agentes do IBGE, abordamos o tema da ciência de dados. Especificamente para os estatísticos entrevistados, perguntamos “você se considera um estatístico ou um cientistas de dados”. Analisamos a seguir as percepções e posicionamentos dos agentes sobre este novo campo de conhecimentos e práticas.

4.9.1 Estatísticos de Estado versus Cientistas de Dados.

Beta é um jovem estatístico com mestrado em ciência da computação que tem se destacado entre seus pares no IBGE como uma referência na área de ciência de dados, fato que foi apontado por outros agentes entrevistados. Conforme vimos, Beta tem participado ativamente do projeto da ONU Big Data e está liderando a rearticulação do grupo de ciência de dados no IBGE. Notadamente, ao ser confrontado com nossa pergunta (“você se considera um estatístico ou um cientistas de dados”), Beta se identificou como um estatístico e fez críticas ao termo ‘cientista de dados’.

Na verdade não, a minha essência continua na estatística. Tem muita coisa da estatística, principalmente dentro da heurística e meta-heurística, em que a estatística pode ajudar a computação e essa via também volta, a computação tem um poder enorme. Na minha opinião não se fala em estatística hoje em dia sem associar à computação. Então não tem como desconectar as duas coisas mas eu não me considero um cientista de dados, para ser muito honesto eu não me sinto muito à vontade com essa terminologia, eu não sei exatamente o que ela quer dizer, mas o que eu faço tem um foco na estatística e eu uso ciência da computação pra fazer isso, se a gente combina que isso é o ‘cientista de dados’, ok (Estatístico Beta, entrevista virtual em 15 de Julho de 2022).

É interessante observar a seguir como a justificação de Beta à sua crítica aos cientistas de dados, revela um ajustamento estratégico de sua conduta às disposições sancionadas pelo campo estatístico oficial e em particular pelos demais colegas no IBGE.

Essencialmente porque eu trabalho no IBGE, e realmente, por essa questão da estatística oficial ter a necessidade de preservar métodos consagrados, de você ter ali esse compromisso, que é uma coisa ruim para conseguir avançar a modernização, existe uma resistência velada que se apoia nisso, mas esse argumento é muito necessário. Você não pode: 'olha o método novo aqui vamos usar' e de repente daqui a seis meses descobrir que aquele método inseriu um viés no índice de preços e agora como é que fica? Eu acho que muito por estar inserido nesse contexto, eu fico um pouco receoso de assimilar essa nomenclatura não tão estabelecida. Eu me considero um estatístico que usa a ciência da computação pra fazer estatística. Mas eu acho que no mercado, fora da área de estatística oficial, esse rótulo é dominante, assim quem faz algo similar ao que eu faço é cientista de dados, eu diria que sim (Estatístico Beta, entrevista virtual em 17 de Maio de 2022).

Vemos que o pertencimento de Beta ao campo estatístico requer que ele assuma uma determinada posição, um 'compromisso com os métodos consagrados', mesmo que considere tal compromisso 'uma coisa ruim', na medida em que implica em uma 'resistência velada' à 'modernização', que ele propriamente defende. Notadamente, embora valorize suas habilidades híbridas (estatísticas e computacionais) que lhe conferem capitais técnicos relevantes, Beta estabelece uma distinção entre o cientista de dados como um profissional do mercado e o estatístico de Estado como um agente público investido de um outro conjunto de valores o que, por outro lado, lhe garante a aquisição de capitais simbólicos que são sancionados e reconhecidos pelo campo estatístico.

Em uma direção semelhante, Pedro Nascimento, reconhecido nacionalmente e internacionalmente como um dos mais prestigiados estatísticos brasileiros, também se posicionou criticamente ao uso da nova nomenclatura. Nascimento diz se considerar um 'estatístico que sempre fez ciência de dados', na medida em que desde o início de sua carreira no IBGE manejou grandes quantidades de dados e fez uso de tecnologia computacional aplicada à estatística¹⁰⁸.

Esse tema da ciência de dados, emergiu como um novo nome, uma questão nova, recente, mas eu digo que sou formado em estatística e sempre fiz

¹⁰⁸ É digno de nota que após a entrevista Nascimento me encaminhou um texto do estatístico David Donoho (2017), professor da Universidade de Stanford, com teor crítico à nova onda da ciência de dados nas universidades norte-americanas.

ciência de dados. Se você parar pra pensar como eu trabalhei ao longo da minha vida, eu sempre fui um estatístico que teve que pôr as mãos na massa no sentido de processar os dados que eu precisava para fazer o trabalho, e no IBGE, em particular, eu trabalhei com pesquisas de grande porte [...] enfrentando todas as questões que os cidadãos que se chamam cientistas de dados hoje enfrentam, ter que processar grandes volumes, ter que extrair significado e automatizar processos e tudo isso foi feito ao longo da minha trajetória, só que se chamava estatística, não tinha esse ângulo, essa visão nova que surgiu mais recentemente. (Pedro Nascimento, entrevista em 21 de Junho de 2022).

Nascimento, entretanto, reconhece que o processo contemporâneo de dataficação de fato configura uma importante mudança nas práticas estatísticas tradicionais: “Eu trabalhei a minha vida inteira no IBGE usando fontes primárias e próprias. A principal mudança atual são as novas fontes que hoje passaram a ser consideradas potencialmente viáveis pra produzir informação pra estatística pública” (Pedro Nascimento, entrevista em 21 de Junho de 2022).

Nós temos ainda uma utilização um tanto tímida dessas fontes novas. É difícil apontar uma razão, é mais uma questão estrutural em organizações como o IBGE, no mundo inteiro, os órgãos de estatísticas são organizações avessas a risco, não é que eles não corram risco mas eles se protegem bastante em relação a situações de risco. E algumas dessas fontes que são exploradas assim de forma intensa no mercado, no mundo privado, pra produção de estatística pública envolvem riscos e dificuldades importantes. No mundo privado há uma avidez por explorar oportunidades que possam resultar em lucro para aquela empresa, na estatística pública é muito delicado você publicar uma estatística por exemplo que fale da população só dos que tem acesso à internet, isso é delicado (Pedro Nascimento, entrevista virtual em 21 de Junho de 2022).

Como vemos, o graduado estatístico estabelece uma distinção entre o uso de fontes de big data pelos agentes do setor privado para a exploração de novas oportunidades de mercado, e a sua incorporação cautelosa pelos estatísticos de Estado para a produção de estatísticas oficiais. Enquanto os primeiros são abertos à experimentação e orientados para a obtenção de lucros, os estatísticos de Estado são estruturalmente avessos a riscos e orientados por uma lógica pública, que se manifesta, por exemplo, na necessidade de garantir a representatividade do conjunto da população.

Beta e Nascimento são estatísticos que representam gerações distintas no IBGE, as similaridade em seus discursos, são, portanto, um dado relevante. Neste sentido, as críticas às novas terminologias (ciência de dados e cientistas de dados) indicam uma identidade e um sentido de pertencimento ao campo estatístico, os quais

estão associados a um conjunto de valores (compromisso com métodos consagrados, aversão a riscos, serviço público, representatividade da população) que são reconhecidos como distintos dos valores do mercado (inovação e experimentação orientadas para a obtenção de lucros), onde a nova terminologia se originou e é amplamente reconhecida. Finalmente, a defesa da ‘estatística’ frente à ‘ciência de dados’ é reveladora da incorporação de disposições sancionadas, valorizadas e reconhecidas entre os estatísticos de Estado e, particularmente entre os estatísticos do IBGE.

4.9.2 O habitus ibgeano e o obstáculo epistemológico.

De acordo com Nelson Senra, características que remontam à fundação do IBGE, reforçadas por uma estrutura organizacional autoritária e hierárquica, contribuíram historicamente para a internalização de determinadas disposições em seus quadros, as quais intitula de ‘tradição renovada’ e ‘mística ibgeana’.

O IBGE, a gente conhece bem é uma instituição presidencialista. As decisões são de cima pra baixo. Não é que a gente não possa ponderar [...] mas é uma instituição de cima pra baixo. No entanto, Teixeira de Freitas nos impregnou desse gentílico. Não sei se a nova geração realmente ainda tem. Mas nós que nos impregnamos de IBGE nós chegamos a ter esse gentílico, nós somos ibgeanos. Teixeira de Freitas tinha o lema para o IBGE de tradição e renovação. Esse era o lema que ele praticava na casa. Teixeira de Freitas quando lançou o primeiro número do Anuário Estatístico sob a sigla IBGE, ele o lançou como número dois mas era o primeiro. Por que isso? Porque ele disse que o primeiro anuário do Brasil havia sido lançado por Bulhões Carvalho na primeira república e na diretoria geral de estatística. Ele só estava dando continuidade. Ele respeitava a tradição do seu grande mestre Bulhões Carvalho. Isso é tradição. É uma tradição renovada por quê? Porque nós temos condições de renovar mas temos a renovação com respeito à tradição. Pois bem, o corpo de funcionários do IBGE pelo menos aqueles da nossa geração e talvez ainda de uma geração seguinte, tem esse espírito gentílico de Ibgeano e Ibgeana, a mística Ibgeana. E é tão impregnado na casa a ideia de tradição e renovação, como dizia Teixeira de Freitas que um presidente quando chega, por mais que a casa seja presidencialista e ele queira fazer e acontecer, ele se limita. Ele se sente tolhido, as pessoas são capazes de lhe fazer perceber que não dá pra embarcar em aventuras (Nelson Senra, entrevista em 16 de Maio de 2022).

Conforme vimos em nossa análise história, o delicado processo de construção da legitimidade dos INEs e da confiança pública nas estatísticas oficiais, apoiado na autoridade do Estado e da ciência, contribuiu para moldar o habitus do estatístico de Estado. Vimos ainda, como a introdução de determinadas inovações no Séc. XX,

como a amostragem probabilística, envolveu resistências e lutas internas à sua incorporação nos INEs. Podemos, portanto, compreender o que Senra intitula de ‘tradição renovada’ como um elemento desta disposição relativamente conservadora, ou seja, uma resistência à disrupção que não é, entretanto, completamente fechada à inovação, mas que submete constantemente a inovação à prova dos valores e técnicas tradicionais. Neste sentido, a ‘impregnação’ pelo ‘gentílico’ e ‘mística ibgeana’, podem ser compreendidos como a efetiva incorporação de um ‘habitus ibgeano’ que corresponde ao estatístico de Estado e que se traduz em um conjunto de valores internamente aceitos, sancionados e reforçados na instituição, os quais se refletem nos discursos e estratégias dos agentes. De acordo com Senra, esta disposição tem funcionado historicamente no IBGE como um eficiente contrapeso do corpo técnico, para resistir a movimentos disruptivos de agentes externos, inclusive aos presidentes da casa.

Pudemos verificar em nossa pesquisa de campo, que este ‘habitus ibgeano’, ou esta disposição relativamente conservadora dos estatísticos de Estado, efetivamente se manifesta nos discursos dos quadros veteranos e mais novos da instituição, especialmente no que diz respeito à incorporação de novas fontes de dados e métodos. Diferentemente do enquadramento disruptivo e revolucionário que tem caracterizado o discurso da ‘revolução de dados’, observamos, em contraste, um tratamento bastante sóbrio e cauteloso do tema entre os agentes do IBGE.

De acordo com Nascimento, por exemplo, não é possível se falar em substituição das fontes primárias neste momento. O estatístico ilustra a sua posição apontando para a importância dos inquéritos do IBGE na mensuração de fenômenos não captados pelos dados administrativos ou por fontes de big data, como, por exemplo, o mercado de trabalho informal:

Não dá pra imaginar jogar tudo que tem fora e dizer que vai colocar uma coisa nova no lugar. Não tem a coisa nova pronta pra colocar no lugar. Talvez daqui a algum tempo a construção da coisa nova permita pegar aquela coisa antiga e deixar pra trás. Eu vou citar uma área por exemplo interessante em que o Brasil tem bons registros administrativos mas que são claramente insatisfatórios para representar a nossa realidade. O setor do trabalho. O Brasil tem bons registros administrativos na área do trabalho [RAIS e CAGED] mas são fontes que só te permitem enxergar parcialmente o que acontece no mundo do trabalho formal. E o mundo do trabalho informal que no Brasil é gigantesco, não é possível enxergá-lo através dos registros administrativos [...] O IBGE tem fornecido nessa área ao longo das últimas décadas resultados muito bons com as suas pesquisas domiciliares [...] Então a minha visão do que a gente deve ter como processo evolutivo é a incorporação de outras fontes, ampliação dos temas, eventualmente a ampliação de

frequência de produção de resultados que pode ser possível com a apropriação de alguns registros e fontes novas, mas substituição do que está aí, tão cedo não temos perspectivas (Pedro Nascimento, entrevista em 21 de Junho de 2022).

Em direção semelhante, o estatístico Gama apontou que inicialmente a ideia do grupo da ONU Big Data era de ‘substituir tudo’ e que atualmente o projeto está ‘evoluindo’ em direção ao uso complementar das novas fontes, juntamente com as fontes tradicionais, o que julga ser uma posição mais acertada.

Eu acompanho o grupo de big data da ONU desde dois mil e quinze, e realmente começou com a ideia de substituir tudo. Não é assim, e eu acho que está evoluindo pra coisa certa, pra tratar o big data como mais uma fonte de dados e eu acho que tanto o grupo quanto o Hub trabalham também nessa direção [...] Essas fontes vão ser incorporadas, o que a gente precisa fazer é estudar qual a melhor forma de incorporar essas fontes dando os resultados com a qualidade que todos os institutos oficiais procuram na hora de produzir estatísticas.

Quando discutimos a introdução das fontes de big data no campo estatístico europeu (capítulo 2), vimos que os problemas de representatividade destas fontes impuseram obstáculos aos métodos tradicionais da estatística frequentista e da teoria da amostragem. No caso europeu, estas limitações passaram a ser contornadas com uso de modelos bayesianos e métodos de aprendizado de máquina. Pudemos verificar com alguns estatísticos do IBGE que problemas semelhantes de representatividade das fontes de big data também passam a ser identificados no Brasil.

De acordo com o estatístico Delta, o uso de novos métodos para a incorporação de big data nas estatísticas oficiais está em discussão e ainda não há uma proposta consolidada:

A gente sabe fazer amostra em pesquisa, amostra baseada em cadastro que a gente sabe o tamanho e conhece as unidades. Mas a gente não sabe fazer amostra em big data. Estimativa é uma área dentro da amostragem. Se você não sabe fazer uma amostra você não sabe estimar também. Então o que sobra pra gente? Estimar com Big Data é simplesmente falar o que a gente encontrou. Isso não é uma coisa que se sustente, porque Instituto de Estatística não pode fazer estimativa sem saber falar da qualidade da estimativa, sem saber falar de uma margem de erro que é a coisa mais básica da estimativa. Então ter métodos pra usar big data é fundamental e a gente ainda não tem. A gente está discutindo mas ainda não tem, pelo menos não tem consolidado. (Estatística Delta, entrevista virtual em 30 de Junho de 2022)

O depoimento de Delta é uma evidência de que o obstáculo epistemológico que tem conduzido à introdução de novos métodos, como os modelos bayesianos e de aprendizado de máquina, nos INEs europeus passa a emergir igualmente no IBGE, na medida em que os estatísticos são instados a lidar com fontes de big data. Questão semelhante foi apontada por Gama, para quem as melhores estratégias para superar os problemas de representatividade das fontes de big data passam pela agregação de outras fontes de dados estruturados com populações conhecidas disponíveis nos INEs, como surveys e registros administrativos.

A produção de estatísticas que sejam representativas depende de você identificar quem você está representando de forma correta e isso a gente ainda está estudando. A gente tem feito algumas tentativas de trabalhar com coleta via webscraping na internet, tem dados interessantes mas um problema no caso de usar big data ou dados que não são gerados pra produção de estatística oficial e pública é que é difícil às vezes delimitar qual a população que é representada por aquelas informações. Então normalmente o que a gente tenta fazer é misturar essa base com uma outra base que a gente conhece a população que representa. Todos os estudos que a gente tem feito são nesse sentido de juntar as duas coisas, não é substituir uma coisa pela outra mas agregar mais informação e mais qualidade e melhorar a informação que você já tem (Estatístico Gama, entrevista virtual em 09 de maio de 2022).

Gama teceu críticas à utilização de métodos e modelos preditivos e complexos na produção de estatísticas oficiais. De acordo com o estatístico: “eu já utilizei vários métodos que hoje eles chamam de ciência de dados¹⁰⁹. Esses métodos mais novos [redes neurais e florestas] eu ainda não usei, eu já tentei mas eles nunca forneceram um resultado que fosse melhor o suficiente que justificasse algo tão complexo se com algo mais simples eu chego num resultado praticamente igual”. Para Gama, modelos preditivos complexos não são adequados para as estatísticas oficiais. O estatístico destacou a importância da explicabilidade dos modelos e da compreensão dos vieses que podem ser introduzidas pelas fontes de big data.

O problema é o seguinte. A maioria dos algoritmos, seguem dois tipos de modelagem. Uma cujo objetivo é prever e outra cujo o objetivo é entender [...] Quando você faz uma regressão, uma logística, com poucas variáveis, você consegue explicar que variáveis daquele modelo, explicam a probabilidade. Quando você faz uma rede neural, você não tem essa lógica. Ela é não linear. Então você não consegue efetivamente explicar o que está acontecendo. Se o seu objetivo é prever alguma coisa, se você prevê bem não tem problema. Se seu objetivo é explicar você não pode usar esses modelos não paramétricos que você não sabe em que cada variável está contribuindo e

¹⁰⁹ Entre os métodos citados por Gama constam: “árvore de regressão e classificação, análise de cluster, análise fatorial, principais componentes, regressão logística, modelos lineares, mistos, simples”

como está contribuindo. Esse é o problema hoje em dia dos modelos ou dos algoritmos que são muito complexos e cujo o objetivo é prever... [por exemplo] pro dono do banco interessa ter uma previsão boa, pra quem empresta e pra quem não empresta, mas pra quem está pedindo empréstimo, interessa saber porque ele está ou não está recebendo, então são duas coisas que não conversam [...] e tem outra coisa, se existe um viés na base de dados que está sendo usada pra treinar, porque o algoritmo vai olhar as decisões passadas pra tirar a decisão dele. Ele não vai decidir nada. O algoritmo não decide nada. Ele só repete o processo decisório do passado [...] mas aí você tem que entender que vieses são esses. E aí esse é um dos maiores problemas com big data. A gente sabe que tem viés e às vezes a gente sabe mais ou menos qual é, e às vezes a gente não faz a menor ideia de qual é esse viés. (Estatístico Gama, entrevista virtual em 09 de maio de 2022)

Conforme podemos observar, os depoimentos de Nascimento, Delta e Gama, estatísticos experientes, demonstraram um comportamento cauteloso na abordagem à inovação e incorporação das novas fontes de dados e métodos. Verificamos posicionamentos semelhantes junto à nova geração de estatísticos do IBGE, que estão diretamente envolvidos nos casos de uso de big data no órgão, que discutimos anteriormente.

Épsilon, que participou do projeto exitoso na utilização de web scraping para estatísticas de preços, demonstrou uma abordagem muito sóbria e cautelosa sobre o uso de big data em outras áreas do IBGE, destacando a importância do planejamento e da gestão de riscos.

Qualquer processo de mudança, de transição tem que ser bem pensado e estruturado. Então em geral você precisa ter fases que vão desde o desenho, do que você pretende fazer, dos estudos, uma fase mais de implementação até você conseguir efetivamente chegar na fase de produção, pra justamente minimizar os riscos que você possa ter, algum problema metodológico que você não avaliou bem ou de um acesso que você conseguiu durante um tempo e ele vai ser restrito depois, então isso tudo tem que ser bem planejado. Se não for, os riscos aumentam. Isso é uma fonte de dado que você vai pegar de uma forma massiva e que depende de um terceiro [...] isso tudo tem que ser bem avaliado, [na coleta tradicional] o dado ainda é do informante, só que você tem que ir lá e pegar, é mais pulverizado e a quantidade é menor. Quando você vai em uma escala muito mais massiva o impacto é maior, uma coisa é você perder um informante que te dava vinte informações, outra coisa é o informante que dava um milhão de informações. Então isso tudo tem que ser bem planejado (Estatístico Épsilon, entrevista em 20 de Maio de 2022).

