



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

**MORGANA RAMOS ALBUQUERQUE**

**A CURADORIA EM UM REPOSITÓRIO INSTITUCIONAL: UMA ANÁLISE SOB A  
ÓTICA DO CICLO DE VIDA DOS DADOS DE SANT'ANA**

**FORTALEZA-CE  
2018**

MORGANA RAMOS ALBUQUERQUE

A CURADORIA EM UM REPOSITÓRIO INSTITUCIONAL: UMA ANÁLISE SOB A  
ÓTICA DO CICLO DE VIDA DOS DADOS DE SANT'ANA

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

Linha de Pesquisa: Representação da Informação e do Conhecimento e Tecnologia.

Orientadora: Prof<sup>a</sup>. Dra. Andréa Soares Rocha da Silva

Coorientação: Prof<sup>a</sup>. Dra. Virgínia Bentes Pinto

FORTALEZA-CE  
2018

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)

---

A311c Albuquerque, Morgana Ramos.

A curadoria em um repositório institucional : uma análise sob a ótica do ciclo de vida dos dados de Sant'Ana / Morgana Ramos Albuquerque. – 2018.

85 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Humanidades, Programa de Pós-Graduação em Ciência da Informação, Fortaleza, 2018.

Orientação: Profa. Dra. Andréa Soares Rocha da Silva.

Coorientação: Profa. Dra. Virgínia Bentes Pinto.

1. Curadoria Digital. 2. Ciclo de Vida. 3. Gestão e Preservação de Dados Digitais. 4. Preservação Digital. 5. Repositórios Digitais. I. Título.

CDD 020

---

MORGANA RAMOS ALBUQUERQUE

A CURADORIA EM UM REPOSITÓRIO INSTITUCIONAL: UMA ANÁLISE SOB A  
ÓTICA DO CICLO DE VIDA DOS DADOS DE SANT'ANA

Dissertação apresentada ao Mestrado Acadêmico em Ciência da Informação do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Ceará, como requisito para a obtenção do título de mestre em Ciência da Informação.

Linha de Pesquisa: Representação da Informação e do Conhecimento e Tecnologia

Aprovada em: 20/09/2018.

BANCA EXAMINADORA

---

Prof<sup>a</sup>. Dra. Andréa Soares Rocha da Silva (Orientadora)  
Universidade Federal do Ceará (UFC)

---

Prof<sup>a</sup>. Dra. Thiciane Mary Carvalho Teixeira  
Universidade Federal do Ceará (UFC) / Universidade Estadual do Ceará (UECE)

---

Prof<sup>a</sup>. Dra. Sandra de Albuquerque Siebra  
Universidade Federal da Paraíba (UFPE)

FORTALEZA -CE  
2018

Ao Alano Regis, meu companheiro, melhor amigo e parceiro na trajetória da vida.  
Dedico.

## **AGRADECIMENTOS**

Agradeço à e a instituição participante desta pesquisa que abriu suas portas tornando possível a construção deste trabalho, colocando à disposição uma equipe que dedicou gentilmente seu tempo.

Sou grata a todos os colegas de Mestrado, meus queridos parceiros. Estivemos juntos nessa caminhada árdua do qual muitos, como eu, acumularam ainda o cansaço do trabalho, mesmo com todas as dificuldades vencemos. Em especial agradeço à Natanna Santana pela solidariedade em me auxiliar nas atividades, aprendi muito com você Natanna. Ao Igor Girão, meu novo amigo, que me ensinou sobre superação e gratidão, tenho muito carinho por você Igor. À Joana Darc, pessoa linda, inteligente e dotada de uma força que me inspira.

Aos colegas e minha equipe de trabalho (Gildênia, Natanael, Rodrigo, Valdeci, Aires, Luiz e Pâmela), agradeço por me escutar e compreender durante esses últimos anos. Também agradeço à Mírian Cris e Gabriela Alves pela valiosa amizade que me deu força e acreditando em mim quando eu já não estava exausta. Agradeço ainda à minha gestora Leonilha Lessa pelo apoio e incentivo.

Agradeço ao Alano Regis, meu marido, melhor amigo e companheiro de vida. Seu amor, paciência e apoio foram essenciais nesses dias turbulentos. Você é meu presente de Deus. Aos meus familiares que sempre torcem por mim em todas as conquistas, Minha mãe Mônica, meus irmãos Gabriel e Jordana, minha avó Rosa, minha tia Regina e meu querido primo Matheus. À minha amiga Camila Regina, por me entender e acreditar que eu sempre posso ser melhor.

Agradeço a todos os professores do PPGCI pelo lindo trabalho para formação de futuros docentes. À Coordenadora, prof. Lídia Eugênia, pela seriedade e comprometimento, e também à Veruska pela amizade, disponibilidade e auxílio em todos os momentos.

Agradeço às professoras Sandra Siebra e Thiciane Carvalho, pela disposição em fazer parte desta etapa. À minha coorientadora Virgínia Bentes, pessoa da qual já faz parte da minha vida desde a graduação sempre me inspirando com sua energia e inteligência. Finalmente, agradeço a minha orientadora Andréa Soares por sua sagacidade, gentileza e serenidade em conduzir esse processo de uma forma tão leve e séria. Tive muita sorte em tê-la junto a mim nesse desafio.

Do avesso desta noite que me encobre,  
Preta como a cova, do começo ao fim,  
Eu agradeço a quaisquer deuses que  
existam, pela minha alma inconquistável.

Na garra cruel desta circunstância,  
Não estremei, nem gritei em voz alta.  
Sob a pancada do acaso,  
Minha cabeça está ensanguentada,  
mas não curvada.

Além deste lugar de ira e lágrimas  
Avulta apenas o horror das sombras.  
E apesar da ameaça dos anos,  
Encontra-me e me encontrará destemido.

Não importa quão estreito o portal,  
Quão carregada de punições a lista,  
Sou o mestre do meu destino,  
Sou o capitão da minha alma.

(Invictus, William Ernest Henley)

## RESUMO

O atual crescimento exponencial na produção de dados é resultado do desenvolvimento e uso de Tecnologias digitais de Informação e de Comunicação. Neste cenário, o fenômeno conhecido como *Big Data*, promove a produção de objetos digitais em uma escala nunca antes vista, trazendo à tona o desafio de manter e preservá-los. Diante disto, a curadoria digital propõe ações para cuidar e salvaguardar dados em sistemas automatizados oferecendo condições para seu reuso e acesso futuro. Em um conceito ressignificado, a curadoria com a predicação digital atua por meio de um ciclo de vida com o objetivo de garantir que as ações necessárias à gestão e preservação sejam devidamente implementadas. Dentre os ambientes digitais, o repositório é apresentado como um sistema de informação com estrutura capaz de armazenar e disseminar os dados de forma eficiente. Diante disso, o estudo se propôs a identificar e avaliar o processo de curadoria digital de documentos acadêmicos produzidos por uma Instituição de Ensino Superior, localizada em Fortaleza, sob a ótica do ciclo de vida dos dados (CVD-CI) de Sant'Ana. A metodologia empregada, pauta-se na abordagem qualitativa, tendo sido feito um estudo de caso da gestão e preservação de documentos digitais produzidos na instituição pesquisada. Os instrumentos utilizados na coleta de dados foram entrevistas, análise documental e observação. As entrevistas foram aplicadas em um grupo focal com bibliotecários realizadas em duas fases, a primeira sobre o ambiente digital e segunda sobre os objetos digitais. De forma separada, também foi realizada uma entrevista com o analista de tecnologia para verificação da infraestrutura tecnológica e o processo de implantação do DSpace na Instituição. Os encontros foram registrados em áudio e com anotações complementares em um caderno de pesquisa. Os *corpora* da pesquisa foram: Trabalhos de Conclusão de Cursos da Graduação, Dissertações e Teses, e como ambiente, o sistema de gestão da biblioteca e o DSpace. A análise dos dados apresentou as categorias de coleta, armazenamento, recuperação e descarte, orientadas pelo ciclo de vida de Sant'Ana, para identificar quais foram ações realizadas pela instituição e se necessitavam de melhorias. Os resultados foram embasados pela literatura e permitiram a elaboração de sugestões como, por exemplo, a necessidade de utilizar um mecanismo de coleta automatizado. Como conclusão, apresentaram-se algumas estratégias de soluções para a melhoria da curadoria dos documentos acadêmicos digitais na instituição, dentre elas a mudança do fluxo de catalogação da instituição; utilização de um software para apoio na Preservação Digital e a criação de uma política de conteúdos digitais.

**Palavras-chave:** Curadoria Digital. Ciclo de Vida. Gestão e Preservação de Dados Digitais. Preservação Digital. Repositórios Digitais.

## **ABSTRACT**

The current exponential growth in data production is result of the development and use of Information and Communication Technologies. The phenomenon, known as Big Data, promotes a production of data in a scale never seen before brings to surface the challenge to maintain and preserve these digital objects. In this context the digital curatorship proposes actions to take care of and protect data in digital systems offering conditions for its use and future access. In this regard the curatorship acts upon the steps of a life cycle with the objective of guaranteeing that the necessary actions for management and preservation to be properly implemented. Within the digital environments, the warehouse is presented as an information system with framework capable of storing and disseminating the data in an efficient way. In view of that, the study achieved to identify the process of digital curatorship of academic documents produced by a Higher Education Institute from the standpoint of data lifespan of Sant'Ana(CVD-CI). Through the specific objectives, it was possible to verify the management (technology, processes and team) of academic documents of the chosen institution; verify the framework and planning of digital environments and analyse actions of the curatorship under the perspective of the lifespan steps chosen. The methodology used, of qualitative approach, consisted in a study case of management and preservation of digital documents produced in a Higher Education Institute located in Fortaleza, Ceará. The analysis' focus were the data and digital environment of the institution, in this way, as derived data were verified the steps of the Sant'Ana cycle regarding the final graduation works, dissertations and thesis, and the library management system and DSpace as environment. The data analyses presented the gathering, storage, recovery and disposal, led by the Sant'Ana lifespan, to identify that actions were taken by the institution or needed improvement. The results were based by the literature and presented suggestions for the management and preservation of the documents, as the necessity of an automatic collection for exemple. As conclusion the results obtaind presented solutions to improve the curatorship of digital academic documents in the institution, among them the creation of a digital contents policy.

**Keywords:** Digital Curation. Information Lifecycle. Management and Preservation of Digital Data. Digital Preservation. Digital Repository.

## LISTA DE FIGURAS

Figura 1 – Ciclo de Vida do DCC.....	24
Figura 2 – Processo de Curadoria Digital do DCU.....	25
Figura 3 – Ciclo de Vida estendido CDD&DCU.....	27
Figura 4 – Ciclo de Vida do UK Data Archive.....	28
Figura 5 - Ciclo de Vida do DataOne.....	30
Figura 6 - Ciclo de Vida dos dados (CVD-CI).....	31
Figura 7 – Interface de apresentação de um protocolo no Repository of Science Methods.....	40
Figura 8 – Interface de apresentação de um resumo em anais de congresso no Alice.....	41
Figura 9 – Softwares para Repositórios.....	44
Figura 10 – Processamento de teses e dissertações com o TEDE2.....	57
Figura 11 – Esquema de metadados Dublin Core.....	57
Figura 12 – Comunidades e coleções acadêmicas criadas pela universidade no DSpace.....	58

## LISTA DE QUADROS

Quadro 1 – Modelos de ciclo de vida comprado.....	33
Quadro 2 – Textos referenciais nas temáticas abordadas.....	49
Quadro 3 – Estimativa inicial da produção anual de conteúdo da universidade.....	55

## LISTA DE ABREVIATURAS E SIGLAS

ALICE	Acesso Livre a Informação Científica da Embrapa
BDTD	Biblioteca Digital de Teses e Dissertações
CVD-CI	Ciclo de Vida dos Dados
DCC	Digital Curation Center
DCU	Digital Curation Unit
DataOne	Rede de Observação de Dados para a Terra
OAIS	Open Archival Information System

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>12</b>
<b>2 A CURADORIA NO AMBIENTE DIGITAL.....</b>	<b>17</b>
2.1 A gênese da Curadoria Digital.....	17
2.2 Abordagens de uma Curadoria em meio digital.....	20
2.3 Ciclo de vida da Curadoria.....	21
2.3.1 O Ciclo de Vida do Digital Curation Center (DCC).....	23
2.3.2 O Ciclo de Vida Digital Curation Unit (DCU).....	25
2.3.3 O Ciclo de Vida do UK Data Archive.....	28
2.3.4 O Ciclo de Vida do DataOne.....	29
2.3.5 O Ciclo de Vida dos Dados (CVD-CI).....	31
2.3.6 Síntese dos Modelos.....	32
2.4 Objetos de curadoria: dados brutos e derivados.....	34
<b>3 AMBIENTES DIGITAIS DE CURADORIA.....</b>	<b>36</b>
3.1 Repositórios Digitais.....	37
3.1.1 Planejamento.....	42
3.1.2 Implantação.....	44
3.1.3 Funcionamento.....	45
<b>4 METODOLOGIA.....</b>	<b>47</b>
4.1 Locus do Estudo de Caso.....	49
4.2 Coleta de dados.....	50
4.3 Análise de dados.....	52
<b>5 DESCRIÇÃO DO CASO.....</b>	<b>53</b>
<b>6 ANÁLISE DE DADOS E DISCUSSÃO DOS RESULTADOS.....</b>	<b>60</b>
6.1 Coleta.....	61
6.2 Armazenamento.....	65
6.3 Recuperação.....	67
6.4 Descarte.....	69
<b>7 CONCLUSÃO.....</b>	<b>71</b>
<b>REFERÊNCIAS.....</b>	<b>75</b>
<b>APÊNDICE.....</b>	<b>81</b>

## 1 INTRODUÇÃO

A informação é fator determinante para construção da sociedade e sua sobrevivência, capaz de modificar a consciência e o comportamento humano. Com o passar dos anos linguagens e suportes foram criados para transmissão de informação, mas apenas a partir da imprensa de Gutemberg foi possível a democratização real do conhecimento. Nesse ponto podemos traçar um contexto entre escassez informacional e abundância.

A era da escassez é caracterizada por todo período que antecede a invenção da imprensa, momento em que a aptidão literária e aprendizagem estavam a serviço de elites dominantes (Estado e Igreja). O conhecimento era copiado a mão e armazenados em palácios imperiais ou monastérios tornando a difusão dessas informações lenta, controlada, mediada por privilégios e posições sociais (GANDEL; KATZ; METROS, 2004).

Séculos após a imprensa, nos deparamos com outra problemática. A chamada Era da informação que trouxe benefícios e desafios, revela seu excesso um fator inquestionável, chegando ao ponto que quantificá-lo não parece ser possível. Esse efeito foi provocado pelo acesso em rede à conteúdos digitais, o que rapidamente permitiu a ampliação do processamento de *bytes* e *bits* dando origem ao *Big Data*, temática que ilustra o contexto informativo atual.

O *Big data* é um fenômeno em que o armazenamento e processamento de dados tornaram-se grandes demais para uma infraestrutura única. Seus fatores de sustentação, demonstram novas características de informação com a grande produção de dados em massa, a diversidade de recursos, a rapidez na transmissão de dados e a promoção de qualidade dos dados para reuso (SILVA, DORNELAS, 2017).

Essa abundância informacional é ressaltada por Constantopoulos et al (2009, p.2. Tradução nossa) o autor revela como sendo “a adaptação das tecnologias [...] que levou a um volume crescente de objetos digitais ou nascidos digitalizados que não só ajudam a dinamizar as operações e serviços diários, mas, também, a experiência e identidade sociocultural”.

O crescimento exponencial da produção de dados pode ser observado na previsão de Gantz e Reinsel (2013), ao afirmarem que o universo digital ampliará de

130 exabytes para 40.000 exabytes entre 2005 a 2020, ou seja, nessa estimativa, dados digitais duplicarão a cada dois anos equivalendo a 40 trilhões de gigabytes. A proporção de dados que necessitam de proteção está crescendo mais do que universo digital em si, e vai de menos de um terço em 2010 para mais de 40% em 2020.

Aspectos como a 'abundância da informação' são abordados no texto "A fadiga da carne", autoria de Gandel, Katz e Metros (2004). Para os autores a atual e prospectiva Era da abundância informacional desafiará muitas suposições básicas e práticas sobre a salvaguarda, proteção, filtragem, preservação, avaliação, purificação, descrição, catalogação e tratamento da informação para propósitos educacionais, de aprendizagem e acadêmicos.

Importante ressaltar que o contexto de abundância, também remetida à avalanche informacional ou explosão informacional, traz a ilusão de que dados em rede permanecerão acessíveis a todo momento. Porém quando verifica-se a fragilidade dos objetos digitais percebe-se o quanto este pensamento é equivocado.

Informações produzidas sem um processo adequado de criação, registro e disseminação correm o risco de perda irrecuperável. Para solucionar isso é imprescindível iniciativas capazes de cuidar das fragilidades e excessos, tornando possível o acesso, a integridade e a fidedignidade da informação. É, pois, nesse contexto que desponta a Curadoria Digital.

Surgindo relacionado à gestão de dados científicos, a Curadoria de Digital emerge como uma área de pesquisa interdisciplinar que abrange atividades em diversas profissões, instituições, atores e setores. Passou a ser trabalhada na Ciência da Informação e na Ciência da Computação, impulsionada pelo crescimento da produção de informação digital (LEE; TIBBO, 2011; SIEBRA et al., 2017)

Corroborando com essa ideia, Sayão e Sales (2012, p.189) ressaltam que a Curadoria surge como uma área de práticas e de pesquisa, dialogando com várias disciplinas e diversos gêneros de profissionais. Isto "[...] porque, como se trata de uma área que só recentemente despontou como crítica para a pesquisa, ainda restam muitas lacunas práticas e teóricas a serem equacionadas, orientadas, preferencialmente, por uma abordagem multidisciplinar".

Iniciativas foram criadas buscando reunir equipes multidisciplinares em esforços de cuidado e preservação de objetos digitais. Desta forma, o *Digital Curator Center* (DCC) destacou-se como renomado centro de estudo em Curadoria Digital

sendo referência sobre o assunto. Para o DCC, a curadoria digital envolve a manutenção, preservação e agregação de valor durante o ciclo de vida.

Giaretta (2004) ratifica o contexto do DCC ao falar que a curadoria digital: cuida e agrega valor aos dados digitais. Isso provavelmente implica a criação de novos dados a partir dos existentes, a fim de torná-los úteis e adequados à necessidade. Através da gestão dos dados seria possível diminuir possíveis ameaças ao valor da pesquisa (com ações para acesso e uso) e minimizar o risco de obsolescência (com ações de preservação). Materiais devidamente preservados reduzem a necessidade de retrabalho, facilitando sua reutilização e disponibilidade para novas pesquisas.

Em síntese, entende-se que a Curadoria Digital assegura a sustentabilidade da documentação digital e, por consequência, de informações para o futuro. Ao momento que atribui valor, há o cuidado e preservação. Os recursos estratégicos, metodológicos e as tecnologias envolvidas facilitam o acesso a dados digitais por meio da melhoria da qualidade desses dados, do seu contexto de pesquisa e da checagem de autenticidade (SANCHEZ; VIDOTTI; VECHIATO, 2017).

A fim de resguardar as etapas necessárias ao planejamento, uso e preservação, os ciclos de vida são estabelecidos para o devido cuidado dos dados digitais. Os modelos são elaborados para adaptar conceitos de curadoria digital às necessidades das instituições e envolvem diversas etapas do fluxo de informação.

As ações são implementadas em ambientes digitais que necessitam da preservação e gestão de informações. Dentre os ambientes, o Repositório é um sistema informacional que reúne e disponibiliza informações, com o objetivo promover a visibilidade. Considerando o contexto da problemática sobre o excesso de informação, o surgimento da Curadoria Digital e a necessidade de gestão de informações em repositórios, fazemos o seguinte questionamento:

**Como se dá o processo de curadoria digital em um repositório institucional?**

O **objetivo geral** desta pesquisa é identificar e avaliar o processo de curadoria digital de documentos acadêmicos produzidos por uma Instituição de Ensino Superior localizada na cidade de Fortaleza, Ceará, sob a ótica do ciclo de

vida dos dados (CVD-CI) de Sant'Ana. Desdobram-se deste, os seguintes objetivos específicos:

- 1) Verificar a gestão (tecnologia, processos e equipe) de documentos acadêmicos digitais de uma instituição de ensino superior;
- 2) Averiguar a estrutura e o planejamento dos ambientes digitais da instituição de ensino superior para armazenamento dos documentos acadêmicos digitais;
- 3) Analisar ações de curadoria sob a perspectiva das etapas do ciclo de vida de dados (CVD-CI) de Sant'Ana;
- 4) Definir ações de melhoria para a coleta, armazenamento, recuperação e descarte dos documentos acadêmicos digitais.

