



**UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE HUMANIDADES
DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA INFORMAÇÃO**

DACLES VÁGNER DA SILVA

**A REPRESENTAÇÃO E A RECUPERAÇÃO DA INFORMAÇÃO: BASES,
DIÁLOGOS E CONTRIBUIÇÕES PARA O FAZER ARQUIVÍSTICO**

FORTALEZA

2019

DACLES VÁGNER DA SILVA

A REPRESENTAÇÃO E A RECUPERAÇÃO DA INFORMAÇÃO: BASES,
DIÁLOGOS E CONTRIBUIÇÕES PARA O FAZER ARQUIVÍSTICO

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Mestre em Ciência da Informação.

Área de Concentração: Representação e Mediação da Informação e do Conhecimento.

Orientador: Prof. Dr. Osvaldo de Souza

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- S579r Silva, Dacles Vágner da.
A REPRESENTAÇÃO E A RECUPERAÇÃO DA INFORMAÇÃO : bases, diálogos e contribuições para o fazer arquivístico / Dacles Vágner da Silva. – 2019.
104 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Humanidades, Programa de Pós-Graduação em Ciência da Informação, Fortaleza, 2019.
Orientação: Prof. Dr. Osvaldo de Souza.
1. Recuperação da Informação. 2. Representação da Informação. 3. Ciência da Informação. 4. Ciência da Computação. 5. Arquivologia. I. Título.

CDD 020

DACLES VÁGNER DA SILVA

A REPRESENTAÇÃO E A RECUPERAÇÃO DA INFORMAÇÃO: BASES,
DIÁLOGOS E CONTRIBUIÇÕES PARA O FAZER ARQUIVÍSTICO

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Mestre em Ciência da Informação.

Área de Concentração:
Representação e Mediação da Informação e do Conhecimento.

Aprovado em: 31/01/2019.

BANCA EXAMINADORA

Prof. Dr. Osvaldo de Souza (Orientador)
Universidade Federal do Ceará (UFC)

Prof.^a Dra. Andrea Soares Rocha da Silva
Universidade Federal do Ceará (UFC)

Prof. Dr. Henry Poncio Cruz de Oliveira
Universidade Federal da Paraíba (UFPB)

A Helena, Carol, Juca e meus pais Severino e Graça.

AGRADECIMENTOS

A Deus, regendo o universo e nos colocando em caminhos a serem descobertos;

A minha companheira Anna Carollyna, pelo grande incentivo, correções, orientações, paciência e ótimos debates. Por também ter me dado a imensa felicidade de ser o pai da Helena, nossa maior riqueza;

Aos meus pais, Severino e Graça, que enfrentaram grandes desafios nesses últimos tempos e demonstraram a força como maior exemplo para superar adversidades;

Ao professor Osvaldo de Souza, pelas orientações e conhecimento adquirido nessa trajetória do mestrado;

Ao professor Washington de Medeiros, pelas valiosas dicas que fortaleceram o caminho de escrita do presente trabalho;

Aos colegas de mestrado, pelas conversas, trocas de mensagens e grandes incentivos;

A todos os servidores (técnico-administrativos e docentes) e colaboradores do Programa de Pós-Graduação em Ciência da Informação/DCINF/UFC, com atenção especial a Veruska Maciel, sempre muito atenciosa e prestativa;

Aos colegas de trabalho, da Divisão de Protocolo Central, da Coordenadoria de Administração e Patrimônio/UFC, pelo incentivo, flexibilidade e compreensão.

Não adianta tirar de onde não tem
Nem tentar encaixar onde não cabe
Sem saber alguém tenta, e quando sabe
Já não dá nem um passo mais além
Pois de trás para frente nada vem
O que foi já não é e nem será
E da frente pra trás, ninguém irá
Desfazer o que fez, certo ou errado
Vou deixar este canto abandonado
Para sempre do jeito como está

“*Qasida*”, Siba

RESUMO

A Recuperação da Informação (RI) e a Representação da Informação têm sido temas de estudo de diversas áreas. Entre essas áreas, podemos destacar a Ciência da Informação (CI), a Ciência da Computação (CC) e a Arquivologia. Diante da assertiva necessidade de representação de documentos e com o estabelecimento cada vez mais frequente de sistemas computacionais para manuseio desses registros documentais, gerados ou recebidos no andamento de suas respectivas atividades, se evidenciou a partir daí a importância da representação para posterior recuperação da informação. Assim, é preciso delimitar a importância do presente estudo como iniciativa que visa examinar fatores que permeiam a Representação e a Recuperação da Informação na Arquivologia, tomando como áreas de diálogo a CI e a CC. Dessa maneira, o objetivo geral foi examinar como se delinea a Representação da Informação e a RI no contexto arquivístico, levando em conta a atuação de áreas como a CI e a CC nesses temas. Os objetivos específicos se expuseram em: a) Mapear aspectos em comum da Representação da Informação e RI da Arquivologia com áreas correlatas, como a CI, bem como a CC; b) Analisar as ideias centrais da Representação da Informação e RI na área arquivística; c) Discutir como a relação entre Representação da Informação e RI se evidencia na Arquivologia; e d) Refletir questões de ordem técnica/teórica da Representação e da RI na área arquivística. Metodologicamente foi utilizada a técnica de análise do conteúdo na busca pela compreensão dos conceitos discutidos no presente estudo. Evidenciou-se que a Representação da Informação se nota a partir da ação administrativa, segmentar, intelectual e descritiva, com elaboração de instrumentos para o fornecimento de acesso. No caso da RI, ela se notabiliza como um processo informatizado, com enfoque em suprir necessidades informacionais.

Palavras-chave: Recuperação da Informação. Representação da Informação. Ciência da Informação. Ciência da Computação. Arquivologia.

ABSTRACT

Information Retrieval (IR) and Information Representation have been themes of study in several areas. Among these areas, we can highlight the Information Science (IS), Computer Science (CS) and Archivology. Faced with the assertive need for document representation and with the increasingly frequent establishment of computer systems to handle these documentary records, generated or received in the progress of their respective activities, the importance of representation for later retrieval of information was evidenced. Thus, it is necessary to delimit the importance of the present study as an initiative that examines factors that permeate the Representation and the Information Retrieval in the Archivology, taking as areas of dialogue the IS and the CS. In this way, the general objective was to examine how Information Representation and IR is delineated in the archival context, taking into account the performance of areas such as IS and CS in these themes. The specific objectives were set out in: a) Mapping common aspects of the Information Representation and Archives IR with related areas, such as IS, as well as CS; b) Analyze the central ideas of Information Representation and IR in the archival area; c) Discuss how the relationship between Information Representation and IR is evident in Archivology; and d) Reflect technical / theoretical issues of Representation and IR in the archival area. Methodologically, the technique of content analysis was used in the search for an understanding of the concepts discussed in the present study. It was evidenced that the Information Representation is noted from the administrative, segmental, intellectual and descriptive action, with elaboration of instruments for the provision of access. In the case of IR, it is known as a computerized process, with a focus on meeting information needs.

Keywords: Information Retrieval. Information Representation. Information Science. Computer Science. Archivology.

LISTA DE QUADROS

Quadro 1 – Tipos básicos de instrumentos de pesquisa da Arquivologia	47
Quadro 2 – Modelos clássicos de Recuperação da Informação	61
Quadro 3 – Obras analisadas que referenciam a Representação e a RI nas áreas de CI, CC e Arquivologia	72
Quadro 4 – Definições do conceito de Classificação arquivística	81
Quadro 5 – Definições do conceito de Indexação arquivística	83
Quadro 6 - Definições do conceito de Descrição arquivística	84
Quadro 7 - Definições do conceito de Recuperação da Informação arquivística	85

LISTA DE FIGURAS

Figura 1 - Camadas da Web Semântica	41
Figura 2 - Relacionamento entre recursos	42
Figura 3 - Processo de Representação nos arquivos.....	45
Figura 4 - Uma taxonomia de modelos de RI	61
Figura 5 - Modelo conceitual dos sistemas arquivísticos de recuperação da informação.....	68
Figura 6 - Interfaces entre a CC, CI e Arquivologia na Representação da Informação	77
Figura 7 - Interfaces entre a CC, CI e Arquivologia na Recuperação da Informação	79

LISTA DE ABREVIATURAS E SIGLAS

ASK	<i>Anomalous State of Knowledge</i>
CC	Ciência da Computação
CI	Ciência da Informação
CIA	Conselho Internacional de Arquivos
DBTA	Dicionário Brasileiro de Terminologia Arquivística
IA	Inteligência Artificial
ISO	<i>International Organization for Standardization</i>
ISAAR(CPF)	<i>International Standard Archival Authority Record for Corporate Bodies, Persons and Families</i>
ISAD(G)	<i>General International Standard Archival Description</i>
ISDF	<i>International Standard for Describing Functions</i>
JSON	<i>JavaScript Object Notation</i>
KWIC	<i>Key Word in Context</i>
NOBRADE	Norma Brasileira de Descrição Arquivística
RDF	<i>Resource Description Framework</i>
RI	Recuperação da Informação
SAA	<i>Society of American Archivists</i>
SIGAD	Sistema Informatizado de Gestão Arquivística de Documentos
TTD	Tabela de Temporalidade de Documentos
UNISIST	<i>United Nations International Scientific Information System</i>
URI	<i>Uniform Resource Identifier</i>
WWW	<i>World Wide Web</i>
XML	<i>eXtensible Markup Language</i>

SUMÁRIO

1	INTRODUÇÃO	14
2	METODOLOGIA	19
3	FUNDAMENTAÇÃO TEÓRICA	23
3.1	Arquivo: histórico, elementos e noções	23
3.1.1	Breve histórico	23
3.1.2	Documento de arquivo: conceitos e definições	27
3.1.3	A necessidade informacional e documental no campo arquivístico	29
3.2	Representação da Informação	34
3.2.1	Representação da Informação: a visão da Ciência da Informação	34
3.2.2	Representação da Informação: aspectos representacionais na Ciência da Computação	38
3.2.3	Representação da Informação: ponto de vista arquivístico	44
3.3	Recuperação da Informação	50
3.3.1	Breve histórico	50
3.3.2	Recuperação da Informação: ponto de vista da CI	55
3.3.3	Recuperação da Informação: ponto de vista da CC	59
3.3.4	Recuperação da Informação: ponto de vista arquivístico	63
4	RESULTADOS E DISCUSSÕES	72
4.1	Pontos de contato da Representação e RI entre a CI, a Arquivologia e a CC	72
4.2	Ideias centrais da Representação da Informação e Recuperação da Informação (RI) no campo arquivístico	81
4.3	Relações da RI com a Representação no campo arquivístico	87
4.4	Questões de ordem técnica/teórica da Representação e Recuperação da Informação no campo arquivístico	88
5	CONSIDERAÇÕES FINAIS	93
	REFERÊNCIAS	95

1 INTRODUÇÃO

A Recuperação da Informação (RI) unida à Representação da Informação, no contexto de frequente dinamicidade tecnológica e grande produção de dados/informações, são considerados tópicos de relevante interesse de áreas como a Ciência da Informação (CI), a Arquivologia e áreas correlatas, como a Ciência da Computação (CC), sendo justamente essa relevância o fator motivador ao presente estudo. Como contextualização inicial, podemos compreender a RI, independentemente da área que a investiga, como a capacidade de suprir demandas informacionais mediante o emprego de expressões de busca (sequências de caracteres) em consultas a Sistemas de Recuperação da Informação (SRI), geralmente operacionalizados em ambientes computacionais. De acordo com Cardoso (2004), a Recuperação da Informação tem demonstrado significância maior diante dos desafios postos a cada dia, percebidos pela necessidade de gestão de um volume cada vez maior de documentos, com correspondente volume de consultas.

Assim, o uso intenso de Tecnologias Digitais de Informação e Comunicação (TDIC) no cotidiano das organizações impulsiona o desafio da RI em ambientes institucionais, principalmente, no uso potencial de informações em operações de ordem administrativa. Nessa perspectiva, a RI depende impreterivelmente de outro tema de frequente estudo: a Representação da Informação. A todo esse contexto de contínua e intensa produção de informações – evidenciado aqui a partir de documentos – aliado a outros desafios que vão sendo postos às disciplinas documentais, como a Arquivologia e áreas afins (ARAÚJO, 2013b) tais como a CI, estabelecem problemáticas que se examinam tanto por estudiosos de uma, como da outra área.

O primeiro desses desafios é colocado pelas discussões que emanam do objeto “informação”, em que invariavelmente se envolvem temáticas que (in)diretamente se relacionam com a sua representação e recuperação da informação. Logo, ao se visualizar o conceito de informação no contexto da CI, percebe-se a existência de uma variedade de discussões, entre as quais uma entende a informação como informação-como-processo, informação-como-conhecimento e informação-como-coisa (BUCKLAND, 1991), bem como a que

percebe e discute a informação a partir dos paradigmas físico, cognitivo e social (CAPURRO, 2003).

Desse modo, a existência dessas discussões sobre a informação corrobora para ampliar o leque de reflexões para além, por exemplo, de questões de natureza tanto objetivista como subjetivista (SILVA *et al*, 2017) em temas de estudo relacionados à informação, a exemplo da RI. Assim, ao incorrer no exame desse conceito, no contexto da Arquivologia, percebe-se que a definição vivencia influências de estudiosos e teóricos da CI, compartilhando da noção de contexto social e da informação como objeto de interesse tanto para arquivistas quanto para cientistas da informação (MARQUES *et al*, 2013). Ademais, cientistas da informação e arquivistas vislumbraram também na representação da informação uma ponte como um fator de forte relação para consecução da RI na posteridade (MEDEIROS; BRÄSCHER; VIANNA, 2017).

Diante da existência de paradigmas, conceitos e questões advindas da CI, a Representação da Informação transparece na Arquivologia em funções como a classificação, a indexação e também naquela que é mais aparente em termos de conhecimento da prática arquivística, denominada de “descrição documental” ou “descrição arquivística”. Tais práticas levam em conta o uso de normativas, instrumentos e fazeres que delineiam a sua atuação, como a classificação, a descrição documental e a indexação. No caso da classificação, a partir do uso de um plano. No caso da descrição documental, é perceptível a adoção de uma norma que estabelece como se representa uma unidade documental (item documental, série, subsérie, fundo etc.), levando em conta a configuração de acordo com questões contextuais e de respeito a princípios arquivísticos. Ainda que algumas dessas normas, a exemplo da ISAD(G)¹ (*General International Standard Archival Description*), tenham sua aplicação naturalmente focada nos arquivos permanentes, há de se estabelecer que a descrição também possa ser efetuada nas primeiras fases documentais – aqui, referindo-se à fase corrente – como Heredia Herrera (1991, p.300, tradução nossa)² salienta: “A descrição é necessária tanto em um arquivo administrativo quanto em um arquivo histórico, embora sua prática possa apresentar algumas variantes”.

¹ Norma geral internacional de descrição arquivística, aprovada pelo Conselho Internacional de Arquivos (CIA).

² “La descripción es necesaria tanto en un archivo administrativo como en un archivo histórico, aunque su práctica pueda presentar algunas variantes”.

Em vista disso, com a necessidade de representação de documentos e estabelecimento cada vez mais frequente de sistemas computacionais para manuseio desses registros documentais, gerados ou recebidos no andamento de suas respectivas atividades, se evidencia a importância do uso de normas arquivísticas de descrição para representação e posterior recuperação da informação – ou de documentos. Soma-se a esse cenário a diversidade de tipos de usuários existentes nos diferentes contextos e lugares de uma organização, com experiências linguísticas diversas; além também da rápida introdução de novas tecnologias no cotidiano organizacional. Assim, a partir dessas colocações, lança-se mão da seguinte questão: como se delinea a Representação da Informação e a Recuperação da Informação na Arquivologia, levando em conta a atuação de áreas como a CI e a CC nessas temáticas?

Dessa maneira, para o estabelecimento dos objetivos do presente estudo, se apresentaram as noções de Recuperação e RI nas áreas da CI, CC e Arquivologia, evidenciando-se como as duas primeiras áreas podem compor em tais temas possíveis diálogos, além de frisar questões peculiares desses temas no campo arquivístico. Assim, elegeu-se como objetivo geral neste trabalho examinar como se delinea a Representação da Informação e a Recuperação da Informação no contexto arquivístico, levando em conta a atuação de áreas como a CI e a CC nesses temas.

Para o alcance do objetivo geral, desdobramos a investigação nos seguintes objetivos específicos:

- a) Mapear aspectos em comum da Representação da Informação e RI da Arquivologia com áreas correlatas, como a CI, bem como com a CC;
- b) Analisar as ideias centrais da Representação da Informação e RI na área arquivística;
- c) Discutir como a relação entre Representação da Informação e RI se evidencia na Arquivologia;
- d) Refletir questões de ordem técnica/teórica da Representação e RI na área arquivística.

A motivação do presente estudo se deu a partir da necessidade de discutir como a CI, ao estabelecer vieses em comum de atuação com a Arquivologia, principalmente nos aspectos informacionais de documentos, dialogava com outras

áreas do conhecimento no tocante às temáticas de representação e RI. Na possibilidade de existência desse diálogo, buscou-se entender como este se consolidava ou se expressava nos estudos da área arquivística. Dessa maneira, incorreu-se na escolha da CI, pela relação estabelecida ao longo do tempo, com o oferecimento, por parte dessa área, de possibilidades de construção de conhecimentos científicos na Arquivologia (ARAÚJO, 2011) e pelo fato de a CI ter investigações científicas na Representação e RI desenvolvidas ao longo de sua trajetória. Com relação a CC, verificou-se que tal área já possuía um desenvolvimento da temática de RI, e sendo o uso da tecnologia recurso comum às práticas da Arquivologia, bem como da CI, verificou-se que esta poderia constituir outra área para diálogo.

A partir de tais colocações, é preciso demarcar a importância do presente estudo como uma iniciativa que visa compreender fatores que permeiam a Representação e a RI, tomando como referenciais teóricos da área da Ciência da Informação, como também aspectos da Ciência da Computação, que podem proporcionar reflexões ao contexto arquivístico; além de buscar entender como podem se estabelecer eventuais relações de tais áreas entre si. A investigação busca também compreender as noções de RI e de Representação da Informação na Arquivologia, tendo como planos de fundo áreas correlatas como a Ciência da Computação e Ciência da Informação, na pretensão de evidenciar possíveis interfaces entre tais áreas. Assim, o referido trabalho busca, principalmente na CI, contribuir com o arcabouço conceitual, no intuito de ampliar o conhecimento de temas como a Representação e a RI em arquivos, visando estabelecer os elementos que as permeiam, as constituem e as influenciam.

Alinhando-se aos objetivos pretendidos, o presente trabalho se apresenta com a seguinte estrutura: 1) Introdução, contendo a contextualização dos temas estudados, os objetivos, o problema da pesquisa e a devida justificativa; 2) Metodologia, explicitando as razões e diretrizes metodológicas utilizadas para a consecução do trabalho; 3) Fundamentação teórica, organizada com os seguintes capítulos: “Arquivo: histórico, noções e elementos”, com explanação do contexto histórico, dos elementos e definições que constitui o conceito de arquivo e das razões de sua existência; “Representação da Informação”, em que são interpostas tanto as visões de áreas como a CI e a CC, bem como da Arquivologia no que toca ao conceito de Representação da Informação; e “Recuperação da Informação”, em

que seus subcapítulos apresentam as concepções de áreas como a CI, a CC e a Arquivologia referentes ao tema de mesmo nome do capítulo; 4) Resultados e discussões, no qual são expostos os dados que subsidiaram o estudo com posteriores análises; e por fim, Considerações finais, em que são observadas questões que o estudo demonstrou e possíveis caminhos para melhor investigação dos fenômenos analisados.

2 METODOLOGIA

O percurso metodológico do presente trabalho se consolidou, inicialmente, quanto aos objetivos como uma pesquisa exploratória. Por conseguinte, quanto aos procedimentos, a investigação foi definida como uma pesquisa bibliográfica. Para Gil (2008, p.27), as pesquisas exploratórias “têm como principal finalidade desenvolver, esclarecer e modificar conceitos e idéias, tendo em vista a formulação de problemas mais precisos ou hipóteses pesquisáveis para estudos posteriores”. O autor ainda ressalta que a pesquisa bibliográfica

[...] é desenvolvida a partir de material já elaborado, constituído principalmente de livros e artigos científicos. Embora em quase todos os estudos seja exigido algum tipo de trabalho desta natureza, há pesquisas desenvolvidas exclusivamente a partir de fontes bibliográficas. Parte dos estudos exploratórios podem ser definidos como pesquisas bibliográficas, assim como certo número de pesquisas desenvolvidas a partir da técnica de análise de conteúdo. (GIL, 2008, p.50)

Com a intenção de fornecer um corpo metodológico à presente pesquisa, elegemos como método de investigação, para execução dos objetivos, a Análise de Conteúdo, na perspectiva de Laurence Bardin (1977). Para a autora, a análise do conteúdo se define a partir de técnicas de análise das comunicações e uso de procedimentos, sistemáticos e objetivos de descrição do conteúdo das mensagens. Há a definição de indicadores (quantitativos ou não) que permitem a inferência de conhecimentos relativos à produção/recepção (variáveis inferidas) dessas mensagens (BARDIN, 1977).

Na perspectiva acima, a Análise de Conteúdo possui diferentes fases, que se “organizam em torno de três pólos cronológicos: 1) a pré-análise; 2) a exploração do material; 3) o tratamento dos resultados, a inferência e a interpretação” (BARDIN, 1977, p.95). A primeira dessas fases é a pré-análise e pode ser compreendida como uma etapa de organização, correspondendo a um período de intuições, sistematizando as ideias iniciais, com intuito de conduzir a um esquema preciso de desenvolvimento de sucessivas operações, num plano de análise (BARDIN, 1977). Essa fase possui atividades a serem desempenhadas. São elas: 1) a leitura flutuante; 2) a escolha dos documentos;

3) a formulação das hipóteses e dos objetivos; 4) a referenciação dos índices e a elaboração de indicadores; e a 5) preparação do material.

A leitura flutuante se desdobrou em livros e artigos científicos da Ciência da Informação, da Arquivologia e da Ciência da Computação que referenciasse a Representação e a RI. Dessa forma, sucedeu-se a próxima atividade, que foi a escolha dos documentos. Nesta, selecionou-se obras ligadas aos conceitos de Representação e RI nas áreas aqui referenciadas. Empreendeu-se buscas em periódicos e livros, tomando como critério para escolha dos textos a regularidade em que são referenciados por pesquisadores da CI, CC e Arquivologia. Ao total, foram coletados dezenove (19) textos, datados de 1976 a 2018. Posteriormente, houve a referenciação dos índices, que são marcações ocorrentes em determinado recorte textual. Nesse caso, optou-se por escolher índices denominados “Representação da Informação”, “Classificação”, “Descrição”, “Indexação” e “Recuperação da Informação”.

Após a escolha dos índices, foi efetuado o recorte do material, visando explorar os conceitos e as definições das temáticas investigadas no presente estudo. A partir do recorte, houve o procedimento de definição de unidade de registro e de unidade de contexto. Quanto à primeira, foi escolhido o tema. Já a segunda, como forma de desambiguação, se consolidou pelo parágrafo, respectivamente. Para melhor compreensão, podemos entender a unidade de registro como a “unidade de significação a codificar e corresponde ao segmento de conteúdo a considerar como unidade de base, visando à categorização e a contagem frequencial” (BARDIN, 1977, p.104). No que concerne à unidade de contexto, esta se entende como a

unidade de compreensão para codificar a unidade de registo e corresponde ao segmento da mensagem, cujas dimensões (superiores às da unidade de registo) são óptimas para que se possa compreender a significação exacta da unidade de registo. (BARDIN, 1977, p.107)

Definidas a unidade de registro e de contexto, executou-se a atividade de dispor os dados coletados em figuras e tabelas, como forma para tabulação dos dados. A partir daqui, houve o início da fase de exploração do material, com leitura e análise tanto das abordagens de Representação e RI na CI, CC e Arquivologia, bem como das definições dessas temáticas. Posteriormente, foi feito o tratamento dos resultados, a inferência e a

interpretação, que ocorreram quase que simultaneamente à fase de exploração do material. Logo, com os dados organizados, empreendeu-se os processos de inferência e interpretação dos conceitos e definições. Como maneira de analisar esses elementos, optou-se por uma abordagem da Análise do Conteúdo chamada de análise temática, corroborando com a escolha da unidade de registro. Para Bardin (1977, p.105), “Fazer um análise temática, consiste em descobrir os «núcleos de sentido» que compõem a comunicação e cuja presença, ou frequência de aparição podem significar alguma coisa para o objectivo analítico escolhido”.

Nesse sentido, o empreendimento da análise temática fez emergir as categorias *a posteriori*, a partir do exame dos conceitos e definições de Representação e RI provenientes da CI, CC e Arquivologia. Tais categorias surgiram com base na leitura e interpretação dos achados teóricos das três áreas e das definições encontradas no contexto arquivístico. A análise temática se deu, portanto, a partir da visualização dos sentidos mais aparentes dispostos nas unidades de codificação (registro e contexto). No caso das definições, com aquelas que se notabilizavam pelo mesmo sentido temático de outras, houve a união destas numa mesma categoria.