Em direção similar, o jovem estatístico Zeta, que participa do projeto de web scraping para preços de hospedagem, apesar de se auto definir como parte de um grupo de pessoas dispostas à inovação no IBGE, defende que as mudanças ocorram devagar com processos robustos para garantir a manutenção da qualidade.

A gente vai avançar devagar. Não vai ser nada de um dia pro outro ou super revolucionário. Eu acho que as transformações vão acontecer, eu percebo que as pessoas estão dispostas, não sei se estou com um olhar enviesado porque já estou nesse meio de pessoas dispostas, mas a percepção é que boa parte entende como uma oportunidade, mas acho que a gente vai evoluir devagar, até pela experiência, porque a gente não tem tanta experiência nessa nova realidade, nessas novas ferramentas. Então também não dá pra sair mudando tudo sem ter processos que são robustos o suficiente pra manter a qualidade (Estatístico Zeta, entrevista virtual em 26 de Maio de 2022).

Em resumo, o conjunto de evidências aponta para posicionamentos cautelosos e relativamente conservadores dos estatísticos do IBGE, que não são refratários à inovação, mas defendem a incorporação gradual, planejada, controlada e complementar das novas fontes e métodos nas práticas estatísticas, juntamente com a preservação das fontes e métodos tradicionais. Tudo leva a crer que esta é uma disposição coletivamente sancionada no âmbito institucional e adequada às estratégias dos agentes. Se, por um lado, o acesso às novas fontes de dados e o domínio das inovações tecnológicas e metodológicas da ciência de dados, são importantes capitais em disputa pelos agentes do campo, ao mesmo tempo, a preservação da disposição relativamente conservadora do estatístico de Estado é valorizada institucionalmente e garante o acesso a capitais simbólicos sancionados pelo campo.

Finalmente, esta disposição relativamente conservadora dos estatísticos de Estado, que identificamos junto aos estatísticos do IBGE, representa uma importante proteção dos sistemas estatísticos nacionais contra investidas disruptivas de governos e do mercado. Neste sentido, constitui-se em um alvo das iniciativas modernizadoras e pró-mercado que defendem uma mudança cultural no campo estatístico, entre outros, pela via da gestação de uma nova geração de cientistas de dados mais ‘empreendedores’ e favoráveis à inovação, e mais permeáveis à incidência dos agentes de mercado.

Conclusão: A disputa pelo Sistema Estatístico Nacional na transição para o regime de dataficação

Neste capítulo, vimos como em um contexto de fragilização institucional do IBGE durante o governo de extrema direita de Jair Bolsonaro deu-se a implantação

do Hub Regional de Big Data no Brasil. Neste período, verifica-se, por um lado, a desarticulação de iniciativas que visavam o fortalecimento institucional do IBGE para a coordenação do sistema estatístico nacional, e, por outro lado, a promoção de ações de ‘modernização’ voltadas à introdução de big data e ciência de dados no órgão, orientadas pela cooperação internacional.

Avaliamos que os movimentos de ‘modernização’ orientados por acordos internacionais visaram promover uma inserção subordinada do IBGE no regime de dataficação, ao mesmo tempo em que tentaram cooptar o órgão para a promoção da pauta de parcerias público-privadas no campo estatístico latino-americano. Notadamente, verificamos como apesar do IBGE ser o principal financiador do hub regional, na falta de uma orientação estratégica nacional e em meio a uma conjuntura de desestabilização institucional, as ações do projeto assumiram uma orientação autônoma, pautada pelo Projeto ONU Big Data. Isto envolveu, por exemplo, a abertura de espaços para a incidência comercial de empresas privadas junto aos INEs latino-americanos, a participação de técnicos do IBGE em grupos de trabalho com especialistas de Big Techs e outras empresas privadas, e a discussão de modelos de acesso aos dados por meio de parcerias público-privadas.

Desdobrando controvérsias na implantação e atividades do hub regional e em casos de uso de fontes de big data no IBGE, verificamos que há diferentes entendimentos e posicionamentos entre os agentes do IBGE sobre o projeto da Plataforma Global da ONU, sobre a incidência de empresas privadas no campo estatístico e sobre os modelos de acesso aos dados. Entre os quadros mais antigos e egressos da instituição, identificamos um posicionamento mais ativo quanto à defesa da autonomia e soberania estatística nacional e mais crítico quanto à inserção das empresas privadas. Entre os quadros mais novos, verificamos em alguns casos uma postura de maior subordinação na relação com os agentes internacionais e de normalização das relações público-privadas no campo estatístico.

No que se refere aos modelos de acesso aos dados, identificamos uma divisão de posicionamentos. Alguns quadros são explicitamente críticos às parcerias público-privadas e defendem um modelo pautado na regulamentação do acesso compulsório aos dados pelo IBGE, enquanto outros quadros são mais abertos às PPPs. Do ponto de vista das modalidades de acesso, os posicionamentos também se dividem em termos da defesa de um modelo de acesso que assegure a centralização dos dados brutos no IBGE, e um segundo modelo descentralizado de acesso aos dados com uso

de computação distribuída e tecnologias de ampliação de privacidade, que preservariam a ‘propriedade’ das bases de dados pelos seus detentores originais.

Finalmente, as entrevistas com os estatísticos do IBGE revelaram uma identidade e sentido de pertencimento ao campo estatístico. Isto é evidenciado na incorporação de disposições características dos estatísticos de Estado — pautadas pelo compromisso com métodos consagrados, cuidados com a qualidade e representatividade dos dados, aversão a riscos e ética do serviço público. Estas disposições se manifestam em um conservadorismo relativo na introdução de novas fontes de dados, métodos e tecnologias e em uma resistência crítica na adoção do novo jargão profissional de ‘cientista de dados’. Vimos que os estatísticos do IBGE se consideram distintos destes profissionais, aos quais associam um perfil de mercado em que a inovação e a experimentação são voltadas à obtenção de lucros.

Esta resistência dos estatísticos ibgeanos às práticas disruptivas da ciência de dados, entretanto, é apenas relativa, na medida em que as novas fontes, métodos e tecnologias constituem capitais valiosos que são disputados pelos agentes e passam a ser objeto de novos projetos em curso na instituição. Por outro lado, a abordagem sóbria e cautelosa dos estatísticos acerca das inovações é coerente com disposições sancionadas pelo campo estatístico, e em particular pelo IBGE. Tais disposições, contribuíram historicamente para conformação de um habitus ibgeano, avesso à disrupção e favorável a uma conciliação processual entre tradição e renovação.

Tudo parece funcionar como se a característica pública, estatal e nacional dos sistemas estatísticos da América Latina e a predominância do habitus do estatístico de Estado nos INEs da região constituíssem uma importante barreira à disseminação de projetos que visam a ampliação da participação do setor privado no campo estatístico, a substituição de fontes e métodos tradicionais e a criação de mercados de dados para estatísticas oficiais. Neste sentido, os argumentos pela necessidade de uma ‘mudança cultural’ e os investimentos na criação de espaços público-privados de articulação e na formação de uma nova geração de cientistas de dados constituem estratégias que visam ajustar as disposições dos estatísticos de Estado, de modo a torná-los mais susceptíveis e permeáveis à associação com os agentes de mercado. Trata-se, portanto, de um estratégia que visa facilitar no médio e longo prazo a incorporação de um novo habitus algorítmico e empreendedor no campo estatístico oficial, que reproduza disposições características das organizações privadas da área de tecnologia e dados.

Conforme vimos, entretanto, tanto a concepção ontológica dos dados (bens públicos e mercadorias), quanto a epistemologia dos modelos (explicativos e preditivos) estão em disputa. Isto se reflete em diferentes estratégias dos agentes, caracterizando um duplo movimento. Por um lado a abertura às parcerias público-privadas e a internalização do perfil profissional dos cientistas de dados, e, por outro, a proteção dos sistemas estatísticos nacionais pela via da regulação e da preservação dos métodos e fontes tradicionais. Neste sentido, o posicionamento ativo do IBGE e dos INEs latino-americanos neste debate, bem como sua capacidade institucional e política de importar, adaptar e desenvolver soluções adequadas aos seus países, que levem em conta a finalidade pública das estatísticas oficiais e a soberania de dados, são elementos que definirão o grau de autonomia ou dependência que assumirão na transição para um regime estatístico-algorítmico de dataficação.

CONCLUSÃO

Neste trabalho, testamos a hipótese de que a introdução de novas fontes de dados e métodos no processo de produção das estatísticas oficiais conduz a uma interseção entre os campos de práticas estatísticas e algorítmicas e a uma transição para um novo regime estatístico-algorítmico de dataficação. Isto foi realizado por meio do levantamento de evidências históricas e empíricas para sustentar a nossa tese de que o processo de transição para um regime de dataficação é caracterizado por dois obstáculos — ontológico (relativo aos dados) e epistemológico (relativo aos métodos) — que conduzem a um duplo movimento no campo estatístico — movimento pró-mercado e contramovimento protetivo. O conjunto de evidências indica que esta transição é um campo aberto de lutas e possibilidades, cujas principais dinâmicas e tendências buscamos caracterizar.

Apoiada na intuição original de Desrosières (2014), nossa análise norteou-se pela noção de que as crises que fundam e transformam o modo de produção e o regime de acumulação capitalista (Boyer, 1990; Chesnais, 1996, Coriat, 2000; Anglietta, 2015) implicam em crises de governamentalidade (Foucault, 2008, 2008a), que tendem, por sua vez, a reconfigurar os regimes estatísticos (Beaud e Prévost, 2011). Este processo ocorre na medida em que a estatística, enquanto ferramenta de prova e de governo, é mobilizada para assegurar a legitimidade do Estado e a reprodução do capital nos contextos de crise. Conforme o quadro analítico a seguir, identificamos a configuração de quatro regimes estatísticos ao longo da história — contabilização, enumeração, precisão e comensuração — para finalmente caracterizar o atual processo de transição para um regime estatístico-algorítmico de dataficação.

Por meio deste quadro analítico, e apoiados no instrumental sociológico do estruturalismo genético/gerativo de Bourdieu (1986, 1990, 1998, 2008), efetuamos uma investigação abrangente de como se estruturam e se transformam os campos estatísticos transnacional e nacional, à luz das mudanças de regimes estatísticos. Conferimos especial atenção aos períodos de crise e transição. Ao mesmo tempo, buscamos traçar uma análise comparativa destes processos no Norte-Global, em particular na Europa Ocidental e Estados Unidos, com o caso brasileiro. Para tanto, nos apoiamos no referencial da teoria latino-americana da dependência (Bambirra, 2013; Cardoso; Faletto, 1970; Frank, 1969; Furtado, 1961, 1968; Marini, 2008; Prebish,

1949; Rangel, 2005; Santos, 2011, Tavares, 1982, 1992) e na historiografia estatística do Brasil (Penha, 1993; Botelho, 1998; Senra, 2006, 2008, 2009, 2016; Camargo, 2009, 2009a, 2016, 2018; Grabois, 2021; Gil, 2023; Pereira, 2023). Com isso demonstramos que, embora as macro determinações globais sejam relevantes em ambos os casos, o desenvolvimento desigual e combinado dos países periféricos, implica tanto em dilatações cronológicas na transição dos regimes, como em diferenciações na estruturação de seus campos estatísticos.

Tabela 6 - Quadro Analítico dos Regimes Estatísticos

Crises do regime de acumulação	Crise de Governamentalidade	Racionalidade Econômica	Regime Estatístico	Ferramentas Estatísticas
Desarticulação do modo de produção Feudal (Séc. XVI)	Crise dos poderes feudais. Unificação do poder central do Estado Territorial.	Mercantilismo	Contabilização	Taxonomia Tabelas Registros Descrições
Crise Agrícola (Sec. XVII) Revolução Industrial (Séc. XVIII-XIX)	Crise de Legitimidade da dominação dinástica. Revoluções burguesas e anticoloniais. Formação do Estado Nação.	Liberalismo	Enumeração	Estatística descritiva Censos nacionais
Longa Depressão (final do Séc.XIX) e 1ª.Guerra	Crise dos dogmas liberais. Contramovimentos protetivos do trabalho. Reformas Sociais do Estado.	Liberalismo Social	Transição	Surveys representativos
Grande Depressão (1929) e 2ª.Guerra	Crise dos dogmas liberais e do mercado autorregulado. Intervenção Estatal na sociedade e economia.	Keynesianismo	Precisão	Estatística inferencial Surveys amostrais Contas nacionais
Choques do Petróleo e Estagflação (1973-1980)	Crise do modelo keynesiano de intervenção planejada do Estado. Reformas gerenciais do Estado.	Neoliberalismo	Comensuração	Indicadores Benchmarkings Rankings
Crises da financeirização (2007-2008-2011)	Crise do modelo neoliberal. Ascensão da extrema direita.	Plataformização	Dataficação	Estatística bayesiana Modelagem Big data IA

Fonte: do autor.

Analizamos como na formação e consolidação dos Estados Nação, por meio de uma série de desapossamentos que constituíram os espaços nacionais de convenções de equivalências, a estatística oficial transmutou-se em um bem público. Apontamos para dois aspectos estruturais desta nova configuração do ponto de vista do Estado: o papel que a estatística oficial passou a assumir como meio geral de cognição do mercado, e como instrumento de governo biopolítico da população. No primeiro caso, a função econômica das estatísticas oficiais, vinculada ao processo geral de extração da mais valia, apoia-se sobretudo na autonomia relativa do Estado

frente ao mercado, o que lhe confere legitimidade para assumir o papel de coordenação dos agentes econômicos. No segundo caso, a função biopolítica é sustentada por construções cognitivas que dotam o Estado de instrumentos para o governo da população — inicialmente baseados na descrição e enumeração e posteriormente na inferência e precisão. Por outro lado, a configuração da estatística oficial como um bem público, ou seja, disponibilizado e acessível ao conjunto da sociedade, conduziu à sua crescente incorporação no debate público e à sua utilização como ferramenta de crítica social em favor da democratização, favorecendo contramovimentos protetivos contra os efeitos do mercado autorregulado, que no caso brasileiro, assumiram o caráter de movimentos nacionais. Em uma palavra, a estatística oficial passou a constituir um instrumento de articulação do duplo movimento do Estado, sendo cada vez mais demandada no curso das sucessivas crises do regime de acumulação do capital e da governamentalidade liberal.

Vimos como a estruturação do campo estatístico e as transições de regimes estatísticos são atravessadas por disputas que têm como pano de fundo contradições vinculadas ao papel da estatística enquanto ferramenta de prova e de governo.

Em primeiro lugar, enquanto ferramenta de governo, a estatística apresenta-se por um lado, como um instrumento da soberania nacional, vinculado ao Estado e à lei, e, por outro lado, como um meio de comparabilidade e competição internacional, associado a uma rede transnacional de instituições e especialistas. Esta dinâmica contraditória é uma característica comum tanto nos países do norte, como do sul-global. Vimos como desde os primórdios da estruturação do campo estatístico transnacional, nos Congressos Internacionais de Estatística do Séc. XIX, as disputas de caráter nacional impuseram restrições aos movimentos internacionalistas. Nos países do Sul-Global, entretanto, a relação dialética entre internacionalismo e nacionalismo estatístico ganha maior relevância, tendo em vista a assimetria de forças destes países no plano internacional.

No caso brasileiro, foram sobretudo os movimentos de caráter nacional que conferiram uma historicidade e organicidade própria à produção estatística nacional. Assim, embora a adoção de novas tecnologias, como o censo demográfico em 1872 e os inquéritos amostrais nos anos 1960, tenham dependido fundamentalmente da importação de inovações desenvolvidas no Norte-Global, é apenas quando os movimentos de caráter nacional são capazes de se apropriar destas tecnologias e lhes conferir sentido e consequência política, que verificamos um efetivo processo de

estruturação do campo estatístico e de transição de regime estatístico no Brasil. Em outras palavras, tudo se passa como se, embora sob a influência do internacionalismo estatístico, o nacionalismo estatístico se impusesse como a principal dinâmica gerativa tanto da estruturação do campo, como da transição de regimes. Este resultado da investigação histórica é uma chave de leitura relevante para a compreensão do contexto contemporâneo.

Em segundo lugar, enquanto ferramentas de prova, os distintos métodos estatísticos quando incorporados em ferramentas de governo no âmbito do Estado, correspondem a determinadas configurações objetivas do campo estatístico. Estas configurações se expressam nas normas, práticas e estruturas que caracterizam o regime estatístico. Neste sentido, mudanças no âmbito epistemológico dos métodos estatísticos relacionam-se a mudanças ontológicas das estruturas estatísticas. Conforme vimos, a transição de um 'regime estatístico de enumeração', amparado na exaustividade dos censos, para um 'regime estatístico de precisão', fundamentado na estatística inferencial e nos inquéritos amostrais, conduziu a uma mudança estrutural no campo estatístico na medida em que os estatistas e administradores generalistas foram substituídos pelos estatísticos de Estado especialistas. Tal transição caracterizou-se por lutas que correspondem à força inercial do habitus incorporado pelos agentes, que tende a resistir a mudanças que possam afetar a lógica e a estrutura do campo. Notadamente, a incorporação de novas práticas pelo campo estatístico relaciona-se, por um lado, com o nível de desenvolvimento cognitivo destas práticas e, por outro, com a sua aderência ao modo de ação do Estado, o que em última instância determina a lógica dos investimentos e a transição para um novo regime estatístico. Esta é outro achado importante da análise histórica para a interpretação do atual processo de transição.

Sob a hegemonia do modelo keynesiano, verificamos o fortalecimento do caráter público e nacional das estatísticas oficiais em todo o mundo. Isto decorreu, por um lado, da intervenção planejada do Estado na economia e sociedade, que passou a demandar cada vez mais números públicos, e, por outro lado, da construção do sistema de governança internacional das Nações Unidas, sob o fundamento da soberania nacional. Verificamos, neste contexto, a consolidação do regime estatístico de precisão, com a incorporação da estatística inferencial, das contas nacionais e dos computadores eletrônicos em Instituições Nacionais de Estatística que passaram a centralizar os sistemas estatísticos nacionais com um maior grau de

profissionalização. Os estatísticos de Estado, detentores de conhecimentos científicos especializados e investidos de um papel público, consolidam-se como a personificação deste regime.

A crise do regime de acumulação keynesiano-fordista e a emergência do neoliberalismo, acompanhada pelos processos de financeirização e digitalização, implicam em uma série de mudanças no regime estatístico. A emergência de novas modalidades de quantificação orientadas pela lógica competitiva e impulsionadas pelas tecnologias de informação e comunicação conduzem à generalização dos indicadores estatísticos, rankings e benchmarkings como instrumentos de governo e à multiplicação de agentes quantificadores fora do âmbito estatal. As Instituições Nacionais de Estatística se tornaram mais permeáveis à lógica de mercado, mais responsivas às inovações tecnológicas e mais susceptíveis à incidência de agentes internacionais e privados. Por outro lado, os INEs se mostraram resilientes e capazes de se adaptar e se reposicionar no novo contexto, preservando a sua autonomia e o caráter público e nacional das estatísticas oficiais por meio de contramovimentos protetivos. No caso brasileiro, a resiliência às reformas neoliberais vinculou-se a um novo posicionamento do IBGE no contexto de redemocratização, reforçado pela organização e resistência de seus quadros. O regime estatístico de comensuração, característico do período neoliberal, configura-se, assim, como um elo de transição para as transformações que emergem com a intensificação dos processos de digitalização, dataficação e algoritmização no Séc. XXI.