A importância deste estudo parte da necessidade de identificar e definir ações que promovam o cuidado, a preservação e a disseminação de materiais digitais, resultando em uma prática de curadoria reproduzível para instituições brasileiras.

Para tanto, a pesquisa é desenvolvida em sete capítulos. Este capítulo introdutório apresenta o contexto do estudo, problematizando-o, fazendo o seu recorte e definindo seu objeto de estudo. Ademais, apresenta os objetivos a serem alcançados e expõe a estrutura lógica da dissertação.

No segundo capítulo é apresentada a evolução conceitual da Curadoria Digital incluindo suas relações com a Ciência da Informação. Também são apresentados modelos de ciclo de vida a fim de elucidar suas etapas e o modelo que será a base de avaliação deste estudo.

O terceiro capítulo aborda os repositórios digitais, suas tipologias, características, planejamento, infraestrutura, tecnologias bem como as etapas necessárias na concepção de repositórios.

No capítulo quarto apresentam-se a descrição de fatos, decisões, informações e detalhes sobre os ambientes digitais de gerenciamento de informações da instituição escolhida para aplicação do estudo.

O quinto capítulo dispõe sobre o percurso metodológico que envolve a avaliação do repositório, evidenciando o método de estudo de caso como norteador deste estudo, além das fontes consultadas, instrumentos de coleta e técnica de análise de dados.

A apresentação e discussão dos dados de acordo com as categorias de coleta, armazenamento, recuperação e descarte do Ciclo de Vida dos Dados (CVD-CI) de Sant'Ana, encontram-se expressados no capítulo sexto. Por fim, o sétimo e

último capítulo sintetiza os resultados encontrados e as considerações sobre o estudo.

As contribuições esperadas é a verificação de forma prática das etapas do ciclo de vida para identificação de pontos de melhoria ao repositório estudado, esta ação propõe evidenciar como pode ser realizada a gestão e preservação de objetos em ambientes digitais.

Desta forma, a Curadoria Digital contribui para a Ciência da informação na criação de ferramentas para o processo de recuperação, representação e disseminação da informação durante o ciclo de vida do objeto digital. Para sociedade, a Curadoria Digital contribui com a manutenção e preservação da memória diante da fragilidade e a obsolescência dos objetos digitais.

## 2 A CURADORIA NO AMBIENTE DIGITAL

Diante do contexto de informação digital, as novas formas de interação e comunicação acompanham a realidade da intensa produção de dados e informações. Essa abundância informacional, tema evocado pela Ciência da informação, consiste em uma problemática contemporânea que causa impactos e traz a necessidade de criar estratégias para selecionar, manter/preservar e organizar informações produzidas.

Este fenômeno pode ser considerado como reflexo da exacerbação de conteúdos publicados na internet causado uma avalanche informativa gerada pela multiplicação de sites, aplicativos e softwares voltados para publicação desses conteúdos.

No atual contexto, em que os desafios da preservação digital não se restringem aos conteúdos digitalizados, mas, de igual forma, ao crescente volume de informação nato-digital (crescente em volume, complexidade e importância), os arquivos e as bibliotecas têm elaborado ações face às ameaças da obsolescência tecnológica e da vulnerabilidade digital [...] (FREITAS, 2017, p.28).

Frente a esse contexto surge uma nova aplicação e definição para a Curadoria, conceito que tradicionalmente sempre foi aplicado a museus e coleções de arte. De acordo com Longair (2015) essência da Curadoria remete ao termo *curare* que significa “cuidado para” expressando a função de custódia e preservação, sendo usando atualmente atrelado a uma gama de atividades no ambiente digital.

Na atualidade, percebemos que a definição do termo Curadoria tornou-se complexa, ou talvez banalizada, passando a ser utilizada em contextos cada vez mais diversos em virtude da sua interdisciplinaridade. Para Siebra, Borba e Miranda (2016) o termo, que trata de ações de cuidado de objetos digitais, ganhou abordagens diferentes: curadoria de dados, curadoria de conteúdo e curadoria digital, causando certa confusão no entendimento de seus limites e relações.

### 2.1 A gênese da Curadoria Digital

Segundo Beagrie (2006), apesar das iniciativas terem se iniciados no por volta dos anos 80 e 90, o termo Curadoria Digital foi cunhado pela primeira vez em

2001, em um Seminário organizado em Londres pela Digital Preservation Coalition e o Centro Espacial Nacional Britânico, visando aumentar a adesão a iniciativa promovida pelo Open Archival Information System Refer-modelo (OAIS) no Reino Unido e compartilhar experiências de bibliotecas digitais, arquivos e da e-science.

Porém apenas a partir de 2003, passou-se a se dar mais atenção à ideia de curadoria de dados dentro do contexto da *e-science* e da infraestrutura cibernética (LEE; TIBBO, 2011). Na realidade, na mesma época o termo curadoria passou a ser trabalhado nas áreas de Ciência da Informação e Ciência da Computação, motivados pelo crescimento exponencial da informação digital, dando origem a noção de Curadoria Digital hoje difundida.

Conforme Beagrie (2006, p. 4 Tradução nossa) o termo foi cada vez mais utilizado para designar estratégias para salvaguardar para as gerações futuras os dados de pesquisas e outros materiais digitais em todo o ciclo de vida. O que inclui não só a preservação digital, mas também segundo autor “todos os processos necessários para uma boa criação e gestão de dados, e a capacidade de agregar valor aos dados para gerar novas fontes de informação e do conhecimento”.

Atualmente grande parte da produção de informações é criada e disseminada através ferramentas digitais. Conforme Siebra, Borba e Miranda (2016, p. 2) “Grande parte da base de conhecimentos, dados e ativos intelectuais de instituições, organizações e indivíduos estão, hoje, cada vez mais em formato digital”.

Ainda segundo as autoras, tais dados fazem parte de diversos tipos de documentos que estão presente no nosso cotidiano, vão desde documentos administrativos até prontuários médicos, objetos de aprendizagem, transações comerciais, bancárias, e-mails, músicas, fotos, vídeos, filmes, notícias, postagens em redes sociais, sites, entre outros.

Na perspectiva de Sayão e Sales (2012, p.189) a Curadoria Digital surge como uma nova área de práticas e de pesquisa, dialogando com várias disciplinas e diversos gêneros de profissionais. “Isto porque, como se trata de uma área que só recentemente despontou como crítica para a pesquisa, ainda restam muitas lacunas práticas e teóricas a serem equacionadas, orientadas, preferencialmente, por uma abordagem multidisciplinar”. Podendo ser considerada como um campo de pesquisa e prática interdisciplinar que reflete uma abordagem holística.

Considerando a formação do Curador, o conhecimento e as aptidões necessárias para os profissionais envolvidos na área deverão ser dinâmicas,

complexas e interdisciplinares. As habilidades, portanto, deverão integrar conhecimentos em Ciência da Computação, Ciência da Informação, Biblioteconomia, Arquivística além de outras disciplinas e domínios relacionados à criação e utilização de dados (NATIONAL RESEARCH COUNCIL, 2015).

Sobre os conjuntos de habilidades similares à curadoria que são usados no trabalho tradicional da biblioteca, os autores Kim, Warga e Moen (2013, p. 68, tradução nossa) afirmam que:

De fato, a curadoria, envolve várias atividades que facilitam a descoberta, o acesso, a disseminação e o arquivamento de informações, coisa que bibliotecários e arquivistas fizeram por centenas de anos. Isso implica que os conjuntos de habilidades similares usados no trabalho de biblioteca tradicional podem ser benéficos para o trabalho de curadoria envolvendo dados e informações digitais.

Os autores acrescentam que competências funcionais exigidas ao Curador envolvem a: criação e edição de diversos tipos de objetos digitais, digitalização, criação de metadados, gestão de repositórios. Além disso conhecimentos ao nível de *workflow* de curadoria, princípios de gestão de dados, arquitetura de repositórios, tecnologia de pesquisa e recuperação, estratégias de preservação, aplicação dos direitos de autor, políticas de investigação e procedimentos relacionados com gestão de dados.

Tendo em vista a relação da temática com outras áreas, Machado et al (2016) faz uma associação da Curadoria com a Gestão da Informação visto que possam ser grandes aliadas no processo de gestão da informação, pois utilizam técnicas para tratamento da informação objetivando seu uso adequado.

A curadoria no ambiente digital representa um importante conceito na teoria e gestão da informação, devido a sua aplicabilidade frente a uma gama de problemas e domínios advindos da necessidade de preservação de ativos digitais, considerando acervos do patrimônio cultural, *e-science*, mídias sociais e a gestão dos registros organizacionais (SAYÃO; SALES, 2012).

Podemos entender que a Curadoria digital reúne habilidades e práticas de várias disciplinas sendo resultado dessa união o seu conjunto de estratégias, abordagens e atividades tecnológicas para gestão de informações digitais.

## 2.2 Abordagens de uma curadoria em meio digital

Entende-se que a abordagem da Curadoria no âmbito digital corresponde ao contexto de objeto, podendo estes serem dados, documentos, bases e etc, considerando também a tipologia, formato, uso e aplicações de cada objeto. Para melhor entendimento ressaltamos neste trabalho o diferencial entre os conceitos de Curadoria de Conteúdo, Curadoria de Dados e Curadoria Digital.

A **Curadoria de Conteúdo**, também conhecida como curadoria de informação, está reacionada a área da Comunicação e uso da informação neste contexto. Por trás deste conceito está o consumo consciente da informação e a geração de materiais frente ao excesso de conteúdos na Web.

Para Santos (2014) a curadoria de conteúdo consiste em coletar, filtrar e classificar informações para um grupo de pessoas, considerando três etapas: A primeira denominada de Pesquisa busca identificar e acompanhar as melhores fontes e geradores de conteúdo; a segunda, conhecida como Contextualização tem o propósito de dar sentido ao conteúdo de acordo com o perfil do público e objetivo da curadoria (adaptando a linguagem; reunindo conteúdos e novas interpretações); a terceira corresponde ao compartilhamento que consiste em oferecer conteúdo de valor para o seu público, neste ponto a curadoria de conteúdo, tem como fator determinante a atuação dos curadores na orientação de fontes confiáveis.

A **Curadoria de Dados** tem como propósito cuidar especificamente dos dados de pesquisa, fomentando a *e-science*. Visto que o uso de diferentes conjuntos de dados cria possibilidades de criação de novos conhecimentos, através deste processo propõe-se a melhoria da qualidade desses dados, do seu contexto de pesquisa e da checagem de autenticidade.

De acordo com Sayão e Sales (2012) dados de pesquisa requerem cuidados específicos em virtude da sua fragilidade de formato e registro. Desta forma, a Curadoria de Dados contribui para assegurar aos dados de pesquisa validade, significando que eles podem ser usados no futuro como evidência.

A **Curadoria Digital** corresponde ao tratamento de objetos digitais em um contexto relacionado ao documento arquivístico e institucional, como é o caso dos objetos depositados em repositórios. Neste caso, enfatizamos que os documentos são dados derivados e portanto possuem formato específico necessitando de tratamento que considere essa perspectiva material /conceitual.

Para Santos (2014), a Curadoria Digital é o processo de estabelecimento e manutenção de um corpo confiável de informação digital dentro de repositórios para uso corrente e futuro por pesquisadores, cientistas, historiadores e acadêmicos em geral. Para a autora suas etapas devem envolver a seleção, preservação, manutenção, coleção e arquivamento de ativos digitais.

Ainda conforme Santos (2014) apesar de diferentes, existe certa semelhança entre a perspectiva de curadoria de dados e da curadoria digital, porém ambas possuem o foco diferente da curadoria de conteúdo, visto que esta última busca atender as necessidades informacionais não considerando em suas etapas a manutenção das informações a longo prazo.

Considerando a área de Ciência da Informação, as autoras Siebra, Borba e Miranda (2016) ratificam que a Curadoria vem sendo usada em especial nas três abordagens:

- a) Curadoria de Conteúdo ou de Informações que consiste na filtragem e seleção de dados na Web;
- b) Curadoria Digital que diz respeito à agregação de valor, gestão ativa e preservação de dados digitais em repositórios;
- c) Curadoria de Dados de Pesquisa que corresponde às técnicas relacionadas à E-science.

Em síntese, podemos resumir que a essência da curadoria em ambientes digitais é tratar a massa de informações resultante do excesso informacional, proporcionando o cuidado necessário para preservação, possibilitando seu acesso e reuso.

### **2.3 Ciclo de Vida da Curadoria**

Conforme já falado, atualmente a comunicação está inserida em um mundo digital onde nossa memória pessoal e coletiva é formada por todos tipos de informação (imagens e vídeos pessoais, documentos de trabalho, planilhas, prontuários, e-books, e-mails, blogs e etc).

O uso das tecnologias de rede, banco de dados e redes sociais nos permitiu produzir, publicar e trocar conteúdos digitais a uma rapidez enorme. Tais tecnologias, incluindo a Web, levaram a um crescente volume de objetos digitalizados ou natos

digitais que não só permitem operações diárias e serviços, mas também definem nossa identidade e experiência sociocultural. Esta vasta quantidade de informações - seja digitalizada ou nascido digital - precisa ser coletada, organizada e preservada a fim de garantir sua disponibilidade no futuro.

Com o objetivo de assegurar a sustentabilidade dos dados para o futuro e conferir valor para seus criadores e seus usuários, podemos considerar que os recursos estratégicos, metodológicos e as tecnologias envolvidas na Curadoria Digital podem facilitar o acesso a dados digitais confiáveis.

Para Constantopoulos et al (2009, p. 2) a Curadoria Digital traz a possibilidade de gestão de conteúdos digitais através do ciclo de vida.

A curadoria digital surgiu como uma nova prática interdisciplinar, prática de comunidade e campo de investigação, que busca encontrar respostas para essas questões. Seu princípio fundamental é que garantir a adequação futura ao propósito da informação digital, à medida que seu contexto de uso evolui, requer o gerenciamento ativo e a avaliação dos ativos digitais durante todo o seu ciclo de vida.

Segundo Pennock (2007) o material digital é suscetível a mudança tecnológica desde o momento da criação em virtude da sua própria natureza. Desta forma, a curadoria e preservação, atividades por vezes negligenciadas, podem ter influência na capacidade de cuidar com sucesso a longo prazo. Portanto, um ciclo de vida busca garantir que todos estágios necessários sejam identificados e planejados para que ações necessárias sejam implementadas na sequência correta.

O ciclo de vida da informação nada mais é do que o modelo de curadoria em ação, sua metodologia para assegurar que todas as atividades necessárias sejam identificadas, planejadas e implementadas de forma a poder garantir a manutenção, autenticidade, confiabilidade, integridade e usabilidade das informações digitais.

Existem inúmeros modelos de ciclo de vida, considerando as particularidades e objetivos de domínios da ciência. Para Sayão e Sales (2016) estes modelos possibilitam aplicar os conceitos de curadoria por meio de uma estrutura que ilustra as atividades de gestão dos dados durante sua vida, garantindo seu uso e proteção.

A criação destes modelos surgem da necessidade de formatar etapas para a gestão de conteúdos desde a responsabilidade de cada ação até a formulação e fluxos de trabalho, podendo variar conforme as políticas institucionais, não sendo necessário sua aplicação completa. Considerando a gestão e acesso contínuo, o

ciclo de vida pode ser entendido como um processo de etapas da informação digital e seu fluxo, desde a criação até a preservação (HIGGINS, 2008).

Desta forma, Hodges (2000) revela que o ciclo de vida geralmente envolve etapas do âmbito da criação, aquisição, catalogação/ identificação, armazenamento, preservação e acesso. Para atingir seu propósito, considerando a multidisciplinaridade aplicada à Curadoria Digital, a diversidade de materiais e domínios existentes, vários modelos de ciclo de vida foram criados no intuito de refletir tais especificidades.

O ciclo de vida pode auxiliar na definição de papéis e responsabilidades, além de criar estruturas de padrões e tecnologia e garantir a documentação devida dos processos e políticas que envolvem o fluxo (HIGGINS, 2008). Tendo em vista que a aplicação do ciclo é essencial para assegurar a continuidade dos materiais digitais, abordaremos alguns dos modelos mais presente na literatura para melhor entendimento de ações de curadoria.

### *2.3.1 Ciclo de Vida do Digital Curation Center (DCC)*

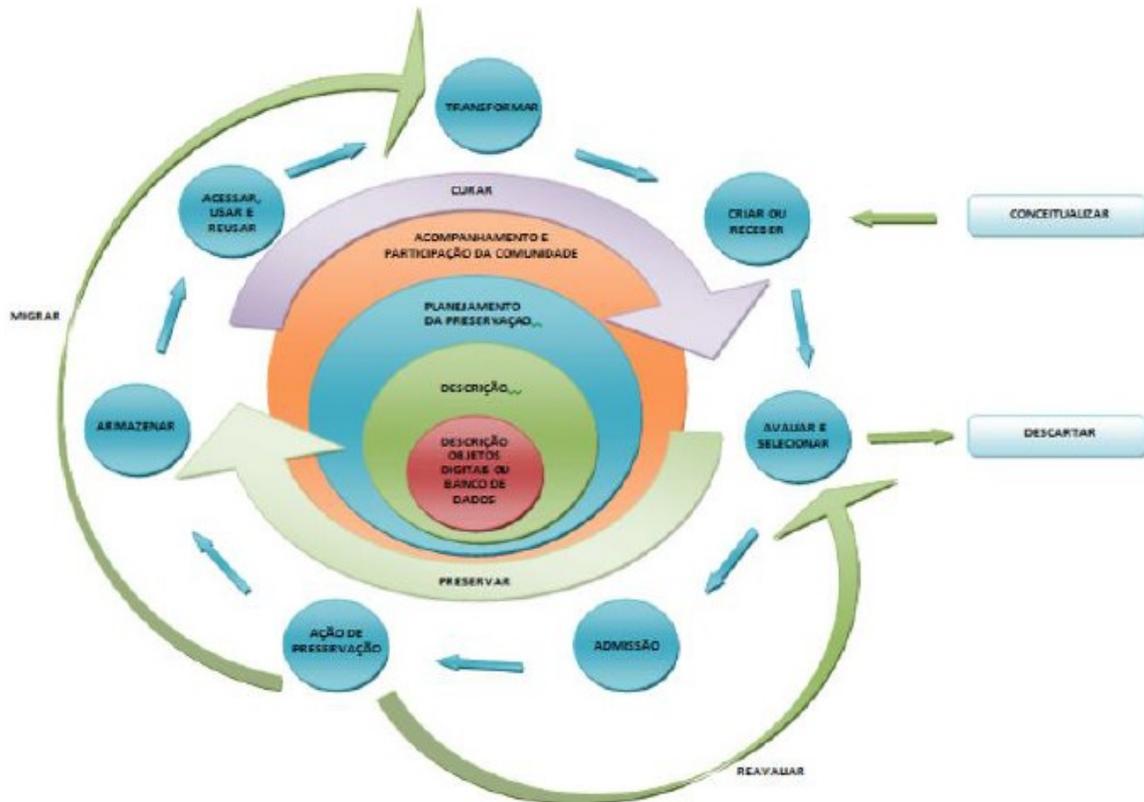
O modelo do Digital Curation Center (DCC), apresentado na figura 1, possui alto nível gráfico de atividades necessárias para uma curadoria bem sucedida. Segundo Higgins (2008) este modelo é genérico e adaptável a diferentes domínios, podendo ser utilizado para planejar atividades de gestão e preservação permitindo a eliminação, seleção e reuso a longo prazo.

De acordo com Constantopoulos et al (2009) o modelo, propõe ações que vão desde a descrição e representação da informação; planejamento da preservação, acompanhamento e participação da comunidade; considera ainda ações sequenciais como a conceitualização, criação e/ou recebimento, avaliação e seleção, arquivamento; e ações de preservação que incluem noções de armazenamento, acesso, uso e reuso e transformação, eliminação/descarte, reavaliação e migração.

Abordagem da curadoria digital neste modelo apresenta as fases de conceitualização; criação ou recepção; avaliação e seleção; processamento; ações de preservação; armazenamento; acesso, utilização e reutilização; transformação; eliminação; reavaliação e migração.

Os elementos chaves do modelo são: dados, objetos digitais e bases de dados. No centro do ciclo de vida da curadoria está especificamente o dado digital (codificado em formato binário). A ideia objetos digitais simples, que são aqueles compostos por um único arquivo, identificador e metadados; e os objetos digitais complexos, que são formados pela combinação de outros objetos digitais, por exemplo, uma página web (SAYÃO; SALES, 2012, p.126).

Figura 1 - Ciclo de Vida do DCC



Fonte: DCC adaptado por Silva e Siebra (2017).

Oferece uma visão geral dos estágios necessários ao processo de curadoria e preservação dos objetos digitais. Higgins (2008) também apresenta o modelo com uma natureza genérica, sendo indicativo e não hermético, ou seja, a instituição/organização poderá escolher se deverá cumprir todos os estágios do ciclo, ou ainda adequá-lo às suas necessidades e realidade.