Conforme a ordem dos objetivos específicos, o objetivo de “a) Mapear aspectos em comum da Representação da Informação e RI da Arquivologia com áreas correlatas, como a Ciência da Informação, bem como com a Ciência da Computação” foi alcançado a partir da reflexão das noções de Representação e Recuperação na Ciência da Informação, Ciência da Computação e Arquivologia, procurando por meio da Análise de Conteúdo, verificar em suas abordagens eventuais aparições de termos ou sentidos em comum entre elas. As noções dessas áreas se basearam naquelas contidas nos subcapítulos 3.2.1, 3.2.2, 3.2.3, 3.3.2, 3.3.3, 3.3.4, que versam sobre entendimentos da Ciência da Informação, Ciência da Computação e Arquivologia, no que concernem os objetos de análise do presente estudo (Representação e Recuperação da Informação).

Para a realização do segundo objetivo específico, que é “b) Analisar as ideias centrais da Representação da Informação e RI na área arquivística”, buscou-se evidenciar o sentido temático envolto nas definições dos conceitos,

com identificação do tema geral resumido em categorias, conforme demonstrado no capítulo 4: “RESULTADOS E DISCUSSÕES”.

Para cumprir o objetivo de “c) Discutir como a relação entre Representação da Informação e RI se evidencia na Arquivologia” se efetuou, a partir das definições postas no objetivo b), na busca por expressões que colocassem a recuperação da informação em definições da área de representação e vice-versa. A partir da ocorrência de tais temas, efetuou-se a discussão com fim ao cumprimento de tal objetivo.

E com vistas ao cumprimento do objetivo de “d) Refletir questões de ordem técnica/teórica da Representação e RI na área arquivística”, lançou-se questões e debates que cercam tais temáticas na área arquivística, no intuito de enriquecer a reflexão sobre tais assuntos, com referências aos aspectos peculiares da Arquivologia.

Assim, desdobrou-se o percurso metodológico do presente trabalho com vistas à compreensão da Representação da Informação e da RI no contexto arquivístico, expondo o estabelecimento, bem como o eventual relacionamento de uma com a outra.

3 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, abordamos um breve histórico dos arquivos, fundamentando o seu contexto por meio dos elementos que o constituem, junto com as noções que se consolidam com a sua evolução e uso. Há também o desenvolvimento do conceito de Representação da Informação e RI, com apresentação de tais temáticas no tocante à Ciência da Informação, à Ciência da Computação e à Arquivologia, respectivamente.

3.1 Arquivo: histórico, elementos e noções

O presente subcapítulo tem a intenção de contextualizar o histórico do que convencionalmente se chama de arquivo, bem como apresentar os elementos que compõe tal conceito. Busca também situar as noções presentes em seu entorno, conforme vistas adiante.

3.1.1 Breve histórico

A produção de registros materiais faz parte da cultura humana e existe desde que houve a necessidade do registro de fatos, sejam eles acontecimentos ou ações sociais (transações comerciais, direitos e deveres, descobertas científicas, literatura). Nesse sentido, a história dos arquivos data em torno de seis milênios a.C e se consolida junto com o nascimento da escrita nas regiões do Crescente Fértil (RIBEIRO, 2005). A ascensão da escrita é, partindo da ideia de necessidade de memorizar fatos da vida social, fator decisivo tanto para a produção de obras literárias, bem como para a administração (GAGNON-ARGUIN, 1998). Com as civilizações da Antiguidade oriental – sumérios, babilônios, egípcios, assírios – os locais utilizados para guarda de documentos possuíam funções exclusivamente atreladas a autoridades. Sobre isso, Bellotto (2002) afirma:

[...] eram arquivos reais, religiosos, diplomáticos, onde se achavam leis, tratados, normas e preceitos, tendo sido encontrados, nas escavações arqueológicas, também documentos privados como contas, receitas terapêuticas etc (BELLOTTO, 2002, p.13).

Na Grécia antiga, o arquivo desempenhara a função de assegurar a autenticidade dos documentos. O *archéion*, localizado no *Métroon*, era a denominação utilizada para nomear os locais de guarda de documentos do Estado. Esses documentos podiam ser acessados por qualquer cidadão, permitindo-se a cópia desses registros (COUTURE; ROUSSEAU, 1998; BELLOTTO, 2002). Da mesma maneira, ocorria em Roma, com a utilização do arquivo como dispositivo que adquiriu grande importância a partir do direito romano, por causa do peso do ato escrito na sua normativa (BELLOTTO, 2002). Era a perpetuação do controle de atividades dos cidadãos pelo Estado. Gagnon-Arguin (1998, p. 30) fortalece a noção de na Antiguidade Clássica – com uma administração mais “democrática” (a exemplo de Atenas) – “[...] os cidadãos têm actividades independentes do governo e onde este dirige por intermédio de instituições que devem, em contrapartida, **inscrever testemunhos das suas acções** (grifo nosso)”.

Na Idade Média, o uso de documentos e, conseqüentemente, das estruturas que conhecemos por arquivos, constituíram-se como lugar onde eram depositados atos ou peças que asseguravam títulos para a posse de terras e de direitos (GAGNON-ARGUIN, 1998). Tais atos tinham a forma de diplomas e findavam no que se chamou de *Trésor des chartes*, importante estrutura de arquivos da coroa Francesa. Para Heloísa Bellotto (2002, p.14),

com a vinda dos povos bárbaros para a Europa ocidental, a decadência da vigência do direito romano e das autoridades centralizadas, os arquivos circunscrevem-se às autoridades feudais. No século XII, passa a haver um certo renascimento das provas documentais escritas, surgindo **os títulos e registros como armas jurídicas, sobretudo relativamente aos direitos sobre as propriedades de terras, dando origem aos arquivos de autoridades feudais** (Grifo nosso).

É nessa época, com o poderio da Igreja, que se acumularam documentos nos arquivos dos mosteiros e instituições eclesiásticas, sob a batuta de monges e sacristãos. Com a grande influência da Igreja na condução de atividades do Estado e da vida social de forma geral, e a partir de relações de reis, príncipes e senhores com autoridades eclesiásticas, tais nobres confiavam à Igreja a guarda de documentos e o registro de contratos sociais (SCHELLENBERG, 2006). Sobre isso, Porto (2013) destaca:

A igreja era uma das privilegiadas nesse sentido, devido ao poder que tinha e ao cargo que ocupava dominava o manuseamento e conservação dos documentos, geria e continha os saberes da época, bem como os documentos guardados em mosteiros e catedrais, denominados por Arquivos Eclesiásticos. Nestes arquivos era feita a guarda e gestão da documentação, não só da Igreja, **mas também de instituições particulares e públicas**. Ficavam a cargo dos religiosos como o Papa, os frades, os abades e monges. (PORTO, 2013, p.17-18, grifo nosso)

É válido lembrar que o acesso aos arquivos eclesiásticos só se concedia a autoridades. A partir da Idade Moderna, com o advento da ideia de estado moderno, passou-se a constituir os grandes arquivos reais (“tesouros do rei”), que continham documentos estatais e arquivos notariais devidamente organizados (BELLOTTO, 2002). Tais iniciativas são, segundo Gagnon-Arguin (1998), elementos que constituirão a noção de arquivos do Estado. É Carlos V, Imperador Romano-Germânico, que inicia o movimento de transferir o tesouro dos diplomas de Castela para Simancas, “constituindo deste modo o núcleo principal dos arquivos do estado de Castela” (GAGNON-ARGUIN, 1998, p. 36). Os arquivos nessa época são utilizados num contexto totalmente jurídico-administrativo, não sendo ainda destinados para a pesquisa histórica (BELLOTTO, 2002).

Adiante, na Idade Contemporânea, têm-se grandes transformações no tocante aos arquivos, enquanto mantenedores de documentos e dispositivos institucionais do Estado. A Revolução Francesa (1789-1799) é um importante marco na história dos arquivos, como explica Bellotto (2002, p.14), segundo a qual “[...] ocorre uma certa abertura dos arquivos públicos aos cidadãos e se procede a reunião da documentação oficial dispersa, em Paris, criando-se um arquivo nacional”. Considerado o primeiro arquivo nacional criado no mundo, o *Archives Nationales de Paris* se notabilizou por guardar documentos da Nova França. Schellenberg (2006) esclarece:

Um decreto de 25 de junho de 1794 estabeleceu, em todo o território nacional, uma administração dos arquivos públicos. Por esse decreto os Archives Nationales passaram a ter jurisdição sobre os documentos dos vários órgãos do governo central, em Paris, os quais mantinham, até então, seus próprios depósitos de arquivos sobre os documentos das províncias, comunas, igrejas, hospitais, universidades e famílias nobres e sobre os arquivos distritais nos quais foram colocados, durante a Revolução, os documentos dos órgãos municipais extintos (SCHELLENBERG, 2006, p.26-27).

O interesse pela pesquisa histórica em arquivos surge a partir do século XIX, em decorrência da influência que o Positivismo exercia na época. Os efeitos perversos que emanaram dessas práticas, que viam fortemente o “teor histórico” que os arquivos possuíam, foram notáveis nas ações de afastamento dos arquivos das administrações produtoras (RIBEIRO, 2011). O crescimento artificial de arquivos com fins históricos exerceu irreparáveis danos às funções administrativas previstas no conceito de arquivo.

A edição do “Manual Holandês”, da autoria de S. Muller, J. H. Feith e R. Fruin (1898), marcou o início de um novo período, em que o predomínio da vertente técnica se vai afirmar definitivamente, libertando a teoria e a prática dos arquivos da dependência directa das concepções historiográficas (RIBEIRO, 2011, p.61).

O manual dos holandeses é considerado outro marco na história dos arquivos, pois foi a partir de tal publicação que se passou a questionar concepções na maneira de organização destes e na forma de pensar a Arquivologia (denominada também de Arquivística) como área de estudos emergente, com enfoque conceitual e técnico sobre a organização de arquivos. Baseados no advento da Arquivística surgiram, ao longo do tempo, inúmeras escolas arquivísticas, concebendo cada uma suas contribuições nos campos conceituais e técnicos e refletindo tais noções nas organizações dos arquivos locais. Velhos e novos desafios foram e continuam sendo postos ao cotidiano dos arquivos, principalmente no tocante a novas formas de comunicação social, incluindo o advento das TDIC, influenciando a reflexão sobre a natureza do documento.

Assim, foi possível observar as noções que permearam o conceito de arquivo ao longo de diferentes épocas e civilizações. No intuito de salvaguardar os registros que atestavam os inúmeros atos e fatos da vida social (incluindo-se aqui documentos institucionais, bem como pessoais), coloca-se, nesse espaço, o protagonismo dos arquivos que passaram a exercer a função de locais onde tais registros eram guardados. Esses registros – denominados no presente estudo por documento arquivístico – servem como produção de um meio com o intuito de salvaguardar dados e ou informações, para usos diversos e comprovação de ações em geral (atividades, funções, eventos etc.). Essa denominação e conceitos pertinentes ao documento arquivístico são o fio condutor neste trabalho.

3.1.2 Documento de arquivo: conceitos e definições

Os arquivos têm a sua noção baseada em um conjunto de registros que consolidam uma ação ou resguardam a memória daquela ao longo do tempo. Os documentos de arquivo serviram, em determinados momentos, de marcações de poder, analogamente, como os gabinetes de curiosidades, da elite inglesa do século XVII (FROHMANN, 2009). Esses registros, denominados no presente estudo de documentos de arquivo, foram e ainda são expressões e veículos de poder, direitos e outras questões, como assevera Gagnon-Arguin (1998, p.32):

Ao longo das épocas e dos regimes, os documentos serviram para o exercício do poder, para o reconhecimento dos direitos, para o registo da memória e para a sua utilização futura. Recordamo-lo aqui porque este papel foi muitas vezes eclipsado pela história dos homens e das sociedades que, no entanto, basearam os seus poderes e a sua perenidade nesses documentos.

Após apresentar sumariamente as razões dos documentos de arquivo possuírem suas significâncias em diferentes tempos, lançamos mão, a partir daqui, do conceito de “documento arquivístico”. Desse modo, ao basear-se no entendimento comum, o senso popular evidencia o documento de arquivo como pertencente ao gênero textual, majoritariamente apresentado nas formas de processos, dossiês e mesmo avulsos, com ampla variedade de tipos documentais. Ainda que documentos arquivísticos institucionais possam pertencer, em sua maioria, ao gênero textual, o conceito é ampliado pelo Dicionário Brasileiro de Terminologia Arquivística (ARQUIVO NACIONAL, 2005, p. 27) que não conceitua diretamente o termo “documento arquivístico”, mas parte do pressuposto de que “arquivo” é o “Conjunto de documentos produzidos e acumulados por uma entidade coletiva, pública ou privada, pessoa ou família, no desempenho de suas atividades, independentemente da natureza do suporte”.

Logo, o documento arquivístico é conceituado de múltiplas maneiras por diferentes escolas e visões presentes na Arquivologia. Couture e Rousseau (1998, p.137), na perspectiva da arquivística canadense, caracterizam o documento arquivístico como aquele que “[...] contém uma informação seja qual for a data, forma e suporte material, produzidos ou recebidos por qualquer

pessoa física ou moral, e por qualquer serviço ou organismo público ou privado, no exercício da sua actividade.”

Duranti (1994), em reflexão sobre a natureza e as características dos documentos arquivísticos, denominados no estudo de “documentos contemporâneos”, descreve que estes são

gerados ou recebidos no curso das atividades pessoais ou institucionais, como seus instrumentos e subprodutos, os registros documentais são as provas primordiais para as suposições ou conclusões relativas a essas atividades e às situações que elas contribuíram para criar, eliminar, manter ou modificar (DURANTI, 1994, p.50)

Cruz Mundet (1994), em seu manual de arquivos, conceitua “documento arquivístico” baseado num grupo de elementos diferenciadores que são:

O caráter serial: os documentos são produzidos um por um e com a passagem do tempo constituem séries (correspondências, minutos...); a gênese: elas ocorrem dentro de um processo natural de atividade, surgem como um produto e refletem as tarefas de seu produtor, não são algo estranho a ele; exclusividade: a informação que contém é raramente encontrada em outro documento com a mesma duração e intensidade, é exclusiva; a inter-relação: como princípio geral peças isoladas (documentos soltos) não têm sentido ou têm muito pouco, sua razão de ser é dada pelo pertencimento a um conjunto – a unidade de arquivo ou arquivo – e pelas relações estabelecidas entre eles³ (CRUZ MUNDET, 1994, p. 99-100, tradução nossa).

Yeo (2008) evidencia o documento arquivístico a partir das ideias de persistência e representação. Na visão do autor, os documentos arquivísticos são “representações persistentes de atividades ou outras ocorrências criadas por participantes ou observadores dessas ocorrências ou por seus procuradores [representantes legais]”⁴ (YEO, 2008, p.136, tradução nossa). A perspectiva de Yeo procura ampliar o conceito de documento arquivístico ao

³ *“El carácter seriado: los documentos se producen uno a uno y con el paso del tiempo constituyen series (correspondencias, actas...); la génesis: se producen dentro de un proceso natural de actividad, surgen como producto y reflejo de las tareas de su productor, no son algo ajeno a él; La exclusividad: la información que contiene rara vez se encuentra en otro documento con idéntica extensión e intensidad, es exclusiva; la interrelación: como principio general las piezas aisladas (documentos sueltos) no tiene sentido o tienen muy poco, su razón de ser viene dada por su pertenencia a un conjunto - la unidad archivística o expediente – y por las relaciones establecidas entre si”.*

⁴ *“persistent representations of activities or other occurments, created by participants or observers of those occurments or by their proxies”.*

ênfatizar a ideia de ocorrências e resumir, a partir de “representações persistentes”, o uso para posteridade do documento.

Das perspectivas lançadas acima, pode-se presumir que o documento arquivístico é parte de um conjunto (organicidade e indivisibilidade), produzido e acumulado (cumulatividade) por uma entidade (proveniência) (ARQUIVO NACIONAL, 2005; BELLOTTO, 2007). Os princípios da organicidade, da indivisibilidade, da cumulatividade e da proveniência, bem como da unicidade (caráter e função únicos que um documento assume em razão do contexto em que foi produzido) são assumidos mais explicitamente (BELLOTTO, 2007; SANTOS, 2011; BELLOTTO, 2016) por algumas definições e mais implicitamente por outras (ARQUIVO NACIONAL, 2005). Assim, por si só caracterizam as peculiaridades dos documentos arquivísticos. Tais princípios estarão intrínsecos também em boa parte das definições da área, ainda que em diferentes escolas arquivísticas.

Fica claro, a partir da exposição de diversos conceitos provenientes de diferentes autores, o caráter orgânico, natural e de conjunto do documento arquivístico, produzido ou recebido no exercício de atividades de uma pessoa ou instituição e guardado como elemento de prova ou de informação. Tais aspectos de prova, informação e testemunho garantem ou justificam a existência e atuação dos arquivos, em diferentes instituições na sociedade.

3.1.3 A necessidade informacional e documental no campo arquivístico

Um aspecto crucial no entendimento da importância dos arquivos diz respeito à capacidade que os registros documentais têm de serem provas primordiais para suposições ou conclusões concernentes a atividades e situações que eles contribuíram para criar, manter, eliminar, dentre outras (DURANTI, 1994). Nessa perspectiva, as razões para consulta, busca e uso em um arquivo transparecem por necessidades tanto informacionais como documentais.

Assim, antes de referenciar a temática de necessidades informacionais, é preciso assinalar que a investigação de tal temática, por estar atrelada ao objeto informação, é feita essencialmente pela Ciência da Informação. Devido à interface existente entre a Arquivologia e a Ciência da

Informação e de ambas perceberem também, segundo Tognoli (2012), a informação-como-coisa (BUCKLAND, 1991), a primeira vem tomando como referência o aporte investigativo e teórico efetuado pela segunda. Dessa maneira, autores como Brenda Dervin, Carol Kuhlthau, Nicholas Belkin, dentre outros, são considerados referências no tocante a “necessidades de informação” e utilizados também em estudos arquivísticos. A título de contextualização, a temática de “necessidades de informação” está inserida no campo macro de estudos de usuários (ARAÚJO, 2013a).

Perante tais assertivas, olhemos à luz da CI para o conceito de necessidades de informação. Dervin e Nilan (1986) mostram algumas definições similares sobre necessidades de informação (*information needs*), elencando diversas definições de diferentes autores.

1) uma incongruência conceitual na qual a estrutura cognitiva da pessoa não é adequada a uma tarefa (FORD); 2) quando uma pessoa reconhece algo errado em seu estado de conhecimento e deseja resolver a anomalia (BELKIN, 1978); 3) quando o estado atual do conhecimento possuído é menor que o necessário (KRIKELAS); 4) quando o sentido interno se esgota (DERVIN, 1977; 1980); e 5) quando há conhecimento insuficiente para lidar com lacunas, incerteza ou conflito em uma área de conhecimento (HORNE)⁵ (DERVIN;NILAN, 1986, p.17, tradução nossa).

Wilson (1981) explica que, na associação de “necessidade” e “informação”, é passada a ideia de necessidade similar a outras necessidades básicas humanas. Na visão do autor, ao examinar a literatura sobre necessidades humanas, há uma divisão por psicólogos em: necessidades fisiológicas, necessidades afetivas e necessidades cognitivas. Nesse sentido, Wilson (1981, p.7-8, tradução nossa) ressalta que

Será rapidamente reconhecido que estas três categorias estão inter-relacionadas: necessidades fisiológicas podem desencadear necessidades afetivas e/ou cognitivas; necessidades afetivas pode originar necessidades cognitivas; e problemas relacionados com a satisfação de necessidades cognitivas (como a incapacidade de

⁵ “‘Information needs’ are similarly defined: 1) a conceptual incongruity in which the person’s cognitive structure is not adequate to a task (FORD); 2) when a person recognizes something wrong in his or her state of knowledge and wishes to resolve the anomaly (BELKIN, 1978); 3) when the current state of possessed knowledge is less than needed (KRIKELAS); 4) when internal sense runs out (DERVIN, 1977; 1980); and 5) when there is insufficient knowledge to cope with voids, uncertainty, or conflict in a knowledge area (HORNE)”.

satisfazer as necessidades ou o medo de revelar necessidades) pode resultar em necessidades afetivas (por exemplo, para reafirmação).⁶

Ainda sobre necessidades de informação, Nicholas Belkin, Robert Oddy e Helen Brooks asseveram que esta é tida como um “estado anômalo de conhecimento” (ASK), em que uma necessidade de informação vem de uma anomalia reconhecida no estado de conhecimento do usuário sobre algum assunto ou situação e que, em geral, o usuário é incapaz de repor o que é necessário para resolver essa anomalia (BELKIN; ODDY; BROOKS, 1982). Baseado nessas concepções é possível compreender a noção de “necessidades de informação” e, assim, fundamentar o entendimento quando relacionado ao campo arquivístico. É necessário antes ressaltar que, conforme Araújo (2011, p.125) relata a respeito da relação da CI com a Arquivologia, “o campo de estudos de usuários de arquivos ainda é pouco desenvolvido, mas pesquisas sobre necessidades de informação foram realizadas por autores como Taylor, Dowle, Pugh e Cox”.

Nesse sentido, ao ocorrer uma transição na abordagem de estudos de usuário centrada no documento para uma abordagem centrada no usuário, um serviço de informação “explicita institucionalmente seus objetivos de atender às necessidades de informação deste. A tomada de decisões relativas ao planejamento e à gestão é orientada sob esta perspectiva” (JARDIM; FONSECA, 2004, p.2). Assim, é recorrente, na prática arquivística, o entendimento que a fase documental (corrente, intermediária ou permanente), no qual o conjunto documental se encontra, determina e/ou evidencia os perfis de usuários desses acervos. Nesse seguimento, à conjuntura dessas fases documentais, dá-se o nome, na Arquivologia, de Teoria das Três Idades. Nessa concepção, Machado e Camargo (1999) definem:

Na **fase corrente**, também conhecida como fase ativa, temos os arquivos imediata e estreitamente ligados às atividades de que se originaram. Por essa razão os arquivos correntes se conservam próximos dos órgãos a que servem, sejam eles responsáveis pelas atividades-fim ou pelas atividades-meio da instituição. A **fase intermediária** corresponde ao período em que os documentos, menos utilizados pelo órgão de origem e já avaliados quanto à sua destinação final, aguardam em depósitos de armazenamento

⁶ “It will be quickly recognized that these three categories are interrelated: physiological needs may trigger affective and/or cognitive needs; affective needs may give rise to cognitive needs; and problems relating to the satisfaction of cognitive needs (such as a failure to satisfy needs, or fear of disclosing needs) may result in affective needs (for example, for reassurance)”.

temporário o cumprimento dos prazos estabelecidos para sua eliminação ou para envio ao arquivo permanente. À terceira fase, chegam os documentos cujas informações são consideradas imprescindíveis para o órgão de origem e para a comunidade que justificou sua existência. O **arquivo permanente** ou histórico – último estágio de guarda, senão definitiva, pelo menos de longa duração – conserva, portanto, os documentos aos quais se atribuiu, no processo de avaliação, a devida relevância, seja para a comprovação de direitos, seja para a pesquisa retrospectiva (MACHADO; CAMARGO, 1999, p.21, grifo nosso).

Com a apresentação do conceito arquivístico de Teoria das Três Idades, é necessário entender os perfis de usuário de cada fase documental. Nas primeiras fases, o público que solicita documentos se limita, geralmente, aos usuários que fazem parte da própria organização (ARQUIVO NACIONAL, 2005). Estes possuem necessidades informacionais correspondentes a: aspectos do próprio exercício laboral ou de atividades, funções e exercícios organizacionais (atividade-meio e atividade-fim); questões que envolvam como interessados a si e a organização no qual é colaborador. Gama e Ferneda (2010, p. 152) enfatizam o valor informacional de documentos na fase corrente, argumentando que,

Durante a vigência do valor primário a informação orgânica está disponibilizada ao produtor, sendo vedado o acesso ao público. Neste contexto a informação registrada nos documentos de arquivo é utilizada para os fins pelos quais foi produzida. O usuário, especialista em uma determinada área de atuação, está inserido no ambiente de produção e conhece a estrutura funcional onde atua.

Antes de situar a fase permanente, é necessário tecer algumas observações. O conceito de fase permanente está constantemente associado aos arquivos estatais. Como se presume, a noção de acesso a tais acervos se consolida – apesar de, nas fases iniciais, existir legislações no Brasil como a lei nº.12.527/2012 (LAI) e exigir o acesso como regra – como público a partir da fase permanente. Entretanto, deve-se refletir e ampliar a aplicação do conceito de “fase permanente” para além dos arquivos do Estado. Isso traz algumas considerações sobre políticas de acesso e grau de restrição que possam ocorrer, afinal, os ativos de uma empresa privada provavelmente não estarão disponíveis para o grande público. Nesse caso, o perfil de usuário continua sendo o das primeiras fases.

Portanto, na fase permanente, no contexto de documentos gerados e recebidos por instituições públicas, os usuários que aí se apresentam são

constituídos por uma ampla variedade de perfis. Isso engloba, por si só, toda a população que, por razões diversas, a fim de atender a demandas de informação próprias, recorre a tais acervos em busca de respostas. Nesse sentido, o caráter probatório e testemunhal (BELLOTTO, 2014), evidenciado nos elementos que compõe o arquivo, se estabelece como motivação para busca e uso de informações em tais lugares. No que toca ao público que busca os arquivos, Bellotto (2014, p.28-29) cita quatro tipos fundamentais de usuários em arquivos públicos, a saber:

1. O administrador, isto é, aquele que produz o documento e dele necessita para sua própria informação, na complementação do processo decisório.
2. O cidadão interessado em testemunhos que possam comprovar seus direitos e o cumprimento de seus deveres para com o Estado.
3. O pesquisador – historiador, sociólogo ou acadêmico – em busca de informações para trabalhos de análise de comportamentos e eventos passados, podendo ser incluído nessa categoria o estudioso em geral.
4. O cidadão comum, aqui não mais o interessado em dados juridicamente válidos, mas o cidadão não-graduado, o aposentado, a dona-de-casa etc. à procura de cultura geral, de entretenimento, campos em que pode haver lugar para o conhecimento da história.