A crise do regime de acumulação financeirizada, evidenciada a partir de 2007/08/11, aponta para uma crise contemporânea da governamentalidade neoliberal. Na esteira da intensificação dos processos de digitalização verifica-se a emergência de uma nova economia orientada aos dados, caracterizada pela convergência entre os processos de financeirização e dataficação. O modelo de extração, cercamento e capitalização dos dados das plataformas digitais conduz à emergência de um regime de acumulação do capital pautado pela ampliação da dominação econômica e política das Big Techs e outras grandes corporações privadas globais. Estas corporações passam a exercer uma governamentalidade algorítmica paralela e análoga aos Estados Nacionais, na medida em que penetram em uma série de setores tradicionalmente estatais.

Nossa investigação revelou que o processo de transição das estatísticas oficiais para um novo regime de dataficação se manifesta em um conjunto de lutas

que emergem na interseção entre os campos estatístico e algorítmico. Conforme vimos, o campo estatístico foi historicamente estruturado ao longo de séculos por meio da relativa autonomização de funções próprias dos Estados Nacionais, orientadas para a acumulação de capital informacional e associadas a um conjunto de ferramentas cognitivas, dentre as quais destacamos as aplicações frequentistas do cálculo de probabilidades. Já o campo algorítmico emergiu estreitamente vinculado ao mercado e a corporações privadas do Norte-Global que se diferenciaram a partir do desenvolvimento das tecnologias de informação e comunicação no Séc. XX e se expandiram com a intensificação dos processos de digitalização, dataficação e algoritmização no Séc. XXI. Associado ao novo campo, verificamos a emergência de novas fontes de dados que conduzem à emergência de novos métodos e modalidades informacionais. Este conjunto de inovações passou a desafiar as práticas tradicionais do campo estatístico, a relevância das estatísticas oficiais e o controle dos Estados Nacionais sobre os novos capitais informacionais.

Na medida em que o novo campo algorítmico passa a ameaçar a hegemonia informacional dos Estados Nacionais e a relevância das estatísticas oficiais, o campo estatístico é pressionado a se ‘modernizar’ e a se engajar com as fontes de big data e com os novos métodos da ciência de dados. Neste contexto, verificam-se dois obstáculos que caracterizam o processo de transição das estatísticas oficiais para o novo regime de dataficação.

Em primeiro lugar, do ponto de vista ontológico, a nova economia política dos dados orientada pelo modelo de plataformização efetua uma transmutação dos dados em mercadorias e ativos, desafiando o monopólio informacional dos Estados Nacionais e a tradicional concepção dos dados para fins estatísticos como bens públicos. Os cercamentos de dados das corporações privadas, condição para a constituição de novos mercados no capitalismo de plataforma, vedam o acesso dos Estados e seus campos estatísticos aos novos capitais informacionais. Ao mesmo tempo, as estatísticas oficiais passam a constituir um potencial nicho de mercado para o campo algorítmico.

Em segundo lugar, do ponto de vista epistemológico, o caráter não representativo das fontes de big data e os novos métodos da ciência de dados — bayesianos, indutivos e baseados na modelagem — desafiam as práticas estatísticas tradicionais — frequentistas, dedutivas e baseadas no desenho de inquéritos e na teoria da amostragem.

Configura-se assim um espaço de lutas na interseção dos dois campos de práticas que caracteriza o processo de transição para o regime de dataficação. Do ponto de vista do campo estatístico estas lutas configuram um duplo movimento. Por um lado, a articulação de um discurso de ‘modernização das estatísticas oficiais’ caracteriza um movimento pró-mercado, que conduz ao estabelecimento de intersecções entre os campos estatístico e algorítmico, por meio de uma agenda de parcerias público-privadas e incentivos à gestação de uma nova geração de cientistas de dados nos INEs. Por outro lado, verifica-se um contramovimento protetivo da autonomia dos INEs e dos métodos estatísticos tradicionais, por meio da regulamentação do acesso às novas fontes de dados como bens públicos, visando a sua incorporação controlada sob a governança dos estatísticos de Estado e dos sistemas estatísticos nacionais.

O discurso de modernização das estatísticas oficiais com big data que emergiu no contexto europeu ganhou uma escala global a partir de 2015, com a sua incorporação na pauta da Comissão Estatística da ONU, no contexto de implementação dos Objetivos do Desenvolvimento Sustentável (ODS) da nova agenda global de 2030. Assim, na medida em que a União Europeia gradualmente se afastou da agenda de parcerias público-privadas em big data e buscou reforçar seus sistemas estatísticos nacionais pela via da regulação, verificamos a disseminação da agenda pró-mercado para o Sul-Global.

A articulação de alto nível entre agentes do campo estatístico e algorítmico transnacional se expressou na estratégia da ‘revolução de dados para o desenvolvimento sustentável’. A revolução de dados passou a instrumentalizar os ODS, a agenda 2030 e a confiança pública na ONU e nas estatísticas oficiais para promover mercados de dados. Isto se deu por meio do incentivo a parcerias público-privadas em big data no setor estatístico, da promoção de novos modelos de financiamento da infraestrutura estatística global e da implantação de mudanças na arquitetura institucional da Divisão Estatística da ONU. Destaca-se a criação de uma ‘parceria global de dados’ que abriu um espaço político para a ingerência direta de corporações privadas globais nos sistemas estatísticos nacionais e global, se revelando ao mesmo tempo uma estratégia de mercado e de governança das Big Tech.

O Projeto ONU Big Data, que resulta desta articulação, constitui um núcleo global de experimentação e disseminação de novas práticas de dataficação e

algoritmização das estatísticas oficiais, com o protagonismo de INEs do Norte-Global e com o financiamento e envolvimento direto de Big Techs e outras corporações privadas. Neste sentido, vimos como organizações globais como a Microsoft e a Nielsen financiam e participam de grupos de trabalho da iniciativa, pautando pesquisas, com a participação de INEs, que tem o potencial de se desdobrar em ações efetivas de penetração destas empresas nas práticas estatísticas, neste caso, por meio de modelos de compartilhamento de dados com tecnologias de ampliação de privacidade e do incentivo ao consumo de dados proprietários pelos sistemas estatísticos nacionais. Vimos ainda como neste contexto, verifica-se o desenvolvimento de um projeto de plataformização das estatísticas oficiais, de acordo com princípios mercadológicos e tecnológicos das plataformas comerciais de dados.

As evidências reunidas demonstram como a Plataforma Global da ONU adota uma concepção neocolonialista de dados, na medida em que intensifica assimetrias tecnológicas, conferindo aos países do Sul-Global um status de provedores de dados e receptores passivos de métodos e tecnologias desenvolvidas por países e corporações do Norte-Global. Destaca-se, neste sentido, a implantação de hubs regionais em países do Sul-Global que funcionariam como estruturas supranacionais intermediárias e facilitadoras da inserção subordinada dos INEs à Plataforma, cuja gestão a partir do Norte-Global é realizada por uma organização privada (GPSDD) financiada por Big Techs (Google e Microsoft).

Nossa pesquisa empírica demonstrou como as parcerias público-privadas e a adesão à premissas de mercado conduziram a uma série de conflitos de interesses no projeto da Plataforma Global. Por outro lado, entrevistas com agentes de alto nível do campo estatístico transnacional revelaram controvérsias sobre as estratégias de apropriação de novas tecnologias e fontes de dados na produção de estatísticas oficiais. Estas evidências confirmam a tese do duplo movimento, sendo possível verificar no nível transnacional, por um lado, o movimento pró-mercado dos promotores das parcerias público-privadas e da comoditização dos dados, e, por outro, o contramovimento da regulamentação e fortalecimento da governança dos sistemas estatísticos nacionais e global, protetivo da soberania nacional e dos dados estatísticos como bens públicos.

Os dados mobilizados e armazenados pelos sistemas estatísticos nacionais e global, o potencial de consumo dos governos nacionais e sistema ONU de novas fontes de dados, métodos e tecnologias, e a confiança e legitimidade das estatísticas

oficiais são elementos-chave para compreendermos o interesse de Big Techs e do campo algorítmico em geral, em seu movimento de incidência e interação com o campo estatístico. Trata-se, por um lado, de construir novos meios de geração de valor a partir dos dados, em novos mercados globais que passam a envolver o setor tradicionalmente público e estatal das estatísticas oficiais. E, por outro lado, da implantação de uma governamentalidade algorítmica que tende a favorecer o poder político e a governança das corporações privadas globais, ampliando ao mesmo tempo seu domínio econômico. Neste sentido, a Plataforma Global é um experimento de vanguarda, na medida em que poderia potencialmente inverter o processo tradicional de compilação de estatísticas oficiais, transferindo o lócus da produção nacional para os níveis supranacional e global.

O alcance global da Plataforma por meio dos hubs regionais pode potencialmente habilitar novas cadeias de valor e mercados de dados, envolvendo, entre outros: i) a padronização global de métodos para a produção de estatísticas oficiais cada vez mais dependentes de dados extraídos e processados por corporações globais; ii) a comercialização de dados guardados por privados para a produção de estatísticas oficiais; iii) a criação de serviços de computação distribuída para a monetização do acesso aos dados; iv) o uso dos dados para treinamento de modelos de aprendizado de IA. Tratam-se de modelos que se encontram em experimentação e cuja implementação encontra resistências vinculadas principalmente ao caráter estatal, público e nacional das estatísticas oficiais.

Os hubs regionais no Sul-Global para a promoção de big data em estatísticas oficiais, podem ser compreendidos, portanto, como uma estratégia para vencer as resistências nacionais à incidência e participação da iniciativa privada nos sistemas estatísticos. Vimos como em um contexto de fragilização institucional do IBGE durante o governo de extrema direita de Jair Bolsonaro deu-se a implantação do Hub Regional de Big Data para a América Latina no Brasil. Neste período, verifica-se, por um lado, a desarticulação de iniciativas que visavam o fortalecimento político e institucional do IBGE na coordenação estatística nacional e o seu reposicionamento estratégico no contexto de dataficação, e, por outro lado, a promoção de ações disruptivas de 'modernização' voltadas à introdução de big data e ciência de dados no órgão, orientadas pela cooperação internacional.

Avaliamos que os movimentos de 'modernização' do IBGE orientados por acordos internacionais durante o Governo Bolsonaro visaram promover uma inserção

subordinada do Brasil no regime de dataficação, ao mesmo tempo em que tentaram cooptar o órgão para a promoção da pauta de parcerias público-privadas no campo estatístico latino-americano. Verificamos como apesar do IBGE ser o principal financiador do Hub Regional, na falta de uma orientação estratégica nacional e em meio a uma conjuntura de desestabilização institucional, as ações do projeto assumiram uma orientação autônoma, pautada pelo Projeto ONU Big Data. Isto envolveu, por exemplo, a abertura de espaços para a incidência comercial de empresas privadas junto aos INEs latino-americanos, a participação de técnicos do IBGE em grupos de trabalho com especialistas de Big Techs e outras empresas privadas, e a discussão de modelos de acesso aos dados por meio de parcerias público-privadas.

O hub regional, por outro lado, tem oportunizado um importante espaço de discussões e debates entre os INEs latino-americanos sobre as novas modalidades de dados e métodos no contexto da dataficação das estatísticas oficiais. Verificamos que nas atividades de intercâmbio e formação do hub a totalidade das experiências com fontes de big data apresentadas e debatidas pelos INEs latino-americanos são realizadas sem intermediação de empresas privadas e em geral por meio do acesso à fontes públicas e abertas de dados. Além disso, a participação dos técnicos do IBGE nos grupos de trabalho do Projeto ONU Big Data tem oportunizado o acesso à novos conhecimentos, métodos e tecnologias para lidar com as novas fontes de dados.

Desdobrando controvérsias na implantação e atividades do hub regional e em casos de uso de fontes de big data no IBGE, verificamos que há diferentes entendimentos e posicionamentos entre os agentes do órgão sobre o projeto da Plataforma Global da ONU, sobre a incidência de empresas privadas no campo estatístico e sobre os modelos de acesso aos dados. Entre os quadros mais antigos e egressos da instituição, identificamos um posicionamento mais altivo quanto à defesa da autonomia e soberania estatística nacional e mais crítico quanto à inserção das empresas privadas no setor. Entre os quadros mais novos, verificamos em alguns casos uma postura de maior subordinação na relação com os agentes internacionais e de normalização das relações público-privadas no campo estatístico. Verificamos ainda um esforço de dissociação do hub regional ao projeto da Plataforma Global, o que indica um possível desvio na tradução dos interesses globais da iniciativa e um contramovimento protetivo local de autonomização, coerente com a lógica pública dos estatísticos de Estado e do campo estatístico regional e nacional.

No que se refere aos modelos de acesso aos dados, identificamos uma divisão de posicionamentos no IBGE. Alguns quadros são explicitamente críticos às parcerias público-privadas e defendem um modelo pautado na regulamentação do acesso aos dados como bens públicos pelo Estado, enquanto outros quadros são mais abertos às PPPs e à concepção dos dados como mercadorias, passíveis de custo financeiro de aquisição. Do ponto de vista das modalidades de acesso, os posicionamentos também se dividem em termos da defesa de um modelo de acesso centralizado dos dados brutos no IBGE, e um segundo modelo descentralizado de acesso aos dados com uso de computação distribuída e tecnologias de ampliação de privacidade (PETs), que preservariam a ‘propriedade’ das bases de dados pelos seus detentores originais. Vimos que os interesses da Microsoft na iniciativa da ONU concentram-se especialmente na difusão do modelo descentralizado. Vimos ainda que há capacidades para o desenvolvimento autônomo deste modelo no IBGE, o que poderia viabilizar o acesso a registros administrativos e outras fontes de dados em que há maiores restrições de sigilo.

Finalmente, as entrevistas com os estatísticos do IBGE revelaram a incorporação de disposições características dos estatísticos de Estado — pautadas pelo compromisso com métodos consagrados, aversão a riscos e ética do serviço público — que se manifestam em um conservadorismo relativo na introdução de novas fontes de dados, métodos e tecnologias, e em uma resistência crítica na adoção do novo jargão profissional de ‘cientista de dados’ — cujo perfil profissional empreendedor e inovador é associado às práticas de mercado. A resistência à inovação e à adoção do big data e ciência de dados entre os quadros do IBGE, entretanto, é apenas relativa, na medida em que as novas fontes, métodos e tecnologias constituem capitais valiosos que são disputados pelos agentes e passam a ser objeto de novos projetos em curso na instituição.

Vimos que novas experiências com dados coletados diretamente na internet (web scraping e web crawling) e com dados de telefonia celular vem sendo realizadas no IBGE. As experiências com big data no órgão são ainda bastante pontuais e iniciais. Apenas uma experiência efetivamente evoluiu para a produção e divulgação de uma estatística oficial (índice de preços de passagens aéreas). Os projetos internos no IBGE revelam, por um lado, capacidades de desenvolvimento autônomo de soluções por meio da cooperação entre diferentes áreas técnicas. Por outro lado,

verifica-se que o acesso aos dados e registros administrativos é a principal restrição ao seu desenvolvimento.

No que diz respeito aos métodos, evidenciamos como o obstáculo epistemológico referente ao caráter desestruturado e não representativo de algumas fontes de big data passa a ser objeto de discussão entre os estatísticos ibgeanos. O que envolve tanto a crítica a determinados métodos prescritivos da ciência de dados, como proposições que apontam para a integração de fontes de big data com fontes tradicionais, estruturadas e representativas, como os registros administrativos e os microdados dos inquéritos.

A abordagem sóbria e cautelosa nos discursos dos estatísticos entrevistados sobre as inovações nos dados e métodos é coerente com as disposições sancionadas pelo campo estatístico, e em particular pelo IBGE. Tais disposições — que incluem o ethos público dos estatísticos de Estado, o compromisso com os métodos consagrados, a manutenção das séries históricas, a confiabilidade dos dados e a precisão e caráter público das estatísticas oficiais — contribuiram historicamente para conformação de um habitus ibgeano, avesso à ruptura e favorável a uma conciliação processual entre tradição e renovação.

Tudo parece indicar que a característica pública, estatal e nacional dos sistemas estatísticos do Brasil e da América Latina, bem como a predominância do habitus dos estatísticos de Estado tanto no IBGE como nos INEs da região constituem uma importante barreira à disseminação de projetos que visam a ampliação da participação do setor privado no campo estatístico, a substituição disruptiva de fontes e métodos tradicionais e a criação de mercados de dados para estatísticas oficiais. Neste sentido, a estratégia da ‘revolução de dados’ e da ‘mudança cultural’ nos INEs com investimentos das Big Techs — na criação de infraestruturas tecnológicas, na promoção de espaços público-privados e na formação de uma nova geração de cientistas de dados — visa ajustar as disposições dos estatísticos de Estado, de modo a torná-los mais susceptíveis e permeáveis à associação com os agentes de mercado. Trata-se, portanto, de facilitar no médio e longo prazo a incorporação de um novo habitus algorítmico e empreendedor no campo estatístico oficial, reproduzindo disposições características das organizações privadas da área de tecnologia e dados que compõem o campo algorítmico.

Conforme vimos, entretanto, no caso do IBGE, tanto a concepção ontológica dos dados (bens públicos e mercadorias), quanto a epistemologia dos modelos

(frequentistas/explicativos e bayesianos/preditivos) estão em disputa. Isto se reflete em diferentes estratégias dos agentes que evidenciam a validade da tese do duplo movimento do campo estatístico no atual processo de transição para um regime de dataficação. Por um lado, a abertura às parcerias público-privadas e a internalização do perfil profissional dos cientistas de dados, e, por outro, a demanda de proteção do sistemas estatístico nacional pela via da regulação e da preservação dos métodos e fontes tradicionais.

Evidentemente, é vital ao campo estatístico brasileiro e latino-americano tanto o acesso às novas fontes de dados como aos métodos que possibilitem traduzí-los em estatísticas oficiais, públicas e confiáveis. Este processo pressupõe tanto a importação de tecnologias produzidas nos centros mais dinâmicos da economia mundial pela via do internacionalismo estatístico, como a sua adaptação ao contexto e necessidades locais e seu desenvolvimento autônomo a partir dos objetivos e capacidades próprias do nacionalismo estatístico, orientado pelos princípios do interesse público, democracia e soberania e por um projeto de desenvolvimento nacional autodeterminado.

A compreensão dos elementos políticos e econômicos encapsulados nas questões técnicas é especialmente relevante em um contexto de transição tecnológica. As estatísticas oficiais são informações públicas fundamentais para o desenvolvimento nacional soberano, tanto no nível do governo, provendo as informações para a sua orientação estratégica, quanto no nível da sociedade civil organizada e população em geral, bem informada para o exercício crítico e ativo da democracia e de sua autodeterminação. Finalmente, o posicionamento ativo e altivo do IBGE e INEs latino-americanos no processo de dataficação, bem como sua capacidade institucional e política de defender soluções adequadas aos seus países, que levem em conta a finalidade pública das estatísticas oficiais e a soberania de dados de suas populações, são elementos relevantes para a definição do grau de autonomia ou dependência que estes países assumirão na transição para um regime estatístico-algorítmico de dataficação.

Para concluirmos, findo no Brasil o período desestruturador do Governo Bolsonaro, verifica-se a partir de 2024 a retomada do processo de discussões no IBGE e no governo brasileiro sobre a necessidade de reestruturação do sistema estatístico e geocientífico nacional pela via da regulação. Este processo, que se dá em um terceiro mandato do presidente Lula e por ocasião da proximidade dos noventa anos

do IBGE, passa a dar especial relevância a pauta da soberania nacional de dados, tema recorrentemente abordado pelo novo presidente da casa, o economista Márcio Pochmann. As discussões atualmente em curso não são objeto desta investigação, mas indicam a abertura de uma nova conjuntura para o campo estatístico brasileiro e latino americano, em um contexto de crescentes disputas na região entre as Big Tech e os Estados Nacionais.