É utilizado para treinamento de curadores do DCC proporcionando os conhecimentos necessários para a abordagem de curadoria e preservação, auxiliando na identificação de etapas adicionais que podem ser exigidas durante o processo. Na visão de Higgins (2008), apesar de proporcionar uma visão geral

gráfica de alto nível, o ciclo de vida do DCC poderá sofrer melhorias sendo assim possível sua evolução. Com este intuito o modelo a seguir foi elaborado.

### 2.3.2 Ciclo de vida do Digital Curation Unit (DCU)

O Digital Curation Unit (DCU) é um centro de pesquisa localizado em Athenas com foco em estudos no campo da ontologia de domínio cultural e demais temáticas que envolvem o contexto tecnológico dos objetos digitais. A fim de elucidar um processo de curadoria digital o DCU organizou suas atividades em duas etapas. Conforme Constantopoulos et al (2009) a primeira parte é denominada de gerenciamento do contexto (Context Management) e a segunda como gerenciamento do ciclo de vida dos recursos digitais (Digital Resources Lifecycle Management), conforme a Figura 2.

Figura 2 - Processo de Curadoria Digital do DCU



Fonte: Silva e Siebra (2017) adaptado de Constantopoulos et al (2009).

O gerenciamento do contexto, como o próprio nome diz, ressalta etapa de contexto de uso da informação, é formado por três processos de apoio: metas e modelos de uso - *goal and usage models* (objetivos do criador e padrões de uso); Modelos de domínio - *domain models* (representações sobre o domínio); Gestão de autoridade - *authority management* (controle de vocabulários).

O gerenciamento do ciclo de vida dos recursos digitais é o ponto em que acontece a curadoria em si e aborda oito processos: Avaliação - *Appraisal* (seleção de recursos); Captura - *Ingest* (importação de recursos); classificação, indexação e catalogação - *Classification, indexing and cataloguing* (descrição dos metadados); Aperfeiçoamento do conhecimento ou adição de valor - *Knowledge Enhancement* (inter relação entre documentos); Apresentação, publicação e disseminação - *Presentation, publication and dissemination* (disponibilização de informação); Experiências de uso - *Usage Experience* (interação entre os usuários); gerenciamento do repositório - repository management (mecanismos de acesso) e Preservação - *preservation* (salvaguarda dos recursos).

Com base nesse processo os pesquisadores do DCU vislumbraram a possibilidade de melhoria do ciclo de vida do DCC através da fusão das etapas de suas metodologias, criando assim uma extensão. Desta forma, o modelo de ciclo de vida estendido da DCU busca aliar as etapas do seu processo (figura 2) com a pretensão de apoiar o planejamento e organização da gestão de conteúdos digitais. Para isso, identifica atividades de curadoria que não estão contempladas no modelo do DCC (CONSTANTOPOULOS et al, 2009).

Os autores partiram de uma conceitualização da prática de curadoria digital e do campo de pesquisa com base nos estudos desenvolvidos, assim foi possível elaborar um ciclo de vida no qual estariam presentes não apenas atividades de avaliação, mas também a experiência de uso dos indivíduos.

Nessa perspectiva, a proposta vai além da noção sintática de transformação do abordada pelo DCC. Porém, apesar das peculiaridades, percebe-se que ambos modelos buscam alcançar a confiabilidade dos recursos, agregando valor por meio de serviços e novos usos para os objetos digitais.

Em resumo, de acordo com Constantopoulos et al (2009) a expansão do ciclo de vida do DCC proposto pelo DCU considera três aspectos:

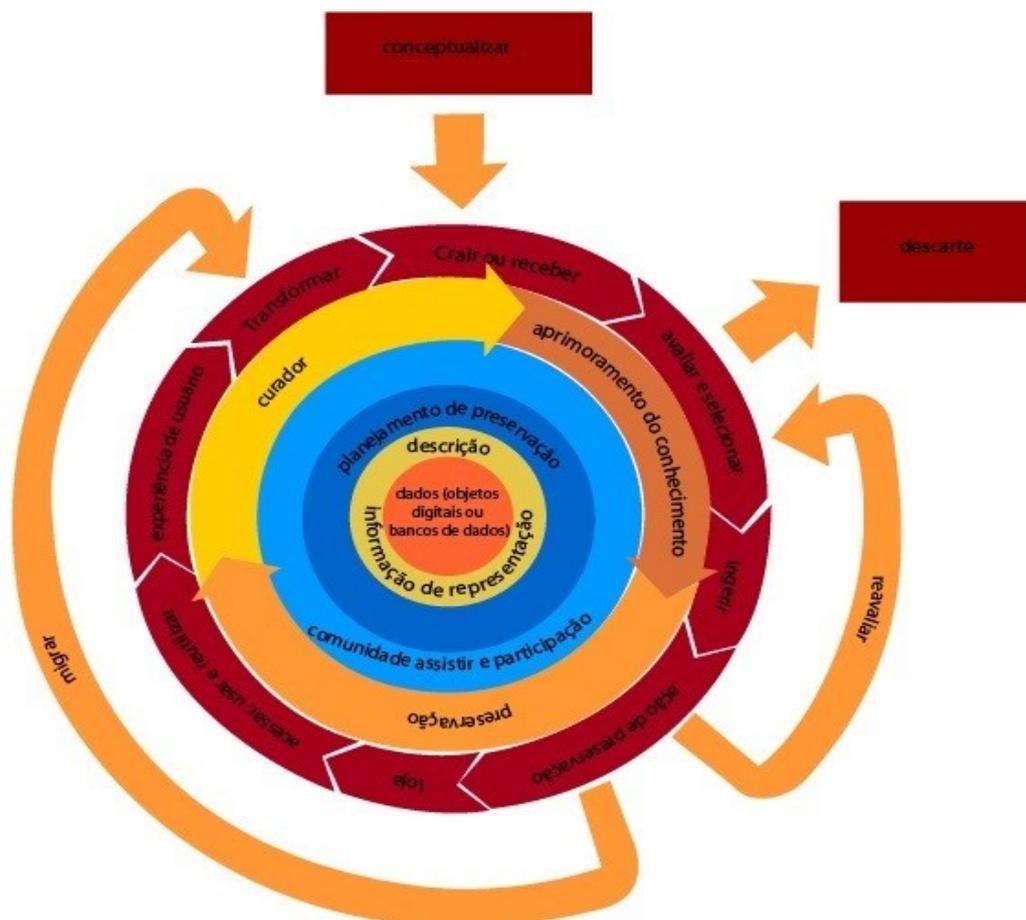
1. **O registro da experiência do usuário no acesso aos dados** - proporcionado com a interação de usuários através da criação de comunidades para criar e modificar ativos de informações. As interações podem ser intensificadas com o uso de recursos como as tags sociais,

anotações e etc. Propõe a adicionar uma ação sequencial após a etapa de “Acesso e Reutilização” na qual registra a experiência do usuário.

2. **Inclusão de vocabulários controlados** - Desta forma seria possível que autoridades evoluíssem como mudanças para o conhecimento da área de domínio. Propõe aumentar a ação ‘descrição e informação de representação’, para incluir informações sobre entidades, conceitos e relações.
3. **Agregar conhecimento aos repositórios** - Considerando que o aprimoramento do conhecimento é um componente essencial do ciclo de vida, a proposta do DCU inclui também aumentar uma ação existente, de “Curar e preservar” para “Preservação, Curadoria e Apoio ao Conhecimento”.

Como resultado, o modelo estendido foi reformulado para visualização gráfica das novas etapas, conforme podemos verificar na figura 3.

Figura 3 - Ciclo de vida estendido DCC & DCU



Fonte: Constantopoulos et al (2009) adaptado pela autora.

Nesta perspectiva, o modelo DCU reconhece que etapas importantes para uma Curadoria de Informações digitais incluem não apenas guarda e organização, mas também a produção, comunicação e como a comunidade acessa os dados para gerar conhecimento.

### 2.3.3 O ciclo de vida do UK Data Archive

O UK Data Archive tem atuado na Curadoria de dados por mais de 40 anos, tornando-se, assim com o DCC, uma referência na preparação e conservação de dados. É um centro especializado e reconhecido internacionalmente na aquisição, curadoria e promoção do acesso a dados de Ciências Sociais e Humanidades.

Para UK Data Archive (2010) Os dados digitais existem em formatos de arquivos específicos e ressalta essa especificidade quando considera a existência de várias formas de representação de dados: textual, bases de dados, geoespaciais, imagens, gravações audiovisuais e dados gerados por máquinas ou instrumentos. A figura 4 apresenta o ciclo de vida composto por seis etapas:

Figura 4 - Ciclo de vida do UK Data Archive



Fonte: UK Data Archive adaptado por Silva e Siebra (2017).

**a) Criação dos Dados:** projetando, planejando o consentimento, coleta e gerenciamento, capturando e criando metadados;

**b) Processamento de dados:** inserir, transcrever, verificar e validar, anonimizar e descrever;

**c) Análise dos Dados:** interpretando, derivando, produzindo saídas e publicação, preparando-se para compartilhar;

**d) Preservação dos Dados:** migração, backup, armazenamento, criação de metadados e documentação, arquivamento;

**e) Acesso a Dados:** distribuição, compartilhamento, controle de acesso e promoção

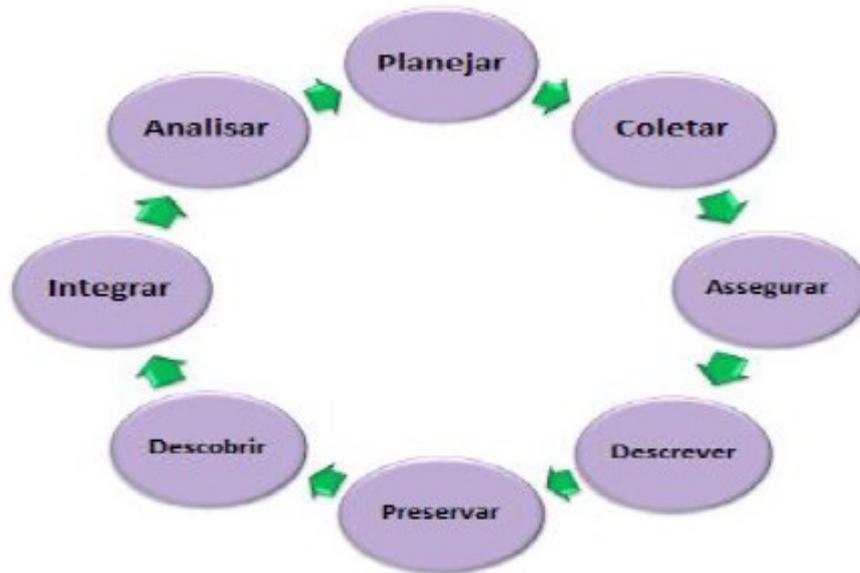
**f) Reutilizando os Dados:** para acompanhamento de novas pesquisas, análises de pesquisa, investigação, ensino e aprendizagem.

Apesar de sua criação ter por objetivo inicial da curadoria de dados de pesquisa, o conceito do modelo permite ser aplicada a diversos tipos de objetos digitais.

#### *2.3.4 O ciclo de vida do DataOne*

A Rede de Observação de Dados para a Terra (DataONE) é uma fundação para inovação das ciências ambientais. Atende às necessidades da ciência para acesso aberto, persistente, robusto e seguro de dados observacionais da Terra. O ciclo de vida dos dados DataONE (figura 5) fornece uma visão geral de alto nível das etapas envolvidas no gerenciamento bem-sucedido e preservação de dados para uso e reutilização.

Figura 5 - Ciclo de vida do DataONE



Fonte: Data One adaptado por Silva e Siebra (2017).

Foi desenvolvido com base no modelo de ciclo de vida apresentado pela National Science Foundation, constituído por oito etapas conforme descrição a seguir:

**a) Plano:** apresenta a descrição dos dados que serão compilados, como os dados serão gerenciados e tornados acessíveis ao longo de sua vida útil;

**Coleta:** as observações são realizadas à mão ou com sensores ou outros instrumentos e os dados são colocados em uma forma digital;

**b) Garantir:** a qualidade dos dados é assegurada através de cheques e inspeções

**c) Descreva:** os dados são descritos com precisão e detalhamento usando os padrões de metadados apropriados;

**d) Preserve:** os dados são enviados para um arquivo apropriado a longo prazo (ou seja, data center);

**e) Descubra:** dados e dados obtidos potencialmente úteis, juntamente a informação relevante sobre os dados (metadados);

**f) Integrar:** dados de fontes diferentes são combinados para formar um conjunto homogêneo de dados que podem ser prontamente analisados;

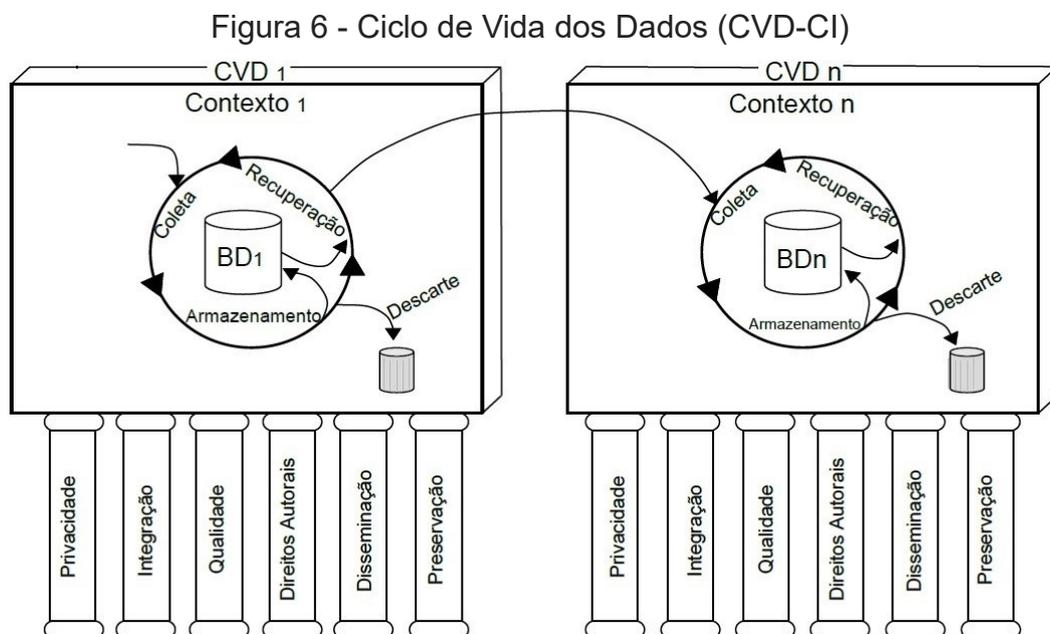
**g) Analisar:** verificação dos dados

Assim como os demais modelos, algumas atividades de pesquisa podem usar apenas uma parte do ciclo. Para a fundação, modelo serve como uma ferramenta de

navegação para o banco de dados do DataONE Best Practices, facilitando os usuários na descoberta dos dados. No intuito de auxiliar educadores e pesquisadores em suas atividades de treinamento e auto-aprendizagem, o DataONE desenvolveu módulos de educação para gerenciamento de dados.

### 2.3.5 O Ciclo de Vida dos Dados (CVD-CI)

O Ciclo de Vida dos Dados (CVD-CI) de Sant'Ana (2013, 2016) foi criado através de análises de modelos similares e nasce como uma proposta voltada para a área da Ciência da Informação. O autor ressalta que a ciência da Informação pode e deve contribuir para que o acesso e uso de dados se desenvolva, buscando identificar e estudar fatores e características que propiciem o equilíbrio no processo de otimização do uso dos dados. A figura 6 apresenta o CVD-CI, ilustrando suas quatro fases (coleta, armazenamento, recuperação e descarte), e incluindo os objetivos da privacidade, integração, qualidade, direitos autorais, disseminação e preservação a cada uma dessas fases.



Fonte: Sant'Ana (2013).

Na análise de Sant'Ana, as quatro fases do CVD-CI resumem as atividades essenciais em um ciclo de vida e agregam mais de uma fase em outros modelos.

Observa-se, ainda, que algumas fases dos modelos estudados foram consideradas como objetivos no CVD-CI. Nessa estrutura o modelo garante o atendimento aos requisitos essenciais da curadoria.

A fase de **Coleta** corresponde às atividades de obtenção e mecanismos de captura dos dados “seja na elaboração do planejamento de como serão obtidos, filtrados e organizados, identificando-se a estrutura, formato e meios de descrição que será utilizado” (SANT’ANA, 2013). A fase de **Armazenamento** envolve ações para persistência dos dados focando em seu uso posterior (reutilização), considerando especificações físicas e lógicas para o registro dos dados. Corresponde a descrição, processamento e tratamento. A fase de **Recuperação** preocupa-se com meios que ampliem os níveis de utilização destes dados como as formas de saída e a interface de visualização dos dados.

Quando os dados não são permanentes e seu limite do tempo previsto é alcançado passa-se a fase de **Descarte**, que pode ocorrer por meio de eliminação total ou parcial do dado. Como estratégia de preservação, dependendo da situação, mesmo no descarte, há possibilidade de geração de dados.

Cada uma das fases se retroalimenta, ou seja, a fase de coleta permite o início do armazenamento, que por sua vez viabiliza a recuperação. Como resultado das ações ocorridas na fase de recuperação, é possível gerar novos dados retomando ações da coleta para situações novas ou para a fase de armazenamento.

### 2.3.6 Síntese dos modelos

Além dos modelos aqui apresentados existem inúmeros outros porém limitamos a apresentação aos modelos com maior repercussão na literatura para compreensão das semelhanças e diferenças entre as abordagens.

Silva e Siebra (2017) enfatizam que considerando as recomendações da ISO sobre a manutenção e acesso dos objetos digitais, os modelos são estruturalmente desiguais tanto na complexidade quanto no detalhamento, porém desempenham funções similares nos procedimentos curatoriais e de preservação. As autoras enfatizam que as ações práticas (o que precisa ser realizado no contexto de cada etapa), bem como quais documentos devem ser produzidos na execução dos

processos/etapas, não são claros, apesar da descrição feita nas etapas dos modelos, fazendo do processo de aplicação do ciclo algo confuso.

Para melhor entendimento, o quadro abaixo propõe reunir os modelos apresentados relacionando quantidades de etapas, ações de preservação e seus formatos.

Quadro 1 - Comparação entre os modelos de ciclo de vida

Modelo	Etapas	Preservação	Formato
<b>DCC</b>	17	Sim, a preservação é planejada e reavaliada periodicamente e continuamente (ação para toda vida) e aplicada no decorrer do ciclo.	Cíclico, com ações sequenciais que ocorrem uma após outra de forma circular e que podem ou não se repetir de acordo com a necessidade do objeto curado.
<b>DDC&amp;U</b>	11	Sim, a preservação é uma etapa anterior ao gerenciamento de repositórios, visto que para manter o objeto acessível a longo prazo deve-se manter formas seguras de acesso e recuperação.	Apesar de no desenho do modelo não haver um sequenciamento explícito, na descrição é colocada uma ordem de execução das ações.
<b>UK DATA ARCHIVE</b>	6	Sim, neste modelo a preservação fica no meio das etapas que devem ser executadas para garantir acesso contínuo aos objetos digitais.	Cíclico, pois as ações de cada etapa podem ser repetidas sempre que preciso.
<b>DATAONE</b>	8	Sim, preocupa-se com a preservação a fim de garantir que os dados possam ser recuperados sempre que possível.	Cíclico com etapas que devem ser aplicadas seguindo a sequência estabelecida no modelo.
<b>CVD</b>	4	Sim, a preservação é apresentada como objetivo a ser considerado em todas as etapas	Apresentado como cíclico, porém a ordenação das etapas é dinâmica. Cada etapa é orientada por objetivos.

Fonte: Silva e Siebra (2017 ) adaptado pela autora.

Dentre os modelos apresentados, o CDV-CI de Sant'Ana sintetiza de forma clara e adequada as fases de processo de curadoria, apresentando um roteiro de questionamentos na orientação de cada etapa. O modelo foi escolhido para a aplicação deste estudo por sua concepção das fases essenciais de curadoria com bases na análise de modelos estrangeiros, mas principalmente pela proximidade com a CI.

## 2.4 Objetos de curadoria: dados brutos e derivados

Considerando o aspecto gerencial da informação é importante ressaltar as características dos dados e objetos digitais, visto que cada um carrega em si um contexto dependente de sua estrutura (materialidade) e aplicação.

Buckland (1991) apresenta características de análise para a informação em seu texto “information as a thing” em que considera 3 formas de uso da informação: a informação-como-processo (ato de informar - intangível); a informação-como-conhecimento (aplicação da informação - intangível) e a informação-como-coisa (representação do conhecimento comunicado).