Há autores que se validam da noção de usuário interno e usuário externo, sendo o primeiro membro da organização e o segundo não pertencente à organização (ALLENDEZ SULLIVAN; PUENTE, 2013; VALERO, 2016), com necessidades distintas. Nesse sentido, ao verificar os tipos de usuário acima, pode-se presumir que as necessidades de usuários observadas no arquivo são caracterizadas tanto pelo viés informacional, quanto pelo viés documental. No sentido informacional, é comum que haja necessidades que podem ser satisfeitas por informações registradas nos documentos e notificadas ao usuário. No entanto, existem as necessidades que se validam a partir da ocorrência da informação registrada em um documento. A primeira, no escopo do presente trabalho, é denominada simplificada de necessidade informacional; a segunda, de necessidade documental.

Dessa maneira, pode-se supor que as razões pelo qual os documentos são consultados no arquivo são: 1) a existência de dados registrados no documento, que podem fornecer subsídios pontuais para suprir necessidades informacionais de quem busca; 2) as provas acerca de direitos concedidos a determinados indivíduos; 3) o fornecimento de relações formadas

pelos eventos registrados a partir do qual inferências podem ser formuladas (MEEHAN, 2006). Logo, os pontos acima convergem a um aspecto em comum na prática arquivística: a existência de elementos informacionais nos documentos que podem constituir importantes pontos de acesso. Assim, a Representação da Informação oferece noções conceituais e práticas que estabelecem importante tema de investigação na Ciência da Informação e de grandes contribuições para a Arquivologia.

3.2 Representação da Informação

A Representação da Informação é uma área de estudos investigada tanto pela CI, como pela Biblioteconomia, esta já possuindo longa trajetória em estudos de representação e organização da informação (ARAÚJO, 2010). Nesse sentido, a Arquivologia também considera a Representação da Informação ao buscar representar os seus conjuntos documentais por práticas como a classificação, descrição e indexação (BARROS, 2016). E no tocante a CC, tal área se consolidou como responsável por investigações sobre o processamento de dados, bem como a automatização desses processos pelo uso de TDIC. O uso de representações é uma prática cotidiana da CC. Dessa maneira, compreender as visões distintas dessas áreas é crucial para compreender como elas se relacionam no tema.

Assim, iniciamos a apresentação da noção de Representação da Informação a partir da 1) Ciência da Informação; depois pelo olhar da 2) Ciência da Computação; e posteriormente com o entendimento da 3) Arquivologia.

3.2.1 Representação da Informação: a visão da Ciência da Informação

A faculdade de representação é uma atividade humana e implica em múltiplas questões, tocando tanto aspectos sociais/culturais, cognitivos, dentre outros, que vão delinear o sentido, significado de um determinado símbolo – frase, ícone, som etc. – para um interpretante.

Etimologicamente, representar significa *fazer alguma coisa presente novamente*. A semiótica, conforme formulada por Peirce, descreve a representação como um processo envolvendo um objeto, alguma coisa que o *representa* e o efeito da representação, na ausência do objeto, na mente de um usuário. Representação é, desta maneira, um processo ocorrendo na mente de alguém, produzindo nesta mente algo distinto do objeto a que se refere. A representação então relaciona o objeto que ela representa com a mente que o percebe (MARCONDES, 2001, p.64).

A ideia de representação junto com as concepções de informação e linguagem, mesmo no plano filosófico, fornecem subsídios para a investigação da Representação da Informação na Ciência da Informação (CI) (KOBASHI, 2007). Inicialmente, a Representação da Informação estabeleceu laços com a Biblioteconomia, principalmente na parte responsável por instrumentos de descrição e classificação bibliográficas (ARAÚJO, 2014). Tal área, nesse contexto, esteve relacionada à RI. Sobre isso, Araújo (2014) destaca:

Suas primeiras manifestações na ciência da informação se deram justamente com a recuperação da informação no campo dos sistemas de informação, ainda na década de 1960, com o desafio de buscar a melhor forma de representar a informação (tanto em termos formais quanto de conteúdo) pensando na otimização da sua recuperação. Os estudos nesse momento conduziram ao surgimento de diversos instrumentos de linguagem controlada e de sistemas de classificação, todos com objetivos de alcançar o máximo de economia de custos, diminuição de ruídos, supressão da redundância, aplicação de princípios lógicos. Os sistemas e linguagens criados competiam, nesse momento, para se ver qual era o melhor, sendo exatamente essa questão — **a busca da melhor linguagem, da melhor representação** — o objetivo inicial deste subcampo da ciência da informação (ARAÚJO, 2014, p. 127, grifo nosso).

A Representação da Informação pode ser compreendida como o fazer que visa substituir uma entidade linguística extensa por uma descrição menor, abreviada (NOVELLINO, 1996). Bräscher e Café (2013, p. 6) entendem tal tema “[...] como o conjunto de atributos que representa determinado objeto informacional e que é obtido pelos processos de descrição física e de conteúdo”. Na CI, tal tema de investigação tem estabelecido os seus significados em torno das áreas de Linguística e Semiótica, da Ciência da Comunicação e da Documentação (KOBASHI, 1996), sendo esta última a que possui maior predominância.

É importante destacar que esta predominância é uma das características existentes na área da CI, em que se visualiza a relação com o documento, evidenciado a partir do surgimento da própria CI, com a iniciativa

de bibliotecários especializados em criar uma associação própria, sendo tal iniciativa considerada o embrião da primeira entidade de ciência da informação no mundo (ARAÚJO, 2014). Assim, a relação inicial da Biblioteconomia e Documentação com a CI se estabeleceu numa abordagem de escola pré-paradigmática, com o triunfo de problemáticas e conceitos trazidos da Biblioteconomia para a nova área. Khun (1975) enfatiza que, frequentemente, com o avançar de uma determinada área, as divergências iniciais tendem a sumir e a desaparecer nas áreas que chamamos de ciência. O triunfo de uma determinada escola pré-paradigmática (com seus preconceitos e crenças) parece ser a explicação de tal fato.

Com noção advinda da Biblioteconomia e Documentação, o documento passa a ser um elemento que envolve as práticas da CI, tendo o conceito de documento evidenciado a partir da ideia de informação registrada num suporte (RODRÍGUEZ BRAVO, 2002). Com base em tal premissa – e proximidade da CI com áreas documentais – a Representação da Informação perpassou pela noção de indexação aplicada a documentos. Pinto (2001), parafraseando Jean-Claude Gardin, denomina, nesse contexto, o conceito de indexação documentária, entendido como um conjunto de atividades que visa: identificar nos documentos os seus traços descritivos ou macro-suposições e extrair os descritores (sintagmas) concernentes ao seu conteúdo. Ainda sobre a representação documentária, Lara (1993, p. 73) explicita a sua visão:

A representação documentária deve ser abordada de forma semelhante. Como produto documentário gerado no processo de Análise Documentária, a representação deve "provocar uma interpretação" – através da proposição de uma cadeia de interpretantes frente a uma relação Sujeito-Objeto – e estabelecer algum tipo de relação com o texto que lhe deu origem. Uma representação documentária desencadeia, conseqüentemente, uma situação de comunicação.

Frente a CI e disciplinas documentais, o advento das TDIC proporcionou grande impacto, principalmente no tocante à produção, uso e armazenamento de informação, além de ter lançado desafios no tocante à representação de documentos (e da informação contida em cada um deles) em meio a uma “nova forma de organização social baseada em redes, ou seja, na difusão de redes em todos os aspectos da actividade na base das redes de comunicação digital” (CASTELLS, 2005, p.17).

Na base dessa preocupação, disciplinas documentais como a Biblioteconomia e a Arquivologia foram atingidas pelo contínuo desenvolvimento tecnológico. Rondinelli (2005, p. 24) afirma que “no mundo do trabalho, por exemplo, os profissionais da informação foram profundamente atingidos e, entre eles, os arquivistas”. À CI, área científica de estreito diálogo com as disciplinas documentais, coube a contribuição investigativa, como forma de subsidiar novos enfoques para essas áreas.

É necessário ressaltar que a CI, ao se propor estudar os fenômenos implícitos à produção, circulação e uso da informação, tem o intuito de criar instrumentos e estabelecer metodologias que viabilizem a busca, recuperação e organização de informações (NOVELLINO, 1996). Na concepção de Novellino (1996), a Representação da Informação envolve desde a análise de assunto de um documento –e a posterior colocação do resultado dessa análise em expressões linguísticas– à atribuição de conceitos ao documento apreciado. A partir dessas ações, evidenciou-se instrumentos para a representação da informação como os vocabulários controlados, os tesouros, as listas de cabeçalhos de assuntos, dentre outros.

As experiências em torno desses instrumentos se notabilizaram no contexto de unidades de informação, a exemplo, principalmente, de bibliotecas. A concepção de instrumentos com viés para o documento esbarrou na transição do foco antes dado ao documento e agora voltado para a informação, naquilo que ficou conhecido como passagem da fase custodial para uma pós-custodial (SANTOS, 2011; SOARES; PINTO; SILVA, 2016). Toda essa confluência impactou na noção de Representação da Informação no campo da CI, evidenciando autores conhecidos pela abordagem que visa o sujeito a partir da cognição (DANIELS, 1986; BELKIN, 1990; INGWERSEN, 1992), além do tocante ao uso e transmissão de informações, contextualizados agora na grande rede de computadores, por exemplo.

Desse modo, as tendências da CI na temática da Representação da Informação têm focado seus esforços na: classificação facetada; vocabulário controlado; uso de ontologias; construção de sistemas de classificação a partir de uma perspectiva sociocultural; uso de teorias computacionais para reflexão da representação; dimensão colaborativa do processo de representação da informação; e interação homem-computador (ARAUJO, 2014). Dessa maneira,

ao se contemplar a relação de Representação da Informação com o computador à luz do processamento automatizado e de todas as concepções que surgiram a partir daí, o estudo engendrará, nesse momento, pelo entendimento e atuação da Ciência da Computação (CC).

3.2.2 Representação da Informação: aspectos representacionais na Ciência da Computação

A CC alicerça o seu trabalho, dentre inúmeras maneiras, a partir do uso de representações. A temática de Representação da Informação não é formalmente um campo de investigação verificado na CC. Entretanto, isso não impede que os conceitos de representação utilizados por tal área não sejam contributos para a presente pesquisa. Nesse sentido, é preciso assinalar que o entendimento da CC sobre a representação alicerça ideias que vão desde suas operações mais básicas até tópicos avançados, como por exemplo: a IA (Inteligência Artificial), realidade virtual, redes neurais, dentre outros. Assim, em operações computacionais, a linguagem binária é a noção utilizada para operação de circuitos computacionais, a partir da representação de zero (0) e um (1) (FOROUZAN; MOSHARRAF, 2012).

No entanto, isso não quer dizer que os computadores operam diretamente com 0 e 1. Tais números são símbolos utilizados no intuito de corresponder à ausência e à presença de voltagem e são chamados de *bit*. Dados como números, caracteres, texto, imagem, vídeo, som são representados, em nível de máquina, a partir de tais valores binários, já que representam pulsos elétricos. Entretanto, a notação binária não é a mais apropriada para entendimento humano, já que gera milhões de sequências com zero (0) e um (1) para representar ações básicas no computador. Dessa maneira, a história da informática é marcada pela abstração crescente de representações, buscando, a partir de um formalismo lógico, a representação do conhecimento (FREITAS, 2003). Um exemplo disso é o uso de tecnologias de banco de dados e seus paradigmas (relacional, orientado a objeto, *NOSQL*, dentre outros), além do XML para organização e JSON (*JavaScript Object Notation*) para troca de dados. Com a expansão da rede mundial de

computadores e surgimento de novas maneiras de interação social, o desafio se consolidou na área. Sobre isso, Freitas (2003) esclarece:

A chegada da Internet e dos sistemas abertos – ambientes desorganizados e ricos, que requerem a existência de agentes inteligentes com forte orientação à cooperação e conseqüentemente à comunicação –, têm demonstrado a necessidade massiva de expressividade e de reuso flexível de conhecimento, provocando uma premente revisão de conceitos (FREITAS, 2003, p.2).

Nessa concepção, uma das iniciativas que visaram organizar estruturalmente e imbuir nos dados uma acepção semântica, capaz de tornar esses dados legíveis por máquinas, foi a Web Semântica. Tema também estudado pela Ciência da Informação (SOUZA; ALVARENGA, 2004; RAMALHO, 2006; RAMALHO; VIDOTTI; FUJITA, 2009; ROBREDO, 2010) e inicialmente concebido pela CC, a Web Semântica não pode ser entendida como uma rede à parte, mas como extensão e melhoria da Web atual (no qual a comunicação é mediada por máquinas que não possuem inteligência para prover sentido aos dados), numa proposta de “ligar” dados e criar sentido em torno deles. Para tanto, a concepção de um instrumento que estabeleça qual o limite de sentido para termos, noções e definições em um dado domínio, e como estes se relacionam hierarquicamente, foi batizado pela área de CC com o nome de ontologia.

Para Freitas (2003, p.4), as ontologias servem para “especificação dos conceitos de um determinado domínio e suas relações, restrições e axiomas, definidos de forma declarativa”. Souza e Alvarenga (2004, p. 137) apresentam as ontologias “como um modelo de relacionamento de entidades e suas interações, em algum domínio particular do conhecimento ou específico a alguma atividade”. Ainda para os autores, o compartilhamento de um vocabulário para troca de informações entre membros de uma comunidade (humanos ou agentes inteligentes) é o principal objetivo da ontologia (SOUZA; ALVARENGA, 2004). Para Noy (2003, p. 175, tradução nossa),

uma ontologia para um domínio enumera e fornece descrições semânticas de conceitos no domínio do discurso, definindo atributos de conceitos relevantes ao domínio e vários relacionamentos entre eles. Por exemplo, uma ontologia que descreve um vinho incluirá conceitos como safras, regiões vinícolas, vinícolas, variedades de

uvas e assim por diante. Ele também irá incluir as relações tais como produzido por, feitos de, cor, ano, e o corpo do vinho.⁷

Desse modo, Tim Berners Lee, principal articulador e personagem crucial no desenvolvimento da *World Wide Web* (WWW), consolidou a estrutura da Web Semântica, conforme figura 1. A sua arquitetura se compõe, dessa maneira, a partir do uso de camadas, com especificações e funções estritas. De início, a camada *Unicode* tem como função garantir o uso padronizado de um mesmo conjunto de caracteres, enquanto a camada *Uniform Resource Indicator*⁸ (*URI*) garante a localização e identificação de páginas de modo unívoco.

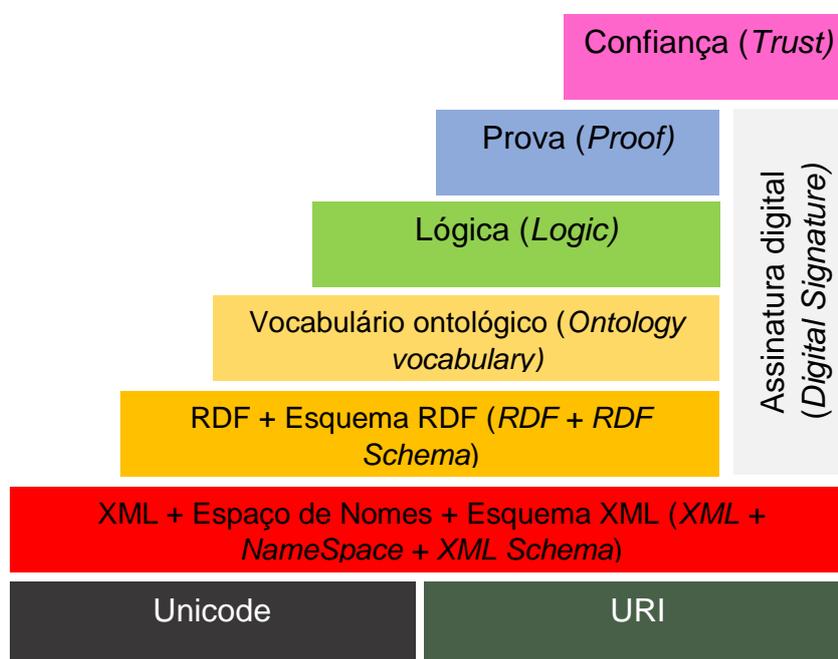
A camada denominada de *XML + Namespace + XML Schema*⁹ tem, entre suas funções, prover a representação dos dados de forma estruturada a partir do uso da metalinguagem *eXtensible Markup Language* (XML), além de fornecer métodos para evitar conflitos de nomes de elementos a partir do uso de um Espaço de Nomes (*NameSpace*). Além disso, ela utiliza um esquema de XML (*XML Schema*), responsável por fornecer uma gramática (ou esquema) expressando restrições sobre documentos XML.

⁷ “An ontology for a domain enumerates and gives semantic descriptions of concepts in the domain of discourse, defining domain-relevant attributes of concepts and various relationships among them. For example, an ontology describing a wines will include such concepts as vintages, wine regions, wineries, grape varieties, and so on. It will also include relations such as produced by, made from, color, year, and body of wine”.

⁸ “indicador uniforme de recursos” (FREITAS, 2003).

⁹ XML + Espaço de Nomes + Esquema XML.

Figura 1 - Camadas da Web Semântica

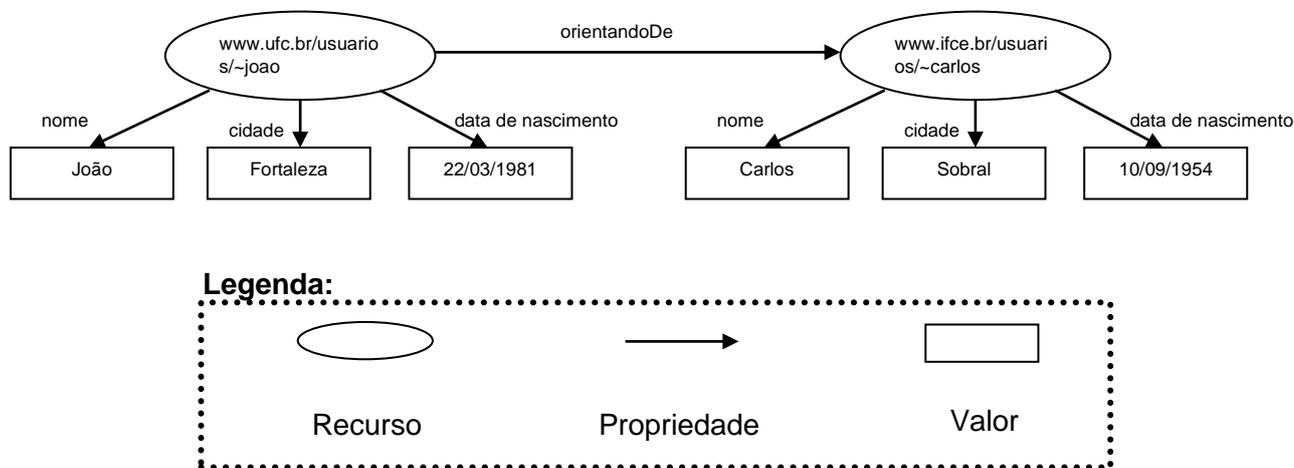


Fonte: <https://www.w3.org/RDF/Metalog/docs/sw-easy>

A camada mais acima denominada de *RDF + RDF Schema*¹⁰ é responsável por descrever recursos na *Web* por meio da ligação desses recursos e suas propriedades. Para tal, é válido contextualizar o conceito de *Resource Description Framework* (RDF), sendo caracterizado por ser um modelo de dados, que descreve os recursos a partir do uso de classes, propriedades e valores (BRICKLEY; GUHA; MCBRIDE, 2014), chamados também de sujeito, predicado e objeto. Essa noção de descrição de recursos constitui o conceito de triplas/declaração. Assim, recurso é tudo aquilo que tem uma *URI* – propriedade que diz respeito a um atributo do recurso (nome, cidade, data de nascimento, orientando de etc) – e valor se refere a uma grandeza (“João”, “Fortaleza”, “22/03/1981”, “www.ifce.br/usuarios/~carlos”) de uma determinada propriedade, podendo ser tanto um valor literal como a *URI* de um recurso. A figura 2 apresenta um modelo simples de descrição utilizando RDF, chamado também de rede semântica.

¹⁰ RDF + Esquema RDF

Figura 2 - Relacionamento entre recursos



Fonte: autoria própria

A camada *Ontology vocabulary*¹¹ é responsável por fornecer uma ontologia que, à luz da CC, possui um significado diferente do termo que foi originalmente criado e usado pela Filosofia. Como já delineado anteriormente, o vocabulário ontológico serve para identificar classes, relações, instâncias e axiomas que permeiam os termos e códigos usados em documentos RDF/XML. O vocabulário ontológico tem como função estabelecer uma convenção de significados entre os agentes que operam os recursos. Assim, tal camada também deve ser capaz de prover a comunicação entre dois domínios distintos, que fornecem termos diferentes para a mesma coisa (Ex.: Unidade Federativa e Estado, cidade e município, carro e automóvel, pessoa e indivíduo, dentre outros), por meio de equivalências entre termos de ontologias.

As camadas *Logic*, *Proof* e *Trust*¹² ainda estão em fase de padronização ou em vias de implementação. Mesmo assim, podemos delinear suas funções a partir das noções trazidas por Koivunen e Miller (2001, s/p, tradução nossa), ao anotarem que, “A camada Lógica permite a gravação de regras, enquanto a camada Prova executa as regras e avalia, juntamente com o mecanismo da camada Confiança, se os aplicativos confiam na prova

¹¹ Vocabulário de ontologia.

¹² Lógica, Prova e Confiança.

determinada ou não”¹³. É válido ressaltar a função da camada Assinatura Digital responsável por detectar alterações em documentos e validar a emissão, por remetentes, de autorias de mensagens (LIMA; CARVALHO, 2004).

A lógica das camadas aqui descritas se consolida a partir da aplicação da camada mais inferior dando suporte à camada imediatamente superior e assim por diante. Portanto, concebe-se que os recursos na *Web* se notabilizam por adotarem um único padrão de caracteres (*Unicode*) e possuam uma identificação exclusiva, uma *URI*. Possuindo tais características, o padrão adotado de estruturação de dados se dá pelo uso do XML. A adoção do XML é fato necessário para funcionamento da camada acima, já que ela usa o modelo *RDF* como fundamental para a ligação de dados. O conceito de ligação de dados, a partir do *RDF*, deverá se validar a partir de um vocabulário ontológico de acordo com o domínio ao qual os dados se referem. As camadas acima – ainda que em fase de implementação – obedecerão à mesma lógica de suporte de uma camada à outra, sucessivamente.

Dessa maneira, apresenta-se os entendimentos por parte tanto da CC e, como feito anteriormente, pela CI. Compreende-se, a partir das perspectivas apresentadas, um sentido de continuidade existente entre a CI em relação a CC, em ambas as direções, no que toca às faculdades de Representação da Informação das referidas áreas. À CI, se valida o entendimento de representações da informação, expressas em registros que constituem as unidades de informação. À CC, a investigação e pesquisa técnica de representações de documentos em ambientes digitais, com a proposta de uso em diferentes contextos e lugares. Assim, Ribeiro (2013, p.536-537) assevera essa relação:

No mundo da Web, classificações e tesouros evoluíram para outras formas de expressão, como as ontologias, as taxonomias, os mapas de conceitos, as folksonomias, ... com vista a tornar a Web semântica uma realidade, sempre ao encontro das necessidades do utilizador, que hoje cada vez menos domina as linguagens herméticas dos bibliotecários e pretende pesquisar na sua linguagem natural, “googlando” e navegando pelo ciberespaço.

¹³ “*The Logic layer enables the writing of rules while the Proof layer executes the rules and evaluates together with the Trust layer mechanism for applications whether to trust the given proof or not*”.

Não é objetivo desse estudo esgotar a discussão ou o paralelismo existente em ambas as abordagens. A visível relação transdisciplinar entre as duas áreas no tocante à Representação da Informação permeia questões abordadas em outras áreas e serve de base para a discussão na Arquivologia. Apesar do perfil específico e da natureza de seus documentos, a Arquivologia possui uma concepção de Representação da Informação a ser abordada no próximo subcapítulo.

3.2.3 Representação da Informação: ponto de vista arquivístico

O documento de arquivo pode ser compreendido a partir do entendimento da sua natureza, esta baseada em princípios postos na Arquivologia que consolidam as suas peculiaridades. A proveniência, a organicidade e unicidade são alguns desses princípios, responsáveis por nortear a prática arquivística. Dito isso, é válido atentar para a variedade de gêneros documentais existentes nos arquivos, ressaltando que são majoritariamente compostos por documentos textuais (PAES, 2004).

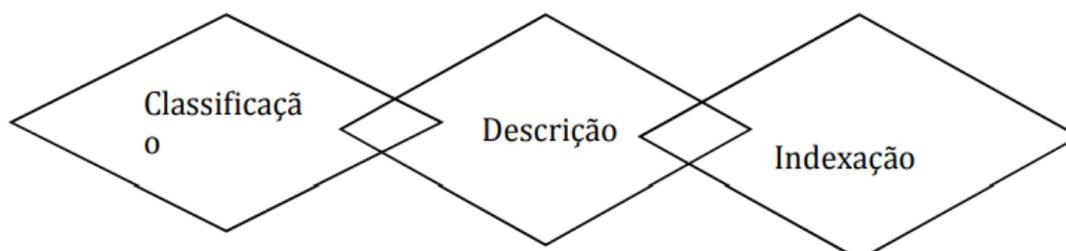
Tais acervos, constituídos no exercício de uma função social, são organizados levando em consideração a estrutura ou funções exercidas pela instituição. Nesse sentido, ao discorrer sobre tais aspectos, evidencia-se a prática arquivística de representação da informação, orientando-se de maneiras específicas tanto nas fases corrente e intermediária, como na fase permanente.