REFERÊNCIAS

- AGGARWAL, C. **Data Mining: The Textbook**. Cham: Springer, 2015.
- ANATEL. Voto No. 30/2020/PR. Processo no 53500.017367/2020-40 Interessado: Fundação Instituto Brasileiro de Geografia e Estatística - IBGE. Brasília: **ANATEL**, 2020.
- ANDERSON, M. *The American Census: A Social History*. New Haven: Yale University Press, 2015.
- ANGLIETTA, M. **A Theory of Capitalist Regulation: The US Experience**. London: Verso, 2015.
- ASSIBGE. Nota da ASSIBGE-SN aos participantes do workshop do IBGE sobre o Censo 2020. Rio de Janeiro: **ASSIBGE**, 2019. Disponível em: <https://assibge.org.br/nota-da-assibge-sn-aos-participantes-do-workshop-do-ibge-sobre-o-censo-2020/>. Acesso em: 12 jan. 2023.
- ASSIBGE. Nota ao Banco Mundial. Rio de Janeiro: **ASSIBGE**, 2019a. Disponível em: <https://assibge.org.br/nota-ao-banco-mundial/>. Acesso em: 12 jan. 2023.
- BACHELARD, G. **A formação do espírito científico: contribuição para uma psicanálise do conhecimento**. Rio de Janeiro: Contraponto, 2005.
- BAMBIRRA, Vânia. **O capitalismo dependente latino-americano**. 2. ed. Florianópolis: Insular, 2013.
- BANBURA, M; REICHLIN, L.; GIANNONE, D. Nowcasting. **CEPR Discussion Papers** 7883, 2010.
- BANDEIRA, L. **Presença dos Estados Unidos no Brasil: dois séculos de história**. Rio de Janeiro: Civilização Brasileira, 1978.
- BEAUD JP.; PRÉVOST JG. The Politics of Measurable Precision: The Emergence of Sampling Techniques in Canada's Dominion Bureau of Statistics, **The Canadian Historical Review**, Toronto, v.79, n.4, p. 691-725, dez. 1998.
- BEAUD JP.; PRÉVOST JG. On the genesis of a new statistical regime: the case of Canada, 1800-2011. **Estatística e Sociedade**, Porto Alegre, n.1, p.26-48, 2011.
- BECKER, H. **Métodos de pesquisa em ciências sociais**. São Paulo: Hucitec, 1993.
- BEER, D. Power through the algorithm? Participatory web cultures and the technological unconscious. *New media & society*, v.11, n.6, p.985-1002, 2009. Disponível em: <https://doi.org/10.1177/1461444809336551>. Acesso em: 03 fev. 2021.
- BENZÉCRI, JP. **Histoire et préhistoire de l'analyse des données**. Partie V L'analyse des correspondances. *Les cahiers de l'analyse des données*, v.2, n. 1, p. 9-

40, 1977. Disponível em: http://www.numdam.org/item?id=CAD_1977__2_1_9_0. Acesso em: 10 out. 2020.

BERNARD, H. R. **Research methods in anthropology: qualitative and quantitative approaches**. Lanham, MD: AltaMira Press, 2005.

BHASKAR, R. **Dialectic: The Pulse of Freedom**. London: Routledge, 2008.

BHASKAR, R. **The Possibility of Naturalism: A philosophical critique of the contemporary human sciences**. London: Routledge, 2014.

BIANCOTTI, C.; ROSOLIA, A.; VERONESE, G.; KIRCHNER, R.; MOURIAUX, F. COVID-19 and official statistics: a wakeup call? (February 12, 2021). **Bank of Italy**, Occasional Paper No. 605. Disponível em: <https://ssrn.com/abstract=3828122>. Acesso em: 13 mar. 2023.

BID. Ideação – Inovação em Gestão Pública. Susana Cordeiro Guerra. Washington: **Inter-American Development Bank**, 2021. Disponível em: <https://blogs.iadb.org/brasil/pt-br/author/susanacordeiro/>. Acesso em: 14 mar. 2023.

BIEVER, C. Twitter mood maps reveal emotional states of America. **New Scientist**, v. 207, Issue 2771, p. 14, 2010. Disponível em: [https://doi.org/10.1016/S0262-4079\(10\)61833-7](https://doi.org/10.1016/S0262-4079(10)61833-7). Acesso em: 03 abr. 2022.

BIG Data and the UN Global Platform by Mark Craddock. London: AWS, 2019. 1 vídeo (38 min). Publicado pelo canal AWS User Group UK. Disponível em: <https://www.youtube.com/watch?v=6OfTYYSI-xk>. Acesso em: 13 set. 2023.

BIRCH, K.; BRONSON, K. Big Tech. **Science as Culture**, v. 31, n.1, p 1–14. 2022. Disponível em: <https://doi.org/10.1080/09505431.2022.2036118>. Acesso em: 21 jul. 2023.

BIRCH, K.; COCHRANE, D.; WARD, C. Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech. **Big Data & Society**, v.8, n.1, p. 1–15, 2021. Disponível em: <https://doi.org/10.1177/20539517211017308>. Acesso em: 17 set. 2022.

BOITO, A. Neofascismo e neoliberalismo no Brasil do Governo Bolsonaro. **Observatorio Latinoamericano y Caribeño**, v. 4, n.2, 2020.

BOTELHO, T. **População e nação no Brasil do século XIX**. 1998. Tese (Doutorado em História) – Universidade de São Paulo, São Paulo, 1998.

BOURDIEU, P. The forms of capital. In: RICHARDSON, J. (org.) **Handbook of Theory and Research for the Sociology of Education**. Westport: Greenwood, 1986. p. 241-258.

BOURDIEU, P. **The Logic of Practices**. Stanford: Stanford University Press, 1990.

BOURDIEU, P. **Practical Reason**. On the Theory of Action. Stanford: Stanford University Press, 1998.

BOURDIEU, P. **Meditações Pascalianas**. Rio de Janeiro: Bertrand Brasil, 2001.

BOURDIEU, P. **A economia das trocas linguísticas**: o que falar quer dizer. São Paulo: EDUSP, 2008.

BOURDIEU, P. **Sobre o Estado**: Cursos no Collège de France (1989-92). 1ª. Edição. São Paulo: Companhia das Letras, 2014.

BOURDIEU, P. **A Distinção**: crítica social do julgamento. Porto Alegre: Editora Zouk, 2017.

BOURDIEU, P.; Chamboredon, JC; Passeron JC. **Ofício de Sociólogo**: Metodologia da pesquisa na sociologia. 8ª. Edição. Petrópolis: Editora Vozes, 2015.

BOYER, R. **A Teoria da Regulação**: uma análise crítica. São Paulo: Nobel, 1990.

BRAAKSMA, B.; ZEELENBERG, K. Re-make/Re-model: Should big data change the modelling paradigm in official statistics?. **Statistical Journal of the IAOS**, v.31, 2015.

BRAMER, M. **Principles of Data Mining**. London, United Kingdom: Springer, 2007.

BRUNO, I.; DIDIER, E. **Benchmarking**. L'Etat sous pression statistique. Paris: La Découverte, 2013.

BUTLER, D. When Google got flu wrong. **Nature**, v.494, p. 155-156, 2013. Disponível em: <https://doi.org/10.1038/494155a>. Acesso em: 14 jan. 2022.

CAMARGO, A. O inventário do passado: Isaac Kerstenetzky e a refundação da memória do IBGE. In: Senra, N. (org.). **História das Estatísticas Brasileiras**. Volume 3. Estatísticas Organizadas (1936-1972). Rio de Janeiro: IBGE, 2009. p.145-167.

CAMARGO, A. Conhecendo os dissidentes: os dados não estruturados do ENDEF e a configuração da pobreza no Brasil. In: Senra, N. (org.). **História das Estatísticas Brasileiras**. Volume 3. Estatísticas Organizadas (1936-1972). Rio de Janeiro: IBGE, 2009a. p.168-212.

CAMARGO, A. **A Construção da Medida Comum**: Estatística e Política de População no Império e na Primeira República. 2016. Tese (Doutorado em Sociologia) - Instituto de Estudos Sociais e Políticos, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2016.

CAMARGO, A. O Censo de 1872 e a utopia estatística do Brasil Imperial. **História Unisinos**, v. 22, n.3, p. 414-428, Set./Out., 2018.

CANUTO, L. Redução no orçamento não afetará qualidade do Censo, diz presidente do IBGE. **Câmara dos Deputados**, Brasília, 04 abr. 2019. Disponível em:

<https://www.camara.leg.br/noticias/561626-reducao-no-orcamento-nao-afetara-qualidade-do-censo-diz-presidente-do-ibge/>. Acesso em: 12 jan. 2023.

CARAM, B.; RESENDE, T. Governo confirma que não haverá Censo do IBGE em 2021. **Folha de São Paulo**, São Paulo, 23 abr. 2021. Disponível em: <https://www1.folha.uol.com.br/mercado/2021/04/governo-confirma-que-nao-havera-censo-do-ibge-em-2021.shtml>. Acesso em: 14 set. 2022.

CARDOSO, F.; FALETO, E. **Dependência e Desenvolvimento na América Latina**, 4 ed. Rio de Janeiro: Zahar Editores, 1970.

CASTELLS, M. **Communication Power**. Oxford: Oxford University Press, 2009.

CASTELLS, M. **A Sociedade em Rede**. São Paulo: Paz e Terra, 2011.

CAVALLO, A.; RIGOBON, R. The Billion Prices Project: Using Online Prices for Measurement and Research. **Journal of Economic Perspectives**, v.30, n.2, p. 151-178, 2016.

CHESNAIS, F. **A Mundialização do Capital**. São Paulo: Xamã Editora, 1996.

CHESNAIS, F. **Finance capital today**: corporations and banks in the lasting global slump. Boston: Brill, 2016.

CHIAPELLO, E.; WALTER, C. The three ages of financial quantification: a conventionalist approach to the financiers' metrology. **Historical Social Research**. v. 41, p. 155–77, 2016.

CHILCOTE, R. Trotsky e a teoria latino-americana do desenvolvimento. **Crítica Marxista**, n.34, p.87-110, 2012.

CHOI, H.; VARIAN, H. Predicting the Present with Google Trends. **Google**, 2009. Disponível em: https://static.googleusercontent.com/media/www.google.com/pt-BR/googleblogs/pdfs/google_predicting_the_present.pdf. Acesso em: 8 ago. 2022.

CHRYSTAL, K.; MIZEN, P. Goodhart's Law: Its Origins, Meaning and Implications for Monetary Policy. In: Festschrift in honour of Charles Goodhart, 15-16 nov. 2001, London: **Bank of England**. Disponível em: https://cyberlibris.typepad.com/blog/files/Goodharts_Law.pdf. Acesso em: 20 out. 2023.

CIURIK, D. Economic Rents and the Contours of Conflict in the Data-Driven Economy. **Policy Brief**. Centre for International Governance Innovation. 25 June, 2020. Disponível em: DOI:10.2139/ssrn.3496025. Acesso em 14 set. 2021.

COPLAN. Grupo de Trabalho para estudar um plano de ampliação e atualização dos levantamentos estatísticos nacionais. **Relatório Final**. Decreto 1283, de 25 de Junho de 1962. Brasília: Comissão Nacional de Planejamento, 1962.

CORIAT, B. **El Taller y el Cronometro**. Ensayo sobre el taylorismo, el fordismo y la producción em masa. Mexico D.F.: Siglo Veintiuno Editores, 2000.

COULDRY, N.; MEJIAS, U. Datafication. **Internet Policy Review**, v.8, n.4, 2019. Disponível em: <https://doi.org/10.14763/2019.4.1428>. Acesso em: 15 out. 2021.

COULDRY, N.; MEJIAS, U. **The Costs of Connection**. Stanford: Stanford University Press, 2019a.

COUMET, E. La théorie du hasard est-elle née par hasard?. **Economies, sociétés, civilisations**, v. 25, n.3, p. 574-598, 1970.

CURVO, R. **Comparação entre as Grandes Crises Sistêmicas do Sistema Capitalista (1873, 1929 e 2008)** Tese (Doutorado em Políticas Públicas, Estratégias e Desenvolvimento) - Instituto de Economia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2021.

DAAS, P.; LOO M. Big Data (and official statistics).In: Meeting on the Management of Statistical Information Systems - **MSIS**, 2013. (Paris, France, and Bangkok, Thailand, 23-25 April 2013). Working Paper, 11, abr. 2023. Genebra: UNECE, 2013.

DAAS, P.; PUTS, M.; BUELENS, J.; VAN DEN HURK,P. Big Data as a Source for Official Statistics. **Journal of Official Statistics**, v.31, n.2, p.249-262, 2015. Disponível em: <https://doi.org/10.1515/jos-2015-0016>. Acesso em: 03 jan. 2021.

DANTZIG, T. **Number: The Language of Science**. Nova York: Pi Press, 2005.

DARDOT, P.; LAVAL, C. **A Nova Razão do Mundo**: Ensaio sobre a sociedade neoliberal. São Paulo: Boi Tempo, 2016.

DASTON, L. **Classical Probability in the Enlightenment**. Princeton: Princeton University Press, 1988.

DEAN, J. Big Data: Accumulation and Enclosure. **Theory & Event**,v.19, n.3, 2016.

DEKKER, A. Computer Methods in Population Census Data Processing. **International Statistical Review**, v. 62, n. 1, p. 55-70, Abr. 1994.

DELBAEN, F.; SCHACHERMAYER, W. A general version of the fundamental theorem of asset pricing. **Mathematische Annalen**. n.300, p. 463–520, 1994. Disponível em: <https://doi.org/10.1007/BF01450498>. Acesso em: 10 mai. 2022.

DEMCHENKO, Y.; GROSSO, P.; DE LAAT, C.; MEMBREY, P. Addressing big data issues in Scientific Data Infrastructure. In: **International Conference on Collaboration Technologies and Systems (CTS)**, San Diego, p. 48-55, 2013. Disponível em: <https://ieeexplore.ieee.org/document/6567203>. Acesso em: 18 ago. 2022.

DEPUTY National Statistician Heather Savory leaving the Office for National Statistics to lead the development of the UN Global Platform on Big Data. **Office for National**

Statiscs, New Port, 2019. Disponível em: <https://www.ons.gov.uk/news/news/deputynationalstatisticianheathersavoryleavingtheofficeforanationalstatisticstoleadthedevelopmentoftheunglobalplatformonbigdata>
Acesso em: 27 out. 2023.

DESA. **Handbook of Statistical Organization**, Third Edition. The Operation and Organization of a Statistical Agency. Department of Economic and Social Affairs. Statistics Division. New York: United Nations, 2007. Disponível em: https://unstats.un.org/unsd/publication/seriesf/seriesf_88e.pdf. Acesso em: 15 set. 2023.

DESA. **Handbook of Statistical Organization**, Fourth Edition. The Operation and Organization of a Statistical Agency. Department of Economic and Social Affairs. Statistics Division. New York: United Nations, 2022. Disponível em: <https://unstats.un.org/capacity-development/handbook/index.cshtml>. Acesso em: 15 set. 2023.

DESROSIÈRES, A. **The Politics of Large Numbers**: a history of statistical reasoning. Cambridge: Harvard University Press, 1998.

DESROSIÈRES, A. **Pour une sociologie historique de la quantification**: L'argument statistique I. Paris: Presses de l'Ecole des Mines, 2008.

DESROSIÈRES, A. **Gouverner par les nombres**: L'argument statistique II. Paris: Presses de l'Ecole des Mines, 2008a.

DESROSIÈRES, A. **Prouver et Gouverner**: Une analyse politique des statistiques publiques. Paris: La Découverte, 2014.

DGINS. **Scheveningen Memorandum**. Big Data and Official Statistics. Luxembourg: Directeurs Généraux des Instituts Nationaux Statistiques, 2013. Disponível em: <https://ec.europa.eu/eurostat/documents/13019146/13237859/Scheveningen-memorandum-27-09-13.pdf/2e730cdc-862f-4f27-bb43-2486c30298b6?t=1401195050000> Acesso em: 18 jul. 2022.

DOUILHET E.; KARANASIOU, A. Legal Responses to the Data in the Era of Big Data: The Paradigm Shift from Data Protection towards Data Ownership. In: SINGH M.; KUMAR D. (Org.). **Effective Big Data Management and Opportunities for Implementation**. Hershey: IGI Global, 2016.

DUNCAN, J.; SHELTON, W. **Revolution in United States Government statistics, 1926-1976**. Washington DC: US Department of Commerce, Office of Federal Statistical Policy and Standards, 1978.

DUNCAN, O. **Toward Social Reporting**: Next Steps. NY: Russell Sage Foundation, 1969.

DYER-WITHEFORD, N.; KJOSEN, N.; STEINHOFF, J. **Inhuman Power**: Artificial Intelligence and the Future of Capitalism. London: Pluto Press, 2019.

EC. **Public consultation on the data act**. Brussels: European Commission, 2021. Disponível em: <https://digital-strategy.ec.europa.eu/en/consultations/public-consultation-data-act>. Acesso em: 13 fev. 2023.

EC. **Data act**. Brussels: European Commission, 2022. Disponível em: ec.europa.eu/commission/presscorner/detail/en/ip_22_1113. Acesso em, 13 jul. 2023. ESGAB. **Eighth Annual Report** by the European Statistical Governance Advisory Board to the European Parliament and the Council on the implementation of the European Statistics Code of Practice by Eurostat and the European Statistical System as a whole. Luxembourg: European Union, 2016.

ELIAS, N. **A Sociedade de Corte**: Investigação sobre a sociologia da realeza e da aristocracia de corte. Rio de Janeiro: Jorge Zahar Ed., 2001.

ESPELAND, W.; STEVENS, M. Commensuration as a Social Process. **Annual Review of Sociology**, v. 24, p. 313-343, 1998.

ESS. Position paper on access to privately held data which are of public interest. **European Statistical System**, Position paper, nov. 2017. Luxembourg: Eurostat, 2017. Disponível em: <https://ec.europa.eu/eurostat/web/european-statistical-system/-/access-to-privately-held-data>. Acesso em: 23 nov. 2022.

ESS. Position paper on the future Data Act proposal. **European Statistical System**, Position paper, jun. 2021. Luxembourg: Eurostat, 2021. Disponível em: <https://ec.europa.eu/eurostat/documents/13019146/13405116/main+ESS+position+paper+on+future+Data+Act+proposal.pdf/37f3b5c7-abfd-5a05-6be2-fdc4b87ee7d2?t=1631695372906>. Acesso em: 23 nov. 2022.

EUROSTAT. Big data – an opportunity or a threat to official statistics? Prepared by Eurostat. In: **Conference of European Statisticians**. Sixty-second plenary session. Paris, 9-11 April 2014. New York: United Nations, 2014. Disponível em: https://unece.org/fileadmin/DAM/stats/documents/ece/ces/2014/32-Eurostat-Big_Data.pdf. Acesso em: 13 mai 2021.

EUROSTAT. **Empowering Society by Reusing Privately Held Data for Official Statistics**. A European Approach. 2022 edition. Luxembourg: European Union, 2022. Disponível em: <https://ec.europa.eu/eurostat/web/products-statistical-reports/-/ks-ft-22-004>. Acesso em: 20 fev. 2023.

EX-PRESIDENTES do IBGE divulgam carta contra cortes no Censo 2020. **Exame**, São Paulo, 15 de Julho de 2019. Disponível em: <https://exame.com/brasil/ex-presidentes-do-ibge-divulgam-carta-contr-cortes-no-censo-2020/>. Acesso em: 13 jun 2021.

FAORO, R. **Os Donos do Poder**: Formação do Patronato Político Brasileiro. Rio de Janeiro: Globo, 1989.

FAUSTO, Boris. **A Revolução de 1930**: história e historiografia. São Paulo: Companhia das Letras, 1997.

FERNANDES, F. **A revolução burguesa no Brasil**. Ensaio de interpretação sociológica. 5ª. Ed. São Paulo: Globo, 2006.