Para o autor, a informação só atinge o objetivo de real quando é palpável, tangível. Portanto, qualquer necessidade informativa propõe uma perspectiva física (materialidade do objeto), correspondendo a informação-como-coisa. Sendo assim, sistemas de informação só existem com base em representações físicas (descrição/dados). Ainda de acordo com Buckland (1991) a informação-como-coisa pode ser entendida como objeto, dados e documentos que possuem caráter instrutivo de dar conhecimento ou comunicar algo.

Dados são fontes (primárias ou secundárias) para a elaboração de documentos como artigos, relatórios e etc, e necessitam de um contexto, ou seja, de parâmetros para sua interpretação (SAYÃO; SALES, 2016). Um exemplo de dados são os números coletados de um pluviômetro, dos quais só fazem sentido se interpretados a partir de parâmetros de referência.

Documentos por sua vez são dados estruturados, descritos e interpretados, relacionados em um contexto específico. São relatórios de pesquisa, relatos de experiência, artigos científicos, livros e etc (SANTOS, 2014). Diferente dos dados brutos, os documentos possuem uma característica hermética e validade científica comprovada, sendo geralmente divulgados em revistas e demais meios de publicação que obedecem uma sistemática própria de direitos autorais.

O documento digital é um objeto conceitual que não necessariamente implica na fixação de dados em unidades físicas e sim em unidades lógicas nas quais forma, conteúdo e suporte são encarados como variáveis a serem consideradas quando da consideração desses para preservação (SANTOS, 2014, p. 68).

‘ A manipulação de dados, no sentido de cuidado e uso, traz a tona sua versatilidade. Desta forma, um dado pode ser analisado e reanalisado inúmeras vezes, o que torna possível sua reutilização para novas interpretações e consequentemente novas pesquisas. Para Sayão e Sales o termo “dado de pesquisa” possui uma amplitude de significados que se modificam conforme seus domínios científicos, objetos de pesquisas e diversas variáveis, podendo ser:

[...] resultado de um experimento realizado num ambiente controlado de laboratório, um estudo empírico na área de ciências sociais ou a observação de um fenômeno cultural ou da erupção de um vulcão num determinado momento e lugar. Dados digitais de pesquisa ocorrem na forma de diferentes tipos de dados, como números, figuras, vídeos, softwares; com diferentes níveis de agregação e de processamento, como dados crus ou primários, dados intermediários e dados processados e integrados; e em diferentes formatos de arquivos (SAYÃO; SALES, 2016, p.93-94).

Percebe-se, diante de suas características, que os dados brutos de pesquisa possuem uma fragilidade maior do que os documentos. Documentos geralmente são publicados em periódicos e possuem uma forma de armazenamento e acesso definidos. Quanto ao acesso e armazenamento de dados digitais de pesquisa, atualmente ainda compete ao pesquisador, ou grupo de pesquisa, a guarda e manutenção dos dados digitais de pesquisa que foram fontes dos documentos por eles publicados. Sua manutenção, portanto, dependerá das estratégias de cuidado aplicadas pelos pesquisadores para guarda dessas fontes.

Considerando a perspectiva de cada objeto digital, sejam dados brutos ou derivados, acreditamos que o ciclo de vida pode ser aplicado para garantia de ações de Curadoria em Repositórios.

### 3 AMBIENTES DIGITAIS DE CURADORIA

A comunicação científica caminhou apoiada na evolução digital. De início informações sobre pesquisas eram divulgadas de forma lenta e custosa. Com o passar do tempo, por volta do século 18, a divulgação ganhou novo formato com a criação dos primeiros periódicos científicos.

Nesse sentido, para Garvey e Griffith (1979) a comunicação científica pode ser entendida como um conjunto das atividades de produção, disseminação e uso da informação. O conceito engloba ações desde o princípio da criação científica, onde as idéias da pesquisa são geradas, até o momento da validação dos resultados pelos pares.

Porém, a intensificação crescente das tecnologias de informação e comunicação põe em evidência uma diversidade de fontes de informação. Na perspectiva científica, em universidades e centros de pesquisa, esse contexto cria possibilidades que favorecem a divulgação de resultados de pesquisa (LYNCH, 2003).

Os avanços das Tecnologias de comunicação e informação (TIC) permitiram o surgimento de redes de comunicação eletrônica, o que revolucionou a troca e uso de informação, seus suportes, formas de acesso impactando também nos meios de divulgação científica.

Foi devido ao surgimento dessas tecnologias que foi possível o acesso a dados, teorias e fontes de pesquisa de maneira mais rápida e menos custosa. Em contrapartida, o aumento intenso da produção literária em período pós-guerra e o uso de novas plataformas de comunicação resultou na “crise dos periódicos”.

Tal efeito foi motivado pela dificuldade das bibliotecas em manter coleções físicas de periódicos, tornando-as inviáveis em virtude da necessidade de espaço e de manutenção onerosa (MULLER, 2006).

Lawrence (2001) revela sobre o impacto de publicações *online* versus *offline*. Seu estudo mostra um crescimento de mais de 300%, em média, nas citações de artigos disponíveis *online* em comparação a artigos *offline* da mesma fonte. O autor acredita que tornar a pesquisa facilmente acessada acelera o progresso científico aumentando o impacto.

Segundo Silva Júnior e Borges (2014), foi a partir da crise dos periódicos que o movimento em favor do acesso aberto conhecido como *Open Access* surgiu permitindo a democratização do acesso. Dentre as primeiras plataformas digitais em acesso aberto podemos citar os repositórios digitais.

As primeiras iniciativas de repositórios de acesso aberto remontam a década de 90, com o ArXiv (1991) (arxiv.org), da Biblioteca Universitária de Cornell, com documentos das áreas de Física, Matemática e Ciências da Computação e o RePec (*Research Papers in Economics*) da área de Economia, mantido por cooperação entre várias organizações profissionais de área (RUMSEY, 2006).

Diante da criação de Repositórios, bases de dados e diversos canais de comunicação científica, é importante ressaltar que em virtude do movimento *Open Access* e da explosão documental a questão não é mais a ausência de informação, mas garantir a seleção de conteúdos com veracidade e objetividade em meio a grande massa documental disponível.

A informação que antes era escassa, agora é abundante. Portanto, validade e visibilidade tornaram-se vitais em um mundo com excesso informacional, ao passo que a relação da sociedade e seu consumo de informação nos coloca em um novo contexto.

### **3.1 Repositórios Digitais**

Na era digital, além da dificuldade de selecionar informações válidas, outra questão vem à tona. O fato é que a grande maioria dos dados gerados diariamente perdem-se na mesma velocidade que são criados, esse efeito é verificado diante da fragilidade de objetos digitais em razão da gradual obsolescência tecnológica.

À medida que os repositórios começam a expandir-se em todas as áreas do globo e se tornam o espelho ou a imagem viva de uma instituição, de uma área científica ou de um país por conterem, ao menos potencialmente, a produção científica, independentemente da sua tipologia, a questão da preservação digital assume-se crucial, sobretudo se tivermos em conta os objetos que existem apenas nos repositórios (SILVA JUNIOR; BORGES, 2014, p. 569).

Como já exposto, verificamos que a comunicação científica mantida e disseminada em meio digital necessitam de ações estratégicas que promovam

visibilidade e preservação dos conteúdos. Nesse sentido, considerando sua função e aplicação, podemos afirmar que os Repositórios Digitais são ambientes ideais para a promoção, manutenção e preservação dos dados digitais.

Para Leite et al (2012, p. 29) eles podem ser conceituados como “sistemas de informação, ou bases de dados, criados com o objetivo maior de reunir, organizar e tornar acessível a produção científica”. Os autores reforçam a importância dos repositórios em um movimento inovador na comunicação científica e na gestão da informação.

Para Crow (2002, p. 53) os repositórios são “coleções digitais que capturam e preservam a produção intelectual da comunidade de uma única universidade ou de uma comunidade multiuniversitária”. Como ambientes digitais eles possibilitam reunir dados e informações de cunho científico, administrativa, técnica, artística e etc., podem ser de acesso aberto ou restrito, de acordo com as políticas adotadas pelas instituições (ABREU; VIDOTTI, 2016).

Porém, diante de tantos sistemas, o conceito de repositório é comumente comparado a outros ambientes digitais. Leite et al (2012) afirmam que não é de se estranhar a comparação entre repositórios com periódicos e sistemas de gerenciamento de bibliotecas.

Sobre isso, é importante ressaltar que periódicos são veículos de publicação pois fazem o registro de propriedade intelectual e possuem o fluxo de certificação (*peer-review*), circulação (plataformas de acesso) e arquivamento (preservação). Repositórios diferenciam-se dos periódicos da seleção à divulgação de conteúdo, portanto não concorrem entre si. Periódicos utilizam em geral a avaliação por pares para seleção de artigos, já a seleção do material de um repositório deve ser definida pela política institucional (LEITE et al, 2012).

Repositórios Institucionais atuam como divulgadores do conhecimento potencializando as funções dos periódicos, mas não são veículos de publicação científica.

[...] os Ris [Repositórios Institucionais] servem para o depósito de material já publicado. É importante frisar que repositórios institucionais são veículos de maximização da disseminação de resultados de pesquisa, dado que tornam seus conteúdos disponíveis e acessíveis amplamente. Do ponto de vista formal, entretanto, a publicação é uma prerrogativa de periódicos e outros tipos de veículos. Esse mal-entendido reforça a falsa ideia de concorrência entre repositórios e periódicos. Um artigo científico é publicado em um periódico e depositado no repositório (LEITE et al, 2012, p.8).

Lynch (2003) reforça a ideia de que a definição de RI não exige um novo papel de publicação acadêmica, mas sim de divulgação de comunicação acadêmica. Através deles é possível gerenciar as mudanças tecnológicas, como a migração de conteúdo digital como parte da organização comprometida com o fornecimento de serviços para sua comunidade.

Leite et al (2012) também ressaltam sobre a confusão no entendimento entre Sistemas de Gestão de Bibliotecas (SGB) e Repositórios. Para os autores entende-se que os SGBs são necessários para a manutenção, desenvolvimento e controle do acervo incluindo livros e publicações que não são institucionais e que estão disponibilizadas na biblioteca como fonte de informação, seja física ou digital. Em contraponto, Repositórios trabalham principalmente com a produção científica e acadêmica em formato digital.

Contudo, segundo Leite et al (2012) é importante ressaltar que apenas a instalação e configuração de um software não garante a denominação do ambiente como Repositório, pois a tecnologia não é suficiente para determinar isso e sim suas ações para cumprimento do objetivo. Na perspectiva técnica, os autores reforçam que a implementação de um repositório envolve etapas de planejamento, implantação e funcionamento, sendo cada etapa é independente e constituídas de atividades necessárias.

Um dos objetivos do Repositório é promover a preservação dos objetos digitais através do gerenciamento de informação, para isso podemos associar a Curadoria Digital como um conjunto de estratégias que auxiliam no cuidado destes materiais. Abreu e Vidotti (2016) reforçam as ações da Curadoria Digital como parte da criação de serviços para o armazenamento de objetos digitais, ratificando o compromisso de preservar a longo prazo documentos e informações que se encontram nos repositórios.

Desta forma, considerando a fragilidade e efemeridade dos objetos digitais, os repositórios podem vir a ser ambientes digitais que promovem o armazenamento e manutenção da produção científica. Enquanto a Curadoria é a estratégia de cuidado, o repositório revela-se ambiente para aplicação do ciclo de vida, ao passo que sua estrutura permite o acesso dos materiais a longo prazo.

Para entendimento das especificidades de um repositório digital, os autores Silva Júnior e Borges (2014) apresentam tipologias em vertentes que podem ou não ser associadas, são elas: Temática, Pesquisa, Nacional e Institucional.

Repositórios Temáticos armazenam conteúdos de assuntos especializados, já repositórios de Pesquisa possuem o objetivo de armazenar e difundir a produção de pesquisas, são patrocinados por agências de fomento ou organizações. Nesse sentido, como exemplo, podemos citar o **Open Access Repository Of Science Methods** (<https://www.protocols.io/>). Na figura 7 verificamos a interface da plataforma que possibilita o cadastro de protocolos de pesquisa e dispõe de ferramentas que auxiliam ao pesquisador na execução dos procedimentos experimentais, registrando as técnicas do processo passo a passo.

Figura 7 - Interface de apresentação de um protocolo no *Repository Of Science Methods*

The screenshot displays the protocols.io website interface. At the top, there is a search bar and navigation links for 'EXPLORE', 'MY LIBRARY', and a user profile for 'Morgana'. Below the search bar, the breadcrumb trail reads: 'Researchers > Aayushi Uberoi > Protocols > Isolation and quantification of MmuPV1 virions from papillomas'. A secondary navigation bar includes 'Steps', 'Abstract', 'Warnings', 'Materials', 'Forks', 'More', and action buttons for 'START EXPERIMENT', 'COPY / FORK', and 'EXPORT'. The main content area shows the protocol title 'Isolation and quantification of MmuPV1 virions from papillomas' by 'Aayushi Uberoi<sup>1</sup>, Paul Lambert<sup>1</sup>'. The author affiliation is 'McArdle Laboratory of Cancer Research', dated 'Oct 31, 2017', with a DOI link. A 'Working' status and a score of 1 are visible. Below this, the protocol steps are listed under the heading 'Harvesting of virus particles from papillomas'. The first step is numbered '1' and contains a list of instructions: '1. Euthanize the animal from which tissue has to be harvested by CO<sub>2</sub> exposure as per IACUC guidelines.', '2. Use a scalpel to excise the papilloma from the ear taking care to not collect the surrounding epithelia (non-papilloma tissue).', and a sub-bulleted list: '• Weigh the papilloma and place in a 15ml round bottom falcon tube. For transportation tube is kept on ice till ready for processing at lab bench', '• For every 10mg wart 700µl DPBS is added to the tube', '2. Papilloma is homogenized for less than one minute using homogenizer and suspension is immediately transferred to a siliconized 1.5ml sterile eppendorf tube.', '3. In a 50ml falcon tube add about 25ml DPBS and wash homogenizer to collect any debris that might have got stuck in the homogenizer blades. Centrifuge falcon tube to collect pellet and aspirate the supernatant. Resuspend pellet in 42.5µl DPBS and combine with above fraction.', '• Add 7.5µl of Triton-X-100 solution (1% Triton-X-100).', '• Add 2µl Benzonase endonuclease and incubate at 37°C for 20 minutes inverting tube every 5 minutes.', '1. Add 2-3mg of Collagenase H and vortex sample and incubate at 4°C', '2. Adjust NaCl concentration of suspension to 0.8M by adding appropriate amount of 5M NaCl (Since suspension is 750µl, add 127.5µl of 5M NaCl).', '3. Centrifuge tube for 5 mins at 5000 x g. Transfer supernatant to fresh siliconized tube. Collect a 20µl aliquot'. On the right side, a blue sidebar titled 'Dear Protocol Author, Aayushi Uberoi' contains a text input field with the prompt 'Ask the author questions, make suggestions for improvements, or share your own experiences with this protocol...'. Below the input field are options for 'Private comment' and a 'POST' button. At the bottom of the sidebar, there are tabs for 'All', 'Step-level', and 'Protocol-level', and a message 'No comments yet.'

Fonte: Uberoi e Lambert (2017).

É através da utilização de identificadores persistentes que há disponibilização de um objeto digital sem de perda do acesso. Neste ponto, a plataforma oferece como serviço a atribuição de DOI garantindo o acesso, disponibilidade de metadados e visibilidade. Os materiais, chamados de protocolos, são dados e não apenas PDFs estáticos, mas possuem interação e compatibilidade de acesso tanto na *web* quanto no celular (*mobile*). Para promover a interação entre os usuários são

ofertados serviços de colaboração como fórum de discussão e melhorias dos protocolos após sua publicação (ZAPPYLAB, 2017).

Lynch (2003) afirma que os primeiros repositórios institucionais surgiram como uma estratégia que permitiu às universidades assumir o papel de editoras, dando autonomia nos processos de publicação e divulgação a produção acadêmica em conteúdo digital. O autor apresenta o repositório institucional como um conjunto de serviços que uma instituição oferta a sua comunidade para gestão e disseminação de materiais digitais criados por ela mesma. Como exemplo nesse sentido, citamos a iniciativa da Embrapa do repositório de **Acesso Livre à Informação Científica da Embrapa (ALICE)**, apresentado na figura 8.

Figura 8 - Interface de apresentação de um resumo em anais de congresso no Alice

Use este identificador para citar ou linkar para este item: <http://www.alice.cnptia.embrapa.br/alice/handle/doc/1078867>

Unidade da Embrapa/Coletânea:	Embrapa Arroz e Feijão - Resumo em anais de congresso (ALICE)
Data do documento:	6-Nov-2017
Tipo do Material:	Resumo em anais de congresso (ALICE)
Autoria:	OLIVEIRA, C. J. P. de GUIMARÃES, C. M. STONE, L. F. SARMENTO, P. H. L. SANTOS, A. J. da S.
Informações Adicionais:	CACIO JOSE PERES DE OLIVEIRA, Faculdade Araguaia; CLEBER MORAIS GUIMARAES, CNPAF; LUIS FERNANDO STONE, CNPAF; PEDRO HENRIQUE LOPES SARMENTO, CNPAF; ANA JULIA DA SILVA SANTOS, UFG.
Título:	Ajuste da população de plantas para a variedade de feijão-comum BRS FC 104.
Fonte/Imprensa:	In: CONGRESSO NACIONAL DE PESQUISA DE FEIJÃO, 12., 2017, Piracicaba. Produtividade e sustentabilidade da cultura do feijão: do campo para a mesa: resumos. Piracicaba: CENA: IAC, 2017.
Páginas:	p. 137.
Idioma:	pt_BR

Fonte: Embrapa (2017).

De tipologia institucional e nacional, a plataforma tem por objetivo de reunir, organizar, armazenar, preservar e disseminar, na íntegra, informações científicas produzidas por pesquisadores da Embrapa a nível nacional. Composto por diversos documentos que vão desde capítulos de livros, artigos em periódicos indexados, artigos em anais de congressos, teses e dissertações, à notas técnicas, dentre outros, produzidos pela instituição.

O Alice utiliza tecnologias padronizadas adotadas pela comunidade científica mundial, sendo interoperável com os demais sistemas de acesso aberto integrando uma rede de informação científica. Contribuindo de forma direta para o aumento do impacto dos resultados de pesquisa, bem como para uma maior visibilidade da instituição (EMBRAPA, 2017). Nesse sentido, pode-se considerar que a forma de armazenamento em um repositório pode contar com diferentes de objetos digitais, porém seu objetivo continuará sendo o de garantir a preservação e acesso ao conhecimento.

Segundo Leite et al (2012), em um repositório institucional, os documentos podem ir de literatura publicada a *papers*, em formato de manuais, cartilhas, trabalhos de conclusão de curso, objetos multimidiáticos como vídeos, áudios e etc.

O repositório Alice da Embrapa, apresentado neste capítulo, tem uma iniciativa de integração com o catálogo da biblioteca institucional. Tal ação propõe centralizar metadados de publicações científicas, de acesso aberto, integrando os conteúdos em uma só interface.

Conforme observamos no capítulo anterior, como estratégia de preservação e cuidado, é possível aplicar conceitos de Curadoria Digital através do ciclo de vida observando suas etapas. Reforçando esse contexto, Abreu e Vidotti (2016) afirmam que a Curadoria Digital determina, mantém e pode garantir qualidade e veracidade das informações nos repositórios digitais, contribuindo com a qualidade de seus serviços. Mesmo assim algumas etapas na sua criação devem ser consideradas.

### 3.1.1 Planejamento

Como etapa primordial do planejamento a elaboração da política para um repositório deve abordar os objetivos, diretrizes, práticas e visão organizacional. Para tanto, exige conhecimento da instituição de maneira a formular ações adequadas aos seus interesses e necessidades.