A atuação nas primeiras fases documentais se volta principalmente para a organização, por meio da função de classificação e em outra frente, mais recentemente, na preocupação com o uso de esquemas/conjuntos de metadados, no contexto tanto de documentos convencionais, mas principalmente, no de documentos digitais. Já na fase permanente, a noção se valida por meio da prática arquivística denominada de descrição de documentos, que intenta representar conjuntos documentais que já cumpriram seus prazos e foram recolhidos para arquivos permanentes.

Barros (2016) tece uma linha contínua de atividades da representação arquivística da informação. Para o autor, tudo começa pela classificação, prossegue pela descrição e se finda com a indexação. A

discussão aqui se delineará em torno dessa sequência, com posterior análise de questões sobre a representação na Arquivologia.

Figura 3 - Processo de Representação nos arquivos



Fonte: Extraído de Barros (2016, p.41)

Inicialmente, discorreremos sobre a classificação. O seu uso se notabiliza a partir da fase corrente, possibilitando os documentos já classificados a serem transferidos para os arquivos permanentes (COUTURE; ROUSSEAU, 1998). Tal atividade consiste em atribuir notações (geralmente numéricas) aos documentos, numa tentativa de conciliar a proveniência do documento de acordo com as funções ou estruturas que compõem uma instituição. Por classificação, entende-se a “Análise e identificação do conteúdo de documentos, seleção da categoria de assunto sob a qual sejam recuperados, podendo-se-lhes atribuir códigos” (ARQUIVO NACIONAL, 2005, p. 49).

A classificação, dentro da conjuntura arquivística, se afirma como uma das primeiras funções que evidenciam a representação da informação. Assim, “Trata-se, antes, do uso da classificação numa perspectiva orgânico-funcional, que tem em vista espelhar a estrutura e a atividade do organismo que produziu a informação” (RIBEIRO, 2013, p.531-532). Assim, consolida-se a partir da análise do documento e de todo o seu contexto orgânico e de proveniência. Em seguida, há o uso de um Plano de Classificação, constituído de códigos que, por sua vez, são atribuídos aos documentos. Esses códigos são os mesmos utilizados em outro instrumento de gestão de documentos, chamado de Tabela de Temporalidade de Documentos (TTD). A TTD indicará

os prazos em cada fase, além da destinação que deve se proceder após a avaliação e o cumprimento de tais prazos.

Na fase permanente, o foco, partindo de uma perspectiva mais clássica da literatura da área, não se notabiliza por preocupações no tocante à gestão documental, mas pela questão de acesso aos documentos que adquiriram importância por outros motivos, diferentes daqueles pelos quais foram criados, sendo dotados de valores que podem assumir vieses historiográficos e/ou memorialistas, bem como de testemunho e prova (BELLOTTO, 2007).

A prática de representação na Arquivologia ainda é invariavelmente enfatizada na fase permanente, no intuito de acesso a documentos por um público costumeiramente mais amplo ou por necessidades informacionais mais variadas. Algumas perspectivas (RODRIGUES, 2003) apontam que a representação em arquivos deve ser enfatizada não só na fase permanente, mas durante todo o ciclo vital dos documentos.

Assim, a prática de representação da informação, denominada de “descrição documental”, pode ser compreendida como o “Conjunto de procedimentos que leva em conta os elementos formais e de conteúdo dos documentos para elaboração de instrumentos de pesquisa” (ARQUIVO NACIONAL, 2005, p. 67). Lopez (2002) ressalta que a organização arquivística não depende apenas das atividades de classificação, mas também das de descrição. Para o autor, a descrição arquivística é a única que garante a vasta compreensão do conteúdo de um acervo, proporcionando tanto o conhecimento como a localização dos documentos que o compõe.

Dessa maneira, a descrição no campo arquivístico toma como referência normas técnicas, como: a ISAD(G), considerada a publicação técnica de referência que fundamentou normas de diversos países e que concebe uma descrição multinível, partindo do fundo até o item documental; a ISAAR(CPF) (*International Standard Archival Authority Record for Corporate Bodies, Persons and Families*¹⁴), que é uma norma internacional de registro de autoridade arquivística para entidades coletivas, pessoas e famílias; a ISDF

¹⁴ Norma internacional de registro de autoridade arquivística para entidades coletivas, pessoas e famílias.

(*International Standard for Describing Functions*¹⁵), que dá diretrizes para a descrição de funções (ampliada também para subfunção, procedimento operacional, atividade, tarefa, transação) de entidades coletivas associadas à produção e manutenção de arquivos; a NOBRADE (Norma Brasileira de Descrição Arquivística), sendo a norma de descrição brasileira baseada na ISAD(G); e o esquema de metadados sugerido pelo e-ARQ Brasil. Assim, a descrição leva em consideração, ainda que indiretamente, a ISO 23081 e ISO 15489.

Vale salientar que os padrões/normas acima elencados são compostos de elementos (ex: título, data, proveniência etc.) que norteiam especificações da unidade documental a ser representada. No contexto do presente trabalho, tais elementos assumem a mesma função do conceito de metadado. Navarro e Ziviani (2013, p. 187) reiteram que

Metadados contêm informação sobre a organização dos dados, seus domínios e relacionamentos. [...] Metadados comumente associados a textos incluem o autor, a data de publicação, a fonte de publicação, o tamanho do documento (em páginas, palavras, bytes) e o gênero do documento (livro, artigo, memorando).

Portanto, com a existência de normas e publicações referentes à descrição de documentos, consolidou-se a produção do que, convencionalmente, chamou-se de instrumentos de pesquisa. Couture e Rousseau (1998) compreendem que tais instrumentos constituem a ponta da lança da arquivística, que “permitem a comunicação da informação orgânica no seio da organização junto dos utilizadores, tanto do seu valor primário como do secundário (administradores, investigadores, cientistas etc)” (COUTURE; ROUSSEAU, 1998, p.137-138). Paes (2004) indica os tipos básicos de instrumentos de pesquisa como Guia, Inventário, Catálogo e Repertório. No quadro abaixo, é exposta a relação de finalidade desses instrumentos.

Quadro 1 – Tipos básicos de instrumentos de pesquisa da Arquivologia

Instrumento	Finalidade
Guia	Orientação dos usuários no entendimento e uso dos fundos que integram o arquivo. Trata-se de um instrumento mais genérico, com

¹⁵ Norma Internacional para Descrição de Funções.

		o intuito de informar sobre a totalidade de fundos existentes no arquivo.
Inventário	sumário	Preza pela identificação e descrição sucinta das unidades de arquivamento.
	analítico	Preza pela identificação e descrição pormenorizada das unidades de arquivamento.
Catálogo		Elaborado segundo um critério temático, cronológico, onomástico ou geográfico, incluindo todos os documentos, pertencentes a um ou a mais fundos, descritos de forma sumária ou pormenorizada.
Repertório		Instrumento de pesquisa que descreve pormenorizadamente documentos previamente selecionados, pertencentes a um ou mais fundos, segundo um critério temático, cronológico, onomástico ou geográfico.

Fonte: Paes (2004).

Visto os instrumentos acima, ressalta-se a existência de outros tipos, ainda que considerados “auxiliares” (PAES, 2004), sendo estes mais detalhados por Bellotto (2007) e Couture e Rousseau (1998). Entre eles, o índice, que aponta nomes, lugares ou assuntos, remetendo o leitor às respectivas notações de localização, podendo ser parte complementar de inventários ou catálogos analíticos, poderá também ter personalidade própria, indexando diretamente os documentos (BELLOTTO, 2007). A partir do caráter próprio, destaca-se a indexação como atividade que constitui a representação da informação arquivística.

Assim, dimensiona-se aqui a discussão para a atividade de indexação, pois autores como Barros (2016) entendem que tal atividade ocorre após a construção dos instrumentos acima. Sousa e Araújo Junior (2017) compreendem que a indexação ocorre paralelamente à classificação, complementando esta última por meio de termos, descritores ou palavras-chave. A concepção do que vem a ser indexação é, de acordo com o DTBA, entendida como o “Processo pelo qual documentos ou informações são representados por termos, palavras-chave ou descritores, propiciando a recuperação da informação” (ARQUIVO NACIONAL, 2005, p. 107). Sousa e Araújo Junior (2017, p.16) entendem a indexação como

atividade de processamento técnico destinada à identificação dos conceitos de que trata o documento, expressando-os na

terminologia utilizada pelo profissional da informação em linguagem natural ou com a ajuda de vocábulos ou termos derivados de linguagens documentárias.

A indexação é vista como atividade crucial à representação da informação em arquivos, visto que a busca por conteúdos/assuntos se consolida como provável ponto de acesso aos usuários. Na visão de autores como Ribeiro (1996 *apud* MEDEIROS; BRÄSCHER; VIANNA, 2017), a representação baseada em estrutura e funções seria apenas um complemento para a busca que comumente é efetuada pelo conteúdo da informação, isto é, por seu assunto. Alinhados a tal ideia e tecendo reflexões sobre a indexação, Sousa e Araújo Junior (2017, p. 14) afirmam que

A classificação possibilita a organização de documentos nos arquivos, por meio de um plano de classificação, além de determinar graus de sigilosidade dos documentos e de seus conteúdos em conformidade com a legislação. **Entretanto, a análise do conteúdo visando identificar a categoria de assuntos (rótulos), por meio do processo de indexação, ainda é uma prerrogativa considerada secundária nas atividades de processamento técnicos nos acervos arquivísticos.** Esta realidade impossibilita o desenvolvimento de instrumentos de busca e recuperação dos documentos em um sistema de arquivo, já que a manutenção do vínculo arquivístico parece ser, entre os profissionais da área, o mais importante a ser feito (Grifo nosso).

Nesse sentido, os desafios quanto à indexação de documentos arquivísticos toca também ao conjunto no qual o registro faz parte, isto é, não devem ser desconsideradas as relações com outros documentos que constituem uma teia orgânica que infere ou informa sobre determinados fatos e atos. Muito difundida no campo da Biblioteconomia, as peculiaridades do meio arquivístico informam que os documentos não devem ser indexados como peças unitárias, mas levando-se em conta o contexto e as relações estabelecidas entre os (variados) documentos (VITAL; BRÄSCHER, 2016) que constituem a gênese de uma ação institucional que se prolonga, se cria, se mantém ou se finda com o registro documental.

Com a percepção de representação do modo arquivístico consolidada, existem também pesquisas que passam a considerar o uso de ontologias como tópico a serem utilizados na área. Na visão de Santos (2010), ao relacionar o uso de ontologias ao plano de classificação, o desenvolvimento de ontologias permitiria adquirir o domínio conceitual no qual o documento se

encontra, a interoperabilidade da informação, além do processamento de documentos a partir da interação de usuários. Luz (2016) também aborda o uso de ontologias na representação arquivística, ao afirmar que uma ontologia cria a conceituação formal de um domínio ou parcela da realidade. Adiante, o autor reafirma a criação de uma relação com a Arquivologia:

[...] pois os arquivos são o resultado das ações das instituições e das pessoas que nelas atuam. Trazem uma representação da realidade das instituições e pessoas e das estruturas que a geraram, expressadas por meio de suas relações orgânicas (o Archival bond), podendo ser considerada uma parcela da realidade. Uma ontologia, portanto, pode manifestar a informação arquivística, o que a torna uma ferramenta de representação da informação orgânica e da estrutura que a geraram e têm sua custódia (LUZ, 2016, p.115).

Assim, o uso de ontologias no domínio arquivístico amplia o caráter representacional na área, já que a ideia de representação no meio arquivístico visa criar uma ampla possibilidade de pontos de acesso, seja ela por um código de classificação, por um assunto ou pelos inúmeros elementos que constituem as normas de descrição. É preciso, a partir desses pontos de acesso, notar a intenção de recuperação da informação em tais iniciativas. Partindo do princípio de representação para posterior recuperação, discute-se adiante a Recuperação da Informação com o mesmo enfoque dado à Representação da Informação.

3.3 Recuperação da Informação

Diante das explanações aqui já explicitadas, traça-se uma perspectiva de exposição semelhante ao capítulo anterior, enfatizando aspectos de áreas do conhecimento abrangidas no presente estudo. Assim, apresenta-se o histórico pormenorizado, visando-se evidenciar a discussão inicial da RI no contexto da CI, para, em seguida, enfatizar o entendimento de tal temática na CC e concluir a exposição com a visão existente na Arquivologia.

3.3.1 Breve histórico

O processo de industrialização traz consigo todo um arsenal de meios para produção de dados, que permitem a inferência de informação e conhecimento por parte do usuário. Desde a invenção da prensa por Gutenberg, do telefone por Antonio Meucci e Alexander Graham Bell (PAVANI, 2007) e da revolução tecnológica que inseriu dispositivos cada vez menores e significativos ao cotidiano social (ROZA, 2017) – a exemplo do microcomputador e do telefone celular – o homem passa cada vez mais a produzir informação e passa também a ser usuário dela.

Ainda que a etapa de produzir dados, visando as tentativas de registro de informação, seja intuitivamente corriqueira, é no procedimento de recuperar o que foi produzido que as entidades (indivíduos e organizações) suprem também suas necessidades informacionais. Sendo uma proeminente área de investigação científica, é importante ressaltar que o tema da Recuperação da Informação teve início na década de 1940, segundo Codina (2002), a partir da atuação do pesquisador e professor norte-americano George Kingsley Zipf, notório estudioso da linguagem, ao qual se atribui a lei de Zipf, pelo fato de que

segundo a qual se estudamos a frequência das palavras em um corpo suficientemente representativo de uma linguagem obtemos essa relação: **Frequência * Ranque = constante** onde a frequência é o número de vezes que uma palavra acontece e o ranque é o número da ordem das palavras na lista de distribuição de frequência. Os estudos de Zipf mostraram que certas propriedades estruturais da massa de informação textual podiam ser descobertas. Por exemplo, acontece que, de acordo com a lei de Zipf, a maioria das palavras acontece algumas vezes e, inversamente, apenas algumas palavras acontecem muitas vezes (CODINA, 2002, p.5, grifo nosso, tradução nossa).¹⁶

Em seguida, a partir de meados de 1952, Mortimer Taube fez experimentos em indexação e uso de operadores lógicos. O bibliotecário norte-americano fez uso da indexação coordenada (aplicação de unitermos) aliada à operação com lógica booleana em contrato com o *United States Armed Services Technical Information Agency* (CHU, 2007).

¹⁶ “según la cual si estudiamos la frecuencia de las palabras en un corpus lo suficientemente representativo de una lengua obtenemos esta relación: Frecuencia * Rango = constante donde la frecuencia es el número de veces que sucede una palabra y rango es el número de orden de la palabra en la lista de distribución de frecuencias. Los estudios de Zipf demostraron que se podían descubrir determinadas propiedades estructurales de masas de informaciones textuales. Por ejemplo, ocurre que, según la ley de Zipf, la mayor parte de las palabras sucede unas pocas veces, y que, a la inversa, sólo unas pocas palabras suceden muchas veces”.

Um artigo publicado por Vannevar Bush, em 1945, chamado “*As We May Think*” (“*Como podemos pensar*”), questionou, de acordo com o contexto da época (caracterizado pela 2ª Guerra Mundial e avanços tecnológicos), a chamada “explosão informacional” no tocante à ciência e à tecnologia. O cientista propôs uma solução, baseada na tecnologia, para encontrar informações relevantes em meio ao problema de explosão informacional. O Memex (*memory + index*) foi uma máquina proposta, porém nunca construída, considerada um ideal e quase utópica, como alguns pensavam (SARACEVIC, 2009).

Quase que paralelamente, ainda em idos de 1948, Hans Peter Luhn, conhecido engenheiro e inventor que trabalhava na IBM (*International Business Machines*), executou trabalhos no tocante à Recuperação da Informação. As contribuições de Luhn se consolidam por meio da construção de um sistema de busca de informação eletrônica, aplicado a estruturas químicas. Luhn desenvolveu também a máquina que aplicava a ideia do KWIC (*Keyword in context*), baseado no conceito de palavras-chave como elementos que representam e recuperam a multiplicidade de facetas do documento; das palavras-chave extraídas dos títulos e tópicos do documento; e do uso de permutações de palavras-chave existentes em títulos e tópicos para ampliação na localização de determinado termo no documento (CHU, 2007).

Nessa linha de sucessão de contribuições, aparece também o cientista norte-americano Calvin N. Mooers, na década de 50. Foi ele que cunhou o termo “recuperação da informação”, inserido no vocabulário da Ciência da Informação, conforme assinala Chu (2007). Garfield (1997 *apud* CHU, 2007, p. 10, tradução nossa) verifica que Mooers entendia a recuperação da informação como “[...] encontrar informações cuja localização ou existência é a priori desconhecida”¹⁷. Nesse sentido, Mooers refletiu sobre desafios de um SRI e estipulou uma lei, que leva o seu nome. A Lei de Mooers diz que “Um sistema de recuperação de informações terá a tendência de não ser usado se é mais irritante e problemático para um usuário obter a informação do que não obtê-la” (DIAS, 2003, p.4). Entre suas invenções, está o *Zatocoding*, um

¹⁷ “[...] *finding information whose location or very existence is a priori unknown*”.

sistema para armazenar um grande número de descritores de documentos em um único cartão especialmente entalhado sobrepondo códigos descritores aleatórios de oito dígitos. O uso do sistema de Zatocoding resultaria em apenas um pequeno, mas tolerável, número de “falhas falsas” na pesquisa bibliográfica (GARFIELD, 1997, apud CHU, 2007, p.10, tradução nossa)¹⁸

Aliado a tais invenções, Gerard Salton, norte-americano, professor da Universidade Cornell, se notabilizou, em meados da década de 60, como um dos mais importantes nomes da área pelo desenvolvimento do trabalho com métodos de processamento de texto completo em computadores, além de ter lançado bases sólidas no tema da Recuperação de Informações. Para Crouch *et al* (1996 apud CHU, 2007, p.11, tradução nossa),

suas contribuições de pesquisa abrangem a gama de recuperação de informação: o modelo de espaço vetorial, ponderação de termo, feedback de relevância, agrupamento, recuperação Booleana estendida, valor de discriminação de termo, construção de dicionário, dependência de termo, compreensão e estruturação de texto, recuperação de passagem – e, é claro, processamento automático de texto usando o SMART.¹⁹

O *System for the Manipulation and Retrieval of Texts* (SMART) foi concebido por Salton como implementação do modelo vetorial. Ferneda (2003, p.32) define o SMART a partir da noção de representação de cada documento por um vetor numérico, em que cada elemento desse vetor possui um valor, que representa, respectivamente, a importância do termo na descrição do documento. Deve-se levar em conta que a atribuição manual desses pesos necessitaria de pessoal especializado executando tal trabalho por um tempo. “No entanto, o sistema SMART fornece um método automático para o cálculo dos pesos não só dos vetores que representam os documentos, mas também para os vetores das expressões de busca” (FERNEDA, 2003, p.42).

Numa referência às contribuições de Salton no que toca ao processamento automático de textos, Lesk (1996) aponta que tal iniciativa trouxe à tona questões referentes à indexação, já que o advento de bancos de

¹⁸ “*system for storing a large number of document descriptors on a single specially notched card by superimposing random, eight-digit descriptor codes. The use of the Zatocoding system would result in only a small but tolerable number of “false drops” in bibliographic search (Garfield, 1997).*”

¹⁹ “*His research contributions span the gamut of information retrieval: the vector space model, term weighting, relevance feedback, clustering, extended Boolean retrieval, term discrimination value, dictionary construction, term dependency, text understanding and structuring, passage retrieval – and, of course, automatic text processing using SMART.*”

dados informatizados poderia elevar o nível de detalhes da indexação, gerando custos elevados para uma indexação adequada. Tal medida fez surgir a ideia de busca por texto livre, pois “poderia haver recuperação completa de qualquer documento usando uma palavra específica e não haveria custos para indexação manual” (LESK, 1996, s/p, tradução nossa)²⁰. Apesar da popularidade dessa ideia, atraiu-se muitas objeções, principalmente

daqueles que apontaram que a indexação não era apenas uma maneira de selecionar palavras, mas era uma maneira de escolher as palavras certas, ou as palavras que eram mais importantes, ou aquelas que tinham sido selecionadas como o rótulo correto para um determinado assunto (LESK, 1996, s/p, tradução nossa).²¹

Havia a intenção de trazer à discussão o uso do vocabulário controlado. Assim, uma série de perguntas foi colocada, questionando se a pesquisa de texto livre seria aceitável e, caso os vocabulários controlados fossem preferíveis, quais seriam as formas de traduzir para eles ou criá-los automaticamente. Tais questões estimularam o desenvolvimento das técnicas de avaliação de SRI, lideradas por Cyril Cleverdon (LESK, 1996), bibliotecário e cientista da computação inglês. Cleverdon desenvolveu, ainda na década de 60, a matemática da revocação e da precisão como medidas de SRI: a revocação diz respeito à fração de documentos relevantes recuperados; e a Precisão refere-se à fração de documentos recuperados que são relevantes. Assim, consolidou-se que a indexação em texto livre foi considerada muito mais eficaz e também mais barata, a partir de experimentos de Cleverdon e seus colaboradores, E. Michael Keen e Frederick Wilfrid Lancaster, além de Gerard Salton. Assim, sucedeu-se que “Uma longa série de experimentos foi realizada e, de fato, as coleções de testes criadas na época ainda estão em uso, embora o maior tenha sido apenas 1400 resumos de documentos aeronáuticos” (LESK, 1996, s/p)²².

Assim, ao delinear avanços nas épocas, destacamos que, durante os anos 60, conceitos como *feedback de relevância* e recuperação multilíngue

²⁰ “There could be complete retrieval of any document using a particular word, and there would be no cost for manual indexing”.

²¹ “those who pointed out that indexing was not just a way of selecting words but was a way of choosing the right words, either the words that were more important, or the ones that had been selected as the correct label for a given subject”.

²² “A long series of experiments was run, and in fact the test collections created at that time are still in use, even though the largest was only 1400 abstracts of aeronautical documents.”

tiveram seus inícios. Ocorreu também, nesse período, o começo de pesquisas referentes às perguntas em linguagem natural e se questionou à época o porquê dos SRI se limitarem à recuperação de documentos. Na década seguinte, isto é, nos idos dos anos 70, houve popularização no uso de computadores, com adoção de sistemas de tempo compartilhado. Daí decorreu a adoção massiva de sistemas comerciais (Dialog, o Orbit e o BRS) e uso dos sistemas de recuperação de texto completo. Na década de 80, tiveram suas proposições, pesquisas sobre a desambiguação dos sentidos, usando dicionários legíveis por máquina, como uma maneira de ajudar a distinguir significados diferentes das mesmas palavras. Ocorreram também trabalhos sobre recuperação bilíngue e parte da atribuição de fala. A década de 90 trouxe a popularização de acesso à rede mundial de computadores, entretanto, com uso dos primitivos métodos de Recuperação da Informação em computadores pessoais (LESK, 1996), enfatizando, dessa maneira, o uso e a pesquisa de métodos simples de modelos quantitativos (booleano, principalmente), em detrimento aos modelos dinâmicos, com os sistemas especialistas, redes neurais, dentre outros.

Diante desse breve apanhado, percebe-se a RI como tema subsidiado por diferentes estudiosos ao longo do tempo. Apesar de temática evidenciada na CC, a RI também é evidenciada em análises e estudos da CI. Nesse sentido, o subcapítulo seguinte aborda questões da CI na RI.

3.3.2 Recuperação da Informação: ponto de vista da CI

A RI pode ser entendida tanto a partir da perspectiva de processo, como da perspectiva de área de estudos. Enquanto processo, se consolida a partir do principal objetivo e desafio de, segundo Baeza-Yates e Ribeiro Neto (2013), recuperar todos os documentos que são relevantes à necessidade de informação do usuário, simultaneamente ao objetivo de recuperar o menor número de documentos irrelevantes. Como área de estudos, é responsável por desenvolver e investigar questões que vão desde a ordem teórica, até a ordem prática, como a medição de índices de revocação, precisão e sucesso nos procedimentos de busca e recuperação, aprimoramento e constituição de algoritmos para Sistemas de Recuperação da Informação (SRI), dentre outras

questões. Ela surge, portanto, como área que incentiva o desenvolvimento de produtos e resolução de problemas comuns a áreas, por exemplo, como a CI e a CC, sendo nesse sentido fato que evidencia a sua base conceitual, multidisciplinar, no objeto informação.