FERREIRA, M.; PINTO, S. A crise dos anos 1920 e a Revolução de 1930. In: FERREIRA, J.; DELGADO, L. (org.). **O tempo do Liberalismo Oligárquico: da Proclamação da República à Revolução de 1930 – Primeira República (1889- 1930)**, 1a ed. Rio de Janeiro: Civilização Brasileira, 2018.

FLORESCU, D.; KARLBERG, M.; REIS, F.; CASTILLO, P.; SKALIOTIS, M.; WIRTHMANN, A. Will 'big data' transform official statistics?. In: **International Conference on Big Data for Official Statistics**. Beijing, China, 28, 30 Oct 2014. Luxembourg: Eurostat, 2014.

FOLEY, D. Rethinking financial capitalism and the “information” economy. **Review of Radical Political Economics**, v.45, n.3, p.257–268, 2013. Disponível em: <https://doi.org/10.1177/0486613413487154>. Acesso em: 8 abr. 2022.

FOUCAULT, M. **As Palavras e as Coisas: Uma arqueologia das ciências humanas**. São Paulo: Martins Fontes, 1999.

FOUCAULT, M. **Segurança, Território, População: Curso dado no Collège de France (1977-1978)**. São Paulo: Martins Fontes, 2008.

FOUCAULT, M. **Nascimento da Biopolítica: Curso dado no Collège de France (1978-1979)**. São Paulo: Martins Fontes, 2008a.

FRANK, A. **Capitalism and Underdevelopment in Latin America**. New York: Monthly Review Press, 1969.

FRANK, A. **El Subdesarrollo del Desarrollo: Um ensayo autobiográfico**. Caracas: Editorial Nueva Sociedad, 1991.

FRANKLIN, J. **Pre-history of probability**. In: HÁJEK, A.; HITCHCOCK, C. (org.). *Oxford Handbook of Probability and Philosophy*. Oxford: Oxford University Press, 2016, p.33-49

FUKUDA-PARR, S. The human development paradigm: operationalizing sen's ideas on capabilities. **Feminist Economics**, v.9, n.2-3, p. 301-317, 2003. Disponível em: <https://doi.org/10.1080/1354570022000077980>. Acesso em: 20 jun. 2023.

FUKUDA-PARR, S.; YAMIN, A.; GREENSTEIN, J. The Power of Numbers: A Critical Review of Millennium Development Goal Targets for Human Development and Human Rights. **Journal of Human Development and Capabilities**, v.15, n 2-3, p.105-117. Disponível em: <https://doi.org/10.1080/19452829.2013.864622>. Acesso em: 20 jun. 2023.

FURTADO, C. **Subdesenvolvimento e Estagnação na América Latina**. 2a. ed. Rio de Janeiro: Civilização Brasileira, 1968.

FURTADO, C. **Desenvolvimento e Subdesenvolvimento**. Rio de Janeiro:

Contraponto, 2009.

GABAY, C.; ILCAN, S. The Affective Politics of the Sustainable Development Goals: Partnership, Capacity-Building, and Big Data. **Globalizations**, v. 14, n. 3, p.468-485, 2017.

GALTON, F. **Probability, the foundation of eugenics**. The Herbert Spencer lecture delivered on June 5, 1907 by Francis Galton. Oxford: Clarendon press, 1907.

GALTON, F. **Essays in Eugenics**. London: The Eugenics Education Society, 1909.
 GANDOMI, A.; HAIDER, M. Beyond the hype: Big data concepts, methods, and analytics. **International Journal of Information Management**, v.35, n.2, p.137-144, 2015. Disponível em: <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>. Acesso em: 3 abr. 2023.

GARCIA, D. Edmar Bacha, pai do real, defende mudança no questionário do Censo. **Folha de São Paulo**, São Paulo 29, jul. 2019. Disponível em: <https://www1.folha.uol.com.br/mercado/2019/07/edmar-bacha-pai-do-real-defende-mudanca-no-questionario-do-censo.shtml>. Acesso em: 14 mar. 2023.

GAWER, A. Digital platforms and ecosystems: remarks on the dominant organizational forms of the digital age. *Innovation*, 24(1), 110–124, 2022. Disponível em: <https://doi.org/10.1080/14479338.2021.1965888>. Acesso em: 27 jun. 2022.

GIL, T. Os Censos e a história do Brasil Colonial. 2022. **O Arquivo Nacional e a História Luso-Brasileira**. Publicado em 30 de dezembro de 2022. Disponível em http://historialuso.an.gov.br/index.php?option=com_content&view=article&id=6619:as-gentes-da-colonia-os-censos&catid=64&Itemid=504. Último Acesso em: 12 de Novembro de 2023.

GILLESPIE, T. The politics of ‘platforms’. *New Media & Society*, v.12, n.3, 347-364, 2010. Disponível em: <https://doi.org/10.1177/1461444809342738>. Acesso em: 07 fev. 2022.

GILLESPIE, T. The Relevance of Algorithms. In: Gillespie, T.; Boczkowski, P.; and Foot, K. (org.). **Media Technologies**. Cambridge: MIT Press, 2013.

GOMES, A. Estado Novo: debatendo nacionalismo, autoritarismo e populismo. In: Ferreira, J.; Delgado, L. (Org.). **O tempo do nacional-estatismo: do início da década de 1930 ao apogeu do Estado Novo: Segunda República (1930-1945)**. Coleção Brasil Republicano. Vol. 2. Rio de Janeiro: Civilização Brasileira, 2018.

GOOGLE. **Research with Google Flu Trends**, Google, 2010. Disponível em: http://blog.google.org/2010/05/researching-with-google-flu-trends_32.html?_gl=1*58xsaq*_ga*MTkwODMwNDExNi4xNzAxNzI5ODkx*_ga_6KQTTW2BX1*MTcwMTcyOTg5MS4xLjEuMTcwMTczMDE0My4wLjAuMA.#:~:text=google%20flu. Acesso em: 15 nov. 2023.

GPSDD. **Global Partnership for Sustainable Development Data: Data For Now** Concept Note. New York: GPSDD, 2019. Disponível em: <https://www.data4sdgs.org/index.php/initiatives/data-now>. Acesso em: 10 out. 2021.

GPSDD. **Global Partnership for Sustainable Development Data: Five-Years Strategy**. New York: GPSDD, 2020a. Disponível em: <https://www.data4sdgs.org/news/our-five-year-strategy-2019-2023>. Acesso em: 10 out. 2021.

GPSDD. **Global Partnership for Sustainable Development Data: Theory of Change**. New York: GPSDD, 2020b. Disponível em: <https://www.data4sdgs.org/resources/global-partnership-sustainable-development-datas-theory-change>. Acesso em: 10 out. 2021.

GPSDD. **Global Partnership for Sustainable Development Data: Five Years Report**. New York: GPSDD, 2020c. Disponível em: https://www.data4sdgs.org/sites/default/files/file_uploads/GPSDD_5YearReport_8.5x11_v9_WebReady.pdf. Acesso em: 10.out. 2021.

GPSDD. **Festival de Dados**. New York: GPSDD, 2023. Disponível em: https://www.data4sdgs.org/sites/default/files/file_uploads/Festival%20de%20Dados%20Supporter%20Info%20Pack.pdf. Acesso em: 04 fev. 2024.

GRABHER, G. Enclosure 4.0: Seizing Data, Selling Predictions, Scaling Platforms. **Sociologica**, v.14, n.3, p. 241–265, 2020. Disponível em: <https://doi.org/10.6092/issn.1971-8853/12107>. Acesso em: 08 set. 2022.

GRABOIS, M. **O IBGE e a Ação Governamental de João Goulart**. Tese (Doutorado em História) — Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2021.

GRADILLAS, M.; THOMAS, L. Distinguishing digitization and digitalization: A systematic review and conceptual framework. **Journal of Product Innovation Management**, p. 1-32, 2023. Disponível em: <https://doi.org/10.1111/jpim.12690>. Acesso em: 03 jan. 2024.

GRAMSCI, A. **Cadernos do Cárcere**. Volume 3: Maquiavel. Notas sobre o Estado e a política. Rio de Janeiro: Civilização Brasileira, 2002.

GROMMÉ, F.; CAKICI, B.; TAKALA, V. Statistician subjects: differentiating and defending. In: RUPPERT, E.; SCHEEL, S. (Org.). **Data Practices: Making up a European People**. London: Goldsmiths Press, 2021, p.237–268.

GUERRA, S. **Carta da Presidente**. Destinatários: lista de servidores do IBGE. Rio de Janeiro, 08 abr. 2021. 1 mensagem eletrônica.

GUIMARÃES, A. Subsídios para a reformulação do sistema estatístico brasileiro. In: COPLAN. Relatório Final. Comissão Nacional de Planejamento. **Grupo de Trabalho**. Decreto 1283, de 25 de Junho de 1962. Brasília: COPLAN, 1962.

HACKING, I. **The Taming of Chance**. New York: Cambridge University Press, 2004.

HACKING, I. **The Emergence of Probability**: A philosophical study of early ideas about probability, induction and statistical inference. New York: Cambridge University Press, 2007.

HAND, D. Official Statistics in the New Data Ecosystem. In: **New Techniques and Technologies in Statistics conference**, Brussels, March 10–12, 2015. Disponível em: <https://croslegacy.ec.europa.eu/system/files/Presentation%20S20AP2%20%20H and%20-%20Slides%20NTTS%202015.pdf>. Acesso em: 3 jun. 2022.

HARVEY, D. **O Neoliberalismo**: história e implicações. São Paulo: Loyola, 2008.

HARVEY, D. **Condição Pós-Moderna**: Uma pesquisa sobre as origens da mudança cultural. São Paulo: Loyola, 2013.

HAYASHI, T; ISHIMURA, G.; OHSAWA, Y. Structural Characteristics of Stakeholder Relationships and Value Chain Network in Data Exchange Ecosystem. **IEEE Access**, vol. 9, p. 52266-52276, 2021, doi: 10.1109/ACCESS.2021.3070017.

HEIDEGGER, M. **Ensaio e conferências**. Petrópolis: Editora Vozes, 2002.

HENDRIKSE, R.; ADRIAANS, I.; KLINGE, T.; FERNANDEZ, R. The Big Techification of Everything. **Science as Culture**, v.31, n.1, p. 59–71, 2022. Disponível em: <https://doi.org/10.1080/09505431.2021.1984423>. Acesso em 23 out. 2023.

HOBBSAWM, E. **Nações e Nacionalismo desde 1870**: Programa, mito e realidade. 4a. Edição. São Paulo: Paz e Terra, 2004.

HOLLY, D. La Commission de statistique de l'ONU. In: Beaud JP.; Prévost JG. (org.). **L'ère du chiffre**: Systèmes statistiques et traditions nationales. Laurier: Presses de l'Université du Québec, 2000.

HUB. Hub Regional de Big Data no Brasil. Rio de Janeiro: Hub Regional de Big Data no Brasil, 2021. 1 vídeo (115 min). Disponível em: <https://www.youtube.com/watch?v=P9TywXA65dE>. Acesso em: 12 jan. 2022.

HUB. **Big data for statistics in Latin America and the Caribbean**. Rio de Janeiro: Hub Regional de Big Data no Brasil, 2022. Disponível em: [//storymaps.arcgis.com/stories/3e42f28257004e4b98b0606518185656](https://storymaps.arcgis.com/stories/3e42f28257004e4b98b0606518185656). Acesso em: 04 fev. 2023.

HUB. Webinar series Road to Punta Del Este Data Festival: Data Science in Official Statistics. Rio de Janeiro: Hub Regional de Big Data no Brasil, 2023. 1 vídeo (63 min). Disponível em: https://www.youtube.com/watch?v=8O_twis2Qd0. Acesso em: 23 out. 2023.

HUB. Webinar: Mobile Phone Data Access for Official Statistics. Tourism Statistics. Rio de Janeiro: Hub Regional de Big Data no Brasil, 2023a. 1 vídeo (63 min). Disponível em: <https://www.youtube.com/watch?v=WmwnHiT7T90>. Acesso em: 16 fev. 2023.

HUB. Webinar: Using Mobile Phone Data for Official Statistics on Population. Rio de Janeiro: Hub Regional de Big Data no Brasil, 2023b. 1 vídeo (68 min). Disponível em: <https://www.youtube.com/watch?v=Z1QEh3UdS8U>. Acesso em: 13 mar. 2023.

HUB. Webinar: Mobile Phone Data Access for Official Statistics. Rio de Janeiro: Hub Regional de Big Data no Brasil, 2023c. 1 vídeo (62 min). Disponível em: <https://youtu.be/pKOi0D46aD4>. Acesso em: 2 ago. 2023.

HUB. Using Mobile Phone Data on Trending Topics. Rio de Janeiro: Hub Regional de Big Data no Brasil, 2023d. 1 vídeo (52 min). Disponível em: https://www.youtube.com/watch?v=_T8d775ls50. Acesso em: 18 ago. 2023.

IANNI, O. **Estado e Planejamento Econômico no Brasil (1930-1970)**. Rio de Janeiro: Civilização Brasileira, 1971.

IBGE: Metade dos jovens do País ainda é subnutrida. **O Globo**, Rio de Janeiro, 17 dez. 1979. Disponível em: <https://oglobo.globo.com/acervo/>. Acesso em: 04 mar. 2023.

IBGE. **Plano Estratégico 2017-2027**. Rio de Janeiro: IBGE, 2017.

IBGE. **Boletim de Serviço 2896**. Publicado em 10/11/2017. Rio de Janeiro: IBGE. 2017a.

IBGE. **Plano Diretor de Tecnologia da Informação e Comunicação 2019/2020**. Rio de Janeiro: IBGE, 2019.

IBGE, **Estratégia Geral de Tecnologia de Informação e Comunicação 2019/2020**. Rio de Janeiro: 2019a.

IBGE. **Relatório de Gestão 2019**. Rio de Janeiro: IBGE, 2020.

IBGE faz parceria com Receita Federal e BID para modernizar contas nacionais. **Agência IBGE Notícias**, Rio de Janeiro, 19 mar. 2021. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/30347-ibge-faz-parceria-com-receita-federal-e-bid-para-modernizar-contas>. Acesso em: 08 fev. 2022.

IBGE. Boletim de Serviço 3074. Despacho – DSP.CD-IBGE no. 10/2021, de 09 de Abril de 2021. **Acordo internacional** a ser firmado com o Departamento de Assuntos Econômicos e Sociais da UNDESA. 2021a.

IBGE sai em defesa do orçamento do Censo 2021. **Agência IBGE Notícias**, Rio de Janeiro, 23 mar. 2021b. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/30350-ibge-sai-em-defesa-do-orcamento-do-censo-2021> Acesso em: 08 fev. 2022.

IRANI, L. **Chasing Innovation: Making Entrepreneurial Citizens in Modern India**. Princeton, NJ: Princeton University Press, 2019.

JONES, K.; MILLARD, F.; TWIGG, L. **Government and Information In Public Policy under Thatcher**. New York: Macmillan Education, 1990.

JOQUE, J. **Revolutionary Mathematics**: Artificial Intelligence, Statistics and the Logic of Capitalism. London: Verso, 2022.

KHAN, N.; ALSAQER, M; SALEHIAN, S. The 10 Vs, Issues and Challenges of Big Data. In: **Proceedings of the 2018 international conference on big data and education**. p. 52-56. 2018.

KITCHIN, R. **The Data Revolution**. Big Data, Open Data, Data Infrastructures & their consequences. Los Angeles: Sage, 2014.

KITCHIN, R. The opportunities, challenges and risks of big data for official statistics. **Statistical Journal of the IAOS**, v. 31, p. 471–481, 2015.

KITCHIN, R. Thinking critically about and researching algorithms. **Information, Communication & Society**, v.20, n.1, p.14-29, 2016. DOI: 10.1080/1369118X.2016.1154087

KLUGER, E. Análise de correspondências múltiplas. **BIB**, v. 86, n. 2, p. 68–97, 2018.
KOEHLER, G. The 2030 Agenda and eradicating poverty: New horizons for global social policy. **Global Social Policy**, v.17, n.2, p.210-216, 2017.

KOVACEVIC, M. Review of HDI critiques and potential improvements. **Human Development Research**, paper 33, p 1-44, 2010.

KUERBIS, B.; MUELLER, M. Exploring the role of data enclosure in the digital political economy. **2022 Workshop on the Economics of Information Security**. 2022.

LAND, K. Social Indicators. **Annual Review of Sociology**, v.9, p.1–26. 1983. Disponível em: <http://www.jstor.org/stable/2946054>. Acesso em: 15 jun 2023.

LANDENFELD, S. Uses of Big Data for Official Statistics: Privacy, Incentives, Statistical Challenges, and Other Issues. In: **International Conference on Big Data for Official Statistics**. Beijing, China, 28-30 Oct. 2014.

LANEY, D. 3D Data Management: Controlling data volume, velocity and variety. **Meta Group, File 949**. 2001. Disponível em: <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>. Acesso em: 14 set. 2023.

LAZER, D.; KENNEDY, R.; KING, G.; VESPIGNANI, A. The Parable of Google Flu: Traps in Big Data Analysis. **Science**. V. 343, n. 14, mar. 2014.

LAZER, D.; KENNEDY, R. What We Can Learn From the Epic Failure of Google Flu Trends. 2015. **Wired**. Disponível em: <https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/>. Acesso em: 6 set. 2022.

LEHDONVIRTA, V. Algorithms that Divide and Unite: Delocalization, Identity and Collective Action in 'Microwork'. In: Flecker, J. (Ed.). *Space, Place and Global Digital Work. Dynamics of Virtual Work*. London: Palgrave, 2016.

LEÓN, L. Property regimes and the commodification of geographic information: An examination of Google Street View **Big Data & Society**, v.3, n.2, p.1-13, 2016. Disponível em: <https://doi.org/10.1177/2053951716637885>. Acesso em: 3 mar. 2023.

LEOPOLDI, M. A economia política do primeiro governo Vargas (1930-1945): a política econômica em tempos de turbulência. In: FERREIRA, J.; DELGADO, L. (org.). **O tempo do nacional-estatismo: do início da década de 1930 ao apogeu do Estado Novo: Segunda República (1930-1945)**. Rio de Janeiro: Civilização Brasileira, 2018.

LOUREIRO, F. Making the Alliance for Progress Serve the Few: U.S. Economic Aid to Cold War Brazil (1961–1964). **Journal of Cold War Studies**, v. 25, n. 1, p. 168–207. 2023. Disponível em: https://doi.org/10.1162/jcws_a_01125. Acesso em: 20 nov. 2023.

LUCAS, R.; RAPPING, L. Price Expectations and the Phillips Curve. **American Economic Review**, v.59, n.3, p. 342-350, 1969.

MACFEELY, S. The Continuing Evolution of Official Statistics: Some Challenges and Opportunities **Journal of Official Statistics**, v. 32, n. 4, p. 789–810, 2016. Disponível em: <http://dx.doi.org/10.1515/JOS-2016-0041>. Acesso em: 13 jan. 2021.

MACFEELY, S.; BARNAT, N. Statistical capacity building for sustainable development: Developing the fundamental pillars necessary for modern national statistical systems. **Statistical Journal of the IAOS**, v.33, p.4, 895–909, 2017.

MACFEELY, S. Big Data and Official Statistics. In: STRYDOM, S.; STRYDOM, M. (org.). **Big Data Governance and Perspectives in Knowledge Management**. IGI Global, p. 25-54. 2019.

MACKENZIE, D.; SPEARS, T. The formula that killed Wall Street: The Gaussian copula and modelling practices in investment banking. **Social Studies of Science**, v.44, p. 393-, 2014.

MACKENZIE, A. Programming Subjects in the Regime of Anticipation: Software Studies and Subjectivity. **Subjectivity**, v. 6, n.4, p. 391–405, 2013.

MANOVICH, L. Database as a symbolic form. **Convergence**, v. 5, n.2, p. 80-99, 1999.

MARINI, R. Dialéctica de la dependência. In: MARTINS, C. (org.). **América Latina, dependencia y globalización**. Bogotá: CLACSO y Siglo del Hombre Editores, p. 107-149, 2008.