Conforme Tomáel e Silva (2007) a política de um repositório institucional deve apresentar os seguintes elementos:

- a) responsabilidade pela criação, implementação e manutenção do repositório;
- b) conteúdo proposto e implementado;
- c) aspectos legais relativos a documentos e licenças de softwares;

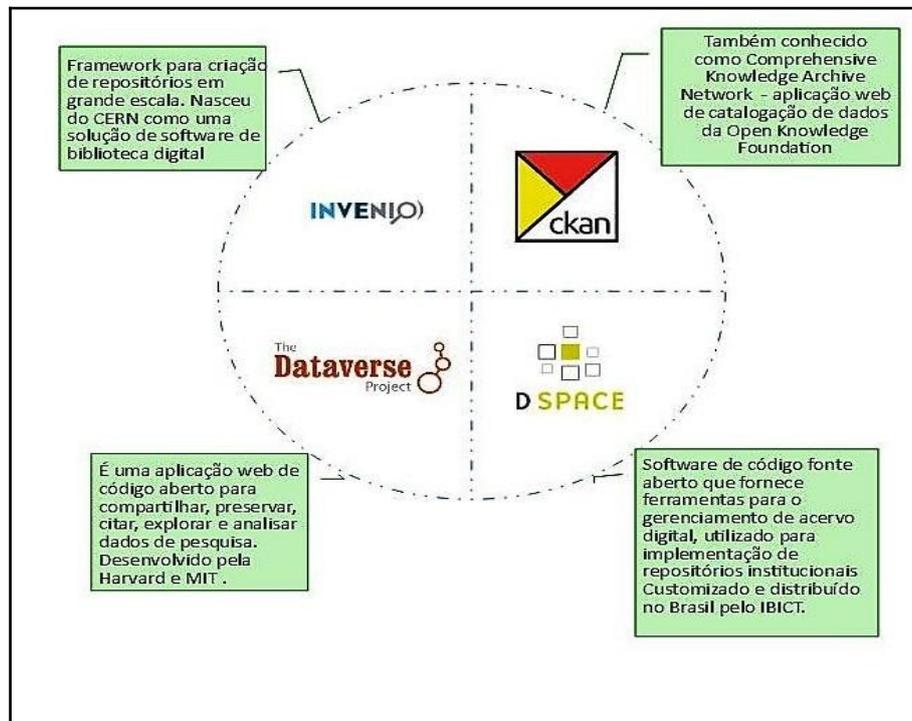
- d) padrões;
- e) diretrizes para preservação digital;
- f) níveis de acesso;
- g) sustentabilidade do repositório.

Para as autoras além dos itens já mencionados, é importante considerar questões como: a segurança dos dados, helpdesk, definição dos níveis de acesso ao usuário e a designação de responsável pela gestão da aquisição de trabalhos. Leite et al (2012) reforçam esses elementos quando afirmam que política deve abordar os objetivos do repositório além de determinar a equipe responsável tanto pela implantação quanto pela sua manutenção. Isso contribui para a definição de um serviço de qualidade na implantação do RI.

A política ainda deverá conter o tipo de material que será depositado ou não, determinando como será feito o depósito com prazos e demais aspectos que auxiliem nesse processo. Após a definição, o registro da política de funcionamento em diretórios internacionais, tais como: ROARMAP e Políticas Melibea, é necessário para garantir a ampliação da visibilidade do repositório. Portanto, o registro nestes diretórios torna-se vital porque buscadores utilizam estes sistemas como fonte para localizar repositórios de seu interesse (LEITE et al., 2012).

Outra ação do planejamento é a definição do software de implantação e gerenciamento do repositório, ou, a tecnologia que será utilizada. Alguns dos mais conhecidos estão presentes na figura 9.

Figura 9 - Softwares para repositórios



Fonte: Elaborado pela autora.

Após a avaliação e escolha do software, é necessária a definição da organização dos conteúdos, ou seja, a arquitetura de informação do repositório. Essa organização pode depender de uma instituição para outra. Segundo Leite et al (2012) em geral, as instituições acadêmicas estruturam suas comunidades de acordo com a hierarquia institucional de faculdades, departamentos ou centros de pesquisa, o que facilita a categorização dos conteúdos pelo criador.

### 3.1.2 Implantação

Elementos chave para no processo de implantação, os metadados possibilita a individualização de cada objeto por meio da descrição para uma recuperação efetiva. Leite et al (2012) recomendam um esquema de metadados para cada tipo de documento (artigo de periódico, livros, teses, dissertações, etc) visto que cada material possui informações específicas. Desta forma, a necessidade elaboração de novos esquemas pode surgir a com o crescimento do repositório e implementação de novos materiais.

Para Abreu e Vidotti (2016) no DSpace podem ser disponibilizados diversos tipos de materiais digitais como livros, apresentações de eventos, relatórios, notícias de jornal, artigos, projetos, softwares, dados bibliográficos, teses e dissertações, entre inúmeros outros.

Para Tomaél e Silva (2007) A definição de padrões de metadados englobam alguns aspectos como:

- a) plataforma tecnológica (escolha de hardware e software)
- b) metadados, que podem ser mais simples e genéricos (Dublin Core e RDF (*Resource Description Format*), ou mais complexos (*Metadata Encoding and Transmission Standards, Encoded Archival Description ou Multimedia Content Description Interface*).
- c) fluxograma de trabalho (inclui tanto a gestão do repositório quanto o fluxo de publicação de documentos);
- d) interface e usabilidade,
- e) interoperabilidade

Ainda sobre esses padrões, Sayão et al (2009) revelam que o *Open Archives Initiative* possibilitou a criação do Dublin Core e o *Open Archives Initiative Protocol for Metadata Harvesting* (OAI – PMH). Essas linguagens auxiliam na coleta automática (*harvesting*), permitindo a interoperabilidade e conseqüentemente o reuso dos metadados dos repositórios abertos.

Além dos metadados descritivos, outra questão que deve ser considerada é o controle de autoridade, ou ainda, os metadados temáticos, pois estes possibilitam estabelecer a base de uniformidade bibliográfica para a descrição de outros documentos.

### 3.1.3 Funcionamento

A etapa final diz respeito ao povoamento dos repositórios, seu mapeamento e seleção dos documentos na prática. Portanto, respeitando a organização e planejamento previamente estabelecidos, é importante armazenar os tipos de documentos que tenham sido publicados por membros da instituição que o mantém e que façam que parte da política.

Questões como licenças e direitos autorais devem ser consideradas nesta etapa. Leite et al (2012) ressaltam que, caso o repositório seja gerenciado pelo software Dspace, é possível inserir no processo de autoarquivamento a autorização para o depósito do documento indicando as licenças *creative commons*. Os autores complementam que além promover a memória institucional, um repositório deverá oferecer serviços à sua comunidade, como por exemplo:

- a) Disseminação seletiva de informação (uso de RSS, QR code e etc);
- b) Digitalização da produção científica impressa;
- c) Dados estatísticos para os autores (métricas de citação, visualização e downloads);
- d) Depósito mediado e autodepósito;
- e) Orientação sobre direitos autorais;
- f) Treinamento e suporte aos usuários;
- g) Garantir o acesso através da atribuição de identificadores persistentes (DOI, URN, PURL, *Handle System* etc).

A manutenção do repositório deverá ser levada seriamente da mesma forma que sua implementação. Estratégias de marketing em redes sociais e cadastro em divulgadores podem ser utilizadas para dar a visibilidade tão esperada ao repositório.

Importante esclarecer que as etapas aqui abordadas tratam de forma técnica a implementação de um repositório, ou seja, quais ferramentas usar e como planejar seu funcionamento. Entende-se porém, que outras ações devem ser respeitadas e consideradas no processo de concepção desses sistemas para garantia efetiva da preservação dos dados. Nesse sentido, este estudo propõe tornar possível a união dos conceitos e aplicações da Curadoria em Repositórios Institucionais.

## 4 METODOLOGIA

Esta pesquisa possui uma abordagem qualitativa, tendo em vista que os objetivos não buscaram medir eventos para análise de dados. Neves (1996) ressalta que um estudo qualitativo tem o objetivo de traduzir e expressar os fenômenos sociais, reduzindo a distância entre teoria e dados, contexto e ação, assim possibilita a obtenção de dados descritivos através do contato direto entre o pesquisador e objeto de estudo. Compreende, portanto, um conjunto de técnicas que visam descrever e decodificar os componentes de um sistema.

A aplicação da perspectiva qualitativa foi feita através de um estudo de caso. De acordo com Yin (2005) o estudo de caso propõe examinar e analisar o “caso” com nível de detalhe, em profundidade, no seu contexto natural. Pode ser entendido ainda como “um método de observação, de construção de raciocínio e de relato de informações que entrelaça teoria com observações de fatos, possibilitando a reflexão e a formulação de hipóteses, abrindo portas para novas descobertas” (CAPITÃO; VILLEMOR-AMARAL, 2016, p. 290).

Caracteriza-se por uma investigação empírica, não experimental, baseada no raciocínio indutivo, aplicada por meio do trabalho de campo, utilizando múltiplas fontes de dados. Explora o fenômeno através de uma coleta de dados envolvendo as mais variadas fontes como: observação, entrevistas, questionários, áudio e vídeo, atas, documentos oficiais, etc. Pode descrever situações ou fatos e produzir conhecimento ao contrastar os efeitos e relações do caso estudado (COUTINHO; CHAVES, 2002).

Para Capitão e Villemor-Amaral (2016, grifo nosso), considerando seus objetivos, o estudo de caso pode: Informar, pois descreve informações sobre determinada situação ou indivíduo; Ilustrar, apresenta o caso como uma experiência; Problematizar, criando hipóteses em relação a um problema, contrastando a teoria e o dados coletados, permitindo uma percepção analítica e sintética do caso; e por fim, Apoiar ou provar hipóteses.

Através do estudo de um caso específico, esta pesquisa verificou ações para melhoria da gestão de dados digitais em um repositório institucional. Para isso, ações identificadas na instituição escolhida foram comparadas às etapas de **coleta**,

**armazenamento, recuperação e descarte** do Ciclo de Vida dos Dados (CVD) de Sant'Ana (2013, 2016).

O ciclo de vida da informação pode ser encontrado sob o olhar da arquivística e da gestão de informação, pois possui a organização e tratamento da informação como objetivos essenciais. Para a Curadoria essa percepção não é abandonada, mas implementada através de ações que garantem o reuso dos dados científicos. Nesse sentido, o ciclo de vida garante que todas as etapas necessárias sejam implementadas corretamente.

Inúmeros modelos foram identificados na literatura, a maioria foi elaborado para instituições com necessidades específicas como o DCC ou UK Archive. Dentre as leituras foi identificado um modelo de ciclo de vida proposto por Sant'Ana (2013, 2016), o CVD, o qual é resultado da análise de vários modelos internacionais, apresenta uma perspectiva da Ciência da Informação, além de abordar todas as ações necessárias em apenas quatro etapas, o que facilitou a categorização de análise dos dados desta pesquisa.

A revisão de literatura possibilitou delimitar abordagens necessárias sobre a temática para uma análise efetiva do ciclo de vida. Segundo Perovano (2016) é através da revisão de literatura que podemos obter informações relevantes e necessárias sobre o problema de pesquisa. Ainda segundo o autor, pode ser definida como um conjunto de ações de identificação, coleta e consulta bibliográfica para embasamento de um estudo.

Neste estudo a revisão foi construída através artigos, livros, dissertações/teses e entrevistas, coletados em bases de dados indexadas no Portal de Periódicos da Capes. Ressalta-se a contribuição dos artigos publicados pela revista *International Journal of Digital Curation*, editada pelo *Digital Curation Center* (DCC), no qual foram publicados documentos considerados referência para área.

No levantamento da bibliografia houve o cruzamento de termos utilizando o operador booleano "AND", considerando os seguintes descritores: curadoria digital, curadoria de conteúdo, curadoria de informação, curadoria de dados, ciclo de vida da informação, repositórios digitais, *e-science*, explosão informacional, comunicação científica eletrônica, ciência da informação, metadados, preservação digital e gestão de dados. Diante das temáticas, ressalta-se (por autor/título) as contribuições dos seguintes documentos:

Quadro 2 – Textos referenciais nas temáticas abordadas

<b>Curadoria Digital</b>	SAYÃO; SALES - Curadoria Digital: um novo patamar para preservação de dados digitais de pesquisa. SANTOS - Curadoria digital: o conceito no período de 2000 a 2013. LEE; TIBBO - Where's the Archivist in digital curation? exploring the possibilities through a matrix of knowledge and skills. HIGGINS - Digital curation: the emergence of a new discipline. DATA CURATOR CENTER - What is digital curation? BEAGRIE - Digital Curation for Science, Digital Libraries, and Individuals.
<b>Ciclo de Vida</b>	SILVA; SIEBRA - Análise de modelos de ciclos de vida para curadoria de objetos digitais. CONSTANTOPOULOS et al - DCC & U: An extended digital curation lifecycle model. SANT'ANA - Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. SANT'ANA - Ciclo de vida dos dados e o papel da ciência da informação.
<b>Ambientes digitais</b>	SAYÃO; SALES - Algumas considerações sobre os repositórios digitais de dados de pesquisa LYNCH - Institutional repositories: essential Infrastructure for scholarship in the digital age. LEITE et al - Boas práticas para a construção de repositórios institucionais de produção científica. SANCHEZ; VIDOTTI; VECHIATO - A contribuição da curadoria digital em repositórios digitais.

Fonte: Elaborado pela autora.

Os textos selecionados mapearam o desenvolvimento conceitual sobre curadoria digital e suas relações com a Ciência da Informação. A leitura e fichamento sobre modelos de ciclo de vida da informação, além do planejamento e estrutura de repositórios como ambientes digitais, subsidiou a escolha das etapas de um modelo teórico para nortear a análise.

#### 4.1 Definição do Caso

Para Coutinho e Chaves (2002), um estudo de caso tem início através da escolha de uma situação, ou um indivíduo específico, que representa algum fenômeno de interesse do pesquisador e na busca pelo entendimento o pesquisador estabelece relações do caso com outros fenômenos e teorias.

O caso tratado neste estudo é o processo de gestão de dados de uma Instituição de Ensino Superior localizada na cidade de Fortaleza, Ceará. A instituição escolhida é uma Universidade privada que está no mercado há mais de 40 anos,

reconhecida pela excelência de seus serviços. A escolha do local se deu pela viabilidade no acesso às informações e pretendo trazer melhorias a um projeto criado pela Instituição.

Inicialmente a análise do processo de gestão seria realizada conjuntamente à análise do planejamento e implantação do repositório institucional, mas durante o processo de coleta de dados verificou-se duas formas de gerenciamento dos dados e diversos tipos de conteúdo institucionais.

Diante disso, optou-se pela verificação do processo de coleta, armazenamento, recuperação e descarte referente aos documentos de Trabalhos de Conclusão de Cursos (TCC) das graduações, além de Dissertações e Teses. O capítulo 5, intitulado “Descrição do Caso”, apresenta uma descrição detalhada dos fatos que justificam a forma de gestão e disponibilização dos materiais pela instituição.

## **4.2 Coleta de dados**

A elaboração dos instrumentos de coleta de dados deve expressar o objetivo do trabalho respeitando as delimitações que foram feitas ao estudo do caso. Diante da necessidade de uma coleta de dados variados foi adotado como conjunto de técnicas a observação participante, entrevista de grupo focal e individual, além de análise documental.

Para compreensão da produção institucional e das decisões que resultaram na atual gestão de dados, foram elaborados dois roteiros de entrevista (Apêndice A e B). O roteiro A abordou o planejamento e ações institucionais para criação de um repositório e o roteiro B teve como foco as categorias de análise definidas pelo CVD (coleta, armazenamento, recuperação e descarte), referente aos TCC, Dissertações e Teses.

A entrevista foi aplicada de forma semiestruturada em um grupo focal de bibliotecários e individualmente com um analista na área de tecnologia da instituição estudada. Para Perovano (2016) a entrevista aplicada de forma semiestruturada facilita a inclusão de novas perguntas que possam surgir durante o processo possibilitando maior aprofundamento dos dados.

O grupo focal é uma técnica de entrevistas com grupos que tem por base a interação. Para Kitzinger (2000) ele reúne de forma colaborativa as informações detalhadas sobre um tópico específico, conduzido por um pesquisador a partir de um grupo de participantes selecionados. Esta técnica foi orientado por um roteiro de entrevistas e permitiu aos participantes a construção das respostas de modo cooperativo. Os roteiros de entrevistas deste estudo foram uma adaptação de perguntas presente nos trabalhos de Sant’Ana (2013, 2016) sobre as etapas do ciclo de vida.

A 1ª fase de entrevistas foi realizada utilizando o roteiro A. O roteiro foi aplicado com um grupo de 6 bibliotecários da instituição que participaram do processo de implantação do software DSpace e demais atividades ligadas ao repositório. O mesmo roteiro também foi aplicado de forma individual com o analista de tecnologia para verificação da infraestrutura tecnológica e o processo de implantação do DSpace na instituição. Na 2ª fase houve a aplicação do roteiro B, desta vez com 3 bibliotecários responsáveis pelo processamento técnico dos documentos institucionais. Os encontros foram registrados em áudio e com anotações complementares em um caderno de pesquisa.

Durante a pesquisa foi possibilitado o acesso à Atas de reunião, comunicações internas, projetos, portarias e políticas institucionais. Os documentos foram organizados e interpretados cronologicamente. A análise documental permitiu verificar as ações realizadas sobre o mapeamento de produção institucional, obrigatoriedade de depósito, fluxos de entrega dos materiais, critérios de aceitação dos documentos, responsabilidades da equipe de trabalho e infraestrutura tecnológica.

A Observação consiste na participação do pesquisador ao vivenciar as atividades cotidianas relacionadas ao objeto de estudo, a fim de identificar os eventos em seu contexto natural (GIVEN, 2008). É uma técnica que auxilia a significação e relacionamento dos dados coletados, geralmente a relação do pesquisador com seu objeto de estudo define o tipo de observação utilizado. Nesta pesquisa foi feita a observação participante, visto que o observador possui envolvimento no contexto social estudado.

Para confrontar as informações coletadas e garantir a credibilidade dos dados foi realizada a triangulação de fontes. Essa técnica permitiu checar as informações coletadas entre os entrevistados e os documentos oficiais da instituição, reduzindo a

possível inferência do pesquisador participante. Para Santos (2014) garante uma análise holística e serve como processo cumulativo de validação, forma de significação e produção de uma imagem mais completa do fenômeno investigado.

### **4.3 Análise de dados**

Nesta fase da pesquisa é apresentado o confronto entre o conceito repositório com a prática identificada no caso, e é onde também tem início a articulação das categorias de estudo com os dados coletados. Para tanto, uma análise de conteúdo categorial foi utilizada como técnica de interpretação dos dados.

A análise de conteúdo constitui uma metodologia de pesquisa usada para descrever e interpretar o conteúdo de toda classe de documentos e textos, bem mais do que uma simples técnica de análise de dados, representando uma abordagem metodológica com características e possibilidades próprias. Essa análise ajuda a reinterpretar as mensagens e a atingir uma compreensão de seus significados além de uma leitura comum (BARDIN, 1977, p. 86).

A análise categorial atua sob o desmembramento do conteúdo em unidades ou categorias, sendo esta delimitação feita de acordo com o material, podendo ser uma palavra, frase ou termo.

Nesse sentido, “a categorização é uma operação de classificação de elementos constitutivos de um conjunto, por diferenciação e, seguidamente, por reagrupamentos segundo o gênero com os critérios previamente definidos” (BARDIN, 1977 p. 119). Assim, as categorias funcionam como classes que agrupam elementos sob um título genérico em razão dos elementos comuns.

As categorias analisadas correspondem às etapas do CVD sendo elas: coleta, armazenamento, recuperação e descarte. A análise de conteúdo descreveu quais ações foram encontradas e quais não foram em cada etapa, permitindo a interpretação relacional entre os documentos, relatos das entrevistas e observação, resultando em uma discussão conceitual e prática.

## 5 LOCUS DO ESTUDO DO CASO

O local escolhido foi uma Instituição de Ensino Superior privada, localizada em Fortaleza, Ceará, que possui mais de 40 anos no mercado, já formou mais de 70 mil profissionais dentre os 40 cursos de Graduação que oferta, e onde atualmente estudam cerca de 28 mil alunos. Além da graduação, possui cursos de Pós-Graduação, Lato (Especialização e MBA) e Stricto Sensu (Mestrado e Doutorado). Dentre os setores desta Universidade, a Biblioteca faz parte das decisões sobre o gerenciamento de documentos acadêmicos, além de ser responsável pela aquisição e manutenção de acervos físicos e digitais para as bibliografias dos cursos.

Com um sistema de gestão de acervos, desenvolvido na própria instituição, a Biblioteca passou a catalogar ao longo do tempo documentos digitais nos mais variados suportes (áudio, vídeo e texto), unindo ao acervo físico os demais conteúdos digitais assinados e produzidos pela instituição. Trabalhos de Conclusão de Curso (graduação), Teses e Dissertações são documentos produzidos pela instituição em suporte digital que aos poucos foram incorporados no catálogo do acervo, transformando o sistema em uma espécie de Repositório.

Dentre estes materiais, as Teses e Dissertações ganham destaque pela obrigatoriedade da legislação educacional, uma vez que, por determinação da Portaria nº 013, de 15 de fevereiro de 2006, a Capes (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) determinou que as instituições que possuam programas de pós-graduação devem manter esses trabalhos de conclusão de seus alunos em arquivos digitais acessíveis ao público via Internet. Assim, os arquivos com as teses ou dissertações são enviados para divulgação na Biblioteca Digital de Teses e Dissertações (BDTD) do Instituto Brasileiro de Informação, Ciência e Tecnologia (IBICT).