Embora algumas das questões postas acima tenham vieses mais técnicos/práticos/experimentais, em que se faz necessário aprofundado conhecimento em noções de lógica, estatística e programação de computadores, a CI já se preocupava, em seus primórdios, com questões referentes à RI. Aliás, é o artigo de Vannevar Bush, já referenciado, que impulsiona o desenvolvimento da CI (SARACEVIC, 2009). Tefko Saracevic (2009, p.1, tradução nossa) alerta para as direções que embasavam as preocupações da CI:

Desde o início, a ciência da informação tinha essas duas orientações: uma que lida com a necessidade de informação, ou mais amplamente o comportamento da informação humana, e a outra que lida com técnicas e sistemas de recuperação de informação.²³

A orientação voltada à necessidade de informação baseou consequentemente a atuação da CI em seus estudos e práticas posteriores. As necessidades do usuário são uma abordagem, segundo Baeza-Yates e Ribeiro-Neto (2013), mais focada pela CI do que pela CC. Corroborando com essa perspectiva, Garcia (2007, p.1) enfatiza que

Para as áreas que lidam com a informação, o sentido é outro, pois para se recuperá-la, há que se selecionar, armazenar, preservar, tratar por um sistema, **para permitir que o desejo sentido ou latente do usuário seja atendido**, avizinhando-nos a um sentido diverso de perdido (Grifo nosso).

Ingwersen (1992), proeminente estudioso dinamarquês do modelo cognitivo da RI, evidencia no livro “*Information Retrieval Interaction*”, o usuário como partícipe do processo aqui discutido. Na visão do autor:

A recuperação da informação está preocupada com os processos envolvidos na representação, armazenamento, pesquisa e busca de

²³ “the outset, information science had these two orientations: one that deals with information need, or more broadly human information behavior, and the other that deals with information retrieval techniques and systems.”

informações relevantes para um requisito de informação desejado por um usuário humano (INGWERSEN, 1992, p. 49, tradução nossa)²⁴

Autores como Capurro (2003), Khapre e Basha (2012) fazem alusão aos paradigmas físico, cognitivo e social para o entendimento do conceito de informação e de como a percepção sobre a RI se consolida nesses casos. Nesse viés de análise, Khapre e Basha (2012, p. 233) são enfáticos ao afirmarem que “no campo da CI, a RI aceitou basicamente a natureza multinível do conceito de informação”²⁵. De acordo com a visão de Capurro (2003), é possível analisar como a RI foi influenciada por cada um dos paradigmas citados anteriormente. No caso do paradigma físico, o autor cita, por exemplo, os experimentos de *Cranfield* e corrobora à sua visão o lugar do sujeito (usuário) no processo.

Torna-se evidente que, no campo da ciência da informação, o que esse paradigma exclui é nada menos que o papel ativo do sujeito cognoscente ou, de forma mais concreta, do usuário, no processo de recuperação da informação científica, em particular, bem como em todo processo informativo e comunicativo, em geral (CAPURRO, 2003, s/p).

Quanto ao paradigma cognitivo, proposto por Bertram C. Brookes e influenciado pela ontologia e epistemologia de Karl Popper, Capurro (2003, s/p) afirma que “Brookes subjetiva, por assim dizer, esse modelo no qual os conteúdos intelectuais formam uma espécie de rede que existe somente em espaços cognitivos ou mentais, e chama tais conteúdos de ‘informação objetiva’” e que, a partir de influências principalmente de Nicholas Belkin,

a teoria dos modelos mentais tem tido impacto no estudo e na concepção de sistemas de recuperação da informação, como mostram as análises empíricas realizadas por Pertti Vakkari com relação à conexão entre estado anômalo do conhecimento e estratégias de busca (VAKKARI, 2003 apud CAPURRO, 2003, s/p).

E quanto ao paradigma social, Frohmann (1995 apud CAPURRO, 2003) alerta quanto ao seu caráter “associal” e idealista. O paradigma social leva em conta a tessitura social onde o indivíduo se situa e não descarta o viés

²⁴ “Information retrieval is concerned with the processes involved in the representation, storage, searching and finding of information which is relevant to a requirement for information desired by a human user”

²⁵ “in the Field of information science, information retrieval has basically accepted the multilevel nature of the concept of information”.

cognitivo. Capurro (2003, s/p) não deixa de citar, ao contextualizar a Recuperação da Informação no paradigma social, o paradigma “social-epistemológico”:

Birger Hjørland desenvolveu, junto com Hanne Albrechtsen (Hjørland 2003, 2003a, 2000, 1998, Hjørland/Albrechtsen 1995) um paradigma social-epistemológico chamado „domain analysis“ no qual o estudo de campos cognitivos está em relação direta com comunidades discursivas („discourse communities“), ou seja, com distintos grupos sociais e de trabalho que constituem uma sociedade moderna. Uma consequência prática desse paradigma é o abandono da busca de uma linguagem ideal para representar o conhecimento ou de um algoritmo ideal para modelar a recuperação da informação a que aspiram o paradigma físico e o cognitivo. Uma base de dados bibliográfica ou de textos completos tem caráter eminentemente polissêmico ou, como o poderíamos chamar também, polifônico. Os termos de um léxico não são algo definitivamente fixo.

Nesse contexto de recuperação da informação, tomando como aporte aspectos próprios do contexto social, a CI empreende pesquisas que levam em consideração, no processo de representação e, conforme discutido aqui, a recuperação e as interações dos próprios usuários da informação nesses processos. Esses empreendimentos investigativos se deram em torno do que se chamou de folksonomia, com fins tanto de representação, bem como de recuperação da informação (ASSIS; MOURA, 2013). Para Viera e Garrido (2011, p.6)

A classificação social ou folksonomia são metadados baseados em hipertexto que formam conjuntos de informação estruturados de forma distribuída. [...] Parte do conjunto de comunidades web 2.0, a folksonomia também pode ser definida como um sistema de organização de informação ou categorização social bottom-up, ou seja, de muitos para muitos, baseada em hipertexto que tem como objetivo recuperar conjuntos de documentos.

Ainda para Assis e Moura (2013), o elemento inovador no estudo da folksonomia aplicado à organização de recursos informacionais diz respeito ao papel dos sujeitos e suas representações na sua gênese, somado a isso a dinâmica das redes sociais. Desse modo, o processo de representação e recuperação da informação é tido como um dos objetivos nesse processo, visto a intenção da folksonomia no desenvolvimento de metodologias e instrumentos para esses fins (ASSIS; MOURA, 2013).

Investigando aspectos da RI, o conceito de relevância e feedback de relevância foram investigados por Belkin e Croft (1992). Tais autores propõem,

a partir dos modelos clássicos de RI, o modelo de filtragem de informações. Por filtragem de informações, entende-se “uma variedade de processos que envolvem a entrega de informações para pessoas que precisam delas” (BELKIN; CROFT, 1992, p. 1, tradução nossa)²⁶. Tal modelo se baseia nas preferências de informações individuais ou de grupo, denominados de perfis (estes em geral representam interesses de longo prazo). Na visão dos autores, “[...] a recuperação de informações e a filtragem de informações são de fato dois lados da mesma moeda” (BELKIN; CROFT, 1992, p. 9, tradução nossa)²⁷. Para Viera e Garrido (2011, p.2),

[...] a noção de relevância, ou seja, recuperar mais documentos que alcancem a necessidade informacional dos usuários e recuperar menos documentos irrelevantes, é o objetivo central da recuperação da informação. No entanto, tudo está interligado: a recuperação efetiva da informação relevante depende tanto das tarefas dos usuários, quanto da visão lógica dos documentos adotados pelos sistemas de recuperação de informação.

Não obstante que o foco da CI se centre nas necessidades e comportamento informacional de usuários, Araújo (2014) reconhece que houve uma intersecção entre CI e CC existente à luz da recomendação de criação da disciplina de RI, proposta por Mooers. Em anos seguintes, a automatização da recuperação da informação “foi encampada dentro do projeto da ciência da informação, chegando mesmo a ser entendida como o ‘núcleo’ da área por diferentes autores, entre os quais Tefko Saracevic, em seu livro *Introduction to information science* de 1970” (ARAÚJO, 2014, p.108). Entretanto, conforme veremos a seguir, devido à natureza de automatização de processos, com enfoque principalmente no tocante a RI, a CC já desenvolvia tal tema de maneira mais aplicada.

3.3.3 Recuperação da Informação: ponto de vista da CC

Nesse entremeio, a CC consolidou seu estudo numa abordagem voltada às técnicas de SRI e algoritmos, com um foco diferente de bibliotecários e da CI, “que adotam uma interpretação do problema de RI centrada no usuário humano” (BAEZA-YATES; RIBEIRO-NETO, 2013, p. 14).

²⁶ “a variety of processes involving the delivery of information to people who need it”.

²⁷ “information retrieval and information filtering are indeed two sides of the same coin”.

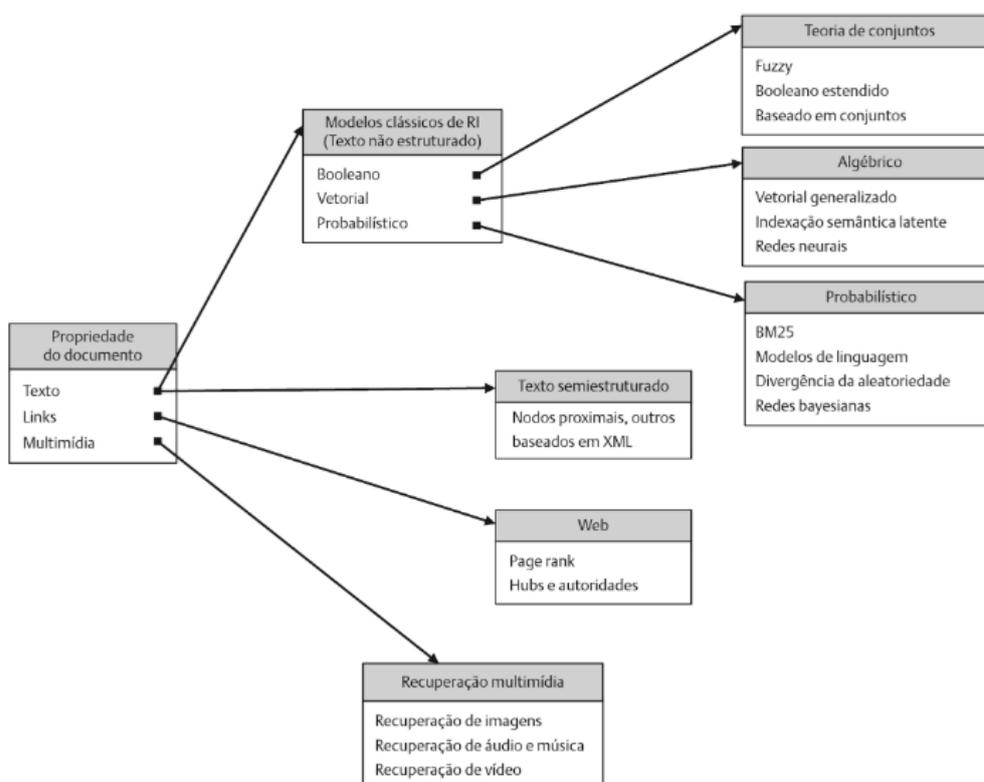
Para Ricardo Baeza-Yates e Berthier Ribeiro-Neto (2013), o entendimento da RI no contexto da CC se consolidou na visão centrada no computador. Nesse viés, [...] a RI consiste principalmente na construção de índices eficientes, no processamento de consultas com alto desempenho e no desenvolvimento de algoritmos de ranqueamento, a fim de melhorar os resultados (BAEZA-YATES; RIBEIRO-NETO, 2013, p.1).

Desse modo, tomaram-se direcionamentos baseados em modelos, denominados de “modelos de Recuperação da Informação”. Um modelo de RI é composto basicamente por uma quádrupla $[D, Q, F, R(q_i, d_j)]$, segundo o qual

1. **D** é um conjunto composto por visões lógicas (ou representações) dos documentos da coleção.
2. **Q** é um conjunto composto por visões lógicas (ou representações) das necessidades de informação dos usuários. Essas representações são chamadas de consultas.
3. **F** é um arcabouço para modelar as representações dos documentos, das consultas e de seus relacionamentos, como conjuntos e relações Booleanas, vetores e operações de álgebra linear, espaços amostrais e distribuições de probabilidade.
4. $R(q_i, d_j)$ é uma função de ranqueamento que associa um número real à representação de uma consulta $q_i \in Q$ e à representação de um documento $d_j \in D$. Esse ranking define um ordenamento entre os documentos em relação à consulta q_i . (BAEZA-YATES; RIBEIRO-NETO, 2013, p.22)

Assim, Ricardo Baeza-Yates e Berthier Ribeiro-Neto (2013), por meio da Figura 4, demonstram modelos de RI baseados nas propriedades do documento, seja ele texto, links ou multimídia. Na perspectiva dos documentos textuais, os modelos podem ser distinguidos em “Texto não estruturado” e “Texto semiestruturado”. No primeiro se consolidaram os modelos clássicos, sendo eles: o Booleano, Vetorial e Estatístico (mais informações no Quadro 2). O texto nessa categoria é concebido como uma simples sequência de palavras. Em relação ao “Texto semiestruturado”, o modelo de Nodos proximais, além de outros baseados em XML são os representativos dessa categoria e levam em consideração elementos estruturais do texto, como o título, (sub)seções e parágrafos, além de incluir texto não estruturado (BAEZA-YATES; RIBEIRO-NETO, 2013).

Figura 4 - Uma taxonomia de modelos de RI



Fonte: Extraído de Baeza-Yates e Ribeiro-Neto (2013, p.24).

No tocante a *Web*, leva-se em conta aspectos referentes ao hiperlink, com modelos como o *Page Rank* e o *Hubs e autoridades*. Quanto a documentos multimídia, Ricardo Baeza-Yates e Berthier Ribeiro-Neto (2013) citam peculiaridades no tocante a tais documentos, afirmando que suas características implicam, na maioria dos casos, a adoção de estratégias de recuperação e a adoção de características de vários modelos de RI, concebendo implementações modernas de SRI (BAEZA-YATES; RIBEIRO-NETO, 2013).

Quadro 2 – Modelos clássicos de Recuperação da Informação

Modelo	Visão básica	Fundamentação
Booleano	Considera que os termos de indexação estão presentes ou ausentes nos	Álgebra booleana

	documentos, a partir de uma abordagem binária e com consulta conectando os termos a partir de operadores lógicos (AND, OR e NOT).	
Vetorial	Atribuem pesos não binários aos termos de indexação das consultas e dos documentos, computando, dessa forma, o grau de similaridade entre cada documento armazenado no sistema e a consulta do usuário.	Álgebra linear
Probabilístico	Calcula, a partir da projeção de pesos binários de termos de uma consulta que estão ou não presentes nos documentos, a probabilidade de documentos serem considerados relevantes em relação àqueles não-relevantes.	Probabilidade

Fonte: autoria própria

A apresentação dos modelos de Recuperação da Informação, conforme o quadro acima, limitou-se aos modelos clássicos, buscando não tornar exaustiva a apresentação dos modelos previstos na Figura 4 e por entender a interface existente da CI em relação aos modelos clássicos desde sua gênese. Dessa maneira, concebeu-se a visão e noções de RI pela CC, compondo uma apresentação que coaduna para evolução da temática aqui evidenciada.

Portanto, a partir da exposição da visão da CI e aqui da CC, recorreremos a algumas reflexões. Khapre e Basha (2012, p.232, tradução nossa) visualizam que “A partir da definição das duas disciplinas de recuperação de informações, é fácil ver que as duas compartilham uma perspectiva comum de recuperação de informações – usuários, tecnologia da

informação e direções de pesquisa”²⁸. Ainda sobre o entendimento da CC sobre a RI e, conseqüentemente, do elemento denominado “informação”, Khapre e Basha (2012, p. 235, tradução nossa) entendem que

[...] pesquisadores de recuperação de informações geralmente tratam a informação como fato e usam informações para fazer algum trabalho, como um índice de informação, codificação de informações, decomposição e análise de informações. Essa visão é baseada na natureza da informação, ou seja, específica, definível e pode ser codificada. A Recuperação da Informação da Ciência da Computação segue o princípio racional e tradicional de que a informação é a existência factual do mundo externo das coisas.²⁹

É válido ressaltar que, apesar das diferenças de abordagem, Calvin Mooers (1951) compreendia que a recuperação da informação é crucial para a documentação e organização do conhecimento, enfatizando, dessa maneira, o estreitamento da relação entre CI e CC. Nesse sentido, como forma de fundamentar o trabalho, buscar-se-á no próximo subcapítulo as noções que têm permeado a temática de RI na área arquivística.

3.3.4 Recuperação da Informação: ponto de vista arquivístico

Conforme visto anteriormente, a RI se define pela confluência de visões advindas de diversas áreas (entre elas a Linguística, Computação, Psicologia, Ciência da Informação, dentre outras), no tocante a entendimentos que vão, desde a visão de máquina, responsável pelo processamento automático de consultas, com uso de algoritmos e otimização dos SRI, àquela voltada ao usuário, seus comportamentos de informação e como os SRI podem conceber e se aproximar do entendimento semântico dessas necessidades informacionais. Partindo dessa premissa, a Arquivologia também se preocupa com aspectos referentes à recuperação da informação em seus acervos,

²⁸ “From the definition of the two disciplines of information retrieval it is easy to see that the two disciplines share a common perspective of information retrieval - users, information technology and research directions”.

²⁹ “Computer science, information retrieval researchers often treat the information as fact, and use information to do some work, such as an index of information, information coding, decomposition and analysis of information. This view is based on the nature of the information, i.e. specific, definable and can be encoded. Computer Science Information Retrieval follow the rational and traditional principle that information is the factual existence of the external world of things”.

sobretudo aqueles de conotação e valor primários, essenciais para a administração de organizações.

Desse modo, em meados da década de 60, Rita Campbell, arquivista e pesquisadora da Universidade de Stanford já se preocupava com aspectos referentes a RI em arquivos, levando em conta o crescente aumento da produção documental à época. Para a pesquisadora,

Arquivos estão sujeitos às pressões opostas de montes de papel e continuamente aumentando as demandas de pesquisa. **A saída do dilema pode ser encontrada na indexação mais intensa de materiais de arquivo por máquina.** O arranjo de arquivos sem qualquer indexação – a organização, seja por proveniência, por dados, por área geográfica, por doador individual ou por alguma combinação deles – não faz parte de uma grande coleção que atenda às demandas de pesquisa de hoje. A necessidade é de indexação detalhada. **A indexação manual detalhada, no entanto, é muito cara, especialmente quando grandes arquivos estão envolvidos** (CAMPBELL, 1967, p.279-280, tradução nossa, grifo nosso).³⁰

De acordo com os trechos destacados acima, nota-se que já haviam preocupações quanto ao extenuante trabalho de indexação de grandes volumes documentais, sugerindo-se soluções como a indexação por máquina. Tal perspectiva se alinha a pesquisas da época, a exemplo daquelas desenvolvidas por Salton sobre o processamento automático de textos, em meados da década de 60.

Assim, a indexação se evidencia por trazer à tona a noção de assunto, embora Campbell (1967) admitisse que, por ser mais fácil encontrar material por assunto nos acervos organizados por proveniência, ainda ocorrem dificuldades, principalmente para o pesquisador, já que “[...] a escolha de um assunto pode não ser fácil de encontrar nas linhas de organização criadas por proveniência” (CAMPBELL, 1967, p.280)³¹. A autora, em seguida, enfatiza que a indexação é “a chave para a recuperação da informação” (CAMPBELL, 1967, p.280, tradução nossa)³².

³⁰ “Archives are subject to the opposing pressures of mounds of paper and continuously increasing research demands. The way out of the dilemma may be found in more intensive indexing of archival materials by machine. Arrangement of archives without any indexing—arrangement whether by provenance, by data, by geographical area, by individual donor, or by some combination of these—does not in a large collection meet today’s research demands. The need is for detailed indexing. Manual detailed indexing, however, is too expensive, especially where large archives are involved”.

³¹ “[...]choice of a subject may not be easy to find within the lines of organization created by provenance”.

³² “is the key to information retrieval”.

Seguindo na perspectiva traçada por Campbell (1967), algumas questões são postas e entendidas como desafios para a área, como se infere da afirmação de Witthus (1992, s/p):

Embora os sistemas de informação de arquivamento possam usar cabeçalhos de assunto do tipo biblioteca, eles dependem para uma recuperação efetiva da compreensão das atividades e funções de indivíduos e organizações e das formas de material geradas por essas atividades. Os arquivistas criaram sistemas de informação arquivística com base na ordem e no contexto dos registros de arquivos, a fim de organizar e recuperar informações geradas pela atividade que criou os registros. O locus de autoridade para a criação de pontos de acesso intelectual aos registros de uma organização ou dos papéis de um indivíduo é a atividade dessa organização ou daquele indivíduo (WITTHUS, 1992, s/p)³³.

O debate sobre o assunto como principal critério (visto, por vezes, a partir de uma perspectiva de complementaridade e, em outras vezes, de oposição) em relação ao conceito clássico de organização arquivística (que leva em conta o contexto orgânico do produtor documental, a partir de suas funções ou de sua estrutura organizacional) encontra em Fox e Wilkerson (1998, p.15, tradução nossa) ênfase nessa questão:

A organização física de documentos em uma coleção de acordo com um sistema de classificação baseado em assunto, como a Classificação Decimal de Dewey ou a Classificação da Biblioteca do Congresso, não é apropriada para materiais de arquivo. Uma organização imposta externamente com base no assunto destrói as relações entre os documentos e os eventos aos quais eles se relacionam e mascara o significado herdado do contexto de sua criação.³⁴

Desse modo, a perspectiva arquivística leva em consideração aspectos peculiares quando debate também a RI, entendendo que os princípios (organicidade, unicidade, proveniência, dentre outros) que envolvem suas práticas devem estar representados nas propostas de tal temática e

³³ *“While archival information systems may use library-type subject headings, they depend for effective retrieval on an understanding of the activities and functions of individuals and organizations and the forms of material generated by those activities. Archivists created archival information systems based on the order and context of archival records in order to organize and retrieve information generated by the activity that created the records. The locus of authority for the creation of intellectual access-points to the records of an organization or the papers of an individual is the activity of that organization or that individual”.*

³⁴ *“Physically organizing documents in a collection according to a subject-based classification system such as the Dewey Decimal Classification or the Library of Congress Classification is not appropriate for archival materials. An externally imposed organization based on subject matter destroys the relationships between documents and the events to which they relate, and masks the meaning inherited from the context of their creation”.*

posterior funcionalidade. Couture e Rousseau (1998), ao falarem sobre a RI em sua obra “Os fundamentos da disciplina arquivística”, corroboram tal temática como fazendo parte de um programa de três fases, “centrado na missão do organismo e integrado na sua política de gestão da informação” (COUTURE; ROUSSEAU, 1998, p.65), a ser desenvolvido na instituição que o implantar.

O primeiro componente desse programa diz respeito à criação, à difusão e ao acesso à informação orgânica. Algumas perguntas são postas para reflexão na instituição, como: “Quem tem acesso à informação? Como é que ela é difundida, classificada e recuperada? Qual o seu ciclo de vida? [...]” (COUTURE; ROUSSEAU, 1998, p.68). No segundo componente do programa estão aspectos referentes à classificação e à Recuperação da Informação. Para Couture e Rousseau (1998, p.68),

As funções e subfunções da empresa são analisadas, esquematizadas e harmonizadas num sistema de classificação e de recuperação válido para todas as suas componentes. São instauradas, e seguidas por todos, normas precisas de classificação e de recuperação da informação, tornando assim caduca a era das pesquisas desorganizadas, com resultados muitas vezes duvidosos. (COUTURE; ROUSSEAU, 1998, p.68)

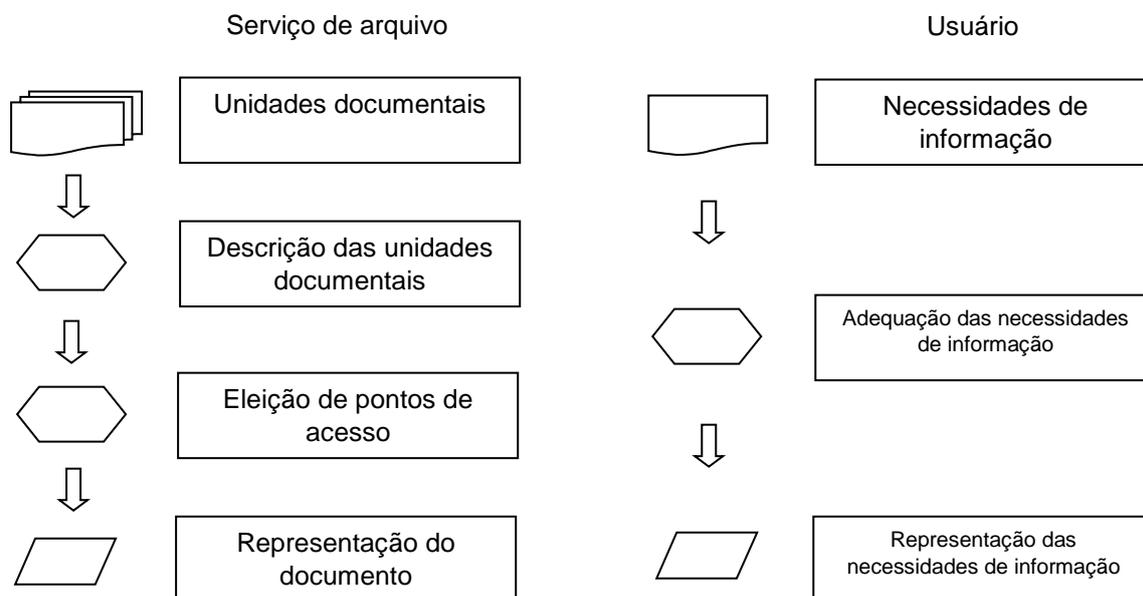
O terceiro componente se centra na proteção e na conservação da informação, com ênfase em práticas relativas a documentos confidenciais e conservação/proteção dos documentos com uma baixa frequência de utilização ou de valores permanentes. Tais práticas constituem elementos cruciais desse tipo de programa (COUTURE; ROUSSEAU, 1998). Assim, a perspectiva de Couture e Rousseau (1998) em torno da RI se evidencia como uma parte integrante de um plano, um programa.