MARX, K. **Grundrisse**. Manuscritos econômicos de 1857-1858: Esboços da crítica da economia política. São Paulo: Boitempo, 2011.

MARX, K. **O 18 Brumário de Luís Bonaparte**. São Paulo: Boitempo, 2011a.

MARX, K. **O Capital**: Livro I. São Paulo: Boitempo, 2013.

MATTHESS, M.; KUNKEL, S. Structural change and digitalization in developing countries: Conceptually linking the two transformations. *Technology in Society*, 63, 2020. Disponível em: <https://doi.org/10.1016/j.techsoc.2020.101428>. Acesso em: 20 ago. 2022.

MAYER-SCHÖNBERGER, V.; CUKIER, K. **Big Data**: A Revolution That Will Transform How We Live, Work and Think. London: John Murray, 2013.

MCCOURT, W.; MINOGUE, M. **The Internationalization of Public management**: Reinventing the Third World State. Cheltenham: Edward Elgar Publishing, 2001.

MCLAUGHLIN, K.; OSBORNE, P.; Ferlie, E. **New Public Management**: Current trend and future prospects. London: Routledge, 2002.

MERRY, S. Measuring the World: indicators, Human Rights and Global Governance. In: **American Society of International Law**. Panel on Indicators, 2009. Disponível em: https://www.researchgate.net/publication/228762025_Measuring_the_World_Indicators_Human_Rights_and_Global_Governance. Acesso em: 10 ago. 2023.

MIYAZAKI, S. Algorhythmic: Understanding Micro-Temporality in Computational Cultures. **Computational Culture**, v.2, Set. 2012. Disponível em: <http://computationalculture.net/algorhythmic-understanding-micro-temporality-in-computational-cultures/>. Acesso em: 10 out. 2021.

MLOPS. UN Global Platform – Mark Craddock. MLOps Community. MLOps Meetup #42, 2020. 1 vídeo (58 min). Disponível em: https://www.youtube.com/watch?v=6oFCqGE_2-U. Acesso em: 14 out. 2023.

MOLLO, M; DOURADO, F.; PARANÁ, E. Financialisation as the development of fictitious capital in developing and developed economies. **Cambridge Journal of Economics**, v.46, n.5, 955–976, 2022.

MOROZOV, Evgeny. **Big tech**: a ascensão dos dados e a morte da política. São Paulo: Ubu Editora, 2020.

MUSIANI, F. Governance by algorithms. **Internet Policy Review**, v.2, n.3, 2013. Disponível em: <https://doi.org/10.14763/2013.3.188>. Acesso em: 14 nov. 2022.

MUSSER, J.; O'REILLY, T. **Web 2.0**: Principles and Best Practices. Sebastopol: O'Reilly Media, 2007.

MUTH, J. Rational Expectations and the Theory of Price Movements. **Econometrica**, v.29, n. 3, p. 315-335, 1961.

NATHAN, M.; ROSSO, A.; GATTEN, T.; MAJMUDAR, P. ; Mitchell, A. **Measuring the uk's digital economy with big data**. London: National Institute of Economic and Social Research, 2013.

NETO, J.; FORTE, C. O Sistema de Contas Nacionais: evolução histórica e implantação no Brasil. **Revista Econômica**, v.18, n1, jun. 2016.

NIELSEN. **NIELSEN IQ**, 2023. Disponível em: <https://nielseniq.com/global/en/>. Acesso em: 7 mai. 2023.

NRC. **Frontiers in Massive Data Analysis**. National Reserach Council of the National Academies. Washington, D.C.: The National Academies Press, 2013.

OLINTO, R. **Carta aos Ibgeanos**. Destinatários: lista de servidores do IBGE. Rio de Janeiro, 01 fev. 2019. 1 mensagem eletrônica.

OLIVER, D.; DREWRY, G. **Public Service Reforms: Issues of Accountability and Public Law**. Pinter, Guildford:1996.

OLSON, D.; KONTY, K.; PALADINI, M.; VIBOUD, C.; Simonsen, L. Reassessing Google Flu Trends Data for Detection of Seasonal and Pandemic Influenza: A Comparative Epidemiological Study at Three Geographic Scales. **PLoS Comput Biology**, v.9, n.10, 2013. Disponível em: <https://doi.org/10.1371/journal.pcbi.1003256>. Acesso em: 13 set. 2023.

ONU. Lessons learned from MDG monitoring from a statistical perspective. **Report of the Task Team** on Lessons Learned from MDG Monitoring of the IAEG-MDG. New York: United Nations, 2013. Disponível em: [https://unstats.un.org/unsd/broaderprogress/pdf/Lesson%20Learned%20from%20MDG%20Monitoring_2013-03-22%20\(IAEG\).pdf](https://unstats.un.org/unsd/broaderprogress/pdf/Lesson%20Learned%20from%20MDG%20Monitoring_2013-03-22%20(IAEG).pdf). Acesso em: 30 ago. 2023.

ORANGE. V-Traffic becomes the first service to incorporate Orange's Floating Mobile Data as traffic information source, **Press Release**, 19 January. Disponível em: <https://www.orange-business.com/en/press/v-traffic-becomes-the-first-service-to-incorporate-orange-s-floating-mobile-data-as-traffic> Acesso em: 14 out. 2023.

OSTROM, E.; HESS, C. **Understanding Knowledge as a Commons: From Theory to Practice**. Cambridge: MIT Press. 2007.

OTERO, H. Socio-history of statistics on Latin America: a review. **Histoire & Mesure**, v. XXXIII, n.2, p.13-32, 2018.

PACHECO, L. Servidores do IBGE entregam cargos contra forma como direção conduz Censo. **Correio Braziliense**, Brasília, 07, jun. 2019. Disponível em: <https://blogs.correiobraziliense.com.br/papodeconcurseiro/servidores-do-ibge-entregam-cargos-contra-forma-como-direcao-conduz-censo/>. Acesso em: 13 ago. 2023.

PANTELIS, K.; AIJA, L. Understanding the value of (big) data. In: **IEEE International Conference on Big Data**, 2013.

PARANÁ, E. **Dinheiro e Poder Social**: Um estudo sobre o Bitcoin. 2018. Tese (Doutorado em Sociologia). Programa de Pós Graduação em Sociologia, Universidade de Brasília, Brasília, 2018.

PARANÁ, E. **Digitalized Finance**: Financial Capitalism and Informational Revolution. Boston: Brill, 2019.

PARANÁ, E. O estágio atual do capitalismo e sua crise (cap. 5). In: TRIGUEIRO, M. **A legitimação no estágio atual do capitalismo**. São Paulo: Contracorrente, 2022, pp. 123-187.

PARANÁ, E. Crise do neoliberalismo ou neoliberalismo de crise? Hipóteses sobre a estranha resiliência da economia neoliberal. In: CASTELO-BRANCO, F.; YAZBEK, A.; CASARA, R. (org.). **Pensar o neoliberalismo**: economia, política, epistemologia. Recife: Ruptura, 2023, p. 215-229.

PARANÁ, E. Financialised Digitalisation, Digitalised Financialisation: the inseparability between technological domination and financial hegemony in contemporary capitalism. In: GATAULINA, I.; KANGAS, A.; POUTANEN, M.; RAJALA, A. VENTOVIRTA, H. (org.). **Rethorising Capitalism**. Tampere University Press, 2024 (no prelo).

PARIS21. **A Road map for a country-led Data Revolution**. Paris: OECD Publishing, 2015. Disponível em: <https://doi.org/10.1787/9789264234703-en>. Acesso em: 20 out. 2021.

PASQUALE, F. (2015). **The black box society**: The secret algorithms that control money and information. Cambridge, MA: Harvard University Press.

PASQUINELLI, M.; JOLER, V. The Nooscope manifested: AI as instrument of knowledge extractivism. **AI & Society**, v.36, p. 1263–1280, 2021.

PENHA, E. **A criação do IBGE no contexto de centralização do Estado Novo**. Rio de Janeiro: CDDI – IBGE, 1993.

PEREIRA, L. Vers l'Atlantide : le Sondage national par échantillon de domiciles (PNAD) et l'autonomisation de la statistique publique au Brésil. **Statistique et société**, 11 fev. 2023. Disponível em: <http://journals.openedition.org/statsoc/721>. Acesso em: 20 nov. 2023.

PETRAKOS, M.; SCIADAS, G.; STAVROPOULOS, P. **Accreditation procedure for statistical data from non-official sources**. Study commissioned by Eurostat. Luxembourg: Eurostat, 2013.

DUFFY, B., POELL, T., NIEBORG, D. Platform practices in the cultural industries: Creativity, labor, and citizenship. **Social Media + Society**, v.5, n.4, p.1-8, 2019. Disponível em: <https://doi.org/10.1177/2056305119879672>. Acesso em: 08 mar. 2022.

POLANYI, K. **A grande transformação**: As origens políticas e econômicas de nossa época. Rio de Janeiro: Contraponto, 2021.

PORTER, T. **The Rise of Statistical Thinking: 1820-1900**. Princeton, New Jersey: Princeton University Press, 1986.

PORTER, T. **Trust in Numbers: The Pursuit of objectivity in science and public life**. Princeton, New Jersey: Princeton University Press, 1995.

PORTER, T. The flight of the indicator. In: ROTTENBURG, R.; MERRY, S.; PARK, S.; MUGLER, J. (Org.). **The World of Indicators: The Making of Governmental Knowledge through Quantification**. Cambridge: Cambridge University Press, 2015.p. 34-55.

POULANTZAS, N. **O Estado, O Poder, O Socialismo**. 1.ed. Rio de Janeiro / São Paulo: Paz e Terra, 2015.

POWER, M. The Audit Society – Second Thoughts. **International Journal of Auditing**, v.4, p. 111-119, 2000.

PRADO, L.; LEOPOLDI, M. O fim do desenvolvimentismo: o governo Sarney e a transição do modelo econômico brasileiro. In: Ferreira J.; Delgado, L. (Org.). **O tempo da Nova República**. Da transição democrática à crise política de 2016. Quinta República (1985-2016). 2018. p. 69-108.

PREBISH, R. O desenvolvimento econômico da América Latina e seus principais problemas. **Revista Brasileira de Economia**, v. 3, n. 3, p. 47–111, 1949.

PRICESTATS **History**. 2023. PriceStats. Disponível em: <https://www.pricestats.com/about-us/history>. 2023. Acesso em: 14 nov. 2023.

RADERMACHER, W. Rethinking the future. **Eurostat Info**. Issue February, 2009.

RADERMACHER, W. The European Statistics Code of Practice as a Pillar to Strengthen Public Trust and Enhance Quality in Official Statistics. **Journal of the Statistical and Social Inquiry Society of Ireland**, v.43, p. 27-33, 2013.

RADERMACHER, W. **Oficial Statistics 4.0**. Cham: Springer Nature, 2020.

RANDERAAD, N. **States and statistics in the nineteenth century: Europe by numbers**. Manchester: Manchester University Press, 2010.

RANGEL, I. Ciclo, tecnologia e crescimento (1969-1981). In: BENJAMIN, César (Org). **Ignácio Rangel: obras reunidas**. Rio de Janeiro: Contraponto, 2005. v. 2, p. 255-408.

RBE. Isaac Kerstenetzky: o reconhecimento da RBE. **Revista Brasileira de Economia**, v.45, n.3, p. 333-482, 1991.

REIMSBACH-KOUNATZE, C. The Proliferation of Big Data and Implications for Official Statistics and Statistical Agencies: A Preliminary Analysis. **OECD Digital Economy Papers**, n. 245. Paris: OECD Publishing, 2015. Disponível em: <https://doi.org/10.1787/5js7t9wqzvg8-en>. Acesso em: 13 ago 2023.

RESENDE, M. O Processo político na primeira república e o liberalismo oligárquico. In: FERREIRA, J.; DELGADO, L. (Org.) **O tempo do liberalismo oligárquico**: da Proclamação da República à Revolução de 1930 – Primeira República (1889- 1930). Coleção O Brasil Republicano. Vol. 1. Rio de Janeiro: Civilização Brasileira, 2018.

RICCIATO, F.; WIRTHMANN, A.; HAHN, M. Trusted Smart Statistics: How new data will change official statistics. **Data & Policy**, v. 2, n. e7, 2020. doi:10.1017/dap.2020.7

RICHTER, H.; SLOWINSKI, P. The Data Sharing Economy: On the Emergence of New Intermediaries. **IIC**, v. 50, p. 4–29, 2019. Disponível em: <https://doi.org/10.1007/s40319-018-00777-7>. Acesso em: 28 out. 2021.

RIKAP, C. **Capitalism, Power and Inovation**: Intellectual Monopoly Capitalism Uncovered. London, New York: Routledge, 2021.

RIKAP, C.; LUNDEVALL, B. Big tech, knowledge predation and the implications for development. **Innovation and Development**, v.12, n.3, p.389-416. 2022. DOI: 10.1080/2157930X.2020.1855825

ROBIN, N.; KLEIN, T.; JUTTING, J. Public-Private Partnerships for Statistics: Lessons Learned, Future Steps. **OECD Working Paper 27 - Paris 21**, 2016.

ROBSBAWM, E. Nações e Nacionalismo desde 1780: Programa, mito e realidade. São Paulo: Paz e Terra, 1990.

ROSA, B. Guedes quer vender prédio do IBGE para fazer Censo e sugere simplificar pesquisa. **O Globo**, Rio de Janeiro, 22 fev. 2019. Disponível em: <https://oglobo.globo.com/economia/guedes-quer-vender-predio-do-ibge-para-fazer-censo-sugere-simplificar-pesquisa-23473491>. Acesso em: 14 jan. 2023.

ROTTA, T.; PARANÁ, E. Bitcoin as a digital commodity. **New Political Economy**, v.27, n.6, p. 1046–1061, 2022.

ROTTA, T.; TEIXEIRA, R. The Commodification of Knowledge and Information. **Greenwich Papers in Political Economy**, Working Paper, n.60, 2018.

ROTTENBURG, R.; MERRY, S. A world of indicators: The making of governmental knowledge through quantification. In: ROTTENBURG, R.; MERRY, S.; PARK, S.; MUGLER, J. (Org.). **The World of Indicators**: The Making of Governmental Knowledge through Quantification. Cambridge: Cambridge University Press, 2015. p.1-33

RUPPERT, E.; SCHEEL, S. Introduction. In: Ruppert, E.; Scheel, S. (Org.). **Data Practices**: Making up a European People. London: Goldsmiths Press, 2021, pp.1–28.

RUVROY, A.; BERNS, T. Algorithmic governmentality and prospects of emancipation. Disparateness as a precondition for individuation through relationships? **Réseaux**, n. 177, p. 163-196, 2013. Disponível em: URL: <https://www.cairn-int.info/journal-reseaux-2013-1-page-163.htm>. Acesso em: 20 fev. 2021.

SADOWSKI, J. When data is capital: Datafication, accumulation, and extraction. **Big Data & Society**, v.6, n.1, 2019. Disponível em: <https://doi.org/10.1177/2053951718820549>. Acesso em: 23 jan. 2022.

SADOWSKI, J. The Internet of Landlords: Digital Platforms and New Mechanisms of Rentier Capitalism. **Antipode**, v. 52, n.2, p.562-580, 2020. Disponível em: <https://doi.org/10.1111/anti.12595>. Acesso em: 19 abr. 2022.

SAIDEL, M. **Neoliberalism Reloaded: Authoritarian Governmentality and the Rise of the Radical Right**. Berlin: Walter de Gruyter GmbH & Co KG, 2023.

SALLUM, B. O Brasil sob Cardoso: neoliberalismo e desenvolvimentismo. **Revista de Sociologia da USP**, v.11, n.2, p.23-47, out. 1999.

SALLUM, B.; GOULART, J. O Estado brasileiro contemporâneo: liberalização econômica, política e sociedade nos governos FHC e Lula. **Revista de Sociologia e Política**, v. 24, n.60, p.115-135, dez. 2016.

SAMUELSON, P. The Pure Theory of Public Expenditure. **The Review of Economics and Statistics**, v. 36, n. 4, nov. 1954, p. 387-389.

SANDBROOK, R. Globalization and the Limits of Neoliberal Development Doctrine. **Third World Quarterly**, v. 21, n. 6, p. 1071-1080, dez. 2000.

SANTOS, T. **Imperialismo y Dependencia**. Caracas: Biblioteca Ayacucho, 2011.

SCHOMM, F.; STAHL, F.; VOSSEN, G. Marketplaces for data: an initial survey. **SIGMOD Rec.** v. 42, n.1, p. 15–26, mar. 2013. Disponível em: <https://doi.org/10.1145/2481528.2481532>. Acesso em: 27 out. 2021.

SCHROECK, M. **Analytics: The real-world use of big data**. Sommers: IBM, 2012. Disponível em: <https://www.bdvc.nl/images/Rapporten/GBE03519USEN.PDF> Acesso em: 15 set. 2023.

SCHWARTZMAN, S. **O Presente e o Futuro do IBGE**. Rio de Janeiro: IBGE, 1995.

SCHWARTZMAN, S. **O IBGE, quatro anos depois**. Documento preparado para discussão com as autoridades federais. 1998. Disponível em: <https://www.schwartzman.org.br/simon/4anos.htm>. Acesso em: 25 nov. 2023.

SCHWEINFEST, S. JANSEN, R. Data Science and Official Statistics: Toward a new data culture. **Harvard Data Science Review**, Issue 3.4, 2021.

SCOTT, L; VARIAN H. Predicting the Present with Bayesian Structural Time Series. Jun. 2013. **SSRN**. Disponível em: <http://dx.doi.org/10.2139/ssrn.2304426>. Acesso em: 3 out. 2022.

SEN, A. **Development as Freedom**. New York: Alfred Knopf, 1999.

SENRA, N. **História das Estatísticas Brasileiras**. Volume 1. Estatísticas Desejadas (1822-1889). Rio de Janeiro: IBGE, 2006.

SENRA, N. **História das Estatísticas Brasileiras**. Volume 2. Estatísticas Legalizadas (1889-1936). Rio de Janeiro: IBGE, 2008.

SENRA, N. **História das Estatísticas Brasileiras**. Volume 3. Estatísticas Organizadas (1936-1972). Rio de Janeiro: IBGE, 2009.

SENRA, N.; FONSECA, S.; MILIONS, T. **O Desafio de Retratar o País**. Entrevistas com Presidentes do IBGE no período de 1985 a 2015. Rio de Janeiro: IBGE, 2016.

SENRA, N. **Tradição e Renovação**. Uma síntese da história do IBGE. 2ª. Edição. Rio de Janeiro: IBGE, 2017.

SHAFER, G. The Unity and Diversity of Probability. **Statistical Science**, Nov., 1990, Vol. 5, No. 4, pp. 435-444.

SHEYVENS, R.; BANKS, G.; HUGHES, E. The Private Sector and the SDGs: The Need to Move Beyond 'Business as Usual'. **Sustainable Development**, v.24, p.371-382, 2016.

SHILLER, R. Narrative Economics. Discussion Paper no. 2069. In: **Presidential address delivered at the 129th annual meeting of the American Economic Association**. Cowles Foundation, 2017.

SILVA, U. O conservadorismo no liberal-desenvolvimentismo de Roberto Campos. **Revista do Instituto de Estudos Brasileiros**, n. 79, p. 18-35, ago. 2021.

SILVA, L.; PARANÁ, E. Da dialética do desenvolvimento à dialética da dependência: o contínuo no descontínuo. **Revista de Ciências Sociais**, n. 47, p. 191-208, 2017.

SIMONSEN: estudo do Bird deveria ser mais crítico. **O Globo**, Rio de Janeiro, 24 out. 1978. Disponível em: <https://oglobo.globo.com/acervo/>. Acesso em: 04 mar. 2023.

SINGER, A., 2012. **Os sentidos do lulismo**: reforma gradual e pacto conservador. São Paulo: Companhia das Letras.