De forma mais específica, a BDTD forma um ambiente com três componentes: a BDTD local (interface de busca alimentada pelo sistema da universidade); o sistema coletor de dados (*Harvesting*<sup>1</sup>, também conhecido como

---

<sup>1</sup>O *harvesting* é o processo que se dá por meio do protocolo *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) que, uma vez implementado no sistema, possibilita a coleta direta os metadados dos documentos armazenados nos sistemas de informação (INSTITUTO BRASILEIRO DE INFORMAÇÃO, CIÊNCIA E TECNOLOGIA, 2014)

TEDE); e a BDTD nacional (portal nacional gerenciado pelo IBICT). Em resumo, um banco local alimenta uma base nacional por meio de *Harvesting*.

Com a instauração da portaria da Capes, a Universidade motivou-se a criar seu próprio repositório, através de uma portaria institucional<sup>2</sup> específica, garantindo, assim, a divulgação local das Teses e Dissertações produzidas em seus cursos. Inicialmente, não se considerou necessária a implementação de um software de gerenciamento de repositórios, e a catalogação foi feita no sistema da Biblioteca para que o *harvesting* fosse enviado à BDTD semanalmente.

Em 2012, foi criado um grupo de trabalho pela equipe da biblioteca para levantar demandas necessárias à criação do repositório. O sistema de gerenciamento de acervos, em virtude das Teses e Dissertações, já era tratado como um tipo de Repositório Digital e passou a armazenar outros materiais acadêmicos da Universidade. Nos anos que seguiram algumas ações foram realizadas para implementar o projeto do repositório institucional, que incluíram reuniões da biblioteca com diversos setores da Universidade. Dentre essas ações pode-se ressaltar:

- a) Elaboração, pela Biblioteca, de uma proposta de Política para o Repositório;
- b) Mapeamento do fluxo e estimativa anual da produção anual de documentos acadêmicos;
- c) Verificação de questões contratuais e legais para disponibilização dos materiais;
- d) Parceria com o Centro de Ciências Jurídicas para implementação do projeto piloto de catalogação dos Trabalho de Conclusão de Curso (TCC) do curso de Direito;
- e) Solicitação ao Núcleo de Tecnologia para desenvolvimento de uma aplicação para visualização dos materiais somente para comunidade acadêmica da Universidade (nível de acesso);
- f) Solicitação ao Núcleo de Tecnologia para desenvolvimento de uma aplicação para depósito por autoarquivamento;
- g) Definição de um fluxo para recebimento dos TCC;
- h) Criação de um modelo de autorização de divulgação de TCC para o Curso de Direito.

---

<sup>2</sup> Portaria Institucional nº 24/2006.

A política institucional elaborada se propunha a estabelecer os materiais que seriam incluídos, quem produziu esses materiais, quais os fluxos de entrega dos materiais, critérios de aceitação dos documentos, responsabilidades e atribuições da Biblioteca para o Repositório. O objetivo definido era centralizar o armazenamento da produção científica e técnica da instituição, permitindo sua preservação, disponibilização e recuperação em um único ambiente, viabilizando o acesso às informações e gestão desses materiais. O quadro 3 apresenta a demanda produzida pela instituição.

Quadro 3 - Estimativa inicial da produção anual de conteúdo na Universidade

Tipo de Documento	Média da produção/ano
Trabalho de Conclusão de Curso de Graduação	3.000 (Concludentes em 2012)
Trabalho de Conclusão de Curso de Pós-Graduação <i>lato sensu</i>	112 (Trabalhos entregues 2012)
Dissertações	116 (Média entre 2009-2012)
Teses	
Encontros Científicos da Instituição	2.695 (Artigos produzidos em 2012)
Revistas produzidas pela Instituição	260 (Média de artigos em 2012)
Vídeos Digitais da Instituição	1.100 (Média entre 2009-2012)
<b>Total</b>	<b>7.283</b>

Fonte: Projeto de Criação do Repositório (2013).

O projeto piloto desenvolvido com o curso de Direito foi efetivado em 2015, com o recebimento das remessas de TCC apresentadas ao Curso em 2014.2. O envio foi feito manualmente por meio de comunicação interna e entrega de CD contendo os trabalhos. A catalogação incluía o arquivo digital em PDF<sup>3</sup> do TCC e Termo de Autorização assinado pelo aluno. O sucesso do projeto iniciou a demanda de outros cursos para inclusão dos TCC de seus alunos no sistema da biblioteca.

<sup>3</sup> Portable Document Format (**PDF**) é um formato de arquivo usado para exibir e compartilhar documentos de maneira compatível, independentemente de software, hardware ou sistema operacional. É um padrão aberto mantido pela International Organization for Standardization (ISO). (ADOBE SYSTEMS INCORPORATED, 2018)

Em 2016, o IBICT comunicou a obrigatoriedade na mudança no processo de *harvesting* das Teses e Dissertações em virtude da obsolescência do TEDE. A forma de envio de metadados precisou ser adaptada e o processo só seria possível se as instituições implementassem o DSpace<sup>4</sup>.

O sistema coletor TEDE tornou-se obsoleto em comparação aos outros sistemas em virtude de evoluções tecnológicas desde a sua criação pelo IBICT em 2002. O IBICT optou em construir um processo de migração para um outro sistema baseado em DSpace, considerando ser este um dos softwares mais utilizados mundialmente para construção de Repositórios Digitais. A tecnologia do TEDE2 foi baseada no software DSpace, customizada com o objetivo de armazenar corretamente as informações das Teses e Dissertações, garantindo a disponibilização de dados para a coleta da BDTD.

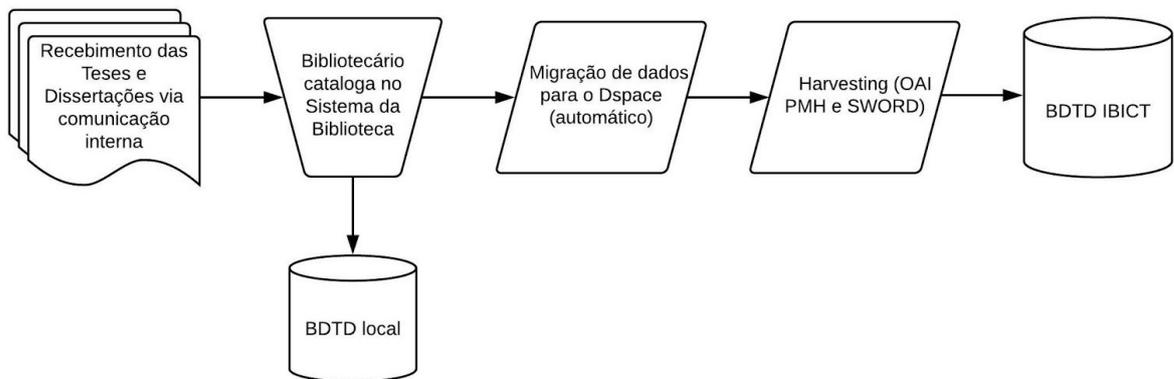
Com o TEDE2 foi implantado o DSpace na Universidade. O processo de configuração, migração e personalização foi implementado entre março e julho de 2016 pela equipe de analistas do Núcleo de Tecnologia e um Bibliotecário. Não houve treinamento sobre o processo de implementação do software, apenas contato com IBICT para esclarecimentos de eventuais dúvidas.

Tendo em vista que os trabalhos são acessados na BDTD Nacional, a visualização da interface institucional do DSpace foi limitada às faixas de IP do Campus da Universidade. Portanto, o software foi implantado com o objetivo de manter o padrão de interoperabilidade com o IBICT. A figura 10 apresenta o fluxo de processamento das Teses e Dissertações com o TEDE2.

---

<sup>4</sup> O DSpace é um software livre que, ao ser adotado pelas organizações, transfere a elas a responsabilidade e os custos com as atividades de arquivamento e publicação da sua produção institucional. (IBICT, 2012).

Figura 10 - Processamento das Teses e Dissertações com o TEDE2



Fonte: Elaborado pela autora.

O esquema de metadados no DSpace é o *Dublin Core*, padrão utilizado nos principais procedimentos que envolvem a apresentação e alimentação do repositório. Como parte da configuração, a biblioteca, junto à equipe do Núcleo de Tecnologia, customizou o esquema padrão conforme a figura 11.

Figura 11 - Esquema de metadados Dublin Core

```

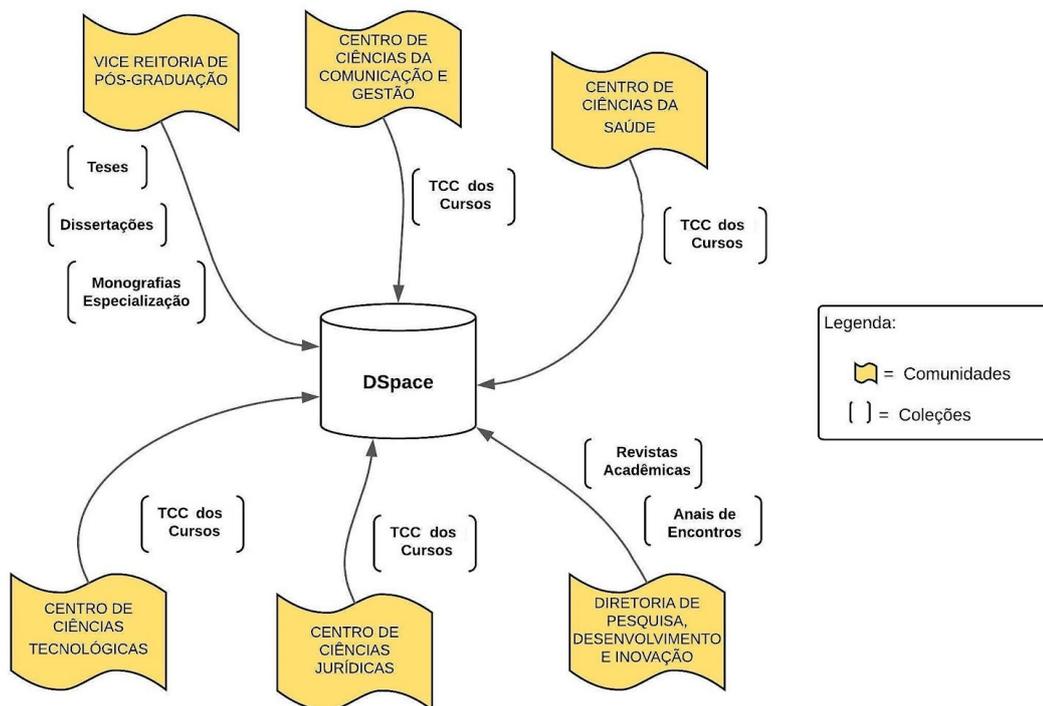
Tipo do documento: <dcvalue element="type" qualifier="none" language="por">Valores : "Tese" ou "Dissertação"
Título: <dcvalue element="title" qualifier="none" language="por">
Título(s) alternativo(s): <dcvalue element="title" qualifier="alternative" language="por">
Autor: <dcvalue element="creator" qualifier="none">
Primeiro orientador: <dcvalue element="contributor" qualifier="advisor1">
Membro da banca: <dcvalue element="contributor" qualifier="referee1">
Membro da banca: <dcvalue element="contributor" qualifier="referee2">
Membro da banca: <dcvalue element="contributor" qualifier="referee3">
Membro da banca: <dcvalue element="contributor" qualifier="referee4">
Membro da banca: <dcvalue element="contributor" qualifier="referee5">
Resumo: <dcvalue element="description" qualifier="resumo" language="por">
Abstract: <dcvalue element="description" qualifier="abstract" language="eng">
Palavras-chave: <dcvalue element="subject" qualifier="none" language="por">
Departamento: <dcvalue element="publisher" qualifier="department" language="por">
Programa: <dcvalue element="publisher" qualifier="program" language="por">
Data de defesa: <dcvalue element="date" qualifier="issued"> formato : YYYY-MM-DD
Data adesão(Cadastro): <dcvalue element="date" qualifier="accessioned"> formato : YYYY-MM-DD
Idioma: FIXO <dcvalue element="language" qualifier="none" language="por">por</dcvalue>
País: FIXO <dcvalue element="publisher" qualifier="country" language="por">Brasil</dcvalue>
Sigla da instituição: FIXO <dcvalue element="publisher" qualifier="initials" language="por">
</dcvalue>
Instituição: FIXO <dcvalue element="publisher" qualifier="none" language="por">
Universidade&#x20;de&#x20;</dcvalue>
Tipo de acesso: FIXO <dcvalue element="rights" qualifier="none" language="por">
Acesso&#x20;Aberto</dcvalue>
  
```

Fonte: Dados da pesquisa (2018)

De acordo com Núcleo de Tecnologia, esquemas de metadados como o *Dublin Core* são flexíveis e permitem a criação de novos campos, caso o modelo padrão não contemple a necessidade de informação. Neste caso ele foi modificado para atender informações essenciais solicitadas pelo IBICT.

Outra questão no DSpace são as comunidades. São elas que determinam a organização de conteúdo dentro do DSpace e são parte da arquitetura de informação do repositório. A Universidade estabeleceu uma estrutura que corresponde ao Organograma e aos tipos de documentos produzidos, tanto acadêmico quanto administrativos. Tais comunidades são representadas em sua hierarquia na figura 12.

Figura 12 - Comunidades e Coleções Acadêmicas criadas pela Universidade no Dspace



Fonte: Elaborado pela autora.

Apesar da criação das comunidades, apenas as Teses e Dissertações continuam a ser migradas do Sistema da Biblioteca para o DSpace em virtude do TEDE2. Atualmente o acervo digital conta com mais de 18.000 documentos.

Desde a implantação do DSpace a Universidade possui a estrutura de um repositório, mas não utiliza o software para essa finalidade. O novo instrumento de avaliação de curso de graduação promete mudar isso. A orientação do MEC é que as universidades disponibilizem os TCC em repositórios institucionais próprios, acessíveis pela internet. Em reunião com a Pró-Reitoria de Graduação, a biblioteca sugeriu a criação de uma portaria regulamentando a obrigatoriedade do envio dos TCC para catalogação.

Observa-se que as ações para criação de um Repositório Institucional na Universidade não foram sequenciais, mas atenderam às necessidades que surgiram em determinados momentos. Tanto o projeto de catalogação de TCC quanto a implementação do TEDE2 contribuíram para o estabelecimento de políticas institucionais, a criação de fluxos informacionais e o desenvolvimento de aplicações.

Os fatos narrados, que corroboram para a compreensão dos eventos e motivos que levaram a implementação do Repositório Institucional deste Caso, demonstram que o planejamento e a execução do referido repositório não foram pautados em ações de Curadoria Digital. Por esta razão, a seguir, realizar-se-á uma análise, com a verificação comparativa entre as ações implementadas com as etapas de Coleta, Armazenamento, Disseminação e Descarte. Nessa análise, serão destacadas ações realizadas e não realizadas, de acordo com o modelo do Ciclo de Vida dos Dados (CVD) de Sant'Ana, apresentando proposições de melhoria como resultado.

## 6 ANÁLISE DE DADOS E DISCUSSÃO DOS RESULTADOS

Desde 2012 a instituição de estudo desta pesquisa reuniu esforços para centralizar em um Repositório toda produção institucional. Os dados coletados permitiram constatar que, para a instituição, tanto o DSpace quanto o Sistema de Gestão de Acervos da biblioteca são considerados como Repositório.

Dentre os documentos analisados, a Política de Divulgação das Produções Científicas da Instituição, considera o repositório institucional como “um banco eletrônico da produção científica e técnica da Instituição, armazenada num formato digital, que permite a busca e a recuperação para seu posterior uso tanto no âmbito nacional quanto internacional”.

Para Leite et al (2012) um repositório institucional constitui em um serviço de informação científica que permite o gerenciamento da produção de uma instituição em uma plataforma digital interoperável. Os autores ainda ressaltam que o ambiente de um repositório deve permitir a reunião, armazenamento, organização, preservação, recuperação e difusão das informações produzidas na instituição.

Em entrevista, a gerente da biblioteca afirmou que considera o sistema de gestão da biblioteca como uma espécie de repositório, pois contém grande parte dos materiais digitais produzidos pela instituição além de proporcionar uma recuperação efetiva das informações.

De encontro a isso, Leite et al (2012) afirmam que sistemas de gestão de biblioteca podem auxiliar na divulgação de conteúdos digitais, porém estes tipos de sistemas tem o objetivo de manter, desenvolver e controlar o acervo como um todo, tanto itens físicos quanto digitais, já os repositórios lidam exclusivamente com a produção em formato digital.

O DSpace, diferente de um sistema de gestão de bibliotecas, é um software que foi desenvolvido para a criação de repositórios digitais e disponibiliza ferramentas para armazenamento, gerenciamento e preservação adequadas para garantir o acesso dos materiais ao longo do tempo. Como resultado seu uso promove a visibilidade acadêmica. Conforme verificamos no capítulo da descrição do caso, o software foi instalado na Universidade para envio de dados das Teses e Dissertações ao IBICT.

Considerando que a estrutura do repositório deste caso não está totalmente definida, e que apenas parte dos materiais está no DSpace, a análise se dará sobre a gestão de TCC, Dissertações e Teses considerando a interface do Sistema de Gestão da Biblioteca e do Catálogo Online. Verificar a prospecção, tratamento e divulgação destes documentos permitirá definir ações para melhoria da gestão, preservação e uso dos materiais produzidos pela instituição.

O ciclo de vida é a Curadoria em ação e o modelo escolhido propõe nortear as fases da Curadoria de modo sequencial garantindo que os elementos essenciais sejam aplicados de forma correta. O Ciclo de Vida dos Dados (CVD) de Sant'Ana (2013) é resultado da análise de inúmeros outros ciclos, propõe uma estrutura voltada para as necessidades de descrição e difusão dos dados sob a perspectiva da CI. Este modelo enfatiza os processos de Curadoria por meio das etapas de **coleta**, **armazenamento**, **recuperação** e **descarte** e cada fase possui ações orientadas pelos objetivos da Preservação, Privacidade, Integração, Qualidade, Direito Autoral e Disseminação.

## 6.1 Coleta

No início do processo de curadoria verificamos a prospecção de dados para alimentação do repositório. A fase de **coleta** no Ciclo de Vida dos Dados (CVD) de Sant'Ana (2013) corresponde a definição das necessidades informacionais auxiliando no escopo, seleção, avaliação, tipologia dos dados, estratégias de localização além de metodologias e ferramentas para captura dos dados. Esta etapa inclui o planejamento para aquisição de dados considerando a viabilidade, escolha e forma de coleta.

Dentre os documentos analisados, a *Política de Divulgação das Produções Científicas da Instituição*, foi considerada como documentação das ações de planejamento do Repositório estudado. A política, elaborada em 2012, tem o objetivo de estabelecer critérios e definições no armazenamento dos materiais acadêmicos produzidos pela instituição. Apresenta os documentos digitais produzidos pela Universidade que integrariam o repositório, dentre eles: os Trabalhos de Conclusão de Curso, Dissertações e Teses. O documento não define claramente uma série de

questões, dentre elas: como é formada a equipe responsável pela gestão do repositório, quais os fluxos e formas de coleta de materiais.

Com relação aos *Trabalhos de Conclusão de Cursos de Graduação* está sendo criada uma portaria institucional com o objetivo de normatizar a divulgação e armazenamento destes documentos para atender as adequações do novo *Instrumento de Avaliação de Cursos do MEC*. A portaria regulará sua forma de depósito no repositório, atribuindo a responsabilidade ao Núcleo de Pesquisa de cada Centro sobre a entrega dos arquivos à biblioteca. A portaria propõe definir ainda o prazo para entrega, o fluxo de envio, forma e formatação dos trabalhos além de propor mecanismos de garantia para submissão prévia do texto em um sistema anti-plágio e revisão ortográfica para garantia da qualidade e direitos do autor.

Com foco nesses objetivos, apesar de não possuir uma política de integridade acadêmica definida, a instituição possui ações para incentivo às boas condutas de pesquisa. A biblioteca oferta cursos sobre Integridade Acadêmica aos docentes incluindo capacitação sobre o uso do *Turnitin*, software antiplágio assinado pela instituição há mais de 10 anos para análise de trabalhos na graduação e pós-graduação.

As Teses e Dissertações já possuem um fluxo definido e portaria institucional. Tem início com o protocolo e entrega do CD pelo autor à Coordenação do Curso, devendo conter todas autorizações devidamente assinadas, dentre elas: a *Autorização para exposição de obra artística*, que garante a cessão gratuita de direitos autorais à instituição para publicação da obra na internet; e a *Declaração de certificação*, documento que garante à biblioteca que o arquivo recebido trata-se de uma cópia fiel da versão final e definitiva, contendo as correções sugeridas pela banca examinadora. A coordenação encaminha o material via Comunicação Interna para a Biblioteca, e esta por sua vez, faz a catalogação e disponibiliza o material nos bancos de dados. Após inserido no sistema da biblioteca, diariamente os dados são migrados de forma automática para o DSpace para o que seja feito o *Havesting*.