Margaret Procter e Michael Cook, em publicação referência da área arquivística (“*Manual of Archival Description*”), datada de 2000, verificaram a existência de estratégias de RI utilizadas pelos usuários. Para os autores, há três estratégias, sendo elas: 1) identificação direta; 2) navegação; e 3) exploração/busca. A identificação direta se dá quando o usuário conhece uma ou mais das características de identificação dos documentos procurados (por exemplo, um nome específico, data ou código de referência). Já a navegação ocorre quando os usuários lêem as páginas de assistência ou auxílio para encontrar informações ou ideias que considerem úteis. Essa estratégia é

empregada de maneira mais efetiva em descrições de texto livre, como aquelas na História Administrativa e Custodial ou nas Áreas de Conteúdo e Estrutura. Na exploração/busca, o objetivo de recuperação do usuário provavelmente está mais ou menos definido. A ação de busca é identificar palavras-chave, nomes, cadeias de caracteres ou referências específicas, observando-os com os auxílios de busca até que eles apareçam. O usuário identificará rapidamente os campos ou espaços na página nos quais os alvos provavelmente aparecerão (PROCTER; COOK, 2000).

Partindo em outra perspectiva, Martín Suquía (2001), ao analisar o uso de sistemas de informação em arquivos, verificou que, em algumas áreas geográficas, sobretudo na Europa latina, os arquivos têm sido pouco ágeis no aproveitamento de tecnologias. Segundo o autor, quando as utiliza, faz exclusivamente em atividades de descrição, gestão de transferência e recolhimento em arquivos administrativos e históricos e nos processos de empréstimo e controle de usuários, com poucos casos de experiências de informatização integral de serviços de arquivo. Na maioria dos casos, as possibilidades de utilização desses recursos têm um caráter de uso interno e, em pouquíssimos casos, os usuários podem utilizá-los como maneira de consulta direta (MARTÍN SUQUÍA, 2001). A perspectiva apresentada pelo autor considera um modelo conceitual dos SRI em arquivos, conforme figura abaixo:

Figura 5 - Modelo conceitual dos sistemas arquivísticos de recuperação da informação



Fonte: Extraído de Suquía (2001, p. 52)

Suquía (2001) corrobora, no elemento “Unidades documentais” (*Unidades documentales*), o desafio ao descrever as relações existentes entre os documentos, enfatizando que “A consideração desta característica condiciona qualquer sistema de recuperação de informação arquivística que pretenda ser verdadeiramente efetiva” (MARTÍN SUQUÍA, 2001, p.53, tradução nossa)³⁵.

Ao falar em “Descrição das unidades documentais” (*Descripción de las unidades documentais*), evidencia-se uma comparação entre a descrição automática e a intelectual/inteligente. Para o autor, a descrição automática é melhor aproveitada em arquivos administrativos e de gestão, enquanto a descrição intelectual/inteligente se fortalece a partir de vários fatores, como “[...] o fato de o conteúdo dos documentos dever ser descrito de modo a extrair a informação evidente e não evidente [...]” (MARTÍN SUQUÍA, 2001, p.54, tradução nossa)³⁶ e dada a existência de “[...] documentos contendo informações não estruturadas, não estritamente atuais, ou documentos de

³⁵ “La consideración de esta característica condiciona cualquier sistema de recuperación de información archivística que pretenda ser verdaderamente eficaz”.

³⁶ “[...] el hecho de que haya que describir el contenido de los documentos de manera que se extraiga tanto la información evidente como la que no lo es [...]”.

acesso para usuários heterogêneos (internos e externos)” (MARTÍN SUQUÍA, 2001, p.54, tradução nossa)³⁷.

Em “Escolha de pontos de acesso” (*Elección de puntos de acceso*), o autor traz inicialmente a discussão sobre pontos de acesso para o texto (com o uso de técnicas de indexação automática, indexação manual e indexação contextual) e, adiante, discute também os pontos de acesso para o contexto, evidenciando que, a partir da norma ISAD (G), os SRI deveriam recuperar as informações no contexto, isto é,

[...] não é suficiente apresentar o usuário com o resultado direto de sua consulta, é necessário que você possa a qualquer momento conhecer a razão do documento que você consulta como parte de um conjunto de documentos, deve dar a conhecer a sua proveniência (MARTÍN SUQUÍA, 2001, p.60, tradução nossa)³⁸

No que concerne ao lado do usuário, Martín Suquía (2001) fala que, no ponto “Necessidades de informação” (*Necesidades de información*), há dois tipos de usuário: o interno, que conhece a organização e sua necessidade de informação é mais exata; o outro usuário, o externo, geralmente possui uma necessidade de informação mais difusa, cabendo ao SRI oferecer respostas satisfatórias ou não, tornando mais concretas tais necessidades.

No próximo ponto, denominado de “Adequação das necessidades de informação” (*Adecuación de las necesidades de información*), o autor questiona se o sistema é capaz de guiar o usuário quando se trata de combinar estratégias de busca (baseada no contexto de busca e na navegação de texto). O autor sugere que os SRI podem ter mecanismos que expandam as expressões de busca para aquelas mais próximas de sua área semântica ou contextual. Para o autor, a navegação deve estar atrelada à seleção de informações feitas, por exemplo, a partir do uso do Plano de Classificação (MARTÍN SUQUÍA, 2001). Finalizando tal ponto,

O estabelecimento de políticas de treinamento de usuários ou o desenvolvimento de poderosas ferramentas de ajuda automatizadas nos sistemas de recuperação podem ser elementos de reforço importantes para garantir a tradução adequada das necessidades do

³⁷ “[...] documentos que contienen información no estructurada, no estrictamente actuales, o documentos de acceso para usuarios heterogéneos (internos y externos)”.

³⁸ “[...] no es suficiente con presentar al usuario el resultado directo de su consulta, es necesario que pueda en cualquier momento conocer la razón de ser del documento que consulta como parte de un conjunto de documentos, debe de dar a conocer su procedencia”.

usuário para a estratégia de busca (MARTÍN SUQUÍA, 2001, p.63, tradução nossa)³⁹

E no último ponto, o autor não se refere diretamente a questões referentes às representações das necessidades de informação, mas indica questões relativas à avaliação da RI. Para tanto, ele inicia enfatizando que, “Não há experiências conhecidas de avaliação dos resultados de pesquisa aplicados a localização das informações nos arquivos” (MARTÍN SUQUÍA, 2001, p.63)⁴⁰, mas que estas devem levar em conta os dois tipos de resultados: o textual e o contextual. O primeiro, “Os resultados textuais, as referências localizadas, devem ser avaliados em termos de sua exaustividade e pertinência” (MARTÍN SUQUÍA, 2001, p.63, tradução nossa)⁴¹. E quanto à segunda,

Como já mencionamos, a necessidade de combinar sistemas de navegação e busca requer uma apresentação de resultados que permita a visualização das referências selecionadas e os links que permitem que essas referências sejam contextualizadas (MARTÍN SUQUÍA, 2001, p.64)⁴²

Giménez-Chornet (2011), ao focar outras questões da RI e daquilo que contextualizou a pesquisa e prática arquivística, em artigo intitulado “La recuperación de la información en los archivos en línea”, reconhece que o foco de investigação se centrou, na década de 80, em aspectos relacionados à classificação, à descrição, à avaliação, a tipologias documentais, à autenticidade e ao ciclo vital dos documentos; entretanto, pouco se centrou sobre a busca ou a RI em arquivos. Para o autor, o motivo da deficiência das pesquisas sobre tal temática em arquivos se deve ao fato de que os pesquisadores da época

[...] se concentraram nos aspectos que mais preocupavam os arquivistas em um ambiente de deficiências tecnológicas com o qual

³⁹ “El establecimiento de políticas de formación de usuarios o el desarrollo de potentes instrumentos automatizados de ayuda en los sistemas de recuperación pueden constituir importantes elementos de refuerzo para asegurar una adecuada traslación de las necesidades del usuario a la estrategia de la búsqueda”.

⁴⁰ “No se conocen experiencias de valoración de resultados de búsquedas aplicados a la localización de información en archivos”.

⁴¹ “Los resultados textuales, las referencias localizadas, deben ser evaluados en cuanto a su exhaustividad y su pertinencia”.

⁴² “Como ya hemos comentado la necesidad de compaginar los sistemas de navegación y búsqueda exige una presentación de resultados que permita la visualización de las referencias seleccionadas y de los enlaces que permitan contextualizar estas referencias”.

a maioria trabalha. [...] Se nos arquivos dos anos 80 e 90 não há capítulo dedicado à recuperação da informação, não é porque o assunto não é interessante, mas porque o ambiente físico e a falta de implementação tecnológica são bastante difundidos (exceto exceções, como é o caso da automação do Arquivo Geral das Índias, e a recuperação da informação é considerada de outra perspectiva” (GIMÉNEZ-CHORNET, 2011, p.1-2)⁴³.

Ainda que seja visto como prática envolta num programa, Silva *et al* (2017) verificaram a RI em arquivos, no Brasil, em publicações de 2013 a 2017, numa perspectiva de conceito envolto em outras temáticas. Para os autores, a RI em arquivos se consolida em frequente diálogo com as seguintes temáticas: 1) Representação da Informação; 2) Busca, acesso e uso da informação; 3) Classificação e organização documental; 4) Gestão da informação; 5) Mediação da Informação; e 6) Necessidades de informação (SILVA *et al*, 2017). Para os autores, a RI:

mostrou-se como área composta de visões dependentes dos entendimentos da Ciência da Computação, bem como pela Ciência da Informação. Evidenciou-se como um processo constituído por outros e que leva em conta – onde parece haver a influência da CI – a relação com o sujeito, a partir de bases lançadas por outras áreas de interesse (SILVA *et al*, 2017, p.106)

Desse modo, buscou-se no presente subcapítulo descrever as noções que se evidenciam na área arquivística concernente a RI. Assim, o próximo capítulo discutirá os resultados da presente pesquisa, enfocando seus pontos em relação aos objetivos estabelecidos no começo da investigação, com vistas aos temas de Representação e da RI.

⁴³ “se centran en los aspectos que más preocupaban a los archiveros en un entorno de carencias tecnológicas con el cual trabajan la mayoría. [...] Si en los manuales de archivística de los años 80 y 90 no hay un capítulo destinado a la recuperación de la información, no es porque no interese el tema, sino porque el entorno físico y la ausencia de implementación tecnológica es bastante generalizada (excepto excepciones, como el caso de la automatización del Archivo General de Indias), y la recuperación de la información se contempla desde otro enfoque”.

4 RESULTADOS E DISCUSSÕES

O presente capítulo se estrutura a partir dos objetivos específicos, dividindo-o em torno de subcapítulos, com vistas ao cumprimento do objetivo geral. Desse modo, a estrutura dos subcapítulos se deu em torno de: “4.1 Pontos de contato da Representação e RI entre a CI, a Arquivologia e a CC”; “4.2 Ideias centrais da Representação da Informação e Recuperação da Informação (RI) no campo arquivístico”; “4.3 Relações da RI com a Representação no campo arquivístico”; e “4.4 Questões de ordem técnica/teórica da Representação e Recuperação da Informação no campo arquivístico”.

4.1 Pontos de contato da Representação e RI entre a CI, a Arquivologia e a CC

A partir da conjuntura exposta sobre a Representação e a RI em arquivos, vê-se a necessidade de discutir como estas podem apresentar aspectos em comum entre as áreas de CI, CC e a Arquivologia. Tais temas, conforme já visto, são estudados pela CI, no tocante principalmente à Representação, e pela CC, no que se refere a RI. Foi utilizado no presente objetivo a Análise do Conteúdo, conforme já definido na metodologia. Adiante, apresenta-se o quadro 3, expondo-se as obras nos quais os entendimentos das áreas foram retirados. Posteriormente, são apresentadas figuras que apresentam as temáticas e os contatos em forma de intersecção existentes entre tais áreas.

Quadro 3 – Obras analisadas que referenciam a Representação e a RI nas áreas de CI, CC e Arquivologia

Título	Área	Tipo de publicação	Ano	Autor
Linguagens documentárias, instrumentos de mediação e	CI	Artigo	1993	Lara

comunicação				
Análise documentária e representação da informação	CI	Artigo	1996	Kobashi
El documento: entre la tradición y la renovación	CI	Livro	2002	Rodríguez Bravo
Uma visão da web semântica	CC	Artigo	2004	Lima e Carvalho
Os fundamentos da disciplina arquivística	Arquivologia	Livro	1998	Couture e Rousseau
Registros documentais contemporâneos como provas de ação	Arquivologia	Artigo	1994	Duranti
A Web Semântica e suas contribuições para a ciência da informação	CI	Artigo	2004	Souza e Alvarenga
Ontologias como domínio conceitual e planos de classificação arquivística	Arquivologia/CI	Dissertação	2010	Santos
Ontologias e a web semântica	CC	Artigo	2003	Freitas
What do we need for ontology integration on the semantic web, position statement	CC	Artigo	2003	Noy
Ontologia Digital Arquivística: interoperabilidade	Arquivologia/CI	Dissertação	2016	Luz

e preservação da informação arquivística em sistemas de arquivo e na Web				
Instrumentos e metodologias de representação da informação	CI	Artigo	1996	Novellino
Arquivo: teoria e prática	Arquivologia	Livro	2004	Paes
ASK for information retrieval: Part I	CI	Artigo	1982	Belkin, Oddy e Brooks
On user studies and information needs	CI	Artigo	1981	Wilson
Information needs and uses	CI	Artigo	1986	Dervin e Nilan
O sentido dos arquivos	Arquivologia	Artigo	2014	Bellotto
Los estudios de usuarios en los archivos	Arquivologia	Artigo	2013	Allendez Sullivan e Puente
Satisfacción de usuarios internos y externos de todos los servicios de una clínica de medicina familiar del sector salud en el estado de México	Arquivologia	Artigo	2016	Valero
Indexação documentária: uma forma de representação do conhecimento registrado	CI	Artigo	2001	Pinto

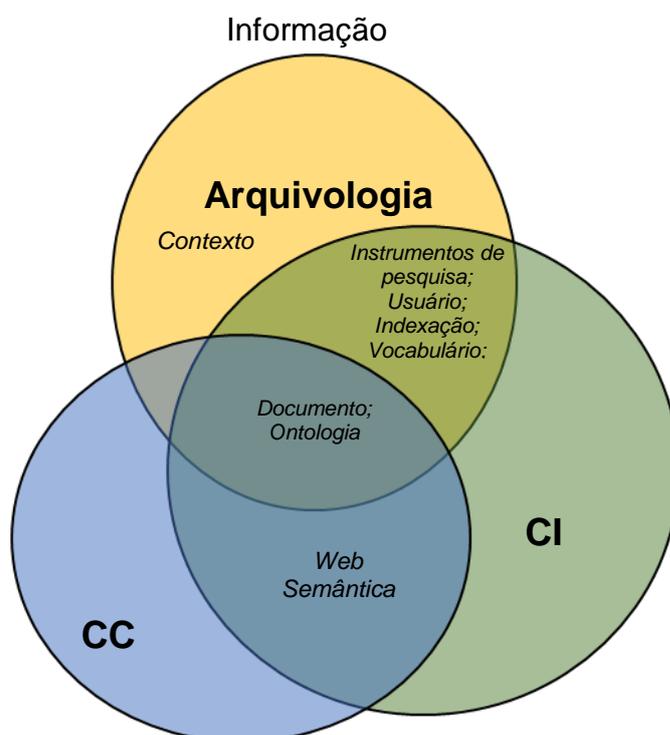
Arquivos permanentes: tratamento documental	Arquivologia	Livro	2007	Bellotto
A Indexação e a Arquivística: aproximações iniciais no universo teórico da organização e representação do conhecimento	Arquivologia	Artigo	2016	Barros
A classificação e o vocabulário controlado como instrumentos efetivos para a RI arquivística	Arquivologia	Artigo	2017	Sousa e Araújo Junior
A indexação de assunto em documentos arquivísticos: convergências e divergências em artigos internacionais	Arquivologia	Artigo	2017	Medeiros, Bräscher e Vianna
Vocabulário controlado: estrutura e utilização	CI	Artigo	2008	Kobashi
Teoria y practica	Arquivologia	Livro	1991	Heredia Herrera
Information filtering and information retrieval: Two sides of the same coin?	CI	Artigo	1992	Belkin e Croft
Folksonomia como uma estratégia para	CI	Artigo	2011	Viera e Garrido

recuperação colaborativa da informação				
Recuperação de Informação: Análise sobre a contribuição da Ciência da Computação para a Ciência da Informação	CI	Tese	2003	Ferneda
Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca	CC	Livro	2013	Baeza-Yates e Ribeiro-Neto
Manual of archival description	Arquivologia	Livro	2000	Procter e Cook
Sistemas de recuperación de la información en los archivos: un análisis de situación y perspectivas	Arquivologia	Artigo	2001	Martin Suquía
La recuperación de la información de los archivos en línea	Arquivologia	Artigo	2011	Giménez-Chornet
Information Retrieval Interaction	CI	Livro	1992	Ingwersen
Information science	CI	Livro	2009	Saracevic
Recuperação da Informação	CI	Artigo	2007	Garcia

Fonte: autoria própria

Diante das propriedades e noções já expostas, evidenciam-se pontos de contato entre as áreas. As figuras 6 e 7 foram produzidas com base nos achados teóricos/conceituais, conforme visto na fundamentação teórica do presente trabalho.

Figura 6 - Interfaces entre a CC, CI e Arquivologia na Representação da



Fonte: autoria própria

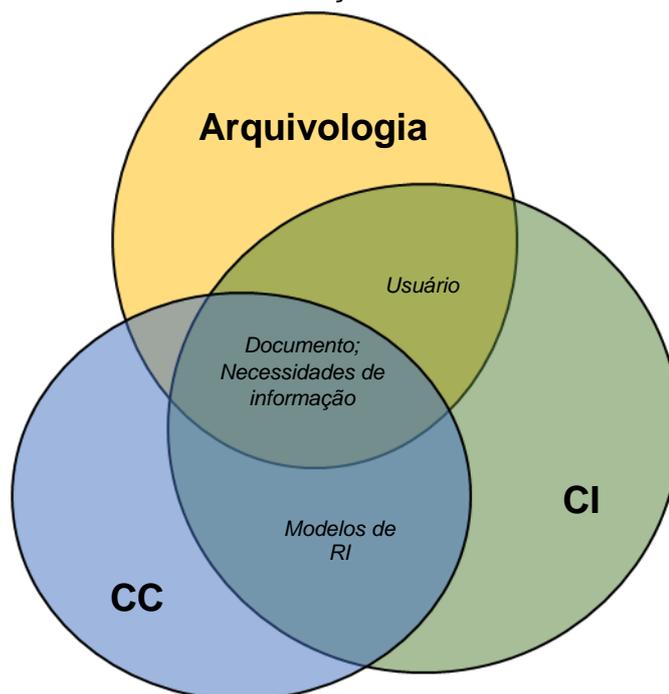
No que toca à temática de Representação da Informação, de acordo com a figura 6, ficaram evidentes, a partir dos achados conceituais das três áreas, visões que corroboram para que exista o uso de conceitos como Documento pela CI, a partir de Lara (1993), Kobashi (1996) e Rodríguez Bravo (2002); Lima e Carvalho (2004), no tocante principalmente a Web Semântica, na CC; Couture e Rousseau (1998) e Duranti (1994) na Arquivologia. É preciso ressaltar que, o conceito de documento, em torno das áreas aqui referenciadas, tem seus usos semelhantes, com o intuito de referenciar dados/informações reunidos em suportes (digitais ou convencionais). Outro assunto discutido pelas áreas é a Ontologia. A CI, a CC e a Arquivologia

possuem posicionamentos alinhados, conforme Souza e Alvarenga (2004), Santos (2010), Freitas (2003), Natash Noy (2003) e Luz (2016).

Referente às áreas de CI e Arquivologia, é possível observar o assunto de “Instrumentos de pesquisa” ao perceber, na primeira, os escritos de Novellino (1996) e, na segunda, os trabalhos de Paes (2004), respectivamente. O mesmo ocorre no tocante ao “Usuário”, com Belkin, Oddy e Brooks (1982), Wilson (1981) e Dervin e Nilan (1986) na CI; e Bellotto (2014), Allendez Sullivan e Puente (2013) e Valero (2016) na Arquivologia. No que se refere à “Indexação”, Novellino (1996) e Pinto (2001) discutem tal assunto em torno da CI, seguindo o mesmo posicionamento do que vem a ser a indexação com os olhares de autores da Arquivologia, como Bellotto (2007), Barros (2016), Sousa e Araújo Junior (2017) e Medeiros, Bräscher e Vianna (2017). O mesmo ocorre com o assunto “Vocabulário” a partir da CI, com os trabalhos de Novellino (1996) e Kobashi (2008). Na Arquivologia, Heredia Herrera (1991) e Medeiros, Bräscher e Vianna (2017) também enfatizam tal assunto.

E entre a CI e a CC há o entendimento do conceito de Web Semântica, que vem sendo investigado por ambas as áreas e que possuem entendimentos consensuais, conforme Souza e Alvarenga (2004), Freitas (2003) e Lima e Carvalho (2004). Quanto à noção de “Contexto”, explorada pela Arquivologia, o conceito tem explorado tal tema de modo exclusivo, em razão de ser um elemento crucial para a representação dos documentos arquivísticos. “Contexto” pode ser entendido aqui como o ambiente em que ocorre a ação registrada no documento, deixando a análise do contexto em focar no documento em si e concentrando sua análise em toda a estrutura que o envolve, seja ela o contexto documental, jurídico-administrativo, de proveniência e tecnológico (CONSELHO NACIONAL DE ARQUIVO, 2014).

Figura 7 - Interfaces entre a CC, CI e Arquivologia na Recuperação da Informação



Fonte: autoria própria

No tocante a RI, a figura 7 demonstra atuações existentes nas e entre as áreas. A primeira delas diz respeito ao uso do “Documento”, num sentido como o visto na análise da figura 6. Referenciam o documento, no processo da RI: na CI, a partir de Belkin e Croft (1992) e Viera e Garrido (2011); na CC, Ferneda (2003) e Baeza-Yates e Ribeiro-Neto (2013); e na Arquivologia, Couture e Rousseau (1998), Procter e Cook (2000), Martín Suquía (2001) e Giménez-Chornet (2011). No tocante a “Necessidades de informação”, enfatiza-se as necessidades existentes em usuários quando utiliza consultas em SRI, conforme Wilson (1981), Belkin, Oddy e Brooks (1982), Peter Ingwersen (1992), Belkin e Croft (1992), Tefko Saracevic (2009), no contexto da CI; na CC, Baeza-Yates e Ribeiro-Neto (2013); e na Arquivologia Martín Suquía (2001).

Um aspecto em comum na RI quando se relaciona à área da CI com a Arquivologia diz respeito ao ponto “Usuário”. No contexto do presente trabalho, é possível observar autores como Saracevic (2009), Garcia (2007), Ingwersen (1992) e Belkin e Croft (1992) revelando o usuário quando se

discute sobre a RI com o enfoque na CI; na Arquivologia, Procter e Cook (2000) investigam estratégias de usuários no processo de RI; enquanto Martín Suquía (2001) evidencia um modelo conceitual dos sistemas arquivísticos de recuperação da informação, tendo foco no usuário em diversas etapas do modelo. E o ponto em comum entre a CI e a CC se evidencia também no tocante a “Modelos de RI”. Sendo um assunto abordado exhaustivamente pela CC, há autores na CI que abordam tal assunto, conforme Belkin e Croft (1992); na CC, Baeza-Yates e Ribeiro Neto (2013).

Assim, é preciso ressaltar algumas observações no que diz respeito aos assuntos conforme figuras 6 e 7. Entre as observações, é preciso ressaltar que o campo de “Usuário”, apesar de não estar explicitamente na área da CC, como demonstra as figuras, não significa que não tenha sido estudado pela mesma área; ou que “Modelos de Recuperação da Informação” não tenham, em algum momento, sido citados por estudos arquivísticos. A figura enfatiza que há interfaces existentes nos assuntos aí expostos, que permeiam a RI e consolidam assim visões existentes nas três áreas. A complementaridade se torna aparente a partir do momento que a Arquivologia ou a própria CI precisa recorrer a Modelos de RI para fazer suas contribuições em tal temática.

Ainda que as figuras 6 e 7 enfatizem elementos (interfaces) que estão dispostos tanto na Representação como na RI, foi necessário empreender uma investigação para entender os elementos em comum existente na CI, CC e Arquivologia ou onde havia esses pontos de contato. As áreas aqui investigadas possuem tópicos que efetivam contribuições de uma para a outra área. É preciso ressaltar que a RI, aqui, leva em conta os aspectos citados por Garcia (2007) e Capurro (2003), considerando o elemento humano na CI, da mesma forma que a Arquivologia também o faz. A visão da CC se consolida por ser mais ligada à máquina, conforme Baeza-Yates e Ribeiro-Neto (2013), a partir do desenvolvimento de fórmulas e técnicas implantadas em SRI. O caminho que liga as três áreas no tocante a RI se efetiva a partir, principalmente, do conceito de documento, necessidades de informação e mais recentemente no estudo de ontologias. As relações entre as três áreas podem ser estabelecidas conforme Kapre e Basha (2012), a partir de elementos como usuários, tecnologias da informação e direções de pesquisa.