SINGER, A. A falta de base política para o ensaio desenvolvimentista. In: Singer, A.; Loureiro, I. (Org.). **As contradições do lulismo** – a que ponto chegamos? São Paulo: Boitempo, 2016.

SINGER, A. The Failure of Dilma Rousseff's Developmentalist Experiment. A Class Analysis. **Latin American Perspectives**, issue 230, v. 47, n. 1, p. 152–168, Jan. 2020.

SKALIOTIS, M. Official Statistics in the era of Ubiquitous connectivity and Pervasive Technologies. **Eurostat Working Paper**. Luxembourg: Eurostat, 2010. Disponível em: https://cros-legacy.ec.europa.eu/system/files/skaliotis_paper_prague.pdf. Acesso em: 1 out. 2023.

SKIENA, S. **The Data Science Design Manual**. Cham, Switzerland: Springer, 2017.
Rubin, E. The Statistical Approach of Karl Marx. **The American Statistician**, v.22, n.2, p.31-33, 1968.

SODRÉ, N. **Formação Histórica do Brasil**. 12 ed. Rio de Janeiro: Bertrand, 1987.

SOGETI. **Report on the review of ethics related to big data**. Services concerning ethical, communicational, skills issues and methodological cooperation related to the use of Big Data in European statistics. Task 1: Ethical review. Deliverable D.1.2. Version 4. Luxembourg: Eurostat, 2017.

SRNICEK, N. **Platform Capitalism**. Cambridge: Polity Press, 2017.

STAHL, F.; SCHOMM, F.; VOSSEN, G. The data marketplace survey revisited. **ERCIS Working Papers**, n. 18, 2014.

STEFANO, F. Ex-presidente do IBGE irá assumir cargo no BID em Washington. **Exame**, São Paulo, 13 set. 2021. Disponível em: <https://exame.com/economia/ex-presidente-do-ibge-ira-assumir-cargo-no-bid-em-washington/>. Acesso em: 10 mai. 2023.

STEINER, C. **Automate this: How algorithms took over our markets, our jobs, and the world**. New York, NY: Portfolio, 2012.

STIEGLER, S. **The History of Statistics: The Measurement of Uncertainty before 1900**. Cambridge: Harvard University Press, 1986.

STORYBOOK. Positium – Financial and Assets. **Storeybook**. Disponível em: <https://ssb.ee/en/10883762-POSITIUM-OU/financial-assets-forecasts>. 2024. Acesso em: 04 fev. 2024.

STREEK, W. **Tempo Comprado: a crise adiada do capitalismo democrático**. Lisboa: Conjuntura Actual Editora, 2013.

STRUIJS, P.; BRAAKSMA, B.; DAAS, P. Official statistics and Big Data. **Big Data & Society**, v. 1, n.1, 2014. Disponível em: <https://doi.org/10.1177/2053951714538417>. Acesso em: 20 set. 2021.

STRUIJS, P.; DAAS, P. Big Data, Big Impact? Integration and management of new data sources. In: Seminar on Statistical Data Collection (Geneva, Switzerland, 25-27 September 2013). **Working Paper for the Conference of European Statisticians**. Geneva: UNECE, 2013.

STRUIJS, P.; DAAS, P. Quality Approaches to Big Data in Official Statistics. In: **International Conference on Big Data for Official Statistics**. Beijing, China, 28, 30 Oct 2014. New York: United Nations, 2014. Disponível em: <https://unstats.un.org/unsd/trade/events/2014/beijing/documents/other/Statistics%20Netherlands%20-%20Quality%20Approaches%20to%20Big%20Data.pdf>. Acesso em: 14 mar. 2021.

SUPLICY, E. A qualidade de vida no Brasil. **Folha de São Paulo**, São Paulo, Opinião, 13 mar. 1977. Disponível em: <https://acervo.folha.com.br/index.do>. Acesso em: 03 jan. 2024.

SUPLICY, E. Com alimento, 64% dos gastos. **Folha de São Paulo**, São Paulo, Economia, 16 jul. 1978. Disponível em: <https://acervo.folha.com.br/index.do>. Acesso em: 03 jan. 2024.

TAVARES, M., Da Substituição de Importações ao Capitalismo Financeiro: Ensaio sobre Economia Brasileira, Rio de Janeiro, Zahar, 1982.

TAVARES, M. Ajustes e reestruturação nos países centrais: a modernização conservadora. *Economia e Sociedade*, v.1, n. 1, p. 21-57, 1992.

THATCHER, J.; O'SULLIVAN, D.; MAHMOUDI, D. Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment and Planning D: Society and Space*, v. 34, n.6, p. 990–1006, 2016.

TROTSKY, L. Projet de programme de l'Internationale: critiques des thèses fondamentales (28 juin 1928). **Oeuvres**. Grenoble: Institut Léon Trotsky, 1988.
UL HAQ, M. **Reflections on human development**. Oxford: Oxford, 1995.

UN. **Fundamental Principles of Official Statistics**. New York: United Nations, 2014. Disponível em: <https://undocs.org/A/RES/68/261>. Acesso em 13 jun. 2021.

UN-BIGDATA. **Annual Plenary meeting of the UN Global Working Group on Big Data for Official Statistics**. New York: UN-Big Data, 2020. 1 vídeo (118 min). Disponível em: <https://www.youtube.com/watch?v=6MojHhYQR1Q> Acesso em: 10 jan. 2023.

UNCTAD. **Technology and innovation report 2021: Catching technological waves Innovation with equity**. New York: United Nations Publication, 2021.

UN-DESA. Public-private partnerships in use of Big Data for official statistics. New York: **UN-DESA**, 2016. Disponível em: <https://www.un.org/en/desa/public-private-partnerships-use-big-data-official-statistics>. Acesso em: 03 fev. 2023.

UNECE. The fundamental principles of official statistics in the region of the economic commission for Europe. **Economic Commission for Europe**, Annual Report. E/1992/32 .E/ECE/1266. Geneva: UNECE, 1992. Disponível em: <https://unece.org/statistics/FPOS>. Acesso em: 14 fev. 2020.

UNECE. Information systems architecture for national and international statistical offices guidelines and recommendations. Economic Commission for Europe. **Conference of European Statisticians statistical Standards and Studies**, n. 51. Geneva: United Nations, 1999. Disponível em: https://unece.org/fileadmin/DAM/stats/documents/information_systems_architecture/1.e.pdf. Acesso em: 23 set. 2023.

UNECE. Strategic vision of the High-level group for Strategic Developments in Business Architecture in Statistics. **Conference of European Statisticians**. Meeting on the Management of Statistical Information Systems. Luxembourg, 23-25 May 2011. Working Paper n. 24. Luxembourg: UNECE, 2011.

UNECE. What does “Big data” mean for official statistics? **Conference of European Statisticians**. Sixty-first plenary session. Geneva, 10-12 June 2013. Work of the High-level Group on the Modernization of Statistical Production and Services. Geneva: UNECE, 2013.

UNECE. Draft Report. **Meeting on the Management of Statistical Information Systems**, MSIS 2013. Paris, France, and Bangkok, Thailand, 23-25 April 2013. Prepared by the UNECE Secretariat. Geneva: UNECE, 2013a.

UNECE. **Classification of Types of Big Data**. Geneva: UNECE, 2013b. Disponível em: <http://www1.unece.org/stat/platform/display/bigdata/Classification+of+Types+of+Big+Data>. Acesso em: 12 fev. 2023.

UNECE. Guidance on common elements of statistical legislation. **Conference of European Statisticians**, 66th Session, Geneva. 2018. Disponível em: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2018/CES_6_Common_elements_of_statistical_legislation_Guidance_for_consultation_for_upload.pdf. Acesso em: 10 fev. 2023.

UN-HLP. **A New Global Partnership: Eradicate Poverty and Transform Economies Through Sustainable Development**. New York: UN Publications, 2013.

UN-IEAG. **A World that counts: Mobilising The Data Revolution for Sustainable Development**, 2014. Disponível em: <https://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf>. Acesso em: 07 set. 2021.

UN-StatCom. **Fundamental Principles of Official Statistics**. New York: UN Statistical Commission, 1994. Disponível em: <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>. Acesso em: 15 fev. 2020.

UN-StatCom. **Terms of reference and mandate of the Global Working Group on Big Data for Official Statistics**. New York: UN Statistical Commission, 2015. Disponível em: <https://unstats.un.org/bigdata/documents/Terms%20of%20Reference%20-%20UN%20Global%20Working%20Group%20on%20BigData.pdf>. Acesso em: 20 out. 2021.

UN-StatCom. **Cape Town Global Action Plan**, 2017. New York: UN Statistical Commission. Disponível em: <https://unstats.un.org/sdgs/hlg/cape-town-global-action-plan/>. Acesso em: 12 out. 2021.

UN-StatCom. **Global Platform for Data, Services and Applications**. UN Statistical Commission. Forty-eighth session. Item 3 (d) of the provisional agenda Items for discussion and decision: Big Data for Official Statistics. New York: UN Statistical Commission, 2017. Disponível em: <https://unstats.un.org/unsd/statcom/48th->

session/documents/BG-3d-global-platform-for-data-services-applications-E.pdf
Acesso em: 20 out. 2021.

UN-StatCom. **Proceedings of the Third International Conference on Big Data for Official Statistics**. UN Statistical Commission Forty-eighth session. Item 3 (d) of the provisional agenda Items for discussion and decision: Big Data for Official Statistics. New York: UN Statistical Commission, 2017a. Disponível em: <https://unstats.un.org/unsd/statcom/48th-session/documents/BG-3d-Proceedings-of-the-3rd-Conference-on-Big-Data-for-Official-Statistics-in-Dublin-Final-E.pdf>. Acesso em: 21 out. 2021.

UN-StatCom. **Review of Implementation of Data Revolution**. UN Statistical Commission Fifty-first session 3 – 6 Item 3(a) of the provisional agenda Items for discussion and decision. New York: UN Statistical Commission, 2019. Disponível em: https://unstats.un.org/unsd/statcom/51st-session/documents/BG-Item3a_Review-of-Implementation-of-Data-Revolution-E.pdf. Acesso em: 20 out. 2021.

UN-StatCom. **UN Global Platform Business Model Options**. UN Statistical Commission Fiftieth session. Item 4(e) of the provisional agenda. Items for information: Big data for official statistics. New York: UN Statistical Commission, 2019a. Disponível em: <https://unstats.un.org/bigdata/documents/reports/GWG%20Background%20document%20-%202019%20-BG-Item4e-UN-GP-Business-Model-Options-E.pdf>. Acesso em 17 out. 2021.

UN-StatCom. **Sustainable Business Model for the UN Global Platform**. New York: UN Statistical Commission, 2019b. 1 vídeo (78 min). Disponível em: <https://www.youtube.com/watch?v=vJUPtA8qCAI> Acesso em: 17 out. 2021.

UN-StatCom. **Report of the Global Working Group on Big Data for Official Statistics**. UN Statistical Commission Fiftieth-first session. Item 3 (t) of the provisional agenda. Items for discussion and decision: big data. New York: UN Statistical Commission, 2019c. Disponível em: <https://unstats.un.org/unsd/statcom/51st-session/documents/2020-24-BigData-E.pdf>. Acesso em: 13 out. 2021.

UN-StatCom. **Report of the Global Working Group on Big Data for Official Statistics**. Statistical Commission Fifty-second session. Item 3 (j) of the provisional agenda. Items for discussion and decision: big data. New York: UN Statistical Commission, 2020. Disponível em: <https://unstats.un.org/unsd/statcom/52nd-session/documents/2021-14-BigData-E.pdf>. Acesso em: 22 out. 2021.

UN-StatCom. **UN GWG Task Team on Big Data for SDGs**. Global Working Group on Big Data for Official Statistics Annual Meeting. New York: UN Statistical Commission, 2020a. Disponível em: <https://unstats.un.org/bigdata/events/2020/gwg-annual-meeting/>. Acesso em: 15 set. 2021.

UN-StatCom. **Report of the Friends of the Chair group on economic statistics**, 2020b. Statistical Commission Fifty-second. Item 3 (c) of the provisional agenda. Items for discussion and decision: economic statistics. New York: UN Statistical Commission,

2020b. Disponível em: <https://undocs.org/en/E/CN.3/2021/7>. Acesso em: 15 set. 2021.

VANDENBERGHE, F. **Teoria social realista um diálogo franco-britânico**. Belo Horizonte: Editora UFMG, 2010.

VERDEGEM, P. Dismantling AI capitalism: the commons as an alternative to the power concentration of big tech. **AI & Society**. 2022. DOI: 10.1007/s00146-022-01437-8.

WARD, M. **Quantifying the world: UN ideas and statistics**. United Nations Intellectual History Project . Bloomington: Indiana University Press, 2004.

WEBER, H. Politics of ‘Leaving No One Behind’: Contesting the 2030 Sustainable Development Goals Agenda. **Globalizations**, v.14, n.3, p.399-414, 2017.

WIENER, N. **Cibernética e Sociedade: O Uso Humano de Seres Humanos**. 7a. Ed. São Paulo: Editora Cultrix, 2000.

WILLIAMSON, J. The Strange History of the Washington Consensus. **Journal of Post Keynesian Economics**. v. 27, n. 2, p. 195-206. 2005.

WU, A.; TANEJA, H. Platform enclosure of human behavior and its measurement: Using behavioral trace data against platform episteme. **New Media & Society**, v.23, n.9, p.2650–2667, 2021. Disponível em: <https://doi.org/10.1177/1461444820933547>. Acesso em: 14 ago. 2022.

YILDIRMAZ, A. Using Big Data to Decode Private Sector Wage Growth. In: **Bloomberg Data for Good Exchange Conference**. New York, 25 sep. 2016. Disponível em: https://data.bloomberglp.com/company/sites/2/2016/09/paper_67.pdf. Acesso em: 16 ago. 2022.

ZUBOFF, S. **The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power** London: Profile Books, 2019.

APÊNDICE A – TESE E META-TESE: TEXTO DE APRESENTAÇÃO DA DEFESA DA TESE DE DOUTORADO EM 02.07.2024.

Quero aproveitar essa oportunidade de fala introdutória para trazer alguns elementos subjacentes ao texto. Abordar a tese e a meta-tese, ou seja, a experiência pessoal do eu-pesquisador na construção objetiva da pesquisa, trazendo à luz aspectos subjetivos que não foram descritos no texto mas que podem ajudar a banca a compreender um pouco melhor porque esta tese é como é, e está estruturada como está.

Escrever uma tese é um pouco como fazer uma pintura que se constrói em camadas superpostas, ao final alguns elementos que a tornaram possível e a sustentam não são diretamente visíveis ao olhar, restam ocultos, e é importante que assim seja. Acredito, entretanto, que estes elementos constituem a própria essência da obra e sua compreensão pode dar uma nova perspectiva àquilo que se apresenta como obra.

Vou começar pela minha relação com o objeto, as estatísticas oficiais e o IBGE. Ingressei no IBGE em 2010 pela via do concurso público como analista de pesquisas e quando decidi realizar uma investigação de doutorado sobre este objeto, o fiz depois de 9 anos de prática em diversas áreas da instituição, iniciando pelo censo de 2010, depois em pesquisas econômicas, pesquisas geocientíficas, base territorial, cadastro de endereços e pesquisas domiciliares. É importante destacar, entretanto, o aspecto marginal desta atuação, que se dá na unidade estadual do Ceará, distante não apenas geograficamente do centro de poder histórico do IBGE no Rio de Janeiro, mas sobretudo distante hierarquicamente, na divisão que nesta casa se estabelece entre o trabalho metodológico/intelectual da sede e o trabalho de coleta/braçal das unidades estaduais e agências.

Ser um analista em uma unidade estadual do IBGE significa muitas vezes ver sua potencialidade intelectual colocada de escanteio em face desta separação e da excessiva centralização da instituição no Rio de Janeiro. Por outro lado, abre-se para nós que estamos na ponta um universo de possibilidades para o conhecimento empírico e concreto do que significa a produção de estatísticas oficiais, o que se revela em sua plenitude nas atividades diretas de coleta de dados, na pesquisa de campo que se opera no espaço vivo do território nacional. Para mim, se há uma mística ibgeana, ela reside fundamentalmente neste encontro e confronto dialético e diário

entre a população e o agente de Estado, encontro do qual resulta a informação estatística.

Foi justamente em uma dessas experiências de campo do IBGE, que tive uma primeira epifania sobre o que viria a ser o tema desta pesquisa. O trabalho de atualização de setores censitários do IBGE, que gerenciei por alguns anos no Ceará, é uma atividade árdua. Implica em percorrer todo um setor, verificar seus limites, todas suas vias e logradouros e visitar todos domicílios e estabelecimentos. Esse trabalho é essencial para garantir uma boa seleção da amostra nas pesquisas domiciliares.

Em um desses trabalhos de campo no interior do estado, eu e minha equipe deparamos com um carro da Google, que fazia um trabalho muito semelhante ao nosso, percorrendo todas as vias de um setor. Entretanto, enquanto parávamos em cada unidade, tocávamos a campainha ou batíamos palma e nos apresentávamos para confirmar o endereço e a espécie do domicílio junto aos moradores, o carro da empresa apenas passava ao largo, com os vidros fechados, a tudo lenta e silenciosamente registrando com uma câmera de 360 graus no teto da viatura. Ao longo de um dia de trabalho, um tanto intrigados, passamos algumas vezes por aquele sócia espião. A parte cômica desta história foi encontrar no final do dia o carro da Google atolado e sua equipe desalentada. Paramos com o carro do IBGE, atrelamos o carro da empresa ao nosso e o tiramos do atoleiro. O IBGE tirando a Google de um atoleiro, não pude deixar de me perguntar se fariam o mesmo por nós.

Tirando o aspecto lúdico da história, de fato a percepção de que novos agentes, novas tecnologias e novas fontes de dados poderiam implicar em mudanças na forma tradicional de produção e distribuição de estatísticas oficiais começou a se apresentar efetivamente para mim como um possível objeto de pesquisa a partir de minha prática profissional no IBGE. As novas fontes de dados e modalidades de informação, o uso crescente de registros administrativos e o surgimento dentro da casa do termo um tanto difuso 'big data' despertaram a minha imaginação para uma possível mudança de era na atividade de produção de estatísticas oficiais e comecei a me perguntar sobre suas possibilidades e significados.

Quero agora abordar um pouco do caminho acadêmico deste eu-pesquisador, o que pode ajudar a lançar luz sobre as escolhas teóricas e metodológicas desta pesquisa. Para isso vou iniciar por um breve paralelo entre a minha pesquisa de doutorado e minha pesquisa de mestrado. A princípio, parecem tratar de dois temas diametralmente distintos. Em minha pesquisa de mestrado,

investiguei o extrativismo da palmeira carnaúba no Ceará. Entre esta dissertação que defendi em 2004, e a presente tese de doutorado passaram-se 20 anos. Há, entretanto, interessantes similaridades entre as duas pesquisas.

Antes de me dedicar a investigação acadêmica sobre a carnaúba, eu havia passado 6 anos trabalhando com comunidades rurais do sertão do Ceará em projetos de desenvolvimento comunitário com trabalhadores extrativistas da carnaúba. Foi a partir desta experiência prática e da identificação de mudanças contemporâneas na atividade que senti a necessidade de pesquisá-la mais a fundo. Há, portanto, uma similaridade no que diz respeito à relação deste eu-pesquisador com o campo acadêmico, ou seja, trata-se de uma relação que parte de uma prática mas que a extrapola na tentativa de compreender a manifestação atual do objeto desta prática a partir de uma investigação histórica, teórica e empírica.

As similaridades, entretanto, não param por aí. O extrativismo da carnaúba e as estatísticas oficiais são ambos objetos bi centenários e sua 'exploração' no Brasil data de meados do Séc. XIX. Os objetos são evidentemente diferentes, mas o seu enquadramento pelo olhar do eu-pesquisador, guardado o amadurecimento ao longo de 20 anos, também é semelhante. Tratou-se de investigar, em ambos os casos, as mutações destes objetos à luz do desenvolvimento histórico do modo de organização social do capital e do modo de ação do Estado, para então compreender suas manifestações atuais por meio de pesquisa empírica junto aos agentes.