Os mecanismos de coleta não foram estabelecidos em nenhum documento institucional. Não há depósito automatizado nem há padrão do meio envio, neste caso os setores envolvidos utilizam-se de Comunicação Interna, *pen drive*, *e-mail* e até dispositivos de armazenamento em nuvem. O formato de envio é unicamente em *pdf*, não tendo sido identificado outros formatos que pudesse considerar a preservação a longo prazo. Sant'Ana (2013) ressalta que na coleta é importante

definir e adotar formatos adequados à preservação para cada tipo documental, como formatos livres (não-proprietários) diferentes para garantia da preservação e acesso.

Quanto a necessidade de envio destes materiais, conforme verificamos as teses e dissertações possuem portaria institucional estabelecendo ratificando a obrigatoriedade de depósito dos trabalhos tendo em vista as exigências da CAPES/MEC, já os Trabalhos de Conclusão de Cursos da Graduação obedecem às orientações do indicador 1.11 do instrumento de Avaliação de Cursos de Graduação - reconhecimento e renovação, que recentemente passou a orientar a disponibilização dos TCC em repositórios institucionais próprios, acessíveis pela internet.

De todos os conteúdos apenas os Trabalhos de Conclusão de Cursos da Pós-Graduação *Lato Sensu* não foram coletados ou tratados. Esses documentos não estão disponibilizados em formato digital, nem no Sistema de Gerenciamento da Biblioteca ou em outro ambiente digital da instituição. Entende-se que diferente das dissertações, teses e TCC da graduação, a falta de coleta dos trabalhos da Pós-graduação *Lato Sensu* pode ser justificada pela ausência de uma legislação que guarda a manutenção e divulgação destes materiais.

Diante disso, verificamos como pontos para melhoria na etapa de **Coleta**:

- a) Criação de uma política para o repositório que defina a comissão responsável e fluxo de entrega dos documentos;
- b) Implementação de um mecanismo de coleta automatizado;
- c) Inversão do Fluxo de catalogação de Teses e Dissertações
- d) Coleta de TCC da Pós-Graduação *Lato sensu*.

Para orientar o processo de funcionamento do repositório e gestão dos dados é essencial estabelecer documentos que indiquem o escopo da coleção, formas de coleta, equipe envolvida e fluxos de trabalho. Para isso a instituição poderá formalizar uma política, baseada nas decisões estabelecidas em portarias institucionais e na política de desenvolvimento de coleções da biblioteca.

Leite et al (2012, p.10) reforçam essa ideia afirmando que a política deve refletir as decisões tomadas ao longo do planejamento do repositório, recomendando que esteja em concordância as demais da instituição. "A política deve abordar os objetivos do repositório, contribuir para a definição do serviço,

determinar a formação da equipe responsável pela implantação e manutenção além do prazo para o depósito de materiais no repositório”.

Sobre a comissão responsável pelo repositório, Leite et al (2012) também ressaltam a importância a composição de uma equipe multidisciplinar capacitada e comprometida. Para os autores ela deverá ser constituída por bibliotecários, analista de sistemas, profissional de comunicação/marketing no sentido de atender ações de planejamento e execução. A necessidade de recursos humanos pode depender do tipo de projeto, da instituição e recursos financeiros disponíveis. Desta forma, o planejamento e a implementação adequada do repositório requerem uma política que considere equipe envolvida na manutenção do repositório.

Sobre o mecanismo de coleta é possível implementar o autoarquivamento, esse processo permite que o próprio autor faça o depósito do documento no repositório. Para Veiga e Macena (2015) repositórios que não utilizam a função de autoarquivamento podem ter dificuldade no processo de coleta de materiais. O DSpace possui workflow para implementação da ferramenta de maneira fácil, porém, para realizar o processo o autor precisa confirmar através de uma licença semelhante à *Autorização para exposição de obra artística*, documento já utilizado pela instituição do qual garante a cessão gratuita de direitos de publicação da obra na internet.

A utilização do autoarquivamento também poderia trazer melhorias ao fluxo de catalogação, pois a inversão do fluxo atual (Biblioteca/DSpace) poderia evitar problemas de migração de dados durante o *harvesting* da BDTD. Assim, o depósito do material poderia ser realizado no DSpace e os metadados migrados ao sistema da biblioteca.

A última questão abordada nesta fase inclui a ausência de coleta dos Trabalhos de Conclusão de Cursos *Lato Sensu*. Acredita-se que a disponibilização de informação, principalmente em repositórios digitais, proporciona o avanço das pesquisas científicas, permitindo e o desenvolvimento da ciência. Desta forma, caso sejam disponibilizados, os TCC *Lato sensu* poderiam ser valiosas fontes de pesquisa, resguardando ainda a memória institucional.

## 6.2 Armazenamento

A segunda fase do ciclo representa o **Armazenamento** considerando ações para permanência de um dado para reutilização. Envolve especificações físicas e lógicas para o registro destes dados, como: processamento, transformação, inserção, modificação, migração e transmissão de dados.

A primeira etapa de tratamento dos documentos é a checagem dos dados do material pelos bibliotecários. Neste momento são validadas informações como título, autor e orientador em comparação à documentação que os acompanha, a Declaração de certificação no caso de teses e dissertações e listagem no caso dos TCC de graduação. Caso seja verificada alguma inconsistência ou ausência de informações necessárias o arquivo é devolvido para a coordenação via CI com indicação do problema.

Para garantir informações básicas à catalogação, o padrão para apresentação dos trabalhos deve manter os elementos pré-textuais como capa, folha de rosto e resumo, mesmo que em casos de TCC no formato de artigos. Os metadados principais que são descritos no sistema da biblioteca e incluem o nome do autor do trabalho, título, resumo, orientador, CDU, curso, palavras-chave. No caso de TCC, foram criadas séries com os nomes dos cursos com o objetivo de facilitar a recuperação dos trabalhos por curso e semestre. A indexação é feita por meio de um banco de termos pré-cadastrados no sistema da biblioteca, não sendo utilizado *thesaurus* externo como auxílio.

Após a descrição e indexação é feito *upload* do *pdf* e gerado automaticamente um registro que identifica o exemplar digital no sistema da biblioteca. Não há migração de suporte na etapa de armazenamento, mantendo a utilização do *pdf* como formato único.

No caso de Teses e Dissertações os metadados inseridos na catalogação são migrados diariamente ao DSpace para realização da coleta (*harvesting*) pela BDTD nacional. Quanto ao local de armazenamento lógico a instituição utiliza a virtualização de servidores para otimização do *backup* e recuperação de dados. Essa técnica executa um ou mais servidores virtuais em um servidor físico, mantém o isolamento e a segurança dos dados promovendo eficiência na gestão e compartilhamento de objetos digitais.

Para garantias de direito autoral nesta fase, a *Autorização para exposição de obra artística* que garante a cessão gratuita de direitos de autoria para publicação também abrange o armazenamento. No caso dos TCC de graduação a cessão para divulgação e armazenamento é estabelecida em cláusula do contrato educacional, assinado no momento da matrícula do aluno.

Considerando que os dados sensíveis geralmente são anonimizados pelos autores para manter a privacidade dos participantes das pesquisas, tendo em vista que eles não são coletados na sua forma bruta, não foi identificado nesta etapa o tratamento de dados sensíveis. Entende-se como dados sensíveis aqueles cuja divulgação pode trazer a discriminação do seu titular, como por exemplo: opção sexual, origem racial ou étnica, dados genéticos ou relacionados a saúde, convicções religiosas, filosóficas ou morais, ou opiniões políticas.

Diante disso verificamos como pontos para melhoria na etapa de **armazenamento**:

- a) Adoção de estratégias para migração de suportes aos formatos abertos que garantem a preservação;
- b) Controle do vocabulário de termos para indexação.

Acredita-se que a migração de suportes pode ser uma estratégia de preservação a ser considerada neste processo. Arellano (2004) revela que ausência de conhecimento sobre preservação digital pode significar a necessidade de garantir cada vez mais a longevidade dos arquivos digitais e as iniciativas de preservação apontam para o uso de padrões de conversão de documentos nos formatos livres. Essa estratégia garante que os arquivos sejam acessados após a obsolescência dos equipamentos e extensões em que foram criados. O autor enfatiza que o sucesso para preservação dos documentos é determinado pela capacidade de uso futuro do objeto digital isso depende especificamente das soluções tecnológicas adotadas.

O processo de armazenamento deve prever estratégias para uma recuperação de qualidade. Lopes (2002) ressalta que um vocabulário controlado de indexação, se propõe a minimizar os problemas de comunicação entre indexadores e usuários. Nesse sentido a adoção de um vocabulário controlado poderá facilitar a indexação de dos conteúdos além de melhorar a qualidade e a consistência dos metadados, enriquecendo semanticamente as descrições, promovendo a interoperabilidade com outros sistemas.

### 6.3 Recuperação

Fase que representa o acesso e uso dos dados. Incluem etapas de estruturação, filtro, representação, refinamento, interatividade, incluindo a visualização dos dados. As ações buscam ampliar a utilização dos materiais, por meio da apresentação de informações (recurso de visualização) e de formas de saída (disponibilização de download). Nesta fase a forma de disponibilização desses materiais, que foram planejados na coleta e tratados no armazenamento, será efetivada.

Por diversos fatores, alguns documentos podem ter perfis de acesso com formas de restrições para a disponibilização e compartilhamento. Sant'Ana (2016) ressalta que a adoção de Sistemas de Gestão de Bibliotecas Digitais torna-se vantajosa, possibilitando definir o acesso aos dados por atribuição de papéis de usuário.

Para a recuperação dos dados armazenados no sistema da biblioteca o usuário pode acessar o catálogo *online* em duas versões: uma disponível em plataforma institucional com autenticação através de matrícula e senha e a outra disponível no site institucional com livre acesso.

Atualmente há restrição para acesso ao TCC de graduação, sendo permitido *download* apenas para a comunidade acadêmica na plataforma institucional. No site é possível verificar apenas o registro do documento, a mensagem de restrição é apresentada como “documento digital possui acesso exclusivo pelo catálogo com autenticação”. Acreditamos que esta condição em breve será revista considerando as orientações de disponibilização dos TCC em repositórios institucionais. Com relação às Teses e Dissertações, conforme já verificamos, são obrigatoriamente disponibilizados de forma livre e gratuita.

Para a interface de apresentação é importante ressaltar que os recursos disponibilizados, seja para download ou visualização, devem refletir os mesmos aspectos atendendo aos requisitos de interação do usuário. Isso significa objetividade e clareza na apresentação das informações sobre os documentos.

O catálogo online possui duas interfaces: uma de busca simples com campo único para pesquisa por autor, título ou assunto, outra de busca avançada onde é

possível fazer cruzamento de metadados utilizando os operadores booleanos. Os resultados são listados em uma página contínua e filtrados por tipo de base, tipo de obra, suporte, assunto e autor, estes recursos possibilitam visualizar relacionamento entre os documentos. Não foi identificado recursos de acessibilidade para pessoas com baixa visão que auxilie a pesquisa ou análise dos resultados de busca.

Para acesso ao download é necessário verificar a interface de detalhes da obra, o sistema faz a liberação do documento mediante o aceite das condições de uso estabelecida pelo termo *para download*, a mensagem enfatiza que a obra destina-se para uso científico/acadêmico e o autor é o titular dos direitos autorais, proibindo a comercialização sem autorização prévia do autor.

Além de navegação intuitiva a interface de recuperação deve manter controle do mecanismo de busca para garantir o padrão dos resultados. O *Framework* utilizado na recuperação é o *Solr*. Como motor de busca ele possui recursos que incluem pesquisa de texto completo, indexação em tempo real, pesquisa por agrupamento, integração de banco de dados e gerenciamento de documentos (*HTML, Word, PDF*). Por ter código aberto, o *Solr* possibilita a modificação da interface de busca e a definição de relevâncias na pesquisa.

Diante disso verificamos como ações para melhoria da etapa de **Recuperação:**

- a) Acesso externo ao texto completo do TCC de graduação;
- b) Avaliar a implementação de recursos de acessibilidade.

Tendo em vista as orientações do instrumento de Avaliação de Cursos de Graduação para a disponibilização dos TCC em repositórios institucionais acessíveis pela internet, acredita-se que a instituição, ao permitir o acesso ao público externo, garantirá a difusão de conhecimento produzido além de manter a finalidade de armazenar, preservar, garantir o livre acesso, disseminando a literatura produzida pela instituição.

Quanto a acessibilidade, entende-se fundamental a análise de interfaces em sistemas digitais, permitindo uma melhor interação para que os repositórios sejam compatíveis com expectativas e necessidades do usuário. Para Freire (2008) uma interface acessível deverá utilizar uma linguagem simples e clara, além de

mecanismos que facilitem a navegação entre as páginas. Além disso deverá obedecer princípios de design como acesso independente de dispositivos, operacionalidade pelo teclado e possibilidade de emissão automática de voz.

Como melhoria, a interface poderá fornecer atalhos por teclado para links importantes além dispor de elementos textuais de forma linear, evitando o uso de colunas tendo em vista que podem atrapalhar o uso de softwares para leitura de tela.

#### **6.4 Descarte**

Última fase do ciclo, o descarte ocorre quando dados não são mais necessários ou são produzidos acima da capacidade de tratamento eficiente. Pode ocorrer por meio da limpeza (exclusão) ou desativação (ocultamento), podendo ser parcial, como um campo específico; ou total, um registro inteiro ou uma categoria de documentos.

Na instituição desta pesquisa o descarte de trabalhos ocorre de forma pontual e moderada, considerando a obrigatoriedade do depósito dos trabalhos, seja na graduação ou pós-graduação. Nos poucos casos que aconteceram, a retirada foi solicitada pelo autor para publicação do conteúdo em formato de livro. Neste caso o processo é validado pela coordenação e formalizado via comunicação interna, tendo em vista que a biblioteca não tem autonomia para o descarte. A documentação e versões do trabalho são arquivadas no servidor interno do setor para fins de registro e informações para o futuro. Para Sant'Ana (2016) o armazenamento de uma cópia de dados excluídos é importante mesmo quando ele não parece ser útil, já que sempre podem surgir novas necessidades de informação.

Conforme foi verificado, não houve exclusão de campos de catalogação. Em casos de mudança de nome, quando há alteração do cadastro no sistema acadêmico da instituição o metadado é automaticamente modificado no campo de descrição da biblioteca, desta forma estes tipos de alterações ou exclusões não são realizados pela biblioteca. Diante disso verificamos como ponto para melhoria para a etapa de descarte:

- a) Elaborar uma política que indique as formas de registro para descarte, incluindo critérios para anonimizar dados.

As coleções devem refletir e se desenvolver de acordo com as necessidades institucionais, baseadas em uma política de desenvolvimento do acervo, considerando o campo de conhecimento, características dos usuários e ambiente. Para Nascimento (2012) importante salientar que, uma vez avaliado, autorizado e disponibilizado pelo auto, um documento não deve ser retirado do RI, a menos que haja situações específicas que obriguem sua retirada. Dentre os exemplos estão o direito de esquecimento, a constatação de ilegalidades ou ainda conflitos de direito autoral.

Mesmo neste contexto, o processo de descarte deverá verificar se trará impacto na disponibilização de metadados, principalmente os que foram produzidos a partir dos dados que serão descartados. Neste ponto Sant'Ana (2016) afirma que a exclusão de um registro pode impactar no relacionamento entre partes distintas causando uma degradação do valor da base como um todo. Portanto, caso seja imprescindível, os processos de descarte ou anonimidade deverão ser rigidamente documentados, mantendo informações detalhadas dos processos de eliminação, possibilitando a verificar relações entre os contextos afetados em futuras consultas.

## 7 CONCLUSÃO

O acesso a dados vem transformando todas as áreas de atuação humana, com especial crescimento nos últimos anos em função do aumento exponencial de alternativas para coleta, armazenamento e recuperação de dados. Este cenário vem sendo definido superando, inclusive, nossa capacidade para lidar com estas novas perspectivas de volume, variedade e velocidade de acesso a dados.

A resposta a essa problemática, além de indicar novas possibilidades, propõe o uso dos dados visando sua manutenção a longo prazo. Nesse sentido, a gestão de dados científicos, também conhecida como curadoria digital, tem por objetivo assegurar que os mesmos possam ser compreendidos e interpretados ao longo do tempo para a promoção do conhecimento. Trata-se de uma temática que envolve contextos e aplicações das áreas da Computação, Biblioteconomia, Administração e principalmente da Ciência da Informação.

Para entender como se dá o processo de curadoria digital em um repositório institucional, este estudo identificou e avaliou a curadoria de documentos acadêmicos produzidos por uma Instituição de Ensino Superior, sob a ótica do ciclo de vida dos dados (CVD-CI) de Sant'Ana. O *locus* do Estudo de Caso foi uma universidade privada localizada no município de Fortaleza, Ceará e o *corpus* foi definido pelos Trabalhos de Conclusão de Curso de Graduação (TCC), Dissertações e Teses além dos ambientes digitais da instituição, como o sistema de gestão da biblioteca e o DSpace.

Nesse sentido, verificou-se na instituição o processo de gestão dos documentos pertencentes ao corpus da pesquisa. Foram analisadas também possíveis ações das etapas de Coleta, Armazenamento, Recuperação e Descarte, do ciclo de vida de dados (CVD-CI) de Sant'Ana. Diante das análises foi possível definir as seguintes propostas de melhoria:

- a) A implantação de um mecanismo de coleta automatizado auxiliará na padronização de envio garantindo a privacidade e organização dos dados, possivelmente melhorando o processo de coleta.
- b) A adoção um sistema de preservação digital de código aberto, como o Archivematica, permitirá a transformação de suportes dos arquivos em

formatos abertos, garantindo a preservação. Técnicas de transformação de dados deverão ser aplicadas na fase de armazenamento.

c) No tratamento, o uso de um vocabulário controlado temático possibilitará o controle de termos, mantendo a uniformidade bibliográfica e servindo como base para a descrição de outros documentos.

d) Na recuperação, a interface poderá implementar recursos de acessibilidade para inclusão e promoção de acesso às pessoas com deficiência auditiva ou individual.

Além dos pontos citados, foram identificados dois processos de tratamento referente as teses e dissertações. Esses documentos são incorporados ao catálogo, através do sistema da biblioteca e ao DSpace para coleta de metadados pela BDTD nacional. Nesse sentido a utilização do autoarquivamento poderia otimizar esse processo, pois o depósito do material poderia ser realizado no DSpace, os metadados poderão ser validado pelo bibliotecário, logo após seriam disponibilizados na BDTD e automaticamente migrados ao sistema da biblioteca.

O autoarquivamento consiste do depósito de conteúdos pelo próprio autor. Esse recurso agiliza a disseminação da informação, uma vez que esta não dependerá do administrador do repositório para ser coletada, o que maximiza o impacto da pesquisa, agilidade do processo e promove a visibilidade da produção acadêmica.

A estrutura e o planejamento dos ambientes digitais também fizeram parte do *corpus* e foram averiguadas para compreensão dos meios para armazenamento dos documentos. Em síntese, podemos concluir que:

e) Os dados apontam que a implementação do Repositório Institucional não foram inicialmente pautados em um ciclo de vida de Curadoria Digital.

f) O sistema de gestão da biblioteca é utilizado como uma forma de Repositório, promovendo o aceso e armazenamento dos documentos digitais.

Os ambientes digitais de armazenamento viabilizam a criação, tratamento, transmissão e acesso de dados. Desta forma, os repositórios institucionais devem ser idealizados como parte essencial na gestão e preservação. Para isso, o DSpace é apresentado como alternativa, ele é um software aberto que possui uma

arquitetura simples e eficiente que tem por objetivo viabilizar o acesso e a preservação das produções acadêmicas.

Acredita-se que a utilização do DSpace na instituição poderá auxiliar no desenvolvimento de ações de gestão e preservação, fornecendo as ferramentas para uma coleta automatizada, garantindo maior controle e privacidade na obtenção dos dados, auxiliando nas ações de salvaguarda e possibilitando a customização de sua interface para requisitos de acessibilidade.

Além dos pontos tratados, diante do contexto documental propõe-se a criação de uma política de conteúdos digitais considerando os tipos de materiais que podem ser depositados. Esta política deverá incluir a equipe responsável, forma e gestão (captura, descrição, preservação) de diferentes materiais no repositório, apresentando um fluxo de gestão dos documentos de acordo com cada a categoria (Teses e Dissertações/TCC de graduação). Além disso, na política poderá ser definida as diretrizes para o descarte de dados, ressaltando os métodos de registro desse processo.

A pesquisa cujos resultados apresentados, ao mesmo tempo em que atinge seus objetivos permite a elaboração de propostas de melhoria para a gestão, manutenção e preservação dos documentos acadêmicos digitais de um repositório institucional. Diante disso fica claro o alcance dos objetivos propostos, ratificando que a curadoria poderá auxiliar na gestão e preservação de documentos em repositórios institucionais.

Neste estudo a preservação digital é apresentada como um dos objetivos para a implementação de repositórios e considerada como um dos grandes desafios diante do excesso da produção de dados. Neste contexto, como elemento da curadoria, a preservação digital é reafirmada como fundamental para a garantia do acesso, manutenção do conhecimento e a recuperação da informação científica.