4.2 Ideias centrais da Representação da Informação e Recuperação da Informação (RI) no campo arquivístico

Partindo do percurso exposto, é necessário conhecer as ideias centrais da Representação e RI no campo arquivístico, conforme estabelecido nos objetivos do presente trabalho. De início, consideremos a noção de Representação da Informação, que se consolida a partir dos conceitos de classificação, indexação e descrição. Os resultados se desdobraram nos procedimentos de aplicação da análise de conteúdo às definições de tais conceitos, isto é, aplicou-se uma análise do conteúdo com viés temático nas definições e, a partir disso, se efetuou a categorização de tais dados, como disposto abaixo.

É importante ressaltar que foi utilizada, especificamente aqui, a Análise de Conteúdo, com ênfase na análise temática, trabalhando com as diversas definições e com a posterior categorização, levando em consideração o sentido aparente que a definição expunha, como já delimitado na metodologia. Dessa forma, o primeiro conceito analisado foi o de classificação, como consta no quadro 4.

Quadro 4 – Definições do conceito de Classificação arquivística

CLASSIFICAÇÃO		
<u>Definição</u>	<u>Autores</u>	<u>Categorias</u>
O arranjo dos documentos de acordo com um plano projetado para tê-los disponíveis para o uso corrente	SCHELLENBERG (2006)	Ação administrativa
Estabelecimento de classes, de famílias, agrupadas estrutural ou hierarquicamente dentro de um conjunto	HEREDIA HERRERA (1991)	Ação segmentar
Organização dos documentos de um arquivo ou coleção, de acordo com um plano de classificação, código de classificação ou quadro de arranjo	ARQUIVO NACIONAL (2005)	
Processo intelectual de identificação e agrupamento sistemático de itens semelhantes, baseado em características comuns que posteriormente podem ser diferenciadas se a quantidade demandar	COUTURE (1999)	Ação intelectual
Análise e identificação do conteúdo de documentos, seleção da categoria de assunto sob a qual sejam recuperados,	ARQUIVO NACIONAL (2005)	

podendo-se-lhes atribuir códigos		
Ação intelectual de construir esquemas para agrupar os documentos a partir de princípios estabelecidos	SOUSA (2009)	
Ordenação intelectual e física de acervos, baseada em uma proposta de hierarquização das informações referentes a eles	LOPES (2009)	
Operação intelectual e técnica, que se traduz numa categorização/sistematização para fins organizativos e numa representação formal tendo em vista a recuperação da informação	RIBEIRO (2013)	

Fonte: autoria própria

De acordo com o quadro acima, o conceito de classificação se apresenta a partir de ideias que se consolidam em torno da ação administrativa, isto é, no sentido de ato executado nos trâmites necessários ao andamento da administração de organizações. Tal definição evidencia a separação de procedimentos aplicados nos arquivos correntes daqueles aplicados em arquivos permanentes, denominado por autores como Schellenberg (2006) como “classificação” e “arranjo”, respectivamente.

Com o devido prosseguimento nas análises e categorizações, a categoria evidenciada foi a de ação segmentar. A noção de classificar, num sentido genérico, implica o ato de divisão e/ou separação, baseando-se em características dos objetos. No meio arquivístico, o enfoque é dado a fundos e séries documentais, utilizando-se critérios estruturais ou hierárquicos, com uso de classes definidas em razão de instrumentos de classificação.

E a outra categoria visualizada foi a denominada de ação intelectual, por levar em consideração o seu fazer como constituída a partir da natureza humana, expressa em atos de operar, analisar, identificar, processar e ordenar, típicos de um plano subjetivo. A noção intelectual parte da atuação do arquivista, responsável por guiar a construção ou adequação dos instrumentos (planos) de classificação arquivísticos.

É importante observar que a definição de classificação apresenta, em uma das definições (Arquivo Nacional, 2005), a peculiaridade de ser uma ação de organização de acordo com um plano de classificação – isto é, uma

ação segmentar – e de ser também uma ação que implica análise/identificação – ou seja, uma ação intelectual.

Quadro 5 – Definições do conceito de Indexação arquivística

INDEXAÇÃO		
<u>Definição</u>	<u>Autores</u>	<u>Categorias</u>
A ação de descrever e identificar um documento de acordo com seu assunto	UNISIST (1976)	Ação descritiva
Identificação e seleção de termos de índice de assunto a partir da análise de conteúdo de um documento	MACNEIL (1996)	
A identificação dos conceitos de que tratam os documentos, por meio de termos, descritores ou palavras-chave	SOUSA; ARAÚJO JÚNIOR (2017)	
Processo pelo qual documentos ou informações são representados por termos, palavras-chave ou descritores, propiciando a recuperação da informação	ARQUIVO NACIONAL (2005)	Fornecimento de acesso
Operação que visa criar rubricas para acesso, por assuntos, à informação	RIBEIRO (2013)	
Atribuição de descritores, com base no conteúdo do documento, que irão compor um índice de busca ⁴⁴	ROCHA (2011)	Elaboração de instrumentos

Fonte: autoria própria

O quadro 5 enfatizou o conceito de indexação, em que se consolidaram categorias como “ação descritiva”, “fornecimento de acesso” e “elaboração de instrumentos”. No que se refere à “ação descritiva”, esta se delinea em torno da função ou ato de representação em si, a partir da análise ou identificação de documentos, com posterior escolha de termos com relação ao seu assunto.

A categoria “fornecimento de acesso” se consolidou com início em torno do processo representativo e com enfoque para o acesso ou o fornecimento dele. A ideia de fornecimento de acesso ainda é evidente nesse caso, expondo a necessidade básica de representação de documentos e se justificando como atividade matriz na prática arquivística.

E a categoria de “elaboração de instrumentos” aparece com enfoque na noção em compor um índice de busca, a partir da atribuição de descritores.

⁴⁴ Contexto de documentos digitais e no processo de captura de documentos pelo SIGAD.

Apesar de se constituir na elaboração de produtos, a definição existente em tal categoria expressa também a necessidade de atribuição de descritores. Tal categoria vai, portanto, adiante, quando enseja a elaboração de instrumentos, e não só a questão descritiva.

Quadro 6 - Definições do conceito de Descrição arquivística

DESCRIÇÃO		
<u>Definição</u>	<u>Autores</u>	<u>Categorias</u>
É a análise feita pelo arquivista sobre os fundos e documentos arquivados agrupados naturalmente ou artificialmente, a fim de sintetizar e condensar as informações nele contidas para serem oferecidas às partes interessadas	HEREDIA HERRERA (1991)	Ação intelectual
Identificar e explicar o contexto e o conteúdo de documentos de arquivo a fim de promover o acesso aos mesmos	CONSELHO INTERNACIONAL DE ARQUIVOS (2000)	
O processo de analisar, organizar e registrar detalhes sobre os elementos formais de um registro ou coleção de registros, como criador, título, datas, extensão e conteúdo, para facilitar a identificação, o gerenciamento e o entendimento do trabalho	PEARCE-MOSES (2005)	
A criação de uma representação precisa de uma unidade de descrição e suas partes componentes, se houver, capturando, analisando, organizando e registrando informações que servem para identificar, gerenciar, localizar e explicar materiais de arquivo e os sistemas de contexto e registros que o produziram	INTERPARES (2018)	
É a parte da tarefa arquivística que abrange as diversas atividades desenvolvidas pelo arquivista para desenvolver os instrumentos que facilitam o acesso a fundos, isto é, pontos de acesso	CRUZ MUNDET (1994)	Elaboração de instrumentos
Criação de ferramentas de acesso (guias, inventários, ferramentas de busca, registros bibliográficos) ou sistemas (catálogos de cartões, bases de dados bibliográficas, bancos de dados EAD) resultantes dessas	YAKEL (2003)	

atividades		
Conjunto de procedimentos que leva em conta os elementos formais e de conteúdo dos documentos para elaboração de instrumentos de pesquisa	ARQUIVO NACIONAL (2005)	
Consiste na elaboração de instrumentos de pesquisa que possibilitem a identificação, o rastreamento, a localização e a utilização de dados	BELLOTTO (2007)	

Fonte: autoria própria

No quadro 6, em torno do conceito de descrição, se evidenciaram duas categorias: a ação intelectual e a elaboração de instrumentos. A primeira se consolida a partir da noção de análise, com atos de identificação, explicação e registro aplicados aos elementos arquivísticos, em geral, a partir da iniciativa do arquivista. Nesse sentido, tal categoria se evidenciou, conforme já visto no quadro 4, na atividade de classificação.

A categoria de elaboração de instrumentos aparece aqui como no sentido dado anteriormente, a partir da execução de uma ação com o intuito de criação de produtos (instrumentos de pesquisa). Entretanto, a definição aqui se amplia pela atuação explícita do arquivista em tal processo. Evidenciam-se também os dois sentidos conforme as categorias já revelam: ao mesmo tempo em que se tem uma ação intelectual, passível e dependente do trabalho cognitivo do arquivista, se desdobra também o objetivo de produção/elaboração de instrumentos de pesquisa.

Quanto à temática de RI, consolidou-se a sua exposição no quadro 7. Assim, se verificou a tímida quantidade de autores ou obras que procuram definir a recuperação da informação no seio arquivístico. Os autores e obras do quadro se verificaram em torno de autores espanhóis e do glossário da *Society of American Archivists* (SAA), com autoria e organização de Richard Pearce-Moses (2005), bem como do DTBA, publicado pelo Arquivo Nacional.

Quadro 7 - Definições do conceito de Recuperação da Informação arquivística

RECUPERAÇÃO DA INFORMAÇÃO		
<u>Definição</u>	<u>Autores</u>	<u>Categorias</u>
Ação que ocorre mediante o uso de um	ESTEBAN	Processo

instrumento de descrição automatizado, a partir da busca livre em bases de dados ou restritos a um ou vários campos que correspondem aos diversos elementos de informação da ISAD(G)	NAVARRO; GAY MOLÍNS (1998)	informatizado
É o processo de seleção de informações de um fundo documental por meios automáticos ou semi-automáticos, ou seja, com o uso parcial ou intensivo de computadores	CODINA (2002)	
Identificação ou localização da informação desejada	ARQUIVO NACIONAL (2005)	Suprimento de necessidade informacional
Um processo que medeia entre grandes quantidades de informações e usuários, recuperando um subconjunto selecionado das informações que mais correspondem à solicitação do usuário	PEARCE- MOSES (2005)	

fonte: autoria própria

Ao se analisar as definições, verificou-se a existência das categorias de processo informatizado, partindo da premissa de uso de computadores e do processamento automático de consultas mediante o emprego de expressões por parte do consulente. Nota-se o uso da expressão “elementos de informação da ISAD(G)” (ESTEBAN NAVARRO; GAY MOLÍNS, 1998) numa clara referência ao que hoje se denomina de metadados.

A outra categoria, denominada de suprimento de necessidade informacional, manifesta a ideia de atender às necessidades informacionais como um fim para a recuperação da informação. Ainda sim, verifica-se que ficam expressos também ideias gerais da Recuperação da Informação da Arquivologia em relação a outras áreas (Ciência da Informação e Ciência da Computação), a respeito de se estabelecer como um processo e de focar na seleção de informações de um dado conjunto de dados/documentos/informações.

Portanto, diante das definições e categorizações efetuadas ao longo dos quadros anteriores, nota-se que ideias centrais da Representação da Informação e da RI no meio arquivístico se fortalecem a partir de várias perspectivas. No caso da Representação da Informação, notadamente operacionalizada a partir das práticas de classificação, indexação e descrição, suas ideias firmam-se em torno da: ação administrativa, segmentar, intelectual

e descritiva, com objetivo de fornecimento de acesso e posterior elaboração de instrumentos. No caso da RI, ela transparece como um processo informatizado, com claro objetivo de suprimento de necessidades informacionais.

4.3 Relações da RI com a Representação no campo arquivístico

Com as ideias centrais expostas, é necessário estabelecer como ambas (Representação e RI) podem se relacionar entre si. As definições elencadas anteriormente deram uma dimensão geral de como a Representação da Informação junto com a RI são vistas no meio arquivístico, principalmente quando se notabilizam a partir de autores e obras de frequentemente citadas na área. Nesse sentido, a verificação de relações de uma com a outra é admissível, a partir da atenção destinada a essas definições, com possíveis identificações de pontos de contato existentes entre elas.

No geral, é possível observar a partir dos quadros 4, 5 e 6 a existência de termos em definições sobre a representação que corroborem a existência da recuperação como um processo e/ou atividade posterior. No quadro 4, que versa sobre a classificação, há uma definição do DTBA que evidencia a seleção da categoria de assunto com perspectiva para posterior recuperação [da informação, supõe-se] (Arquivo Nacional, 2005). Ainda no mesmo quadro, há a existência de outra definição, quando evidencia a prática representacional do documento com vistas à recuperação da informação (RIBEIRO, 2013). No quadro 5, corrobora-se a mesma lógica do quadro 4, quando evidencia a indexação com vistas à atribuição de termos, palavras-chave ou descritores, para representação de documentos/informações, embasando a recuperação da informação (ARQUIVO NACIONAL, 2005).

Já no sentido inverso, isto é, de RI em direção à Representação da Informação, o quadro 7 apresenta a definição que expõe implicitamente a representação a partir da citação da ISAD(G) (PEARCE-MOSES, 2005). No sentido da Recuperação em relação à Representação da Informação, nota-se uma discreta relação, ficando aparente, nesse relacionamento (bidirecional, por sinal), a ideia de que é primordial se representar primariamente, com foco posterior para recuperação, numa associação de complementaridade. A razão

da representação, quando evidencia e traz a recuperação para o seu contexto, transparece aqui como um objetivo, etapa ou requisito no processo de recuperação da informação. Diante da exposição de noções e do contexto que as circunda, verificou-se a existência de questões/reflexões que se aproximam da Representação e RI no enfoque arquivístico. Nesse sentido, discutiram-se adiante considerações que se relacionaram, em primeiro plano, em torno da Representação da Informação e, em seguida, tendo o mesmo enfoque, mas na RI.

4.4 Questões de ordem técnica/teórica da Representação e Recuperação da Informação no campo arquivístico

Assim, comecemos falando da representação no campo arquivístico. A perspectiva em tal área assevera o uso de estruturas hierárquicas (Proveniência) e de organização (Ordem original) como formas principais de representação da informação. Desse modo, a representação arquivística do conteúdo é tida como secundária e, portanto, não evidenciada com a mesma importância da representação arquivística da procedência/proveniência e da representação arquivística da ordem [original] (ZHANG, 2012). Ainda no tocante à ordem original, Yakel (2003, p.6, tradução nossa) ressalta que tal princípio, na prática arquivística, visa manter ou estabelecer um reflexo da ordem original [interna] existente nos escritórios para os arquivos. Na visão da autora, “os modelos de metadados emergentes para gerenciamento, descrição e preservação de objetos digitais também visualizam metadados como cumprindo essa função de extensão de limites [do escritório e do arquivo]”⁴⁵.

A prática de manutenção do vínculo arquivístico se consolida, entre arquivistas, de fundamental importância na representação, impossibilitando a análise de conteúdo, por meio principalmente da indexação (SOUSA; ARAÚJO JUNIOR, 2017) como potencial atividade. Ribeiro (2013) questiona, nessa direção, diante de uma conjuntura social de acesso à informação como desafios a serem enfrentados pelos arquivistas, somados à investigação do tratamento da informação por assuntos nos ambientes arquivísticos.

⁴⁵ “*Emerging metadata models for digital object management, description, and preservation also view metadata as fulfilling this boundary spanning function*”.

Volvidos cerca de quinze anos, **é legítimo que nos questionemos sobre o modo como evoluiu o tratamento da informação por assuntos nos arquivos** e que nos interroguemos sobre como se tem processado a integração dos arquivos na sociedade digital e em rede em que nos movimentamos. Não se conhecem estudos de síntese que analisem esta problemática, mas o que podemos afirmar é que, nas últimas duas décadas e sob o efeito incontrolável da internet, os arquivistas se vêem pressionados a disponibilizar na Web os seus conteúdos e sofrem a pressão dos utilizadores que não querem ter de se deslocar às salas de leitura dos arquivos para consultar a informação de que necessitam (RIBEIRO, 2013, p.537, grifo nosso).

É preciso assinalar, como já dito, que “As iniciativas de pesquisa continuaram a se concentrar na identificação, autenticação e preservação das relações contextuais dos registros por meio de metadados de manutenção de registros” (ZHANG; MAUNEY, 2013, p. 176), não privilegiando, mais uma vez, a relação entre a descrição do conteúdo documental e o contexto de produção e trâmite existentes. Nesse sentido, reflete-se sobre o caráter representacional que é gerado a partir do uso de padrões de descrição e dos produtos gerados pela atividade de representação da informação no contexto arquivístico. Zhang (2012, p. 48, tradução nossa) reflete sobre o poder da representação, num estudo sobre a prática nos arquivos:

no entanto, os pesquisadores também apontam que a representação nunca é exatamente idêntica ao objeto representado e pode ser uma idealização tendenciosa desse objeto. A representação é sempre influenciada pelo objeto representado e pelo sujeito que faz a representação, e a maneira como as coisas são representadas pode ser influenciada por tradições ou ideologias.⁴⁶

Atentando-se ao contexto que envolve a prática de representação nos arquivos e sob uma perspectiva complementar, posicionando o elemento humano como partícipe e mediador no processo, Hedstrom (2002) nos fornece outras implicações trazidas na prática de representação, evidenciando que o conhecimento das equipes de arquivo não é formalmente representado nos sistemas descritivos. A representação no campo arquivístico se apresenta pela forte interação do arquivista em todo o processo de tratamento documental. Em geral, um dos primeiros aspectos diz respeito à classificação, em que se assegura a atribuição de códigos (relacionados a classes

⁴⁶ “However, researchers also point out that representation is never exactly identical with the object being represented and might be a biased idealization of that object. Representation is always influenced by both the object represented and the subject doing the representation, and the way things are represented may well be influenced by traditions or ideologies”.

funcionais/estruturais) para documentos, bem como a construção de instrumentos que baseiam tal prática, a partir da interlocução de um trabalho consideravelmente intelectual. O mesmo ocorre no processo de indexação, no qual há integral atuação do arquivista, afirmando, mais uma vez, o aspecto humano como crucial no exercício laboral de representação da informação/descrição. Essas considerações –é preciso ressaltar– levam em conta a prática aplicável a acervos documentais com pouquíssimo ou nenhum tratamento.

Um ponto do trabalho arquivístico diz respeito ao advento tecnológico, com suas múltiplas intervenções e conseqüente adoção de soluções para informatização (tanto de procedimentos inerentes à atividade-meio e à atividade-fim), que trouxeram desafios no tocante à descrição de documentos arquivísticos. Como forma de atenuar os impactos de iniciativas tecnológicas – sistemas digitais para produção, uso e gerenciamento de documentos – verificou-se que as soluções perpassam, inicialmente, pela adoção de vocabulários controlados (KOBASHI, 2008), entendidos como “Conjunto normalizado de termos que serve à indexação e à recuperação da informação” (ARQUIVO NACIONAL, 2005, p. 174). Dito isso, as reflexões se ampliam ao se evidenciar a diversidade de expressões de representação em documentos administrativos. A questão se estende quando a representação é efetuada pelos criadores dos documentos ou seus usuários. Na visão de Martín Suquía (2001, p.54, tradução nossa):

Não há consistência na descrição porque é o criador do documento que o descreve, geralmente de uma maneira muito vaga, e como um instrumento de recuperação pessoal, não destinado a usuários “externos”, sem pretender extrair a informação que o criador do documento dá por conhecido.⁴⁷

Com a adoção do vocabulário controlado, hoje se tende a alinhar as preocupações acima com a adoção de esquemas de metadados. Tal perspectiva assegura que, entre algumas das práticas que podem ser adotadas no registro ou captura de um documento, sejam necessárias, por parte de um SIGAD (Sistema Informatizado de Gestão Arquivística de Documentos), a

⁴⁷ “No existe una consistencia en la descripción debido a que es el propio creador del documento quien lo describe, normalmente de una manera muy vaga, y como instrumento de una recuperación personal, no pensada para usuarios “externos”, sin pretender extraer la información que el creador del documento da por conocida”.

adoção de um conjunto ou padrão de metadados que supram as necessidades do domínio onde a instituição atua. Nesse sentido, é preciso ressaltar a importância da ISO 23081, que, dentre seus objetivos, tem a intenção de servir como guia para entender, implementar e usar metadados no âmbito da ISO 15489, que é a norma que especifica a gestão de documentos. A ISO 23081 não define um conjunto de metadados obrigatórios, mas traz princípios para criação de metadados de gerenciamento de registros e possibilidades de avaliação de conjuntos de metadados.

Portanto, do ponto de vista de funcionamento dos arquivos, várias questões são postas e que podem influenciar aspectos da representação da informação. O uso de padrões de metadados é uma das práticas que buscam, ainda que sob outra perspectiva e em outros níveis de atuação, aperfeiçoar procedimentos que influenciarão na própria representação. Assim, esboçam-se as notáveis peculiaridades relativas a questões de ordem técnica e teórica, como a classificação, a descrição de documentos e a indexação. Traz, atrelada às atividades supracitadas, a questão da não primazia pela descrição do conteúdo de documentos, em detrimento à escolha de representação da estrutura hierárquica ou de funções que representam o contexto documental em que o documento foi produzido ou recebido por uma instituição.

Nesse sentido, a área da Representação da Informação fornece subsídios para a busca de informações. No entanto, a Representação da Informação por si só não evidencia etapas e soluções para suprir necessidades informacionais de indivíduos. No tocante a esse aspecto, a RI se apresenta como importante elo com a Representação. A RI, no plano arquivístico, se consolida a partir de práticas tradicionais (e consolidadas, como o uso do modelo booleano), e elementos que se evidenciam de maneira peculiar.

Um dos fatores mais evidentes na RI no contexto arquivístico é, em suma, não haver constatações técnicas de aspectos peculiares da natureza arquivística dos documentos e dos elementos que compõem um acervo arquivístico nos estudos, capazes de trazer diferentes abordagens para o problema de recuperação de documentos/informação. A essas ideias soma-se o que Hedstrom (2002) revela sobre a imprecisão de sistemas descritivos não conceberem o conhecimento existente em equipes de arquivo, o que consolida a mediação humana como meio para sucesso em buscas de usuários.

No sentido disposto acima, a RI aqui se verifica fortemente atrelada pela capacidade de mediação e compreensão da necessidade informacional de um determinado consulente por parte do arquivista. A essa capacidade, acrescenta-se a faculdade de melhorias e utilização de variadas expressões em buscas, além do contínuo aprimoramento no encontrar da informação de acordo com a maneira que um determinado indivíduo expõe sua necessidade informacional e/ou documental.

Partindo do pressuposto de relação, a Representação da Informação e a RI no contexto arquivístico se validam também a partir da noção de que o procedimento de Recuperação é fortemente influenciado pela qualidade da representação (CHU, 2003 apud ZHANG, 2012). Isso pode indicar que, diante da conjuntura laboral de múltiplas atribuições do arquivista, a capacidade de representação e posterior recuperação podem ser afetadas. O aspecto de mediação do profissional de arquivo (FERREIRA; ALMEIDA JÚNIOR, 2013) transparece como elemento primordial, indispensável, partícipe da relação Representação-Recuperação da Informação. A atuação do arquivista, nesse aspecto, evidencia uma intensa relação de Representação e RI no âmbito dos arquivos.

5 CONSIDERAÇÕES FINAIS

Diante das exposições que permearam esse estudo, evidenciou-se que a Representação ao lado da RI são atividades primordiais ao fazer arquivístico. As preocupações de uma área podem ser compartilhadas com outras. As compreensões das áreas aqui evidenciadas nos fornecem o entendimento de que aspectos em comum são conexões existentes entre as áreas, como no caso da Representação da Informação, com as questões de ontologia, vocabulário, usuário, dentre outros, evidentes tanto na Arquivologia, como na Ciência da Informação e Ciência da Computação. O mesmo se compreende da RI, quando compartilha conceitos como análise documentária, indexação, entre outros tópicos.

O entendimento das ideias centrais existentes no meio arquivístico no tocante à Representação da Informação torna claro que, a partir das noções de classificação, indexação e descrição, a compreensão de tal tema se desenvolve em torno da ação administrativa, segmentar, intelectual e descritiva, com elaboração de instrumentos para o fornecimento de acesso. No caso da RI, ela se notabiliza como um processo informatizado, com enfoque em suprir necessidades informacionais. A Representação transparece como uma etapa essencial para a consecução da RI. Há, ao mesmo tempo, uma relação de associação e dependência, o que não pode indicar prejuízos para uma ou outra, mas uma lógica coerente, entendendo que o ato de recuperação é extremamente dependente da qualidade de representação.

Algumas reflexões se verificaram em torno da qualidade de representação do viés arquivístico, com a ênfase no elemento contextual, em detrimento ao conteúdo dos itens documentais. É necessário salientar que a conjunção de uma perspectiva com a outra é um desafio a ser superado por futuras investigações. Não se deve haver prejuízo no processo representacional, incorrendo-se em impacto direto no fazer da recuperação.

Desse modo, compreendem-se as visões existentes da Representação e RI que se revelam em tais temáticas na CC, CI e Arquivologia. Corroboram-se a complementaridade existente entre as áreas, com especial atenção à relação que a Arquivologia possui com a CI e esta com a

CC. A ideia de antagonismo não se explicita, entendendo que a CI pode oferecer à Arquivologia a possibilidade de constituição de práticas científicas, além do aprimoramento e reflexão do campo epistemológico e teórico do fazer e saber arquivísticos.