Para compreender as opções teóricas da presente pesquisa é preciso ainda destacar que sou um outsider no campo sociológico. Fiz graduação em administração na Fundação Getúlio Vargas em São Paulo, mestrado em desenvolvimento e meio ambiente na Universidade Federal do Ceará e finalmente este doutorado em sociologia. A entrada no campo sociológico foi possibilitada pela sorte de alguns bons encontros intelectuais que gostaria de mencionar e agradecer.

Tive a sorte de ter uma socióloga como orientadora de mestrado, a Profa. Monica Martins e de posteriormente colaborar com o cientista político Prof. Daniel Zirker enquanto elaborava o projeto da tese. Na UFC tive a renovada sorte de ter o Prof. Edemilson Paraná como orientador, sempre me incentivando e abrindo caminhos, e de participar de seu grupo de Estudos em Economia, Tecnologia e Sociedade (NETS) onde realizamos estudos coletivos sobre a nova economia política da digitalização e dos dados. Já no curso desta pesquisa, os encontros com o Prof. Alexandre Camargo nos GTs de quantificação da Anpocs onde os resultados parciais

da investigação foram sendo apresentados e debatidos, e os diálogos com a Profa. Cecília Rikap, o Prof. Tomás Rotta e a Profa. Evelyn Ruppert no período de estudos em Londres, foram fundamentais. A todos estes mestres sou muito grato.

Apesar de todos esses apoios, uma tese de doutorado é um processo solitário, permeado por uma série de lutas internas. Novamente, é importante que assim seja. Neste sentido, a construção do aparato teórico e metodológico desta pesquisa de doutorado foi conquistado no próprio curso da pesquisa, em idas e vindas, muitas vezes tateando indeciso entre as diversas escolas do pensamento sociológico.

A princípio, a sociologia interpretativa de Max Weber e seus estudos sobre a burocracia e a racionalidade instrumental me pareceram um caminho possível para ler as estatísticas oficiais. Posteriormente, explorei a problematização desta mesma racionalidade instrumental pela via marxista da Escola de Frankfurt, em seu conceito de reificação. Estas abordagens, entretanto, não pareciam abrir maiores possibilidades para a compreensão das distintas facetas das estatísticas oficiais e de suas transformações na era dos dados, de certo modo a racionalidade weberiana e a reificação são conceitos totalizantes que se encerram em si, dificultando a percepção das nuances entre as distintas formas de dominação instrumental.

Passando para os estudos do marxismo estrutural, trilhei a possibilidade de investigar as estatísticas como aparelhos ideológicos do Estado. Mas de modo similar, tampouco estas lentes me pareceram o melhor instrumento para compreender as mudanças históricas destes aparelhos, bem como para analisar o contexto contemporâneo em que corporações privadas passam a disputá-los. Embora a influência de Nico Poulantzas, no que se refere ao conceito da autonomia relativa do Estado, tenha permanecido entre os alicerces da tese.

A seguir, enveredei nos estudos sociais da ciência e tecnologia, em particular na teoria do ator-rede de Bruno Latour. Me interessaram sobretudo as possibilidades de pesquisa empírica oferecidas por esta abordagem. Seguir os atores, desdobrar controvérsias, abrir caixas pretas, compreender as traduções, me pareceram um bom guia metodológico para estudar meu objeto. E de fato, há traços desta influência na minha investigação. Entretanto, a negação da teoria e da explicação social, uma certa desconsideração dos macroprocessos econômicos e sociais, e a insistência um tanto performativa na ação dos não-humanos me levaram a abandonar esta via teórica.

Foram os estudos sócio históricos da estatística que finalmente me levaram à sociologia da quantificação de Alain Desrosières. Aqui ancorei em um porto

razoavelmente seguro e encontrei um importante suporte histórico e teórico para o estudo das estatísticas oficiais. Afinal, o próprio Desrosières era ele próprio um estatístico de Estado buscando compreender os sentidos desta prática e de suas transformações. A noção da estatística como ferramenta de prova e governo, as distinções epistemológicas entre a estatística frequentista e bayesiana, o entendimento das ferramentas estatísticas como instrumentos imbricados na racionalidade governamental, a aproximação com a noção foucaultiana de governamentalidade, a identificação do neoliberalismo como fator de crise das estatísticas públicas, tudo isso e muito mais, essa pesquisa deve à Desrosières.

É preciso, entretanto, no diálogo com os cânones, compreender os seus limites e os condicionantes de seu tempo. A aproximação de Desrosières com a escola pragmática o levou a um distanciamento da sociologia crítica de Pierre Bourdieu. Assim a sua sociologia tende a um certo institucionalismo e a uma ênfase nos aspectos epistêmicos, de certo modo abrindo mão da investigação do significado estrutural das lutas no curso das transições históricas. Neste sentido, nesta tese busquei reaproximar a sociologia da quantificação ao estruturalismo genético de Bourdieu, utilizando seus conceitos de campo, habitus e capital, como ferramentas analíticas para a leitura do processo histórico de estruturação do campo estatístico e para a compreensão das disputas e do que está em jogo na sua interseção com o campo algorítmico no novo contexto dos dados.

Por outro lado, apesar das pesquisas de Desrosières revelarem a importância dos aspectos econômicos nos processo de transformação das ferramentas estatísticas, ficou evidente a necessidade de um maior aporte da economia política para esta análise. Foi neste sentido que busquei agregar à sociologia da quantificação, além do conceito de governamentalidade de Michel Foucault já implícito no esquema de Desrosières, o conceito de regime de acumulação do capital da escola francesa da regulação. Finalmente, combinando a análise de Desrosières com o conceito de regimes estatísticos de Jean Pierre Beaud foi possível construir um esquema analítico na primeira parte do trabalho, cuja proposição é a de que: crises no regime de acumulação do capital conduzem a crises de governamentalidade que por sua vez implicam na reconfiguração dos regimes estatísticos. Com esse esquema, propus uma estratificação histórica para investigar as características e os processos de transição entre os regimes estatísticos de contabilização, enumeração, precisão, comensuração e finalmente a atual transição para um regime de dataficação. Acredito

que este esquema analítico seja uma importante contribuição da tese para o campo da sociologia da quantificação.

No campo da economia política, a leitura da Grande Transformação de Karl Polanyi acendeu uma nova luz para a pesquisa, a de que tudo parece levar a crer que estamos vivendo uma segunda grande transformação. Seja no processo de cercamento e comodificação dos dados e do conhecimento, nos paralelos entre a crise de 1929 e a crise de 2008, no esgotamento do neoliberalismo, na ascensão de novos fascismos e na conflagração de um novo possível conflito global. A noção de duplo movimento de Polanyi se apresentou então como um outro conceito poderoso para compreender as transformações contemporâneas do processo de dataficação, especialmente do ponto de vista da relação entre o Estado e o mercado.

Ainda na economia política, a teoria do valor de Marx contribuiu com ferramentas para, por um lado compreender as estatísticas oficiais como um importante elemento das condições gerais de produção envolvidas no processo de extração da mais valia e de, por outro, enquadrar a comodificação dos dados como parte de um novo processo de acumulação do capital por meio do cercamento de comuns e extração de rendas de conhecimento.

A releitura dos conceitos marxianos vieram do diálogo com investigações contemporâneas da economia política da digitalização, dados e plataformas. Articular este campo à sociologia da quantificação foi um dos desafios teóricos da pesquisa. O que resta evidente é que estes dois campos tem muito a dialogar. Muitas vezes parece haver uma lacuna no entendimento de que os processos contemporâneos de dominação por meio dos dados e algoritmos guardam conexões com os processos históricos de dominação pelas estatísticas. Embora existam evidentemente distinções neste processo, que é o que justamente procuramos demonstrar nesta tese, a dataficação das estatísticas oficiais toca em um ponto nevrálgico e ainda pouco estudado por ambos os campos. O que diz respeito a própria natureza da autonomia relativa do Estado e da sua tradicional dominação simbólica por meio das estatísticas, que passa a ser disputada pelos interesses políticos e econômicos das grandes corporações de tecnologia. Creio que uma segunda contribuição de minha pesquisa seja a de apontar esta lacuna e a de tentar estabelecer uma ponte entre os dois campos para estudar este fenômeno.

Finalmente, ainda do ponto de vista da teoria mobilizada, outro desafio foi o de investigar o desenvolvimento histórico das estatísticas oficiais e do processo de

dataficação em sua dualidade estrutural, ou seja, como compreender as diferenças do desenvolvimento da estatística oficial no norte e sul-global e em particular no Brasil. Para dar conta dessa complexidade do objeto, a pesquisa se aproxima da tradição do pensamento latino-americano sobre o subdesenvolvimento e a dependência, incorporando criticamente o conceito de colonialismo de dados e se valendo de importantes contribuições da história da estatística e sociologia da quantificação brasileira.

Toda essa mobilização teórica e histórica resultou em achados aparentemente muito simples e que pareciam estar à vista todo o tempo mas que não era ainda possível enxergar de pronto. Quer seja, que o processo contemporâneo de dataficação das estatísticas oficiais implica em duas mudanças estruturais fundamentais. Em primeiro lugar, uma mudança ontológica na economia política dos dados, de sua tradicional concepção de bens públicos para mercadorias. Em segundo lugar, uma mudança epistemológica nos métodos estatísticos, da orientação dedutiva e frequentista das estatísticas oficiais para uma outra indutiva e bayesiana do big data e ciência de dados. Esta é fundamentalmente a nossa tese, mas as coisas não se passam de uma forma assim tão simples. O mundo social não existe no vácuo, há sempre atritos, inércias, resistências, conflitos e lutas. De forma que aquilo que se apresenta como mudança é em si um obstáculo. As forças implicadas na superação e na contenção destes obstáculos são indutoras de movimentos. Neste sentido, destacamos três eixos de vetores de forças que são colocadas em movimento e oposição no processo de dataficação das estatísticas.

Em um eixo vertical, os obstáculos ontológicos relativos às disputas em torno dos dados e os obstáculos epistemológicos das disputas em torno dos métodos.

Em um eixo horizontal, de um lado os movimentos de disrupção associados às forças do mercado e de outro lado os contramovimentos de conservação associados às forças protetivas da sociedade e da vida. Este duplo movimento, que tende a ser absorvido pelo Estado se manifesta no interior do campo estatístico. Neste sentido um elemento fundamental na análise desta oposição é a interseção entre o campo estatístico, tradicionalmente vinculado ao Estado, e o campo algorítmico, umbilicalmente vinculado às corporações de tecnologia, na nova conjuntura dos dados. Aqui entram em oposição o conjunto de disposições científicas e burocráticas que conformaram historicamente o habitus relativamente conservador do estatístico

de Estado em seu encontro e confronto com o novo perfil profissional do cientista de dados e seu habitus empreendedor e disruptivo.

Em um eixo transversal, um segundo conjunto de forças que está continuamente em movimento no campo estatístico diz respeito a uma contradição intrínseca da estatística enquanto ferramenta de governo, quer seja, a força centrípeta da soberania nacional e a força centrífuga da competição entre os Estados. Aqui, um elemento fundamental da análise é a relação dialética entre o internacionalismo e o nacionalismo estatístico. Este eixo de forças é fundamental sobretudo para a compreensão de como se processam as mudanças nos países periféricos como o Brasil.

Este quadro analítico é o cerne desta tese e corresponde à lente sociológica com a qual analisamos os dados empíricos.

Gostaria de tratar agora dos dados empíricos. Para tanto, vou retornar ao eu-pesquisador e ao eu-ibgeano, para abordar os conflitos envolvidos na construção do eu-sociólogo.

Estamos tratando aqui de um Eu-Ibgeano cearense e marginal, desconhecido nas esferas internas de disputa institucional do IBGE, que se desloca para a posição de um Eu-pesquisador, almejando entrar em um campo sociológico do qual é um outsider. Há uma série de riscos neste deslocamento, mas há também algumas vantagens nesta posição. O pertencimento marginal ao campo estatístico facilita a entrada no empírico para a realização de entrevistas com pares e observação de reuniões e atividades. Ser ibgeano abre portas, ser desconhecido facilita a interação e a observação. Por outro lado, ser um outsider no campo sociológico possibilita lidar com maior flexibilidade com os conceitos de uma forma mais livre.

O objeto empírico tal qual o aparato teórico foi construído e conquistado ao longo da investigação. Ao propor o tema da pesquisa em 2018 e ao ingressar no curso de doutorado em 2019, eu não tinha qualquer conhecimento da existência do Projeto ONU Big Data, da Plataforma Global e do Hub Regional de Big Data que seria implantado no IBGE. Até porque, são eles mesmos contemporâneos à própria pesquisa e foram se manifestando, se estruturando e se movimentando ao longo da investigação. Basta ver que a versão beta da Plataforma Global é de 2020 e que a assinatura do memorando de entendimentos do hub regional no Brasil é de 2021. Estudar um objeto no seu nascimento e em constante movimentação foi um grande desafio. No momento que estou aqui apresentando esta tese, estes objetos continuam

em plena ação e transformação. Até que ponto esta própria pesquisa interferiu nestes objetos é uma outra questão a se perguntar, considerando as entrevistas realizadas e a publicação parcial dos seus resultados em um *paper* no início deste ano.

Os acontecimentos empíricos ao longo da pesquisa foram decisivo para os rumos que a investigação tomou. Já não era possível restringir o olhar ao caso brasileiro, nem tampouco observá-lo apenas do ponto de vista do norte-global. A dualidade estrutural da dataficação das estatísticas oficiais entre norte e sul-global estava como que estampada na lógica mercantil, extrativista e neo-colonialista do projeto da Plataforma Global. O envolvimento das Big Techs no financiamento desta estrutura, as suas ações de lobby junto a Comissão e Divisão Estatística da ONU, as ações de soft power na formação dos cientistas de dados e sensibilização dos INEs, a promoção de parcerias público-privadas, etc. Tudo isso foi se apresentando ao longo da investigação.

Ao mesmo tempo, a opção de realizar um doutorado neste exato momento histórico dizia respeito a uma decisão pessoal de não participar da gestão pública durante o governo de extrema direita de Jair Bolsonaro. É preciso observar que no período anterior a esta eleição uma forte mobilização bolsonarista ascendeu entre os quadros do IBGE. A dificuldade de convivência com este fenômeno e com os colegas que a ele aderiram foi um dos fatores que me levaram ao pedido de afastamento para a realização da pesquisa.

A coincidência entre a intensificação da agenda de dataficação no IBGE, fortemente vinculada a interesses externos, e o período de desestruturação institucional durante o Governo Bolsonaro, conforme analisamos na pesquisa, não é fortuita, mas trata-se sobretudo de uma confluência de interesses. Como se torna cada vez mais evidente, há muitas afinidades eletivas entre a governança algorítmica das Big Tech e os partidários da extrema direita.

À medida em que os dados foram vindo a tona em minha pesquisa, o eu-ibgeano foi tomado por muitas angustias. Principalmente ao constatar a adesão do IBGE ao projeto da Plataforma sem um reflexão institucional séria. O mesmo diz respeito à adesão de colegas, muitas vezes encantados pela tecnologia e pela participação em espaços internacionais, que sem questionar seus condicionantes políticos e econômicos, deixaram-se instrumentalizar e cooptar por uma agenda problemática do ponto de vista da independência institucional e soberania nacional.

Foi necessário adotar uma constante vigilância epistemológica para lidar com as crises entre o eu-ibgeano e o eu-pesquisador neste processo.

Por outro lado, as entrevistas revelaram processos autônomos e promissores de uso de novas fontes de dados e tecnologias em curso no IBGE por iniciativa de suas próprias equipes técnicas. Revelaram a demanda entre os agentes por uma nova legislação protetiva do sistema estatístico nacional que garanta o acesso compulsório aos dados guardados por privados como bens públicos, demonstraram que há resistências entre os quadros internos às parcerias público-privadas e ao perfil mercadológico dos cientistas de dados e finalmente indicaram uma postura de proteção das fontes e métodos tradicionais e um tratamento cuidadoso com as novas fontes e métodos. Uma postura que não é fechada à inovação mas que é refratária à disrupção. Como vimos esses contramovimentos protetivos se dão no nível de disposições historicamente adquiridas pelos agentes enquanto ibgeanos e estatísticos de Estado.

Finalmente, os principais resultados demonstraram:

Em primeiro lugar, que há interesse e investimentos de Big Techs, em particular da Google e Microsoft, na construção de um mercado de estatísticas oficiais sob o seu controle político e econômico. Isso foi evidenciado no lobby destas empresas sobre a divisão estatística da ONU e na construção da Plataforma Global como uma estrutura de nuvem orientada para a extração de dados do Sul-Global e para a padronização e comercialização de dados e métodos algorítmicos para estatísticas oficiais por meio de arranjos público e privados.

Em segundo lugar, que a implantação dos hubs regionais da Plataforma no sul-global são parte de uma estratégia voltada à superar resistências dos Estados Nacionais à nova estrutura de dominação, com a introdução de agentes e mecanismos de mercado nos processos de produção e distribuição de estatísticas oficiais no Sul-Global. Isso foi evidenciado nos grupos de trabalho com a participação de representantes dos Institutos de estatística e especialistas da iniciativa privada, em parcerias público-privadas para intermediação e comercialização de dados e métodos e em mecanismos de soft power como festivais de dados e treinamentos.

Este movimento pró-mercado autointitulado uma “revolução de dados pelo desenvolvimento sustentável” instrumentalizou o discurso da modernização das estatísticas oficiais com big data, centrando-o em torno da agenda de parcerias com o setor privado e da promoção de uma mudança cultural no setor estatístico,

simbolizada pela introdução do novo perfil profissional do cientistas de dados nos Institutos Nacionais de Estatística, mais aberto à experimentação e a inovação e mais permeáveis à lógica de mercado. Esta estratégia mobiliza a confiança nas estatísticas oficiais e na ONU e o discurso positivo da sustentabilidade para encapsular e camuflar seus reais objetivos de mercado.

Por outro lado, os resultados também apontaram controvérsias e contramovimentos protetivos tanto a nível da ONU, como na União Europeia e no Brasil. No campo estatístico, estes contramovimentos tem se materializado em demandas por uma nova legislação estatística e mecanismos que garantam o acesso compulsório dos Institutos estatísticos aos dados guardados por privados, bem como a proteção dos dados para fins estatísticos como bens públicos e o fortalecimento dos sistemas estatísticos nacionais. Os contramovimentos também se revelam em resistências internas dos estatísticos de Estado ao ethos empreendedor e aos métodos disruptivos da ciência de dados.

Para finalizar, restringi o recorte temporal da investigação empírica para o período de 2019 a 2023. Afinal, era preciso em algum momento parar de seguir o objeto para descrevê-lo em um texto relativamente coeso que apontasse uma conclusão possível, mesmo que sabidamente transitória, de um processo vivo que segue em movimento.

A partir de 2024, sob um novo governo e uma nova gestão, o IBGE deu início a uma série de diálogos institucionais por ocasião da proximidade de seus 90 anos. Estes diálogos passam a visar a reestruturação do sistema estatístico nacional e a revisão da legislação estatística motivados principalmente pelo novo contexto dos dados e por um discurso de proteção da soberania nacional contra a incidência das Big Tech. De certo modo é como assistir a manifestação do contramovimento protetivo em plena ação no Brasil.

A minha participação neste processo enquanto finalizava este texto, levou a novos conflitos entre o eu-ibgeano e o eu-pesquisador. Até que ponto a minha participação neste processo interferia no meu próprio objeto de investigação. Até que ponto o conhecimento sociológico que é construído em um outro tempo, sob suas próprias determinações e limitações, tem aplicação válida no domínio do real e no tempo vivo da política.

Essas são questões que esta pesquisa não tem como responder, demonstrando que o que se esgota mesmo é o tempo da tese e nunca a tese em si. Muito obrigado.