Assim, nota-se a existência de práticas de coprodução entre a Ciência da Informação e a Ciência da Computação, com eventual colaboração na constituição dos conceitos que as fundamentam. A interface, nesse caso, se consolida a partir da associação em prol do entendimento e aprofundamento de questões comuns e cruciais às áreas.

Ainda que os elementos de áreas como a CI e a CC possam ser utilizados em incursões no meio arquivístico, há a perene necessidade de temáticas, como a Representação e RI, terem maior aprofundamento e integração com a Arquivologia. Pode-se constatar que a Representação da Informação no contexto arquivístico possui estudos que a descrevem e a conceituam. Já a RI apresenta incidência em menor quantidade de publicações que a foquem como elemento central de pesquisas. Sua existência, com maturidade evidente em campos como a CC, não tem encontrado bases para consolidação, crescimento e diálogo mais abrangente nos estudos arquivísticos. Desse modo, ficam evidentes as necessidades de potencializar pesquisas nesses temas, sobretudo na RI.

Assim, a Representação da Informação, a partir dos dados apresentados, apresenta-se como etapa primordial para um melhoramento nas práticas de RI em arquivos. São necessárias pesquisas que aprofundem mais temáticas como, por exemplo, a construção de ontologias, esquemas e conjuntos de metadados para a Arquivologia, explorando a diversidade de gêneros, espécies e tipos documentais existentes. Nesse caminho, indicamos como aporte essencial o direcionamento para o viés da descrição arquivística, que classicamente foca sua atuação numa visão mais estrutural, em detrimento a um direcionamento mais voltado para o conteúdo do documento. Assim, poder-se-á ampliar o campo de investigações em áreas, por exemplo, como a extração automática de textos, *machine learning* (e a *deep learning*), dados abertos, Web Semântica, dentre outras áreas que tem dialogado fortemente dentro da CC e da CI e se fazem necessárias na Arquivologia.

REFERÊNCIAS

ALLENDEZ SULLIVAN, Patricia Mónica; PUENTE, Marcelo de la. **Los estudios de usuarios en los archivos**. Buenos Aires: Consultora de Ciencias de la Información, 2013.

ARAÚJO, Carlos Alberto Ávila. O conceito de informação na Ciência da Informação. **Informação & Sociedade: Estudos**, v. 20, n. 3, 2010.

ARAÚJO, Carlos Alberto Ávila. Ciência da Informação, Biblioteconomia, Arquivologia e Museologia: relações teóricas e institucionais. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 16, n. 31, p. 110-130, 2011.

ARAÚJO, Carlos Alberto Ávila. Correntes teóricas da Arquivologia. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 18, n. 37, p. 61-82, ago. 2013a. ISSN 1518-2924. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2013v18n37p61/25338>>. Acesso em: 14 nov. 2018.

ARAÚJO, Carlos Alberto Ávila. O que é Ciência da Informação? **Informação & Informação**, v. 19, n. 1, p. 01-30, dez. 2013b. ISSN 1981-8920. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/15958>>. Acesso em: 03 mar. 2018.

ARAÚJO, Carlos Alberto Ávila. **Arquivologia, biblioteconomia, museologia e ciência da informação: o diálogo possível**. Brasília: Briquet de Lemos, 2014.

ARQUIVO NACIONAL. **Dicionário brasileiro de terminologia arquivística**. Rio de Janeiro: Arquivo Nacional, 2005.

ASSIS, Juliana de; MOURA, Maria Aparecida. Folksonomia: a linguagem das tags. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 18, n. 36, p. 85-106, abr. 2013. ISSN 1518-2924. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/24015>>. Acesso em: 13 fev. 2019.

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. **Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca**. Bookman Editora, 2013.

BARDIN, Laurence. **Análise de conteúdo**. Tradução de Luis Antero Reto e Augusto Pinheiro. Lisboa: Edições, v. 70, 1977.

BARROS, Thiago Henrique Bragato. A Indexação e a Arquivística: aproximações iniciais no universo teórico da organização e representação do conhecimento. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 21, n. 46, p. 33-44, 2016.

BELKIN, Nicholas J. The cognitive viewpoint in information science. **Journal of information science**, v. 16, n. 1, p. 11-15, 1990.

BELKIN, Nicholas J.; ODDY, Robert N.; BROOKS, Helen M. ASK for information retrieval: Part I. Background and theory. **Journal of documentation**, v. 38, n. 2, p. 61-71, 1982.

BELKIN, Nicholas J.; CROFT, W. Bruce. Information filtering and information retrieval: Two sides of the same coin?. **Communications of the ACM**, v. 35, n. 12, p. 29-38, 1992.

BELLOTTO, Heloísa Liberalli. **Arquivística: objetos, princípios e rumos**. Associação de Arquivistas de São Paulo, 2002.

BELLOTTO, Heloísa Liberalli. **Arquivos permanentes: tratamento documental**. FGV Editora, 2007.

BELLOTTO, Heloísa Liberalli. O sentido dos arquivos. Belo Horizonte. In: **I Ciclo de Palestras da Diretoria de Arquivos Institucionais–DIARQ**, 2014. Disponível em: <https://www.ufmg.br/diarq/anexos/wfd_14012774465385cc06bbb48--fala_bellotto.pdf>. Acesso em 22 abr. de 2017.

BELLOTTO, Heloísa Liberalli. A diplomática como chave da teoria arquivística. **Archeion Online**, [S.l.], v. 3, n. 2, jan. 2016. ISSN 2318-6186. Disponível em: <<http://www.periodicos.ufpb.br/index.php/archeion/article/view/27544/14808>>. Acesso em: 14 nov. 2018.

BRÄSCHER, Marisa; CAFÉ, Ligia. **Organização da informação ou organização do conhecimento?**. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., **Anais...**, São Paulo: ANCIB, 2013. Disponível em: <<http://enancib.ibict.br/index.php/enancib/ixenancib/paper/view/3016>>. Acesso em: 22 abr. de 2018.

BRICKLEY, Dan; GUHA, Ramanathan V.; MCBRIDE, Brian. RDF Schema 1.1. **W3C recommendation**, v. 25, p. 2004-2014, 2014.

BUCKLAND, Michael K. Information as thing. **Journal of the American Society for information science**, v. 42, n. 5, p. 351-360, 1991.

CAMPBELL, Rita. Automation and information retrieval in archives—the broad concepts. The **American Archivist**, v. 30, n. 2, p. 279-286, 1967.

CAPURRO, Rafael. Epistemologia e Ciência da Informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 5., 2003, Belo Horizonte. **Anais eletrônicos...** Disponível em: <http://www.capurro.de/enancib_p.htm>. Acesso em: 04 abr. 2017.

CARDOSO, Olinda Nogueira Paes. Recuperação de Informação. **INFOCOMP**, v. 2, n. 1, p. 33-38, 2004. Disponível em: <<http://www.dcc.ufla.br/infocomp/index.php/INFOCOMP/article/view/46>>. Acesso em: 23 jun. 2017.

CASTELLS, Manuel. A Sociedade em Rede: do Conhecimento à Política. In: CARDOSO, Gustavo; CASTELLS, Manuel. **A sociedade em rede: do conhecimento à ação política**. Imprensa Nacional–Casa da Moeda: Belém-Portugal, 2005.

CHU, Heting. **Information representation and retrieval in the digital age**. Information Today, Inc., 2007.

CODINA, Lluís. Fundamentos de teoría de recuperación de información. In: **Organización y recuperación de la información: documentos de lectura**. Barcelona: Fundació per a la Universitat Oberta de Catalunya, 2002. p. 5-32.

CONSELHO INTERNACIONAL DE ARQUIVOS. **ISAD(G)**: Norma geral internacional de descrição arquivística. 2 ed. Rio de Janeiro: Arquivo Nacional, 2000.

CONSELHO NACIONAL DE ARQUIVO. Câmara Técnica de Documentos Eletrônicos – CTDE. **Glossário: Documentos Arquivísticos Digitais**. Versão 6ª. Rio de Janeiro: CONARQ, 2014.

COUTURE, Carol (Ed.). **Les fonctions de l'archivistique contemporaine**. Quebec: PUQ, 1999.

COUTURE, Carol; ROUSSEAU, Jean-Yves. **Os fundamentos da disciplina arquivística**. Lisboa: Dom Quixote, 1998.

CRUZ MUNDET, José Ramón. **Manual de archivística**. Fundación Germán Sánchez Ruipérez, 1994.

DANIELS, Penny. Cognitive models in information retrieval—an evaluative review. **Journal of documentation**, v. 42, n. 4, p. 272-304, 1986.

DERVIN, Brenda; NILAN, Michael. Information needs and uses. **Annual review of information science and technology**, v. 21, p. 3-33, 1986.

DIAS, Guilherme Ataíde. Revisitando a lei de Mooers : a recuperação de artigos de periódicos científicos eletrônicos. ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMACAO, 5., 2003, Belo Horizonte. **Anais...** . Belo Horizonte: UFMG, 2003.

DURANTI, Luciana. Registros documentais contemporâneos como provas de ação. **Revista Estudos Históricos**, v. 7, n. 13, p. 49-64, 1994.

ESTEBAN NAVARRO, M. A.; GAY MOLÍNS, P. La normalización de la descripción y la recuperación de información en los archivos: vino viejo en odres nuevos. **Boletín de la ANABAD**, Madrid, v. 68, n. 1, 1998.

FERNEDA, Edberto. **Recuperação de Informação**: Análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. 2003, 147 f. Tese (Doutorado em Ciências da Comunicação) – Universidade de São Paulo, São Paulo, 2003.

FERREIRA, Letícia Elaine; ALMEIDA JÚNIOR, Oswaldo Francisco de. A mediação da informação no âmbito da arquivística. **Perspectivas em Ciência da Informação**, v. 18, n. 1, p. 158-167, 2013.

FOROUZAN, Behrouz A.; MOSHARRAF, Firouz. **Fundamentos da ciência da computação**. Cengage Learning, 2012.

FOX, Michael J.; WILKERSON, Peter L. **Introduction to archival organization and description**. Getty Publications, 1998.

FREITAS, Frederico Luiz Gonçalves de. Ontologias e a web semântica. **Jornada de Mini-Cursos em Inteligência Artificial, SBC**, v. 8, 2003.

FROHMANN, Bernd. Revisiting “what is a document?”. **Journal of documentation**, v. 65, n. 2, p. 291-303, 2009.

GAGNON-ARGUIN, Louise. Os arquivos, os arquivistas e a arquivística: Considerações Históricas. In: COUTURE, Carol; ROUSSEAU, Jean-Yves. **Os fundamentos da disciplina arquivística**. Lisboa: Dom Quixote, 1998.

GAMA, Fernando; FERNEDA, Edberto. Mediação da informação nos arquivos permanentes: serviços de referência arquivística no ambiente digital. **Inf. Inf.**, Londrina, v. 15, n. 2, p. 148-169, jul./dez. 2010.

GARCIA, Joana Coeli R. Recuperação da Informação. **DataGramZero**, v. 8, n. 6, 2007. Disponível em: <<http://www.brapci.ufpr.br/brapci/v/a/10284>>. Acesso em: 11 fev. 2018.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. Editora Atlas SA, 2008.

GIMÉNEZ-CHORNET, Vicent. La recuperación de la información de los archivos en línea. In: **BiD: textos universitaris de Biblioteconomia i Documentació**. Universitat de Barcelona, 2011. p. 1-13.

HEDSTROM, Margaret. **Archives, Memory, and Interfaces with the Past**. *Archival Science* 2: 21–43, 2002.

HEREDIA HERRERA, Antonia. Archivistica general. **Teoria y practica**. Sevilla: Diputación de Sevilla, 1991.

INGWERSEN, Peter Emil Rerup. **Information Retrieval Interaction**. Taylor Graham, 1992.

INTERPARES. **Glossary**. 2018. Disponível em: <http://www.interpares.org/ip2/display_file.cfm?doc=ip2_glossary.pdf&CFID=17713477&CFTOKEN=60357453>. Acesso em: 19 dez. 2018.

JARDIM, José Maria; FONSECA, Maria Odila. Estudos de usuários em arquivos: em busca de um estado da arte. **DataGramZero** - Revista de Ciência da Informação, v. 5, n. 5, out. 2004.

KHAPRE, Shailesh; BASHA, MS Saleem. A theoretical paradigm of information retrieval in information science and computer science. **International Journal of Computer Science Issues**, v. 9, n. 2, p. 1694-0814, 2012. Disponível em: <<https://www.ijcsi.org/papers/IJCSI-9-5-2-232-240.pdf>>. Acesso em: 24 mar. 2018.

KHUN, Thomas S. **A estrutura das revoluções científicas**. Tradução Beatriz Vianna Boeira e Nelson Boeira. São Paulo: Ed. 1975.

KOBASHI, Nair Yumiko. Análise documentária e representação da informação. Rio de Janeiro, 1996. p. 5-27. Informare. **Cadernos do Programa de Pós-Graduação em Ciência da Informação**, Rio de Janeiro, v. 2, n. 2, p. 5-27, jul./dez. 1996.

KOBASHI, Nair Yumiko. Fundamentos semânticos e pragmáticos da construção de instrumentos de representação de informação. **DataGramZero-Revista de Ciência da Informação**, v. 8, n. 6, 2007.

KOIVUNEN, Marja-Riitta; MILLER, Eric. W3c semantic web activity. **Semantic Web Kick-Off in Finland**, v. 2, p. 27-44, 2001.

LESK, Michael. **The seven ages of information retrieval**. Occasional paper, International Federation of Library Associations and Institutions - UNIVERSAL DATAFLOW AND TELECOMMUNICATIONS CORE PROGRAMME: Ottawa, 1996.

LARA, Marilda Lopes Ginez de. Linguagens documentárias, instrumentos de mediação e comunicação. **Revista Brasileira de Biblioteconomia e Documentação**, v. 26, n. 1/2, p. 72-80, 1993. Disponível em: <<http://basessibi.c3sl.ufpr.br/brapci/v/a/2164>>. Acesso em: 22 fev. 2018.

LIMA, Júnio César de; CARVALHO, Cedric L de. Uma visão da web semântica. **Relatório Técnico, março**, 2004. Disponível em: <http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-04.pdf>. Acesso em: 18 dez. 2018.

LOPES, Luís Carlos. **A nova arquivística na modernização administrativa**. 2. ed. Brasília: Projeto Editorial, 2009.

LOPEZ, André Porto Ancona. **Como Descrever Documentos De Arquivo: Elaboração De Instrumentos De Pesquisa**. São Paulo: Arquivo Do Estado, Imprensa Oficial, 2002.

LUZ, Charley dos Santos. **Ontologia Digital Arquivística: interoperabilidade e preservação da informação arquivística em sistemas de arquivo e na Web**. 2016. 144 f. Dissertação (Mestrado) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2016.

MACHADO, Helena Corrêa; CAMARGO, Ana Maria de Almeida. **Como implantar arquivos públicos municipais**. São Paulo: Arquivo do Estado, 1999.

MACNEIL, Heather. **SUBJECT ACCESS TO ARCHIVAL " FONDS"**: Balancing Provenance and Pertinence. *Fontes Artis Musicae*, p. 242-258, 1996.

MARCONDES, Carlos Henrique. Representação e economia da informação. **Ciência da informação**, v. 30, n. 1, p. 61-70, 2001.

MARQUES, Angelica Alves da Cunha *et al.* Recuperação da Informação (RI): o foco comum entre a Ciência da Informação (CI) e a Arquivística. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., **Anais...**, São Paulo: ANCIB, 2013. Disponível em: <<http://enancib.ibict.br/index.php/enancib/ixenancib/paper/view/2979/2105>>. Acesso em: 22 dez. 2018.

MARTÍN SUQUÍA, Ramón. **Sistemas de recuperación de la información en los archivos**: un análisis de situación y perspectivas. *Lligall* 17, 2001, p. 47-69.

MEDEIROS, Graziela Martins de; BRÄSCHER, Marisa; VIANNA, William Barbosa. A indexação de assunto em documentos arquivísticos: convergências e divergências em artigos internacionais. **Em Questão**, v. 23, n. 1, 2017.

MEEHAN, Jennifer. Towards an Archival Concept of Evidence. **Archivaria**, n. 61, p. 127-146, 2006. Disponível em: <<https://archivaria.ca/index.php/archivaria/article/view/12538>>. Acesso em: 12 abr. 2018.

MOOERS, Calvin N. Zatocoding applied to mechanical organization of knowledge. **American documentation**, v. 2, n. 1, p. 20-32, 1951.

NAVARRO, Gonzalo; ZIVIANI, Nivio. Documentos: linguagens e propriedades. In: BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. **Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca**. Bookman Editora, 2013.

NOVELLINO, Maria Salet Ferreira. Instrumentos e metodologias de representação da informação. **Informação & Informação**, Londrina, v. 1, n. 2, p. 37-44, 1996. Disponível em:

<<http://www.uel.br/revistas/uel/index.php/informacao/article/view/1603/1358>>. Acesso em: 12 mar. 2018.

NOY, Natasha F. WEB, Semantic. What do we need for ontology integration on the semantic web, position statement. **Proc. of the 1st Semantic Integration Ws**, p. 175-176, 2003.

PAES, Marilena Leite. **Arquivo: teoria e prática**. 3. ed. rev. ampl. – Rio de Janeiro: Editora FGV, 2004.

PAVANI, Ana. A produção científica disponível ao mundo: a tecnologia, a vontade e os acessos. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, n. Especial 1, 2007. Disponível em: <<http://www.redalyc.org/html/147/14720411008/>>. Acesso em: 12 mai. 2018.

PEARCE-MOSES, Richard. **A glossary of archival and records terminology**. Chicago, IL: Society of American Archivists, 2005.

PINTO, Virgínia Bentes. Indexação documentária: uma forma de representação do conhecimento registrado. **Perspectivas em Ciência da Informação**, [S.l.], v. 6, n. 2, nov. 2001. ISSN 19815344. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/423/239>>. Acesso em: 17 set. 2018.

PORTO, Daniela Miguéns. **História e evolução do arquivo: a exemplaridade da Torre do Tombo**. 2013. Tese de Doutorado. Universidade da Beira Interior.

PROCTER, Margaret; COOK, Michael. Manual of archival description. **Hampshire: Gower**, 2000.

RAMALHO, Rogério Aparecido Sá. **Web Semântica: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação**. 2006.

RAMALHO, Rogério Aparecido Sá; VIDOTTI, Silvana Aparecida Borsetti Gregório; FUJITA, Mariângela Spotti Lopes. Web semântica: uma investigação sob o olhar da Ciência da Informação. **DataGramZero**, Rio de Janeiro, v. 8, n. 6, p. A04-1001, 2009.

RIBEIRO, Fernanda. Os arquivos na era pós-custodial: reflexões sobre a mudança que urge operar. **Boletim cultural–Câmara Municipal de Vila Nova de Famalicão**. Vila Nova de Famalicão: Câmara Municipal. p. 0871-3308, 2005.

RIBEIRO, Fernanda. A arquivística como disciplina aplicada no campo da ciência da informação. **Perspectivas em gestão & conhecimento**, v. 1, n. 1, p. 59-73, 2011.

RIBEIRO, Fernanda. O uso da classificação nos arquivos como instrumento de organização, representação e recuperação da informação. In: I Congresso ISKO Espanha e Portugal, 1. **Livro de Atas de Conferência Internacional**, Porto, p.528-539, 2013.

ROBREDO, Jaime. Ciência da informação e web semântica: linhas convergentes ou linhas paralelas?. In: ROBREDO, Jaime; BRÄSCHER, Marisa. (Org.). **Passeios pelo bosque da infomação**: estudos sobre a representação e organização da informação e do conhecimento. Brasília: IBICT, 2010. p. 2-40. Disponível em: <<http://www.ibict.br/publicacoes/eroic.pdf>>. Acesso em: 20 nov. 2018.

ROCHA, Claudia Lacombe C. M.. Captura y Transferencia. In: CRUZ MUNDET, José Ramón *et al.* **Administración de documentos y archivos. Textos fundamentales**. Coordinadora de Asociaciones de Archiveros, 2011.

RODRIGUES, G. M. A representação da informação em arquivística: uma abordagem a partir da perspectiva da norma internacional de descrição arquivística. In: RODRIGUES, G. M.; LOPES, I. L. (Orgs.). **Organização e representação do conhecimento na perspectiva da Ciência da Informação**. Brasília: Thesaurus, 2 , 2003, p.210-230.

RODRÍGUEZ BRAVO, Blanca. **El documento: entre la tradición y la renovación**. Gijón: Ediciones Trea, 2002.

RONDINELLI, Rosely Curi. **Gerenciamento arquivístico de documentos eletrônicos**: uma abordagem teórica da diplomática arquivística contemporânea. 4. Ed. Rio de Janeiro: FGV, 2005, 160 p.

ROZA, Rodrigo Hipólito. Revolução Informacional e os Avanços Tecnológicos da Informática e das Telecomunicações. **Ciência da Informação em Revista**, v. 4, n. 3, p. 03-11, 2017. Disponível em: <<http://www.seer.ufal.br/index.php/cir/article/view/3482>>. Acesso em: 12 mai. 2018.

SANTOS, Silvana Aparecida Silva dos. **Ontologias como domínio conceitual e planos de classificação arquivística**: um estudo de caso sobre instituições federais de ensino superior. 2010, 160 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2010.

SANTOS, Vanderlei Batista dos. **A teoria arquivística a partir de 1898**: em busca da consolidação, da reafirmação e da atualização de seus fundamentos. 2011. 279 f. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2011.

SARACEVIC, Tefko. Information science. In: Marcia J. Bates and Mary Niles Maack (Eds.) **Encyclopedia of Library and Information Science**. New York: Taylor & Francis. pp. 2570-2586, 2009.

SCHELLENBERG, Theodore Roosevelt. **Arquivos modernos**: princípios e técnicas. Tradução de Nilza Teixeira Soares. 6. ed., Rio de Janeiro: Editora FGV, 2006.

SILVA, Dacles Vágner da *et al.* Entre o objeto e o sujeito: composição, diálogo e confluência na RI em arquivos. **Informação em Pauta**, Fortaleza, v. 2, n. 2, p. 92-109, dec. 2017. ISSN 2525-3468. Disponível em: <<http://periodicos.ufc.br/informacaoempauta/article/view/20438>>. Acesso em: 23 abr. 2018.

SOARES, Ana Paula Alves; PINTO, Adilson Luiz; SILVA, Armando Malheiro da. O Paradigma Pós-Custodial Na Arquivística. **Páginas a&b: arquivos e bibliotecas**, p. 22-39, 2016.

SOUSA, Renato Tarciso Barbosa de. A classificação como função matricial do que-fazer arquivístico. In: SANTOS, Vanderlei Batista dos; INNARELLI, Humberto Celeste; SOUSA, Renato Tarciso Barbosa de (Org). **Arquivística: temas contemporâneos**: classificação, preservação digital, gestão do conhecimento. Distrito Federal: SENAC, 2009.

SOUSA, Renato Tarciso Barbosa de; ARAÚJO JÚNIOR, Rogério Henrique de. A classificação e o vocabulário controlado como instrumentos efetivos para a RI arquivística In: **Da produção à preservação informacional**: desafios e oportunidades [en ligne]. Évora : Publicações do Cidehus, 2017. Disponível em: <<http://books.openedition.org/cidehus/2790>>. Acesso em: 22 de Set. 2018.

SOUZA, Renato Rocha; ALVARENGA, Lídia. A Web Semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, v. 33, n. 1, 2004.

TOGNOLI, Natália Bolfarini. A informação no contexto arquivístico: uma discussão a partir dos conceitos de informação-como-coisa e informação orgânica. **Informação Arquivística**, p. 113-122, 2012.

UNISIST. **Indexing principles**. Paris. UNESCO. 1976, 11 p.

VALERO, Carlos Navarrete. Satisfacción de usuarios internos y externos de todos los servicios de una clínica de medicina familiar del sector salud en el estado de México. **CES Salud Pública**, v. 6, n. 2, p. 138-145, 2016.

VIERA, Angel Freddy Godoy; GARRIDO, Isadora dos Santos. Folksonomia como uma estratégia para recuperação colaborativa da informação. **DataGramZero**, v. 12, n. 2, 2011.

VITAL, Luciane Paula; BRASCHER, Marisa. Descrição Arquivística: uma discussão conceitual. **Informação & Informação**, v. 21, n. 1, p. 213-229, 2016. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/22497>>. Acesso em: 23 fev. 2018.

WILSON, T. D. On user studies and information needs. **Journal of Documentation**, 37(1):3-15, 1981.

WITTHUS, Rutherford W.. The Archival Information System as a Model for Retrieval of Interdisciplinary Materials. **Issues in Interdisciplinary Studies**, 1992.

YAKEL, Elizabeth. Archival representation. **Archival Science**, v. 3, n. 1, p. 1-25, 2003.

YEO, Geoffrey. Concepts of record (2): prototypes and boundary objects. **The American Archivist**, v. 71, n. 1, p. 118-143, 2008. Disponível em: <<http://americanarchivist.org/doi/abs/10.17723/aarc.71.1.p0675v40tr14q6w2?code=same-site>>. Acesso em: 2 abr. 2018.

ZHANG, Jane. Archival representation in the digital age. **Journal of Archival Organization**, v. 10, n. 1, p. 45-68, 2012.

ZHANG, Jane; MAUNEY, Dayne. When archival description meets digital object metadata: a typological study of digital archival representation. **The American Archivist**, v. 76, n. 1, p. 174-195, 2013.