Destaca-se, ainda, o papel da Ciência da Informação (CI) no estudo e na proposta de soluções para o processo de Curadoria, oferecendo um arcabouço teórico que contribui no desenvolvimento de estratégias diante da Era dos dados.

A literatura sobre o tema tem despontado fortemente nos últimos dois anos, com a contribuição de autores como Luiz Sayão e Sandra Siebra. Porém as publicações em língua inglesa figuram como fonte referencial, o que muitas vezes dificultou o acesso a livros que apresentavam metodologias de aplicação da curadoria. A percepção de práxis da curadoria digital foi a maior dificuldade sentida

durante o processo de investigação, além da variedade terminológica abordada conceitualmente sobre a temática. Em alguns trabalhos as ações de curadoria foram assemelhadas a gestão de dados e nem sempre relacionadas a um ciclo de vida de forma clara. Ressalta-se neste ponto a ausência de estudos que apresentam análise de ações implementadas de curadoria em repositórios institucionais no Brasil.

Considera-se que esta pesquisa abre várias vertentes para estudos sobre metodologias de uso da curadoria em bibliotecas, e outros ambientes digitais, a fim de elaborar modelos customizados para cuidado da informação digital. Sugere também um estudo conceitual mais aprofundado para delimitar as os tipos de curadoria, suas relações, ressignificações e contextos de atuação.

Acredita-se que este estudo contribui para o desenvolvimento da área de Ciência da Informação, quanto ao desenvolvimento das temáticas de curadoria digital, preservação digital e de repositórios institucionais por meio da compreensão da análise do ciclo de vida de Sant'Ana.

## REFERÊNCIAS

- ABREU, Janice Pereira; VIDOTTI, Silvana Aparecida Borsetti Gregório. Curadoria Digital nos Contexto dos Repositórios Digitais. In: ENCONTRO INTERNACIONAL DE DADOS, TECNOLOGIA E INFORMAÇÃO, 2., 2016. Marília. **Anais...** Marília: UNESP, 2016.
- ARELLANO, Miguel Angel. Preservação de documentos digitais. **Ci. Inf.**, Brasília, v. 33, n. 2, p. 15-27, maio/ago. 2004.
- BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, 1977.
- BEAGRIE, N. Digital Curation for Science, Digital Libraries, and Individuals. **International Journal of Digital Curation**, v. 1, n. 1, p. 3-16, nov. 2006.
- BEAGRIE, Neil. The Digital Curation Centre. **Learned Publishing**, v. 17, n. 1, p. 7-9, 2004.
- BRASIL. Ministério da Educação. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Portaria nº 013, de 15 de fevereiro de 2006**. Brasília, 2006. Disponível em: <[http://repositorio.unb.br/documentos/Portaria\\_N13\\_CAPES.pdf](http://repositorio.unb.br/documentos/Portaria_N13_CAPES.pdf)>. Acesso em: 07 out. 2017.
- BRASIL. Ministério da Educação. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Portaria nº 013, de 15 de fevereiro de 2006**. Institui a divulgação digital das teses e dissertações produzidas pelos programas de doutorado e mestrado reconhecidos. Brasília, 2006. Disponível em: <<https://goo.gl/6NWTrC>>. Acesso em: 18 abr. 20018.
- BRASIL. Ministério da Educação. **Instrumento de Avaliação de cursos de graduação Presencial e a distância: reconhecimento e renovação de reconhecimento**. Brasília: Inep/MEC, 2017. Disponível em: <<https://goo.gl/53DH1Q>>. Acesso em: 18 abr. 2018.
- BUCKLAND, M. K. Information as thing. **Journal of the American Society for Information Science(JASIS)**, v.45, n.5, p.351-360, 1991.
- CAPITÃO, Cláudio Garcia; VILLEMOR-AMARAL, Anna Elisa de. Pesquisa Qualitativa com estudo de caso. In: BAPTISTA, Makilim Nunes; CAMPOS, Dinael de. **Metodologias Pesquisa em Ciências: Análise Quantitativa e Qualitativa**. 2. ed. Rio de Janeiro: LTC, 2016. cap 21.
- CASTILHO, Carlos Albano Volkmer de. **O papel da curadoria na promoção do fluxo de notícias em espaços informativos voltados para a produção de conhecimento**. 2015. 155 f. Tese (Doutorado em Engenharia e Gestão do Conhecimento)- Universidade Federal de Santa Catarina, Florianópolis, 2015. Disponível em: <<https://goo.gl/UEdULk>>. Acesso em: 13 abr. 2016.
- CONSTANTOPOULOS, P. et al. DCC & U: An extended digital curation lifecycle model. **The International Journal of Digital Curation**, v. 4, n.1, 2009.

CONSTANTOPOULOS, P.; DALLAS C. Aspectos de uma agenda de curadoria digital para herança cultural. **IEEE International Conference on Distributed Human-Systems de Máquinas**. Atenas, Grécia: IEEE, 2008.

COUTINHO, Clara Pereira; CHAVES, José Henrique. O estudo de caso na investigação em Tecnologia Educativa em Portugal. **Revista Portuguesa de Educação**, v.15, n. 1, p. 221-24, 2002.

CROW, R. **The Case for Institutional Repositories**: a SPARC Position Paper. Washington, DC: Scholarly Publishing & Academic Resources Coalition, 2002. Disponível em: <<http://www.arl.org/sparc/IR/ir.html>> Acesso em: 15 fev. 2017.

DATA OBSERVATION NETWORK FOR EARTH. **Data Life Cycle**. 2016. Disponível em: <<https://www.dataone.org/data-life-cycle>>. Acesso em: 01 jul. 2017.

DIGITAL CURATOR CENTER. **What is digital curation?** Disponível em: <<http://www.dcc.ac.uk/digital-curation/what-digital-curation>>. Acesso em: 13 abr. 2016.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Repositório Acesso Livre à Informação Científica da Embrapa**. 2017. Disponível em: <<https://www.alice.cnptia.embrapa.br/>>. Acesso em: 01 jul. 2018.

FREIRE, André Pimenta. **Acessibilidade no desenvolvimento de sistemas web**: um estudo sobre o cenário brasileiro. 2008. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, University of São Paulo, São Carlos, 2008.

FREITAS, Cristiana Vieira de. O futuro é hoje: perfis e competências dos profissionais da informação para a curadoria digital. In: PEREIRA, A. A.; RIBEIRO, M.; MEIRELES, P.; PENTEADO, P. **Encontro Curadoria Digital, Estratégias e experiências**: atas. Lisboa: Universidade Nova de Lisboa, 2017. p. 28-39. Disponível em: <<https://goo.gl/39g6qT>>. Acesso em: 13 dez. 2017.

GANDEL, Paul B.; KATZ, Richard N.; METROS, Susan E. Weariness of flesh reflections life mind era abundance . **EDUCAUSE Review**, v. 39, n. 2, p. 40-51, mar./ abril 2004. Disponível em: <<https://goo.gl/ovhrz8>>. Acesso e m: 01 jun. 2017.

GANTZ, J.; REINSEL, D. **The digital universe in 2020**: big data, bigger digital shadows and biggest growth in th far East. 2013. Disponível em: <<https://goo.gl/DfwN79>>. Acesso em: 01 jun 2017.

GARVEY, W. D.; GRIFFITH, B. G. Communication and information processing within scientific disiciplines: emprirical findings. In: GARVEY, W. D. **Communication**: the essence of science. Oxford: Pergamon Press, 1979.

GIARETTA, David. **DCC approach to digital curation – [draft]**. Reino Unido: DCC, 2004. Disponível em: <<https://goo.gl/PtjY6X>>. Acesso em: 02 abr 2017.

GIVEN, L. M. **The Sage encyclopedia of qualitative research methods**. California: SAGE Publications, 2008.

HIGGINS, S. Digital curation: the emergence of a new discipline. **The International Journal of Digital Curation**, v. 6, n. 2, 2008.

HODGES, Gail M. Best practices for digital archiving : an information life cycle approach. **D-Lib Magazine**, v. 6, n. 1, 2000. Disponível em: <<https://goo.gl/VQCKb3>>. Acesso em: 12 out. 2017.

INSTITUTO BRASILEIRO DE INFORMAÇÃO, CIÊNCIA E TECNOLOGIA. **Tede/Tede2**. 2014. Disponível em: <<http://wiki.ibict.br/index.php/TEDE/TEDE2>>. Acesso em: 15 abr. 2018.

KIM, Jeonghyun; WARGA, Edward; MOEN, William. Competencies required for Digital Curation: an analysis of job advertisements. **International Journal of Digital Curation**, v. 8, n. 1, p. 66–83, 2013. Disponível em: <[ijdc.net/index.php/ijdc/article/view/8.1.66](http://ijdc.net/index.php/ijdc/article/view/8.1.66)>. Acesso em: 21 out. 2017.

KITZINGER, J. Focus groups with users and providers of health care. In: POPE, C.; MAYS, N. (Org.). **Qualitative research in health care**. 2. ed. London: BMJ Books, 2000.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. **Fundamentos de Metodologia Científica**. 7. ed. São Paulo: Atlas, 2010.

LAWRENCE, S. Online or invisible? **Nature**, v. 411, n. 6837, p. 521, 2001.

LEE, Christopher A.; TIBBO, Hellen. Where's the Archivist in digital curation? exploring the possibilities through a matrix of knowledge and skills. **Archivaria**, v. 72, p. 123-168, 2011. Disponível em: <<https://ils.unc.edu/callee/p123-lee.pdf>>. Acesso em: 21 out. 2017.

LEITE, Fernando César L.; COSTA, Suely. Repositórios institucionais como ferramenta de gestão do conhecimento científico no ambiente acadêmico. **Perspectiva em Ciência da Informação**, Belo Horizonte, v. 11, n. 2, p. 206-219, maio/ago, 2006. Disponível em: <<https://goo.gl/uM4ESE>>. Acesso em: 12 dez. 2016.

LEITE, Fernando César Lima et al. **Boas práticas para a construção de repositórios institucionais de produção científica**. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2012.

LONGAIR, S. Cultures of Curating: The Limits of Authority. **Museum history journal**, v. 8, n. 1, p. 1-7, Jan. 2015. Disponível em: <<https://goo.gl/UGfeFa>>. Acesso em: 25 jul. 2017.

LOPES, Ilza Leite. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. **Ci. Inf.**, Brasília, v. 31, n. 1, p. 41-52, jan./abr. 2002.

LYNCH, Clifford A. Institutional repositories: essential Infrastructure for scholarship in the digital age. **Libraries and the Academy**, v.3, n. 2, p.327-366, 2003.

MACHADO, Kettuly Costa et al. Avaliação de Desempenho na Gestão da Informação Digital: contribuições da literatura científica. **Inf. & Soc.:Est.**, João Pessoa, v. 26, n.2, p. 155-172, maio/ago. 2016

MACHADO, Kettuly Costa; VIANNA, William Barbosa. Curadoria Digital e Ciência da Informação: correlações conceituais relevantes para apropriação da informação. In: Encontro Nacional de Pesquisa em Ciência da Informação, 17., 2016, Salvador. **Anais eletrônicos...** Salvador: UFBA, 2012. Disponível em:<<https://goo.gl/an7KNP>>. Acesso em: 15 out. 2017.

MUELLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, v. 35, n. 2, ago. 2006. Disponível em: <<http://www.scielo.br/pdf/ci/v35n2/a04v35n2.pdf>>. Acesso em: 10 jun. 2017.

NASCIMENTO, Aline Vieira; SANTOS, Ana Cristina Gomes. Desenvolvimento de coleções em bibliotecas universitárias: o caso dos repositórios institucionais. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 17., Gramado, RS. **Anais...** Gramado, RS: UFRGS, 2012.

NATIONAL RESEARCH COUNCIL. Preparing the Workforce for Digital Curation. 2015. Disponível em: <[http://www.nap.edu/download.php?record\\_id=18590#](http://www.nap.edu/download.php?record_id=18590#)>. Acesso em: 07 out. 2017.

NEVES, J. L. Pesquisa qualitativa: características, usos e possibilidades. **Caderno de pesquisas em administração**, São Paulo, v.1, n.3, 1996. Disponível em:<[https://goo.gl/dVocDe](http://https://goo.gl/dVocDe)>. Acesso em: 27 abr. 2017.

PENNOCK, Maureen. Digital curation and the management of digital library cultural heritage resources. **Local Studies Librarian**, v. 25, n. 2, p. 3-7, 2006. Disponível em: <<https://goo.gl/a3FV8o>>. Acesso em: 23 jan. 2017.

PEROVANO, Dalton Gean. **Manual de metodologia da pesquisa científica**. Curitiba: intersaberes, 2016.

RUMSEY, S. The purpose of institutional repositories in UK higher education: a repository manager's view. **International Journal of Information Management**, v. 26, n. 3, p. 181-186, 2006. Disponível em: <http://eprints.lse.ac.uk/archive/00000800> Acesso em: 20 fev. 2017.

SANCHEZ, Fernanda Alves; VIDOTTI, Silvana Aparecida Borsetti Gregório; VECHIATO, Fernando Luiz. A contribuição da curadoria digital em repositórios digitais. **Rev. Inf. na Soc. Contemp.**, Natal, n. esp., p. 1-17, 2017.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Inf. Inf.**, Londrina, v. 21, n. 2, p. 116-142, maio/ago., 2016.

SANT'ANA, R. C. S. G. A. Ciclo de vida dos dados e o papel da ciência da informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais Eletrônicos...** Florianópolis: UFSC, 2003. Disponível em: <<https://goo.gl/yr3wLi>>. Acesso em: 06 abr. 2018.

SANTARÉM SEGUNDO, J. E. **Representação iterativa**: um modelo para repositórios digitais. 2010. 224 f. Tese (Doutorado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2010. Disponível em: <<https://goo.gl/dD4VsL>>. Acesso em: 26 jul. 2017.

SANTOS, Thayse Natália Cantanhede. **Curadoria digital**: o conceito no período de 2000 a 2013. 2014. 166 f. Dissertação (Mestrado em Ciência da Informação)- Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2014. Disponível em: <<https://goo.gl/hf35E8>>. Acesso em: 13 abr. 2016.

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Inf. & Soc.:** Estudos, João Pessoa, v. 22, n.3. p.179-191, set./dez. 2012. Disponível em: <<https://goo.gl/i5kUDj>>. Acesso em: 25 jul. 2017.

SAYÃO, Luis Fernando (Org.) et al. **Implantação e gestão de repositórios institucionais**: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009.

SAYÃO, Luis Fernando; SALES, Luana Farias. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Inf. Inf.**, Londrina, v. 21, n. 2, p. 90–115, maio/ago. 2016.

SAYÃO, Luis Fernando; SALES, Luana Farias. Curadoria Digital: um novo patamar para preservação de dados digitais de pesquisa. **Inf. & Soc.:Est.**, João Pessoa, v. 22, n. 3, p. 179-191, set./dez. 2012.

SIEBRA, S. A.; BORBA, V. R.; MIRANDA, M. J. K. F. O. Curadoria digital: um termo interdisciplinar. **Revista do Encontro Nacional de Pesquisa em Ciência da Informação**, v. 17, 2016.

SILVA JÚNIOR, L. P.; BORGES, M. M. Preservação digital no Repositório Científico de Acesso Aberto de Portugal. **Rev Eletron de Comun Inf Inov Saúde**, v. 8, n. 4, p. 567-574, out./dez. 2014. Disponível em: <<https://goo.gl/zwNTrG>>. Acesso em: 04 ago. 2017.

SILVA, Faysa de Maria Oliveira; SIEBRA, Sandra de Albuquerque. Análise de modelos de ciclos de vida para curadoria de objetos digitais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017, marília. **anais eletrônicos...** Marília: Unesp, 2017. Disponível em: <<https://goo.gl/CZsduX>>. Acesso em: 07 out. 2017.

SILVA, Edilma Maria Santos; DORNELAS, Jairo Simião. A tecnologia big data no processo decisório: potencialidades e possibilidades. In: INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS & TECHNOLOGY MANAGEMENT –

CONTECSI, 14., São Paulo, 2017. **Proceedings**... São Paulo: Universidade de São Paulo, 2017. Disponível em: <<https://goo.gl/Pcrvhy>>. Acesso em 10 maio 2018.

TOMAÉL, Maria Inês; SILVA, Terezinha Elisabeth da Silva. Repositórios Institucionais: diretrizes para políticas de informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8. , 2007, Salvador. **Anais eletrônicos**... Salvador: UFBA, 2007. Disponível em: <<https://goo.gl/UnchR2>>. Acesso em: 07 out. 2017

UBEROI, Aayushi; LAMBERT, Paul. Isolation and quantification of MmuPV1 virions from papillomas. **Protocols.io**, out. 2017. Disponível em: <<https://goo.gl/tbn1mA>>. Acesso em: 07 out. 2017.

UK DATA ARCHIVE. **Research data lifecycle**. 2010. Disponível em: <<https://www.ukdataservice.ac.uk/manage-data/lifecycle>>. Acesso em: 07 out. 2017.

VEIGA, V.; MACENA, L. G. O Autoarquivamento Nos Repositórios Institucionais Brasileiros: Um Estudo Exploratório. **PontodeAcesso**, Salvador, v.9, n.3, p. 35-47, dez. 2015

YIN, R. K. **Estudo de caso: planejamento e métodos**. 3. ed. Porto Alegre: Bookman, 2005.

ZAPPYLAB. **Open Access Repository Of Science Methods**. 2017. Disponível em: <<https://www.protocols.io/>>. Acesso em: 01 jul. 2017.

## APÊNDICE A – ROTEIRO DE PERGUNTAS SOBRE A CRIAÇÃO E PLANEJAMENTO DO REPOSITÓRIO



**Universidade Federal do Ceará  
Centro de Humanidades  
Programa de Pós-Graduação em Ciência da Informação (PPGCI/UFC)**

### **Roteiro da Entrevista com os profissionais da instituição para coleta de dados**

Prezado(a) Sr. (a)

Estamos realizando uma pesquisa visando a coleta de dados para elaboração da dissertação intitulada “A curadoria em um repositório institucional: uma análise sob a ótica do ciclo de vida dos dados (CVD-CI) de Sant’Ana”, requisito para a conclusão do mestrado em Ciência da Informação, da Universidade Federal do Ceará. O objetivo principal desta entrevista é identificar etapas de curadoria digital nos objetos e ambientes digitais da instituição. Neste sentido, gostaríamos de contar com a sua valiosa colaboração, respondendo a estas perguntas. Garantimos que as informações aqui fornecidas serão de uso exclusivo para o desenvolvimento desta pesquisa.

Desde já agradecemos sua atenção.

Atenciosamente,

Morgana Ramos Albuquerque

### **PERGUNTAS:**

1. O que a instituição entende como sendo um repositório?
2. Quando o repositório foi criado?
3. Que necessidade para sua a criação?
4. Qual o objetivo do repositório?
5. Como foi feito o planejamento de ações para implementação do repositório?
6. Utilizam algum tipo de software para repositórios? se sim, qual foi a razão da escolha?
7. Que equipamentos ou materiais são utilizados para o funcionamento do repositório?
8. Foram mensurados custos (iniciais, de implementação e de longo prazo) para o repositório?
9. Foi realizada durante o planejamento alguma avaliação das necessidades da comunidade acadêmica?

10. Que profissionais fazem parte da equipe de trabalho do repositório?  
Como é feita a distribuição das funções?
11. A equipe foi treinada ou preparada para atuar no repositório?
12. Quem é o administrador do repositório?
13. Qual o padrão de metadados utilizado?
14. Quais comunidades foram estabelecidas no repositório? e porque?
15. Que tipo de materiais digitais são produzidos pela univor?
16. Que materiais digitais fazem parte do repositório? e porque?
17. Existe política institucional que define as ações do repositório  
(responsabilidades, direitos e deveres)?
18. Como foi construído o esquema de metadados? qual a relevância dos  
campos estabelecidos?
19. Para implantação foi feito algum projeto piloto? se sim, como foi?

## **APÊNDICE B – ROTEIRO DE PERGUNTAS SOBRE AS CATEGORIAS DO CICLO DE VIDA DOS DADOS (CVD)**



**Universidade Federal do Ceará  
Centro de Humanidades  
Programa de Pós-Graduação em Ciência da Informação (PPGCI/UFC)**

### **Roteiro da Entrevista com os profissionais da instituição para coleta de dados**

Prezado(a) Sr. (a)

Estamos realizando uma pesquisa visando a coleta de dados para elaboração da dissertação intitulada “a curadoria em um repositório institucional: uma análise sob a ótica do ciclo de vida dos dados (CVD-CI) de Sant’Ana”, requisito para a conclusão do mestrado em Ciência da Informação, da Universidade Federal do Ceará. O objetivo principal desta entrevista é identificar etapas de curadoria digital nos objetos e ambientes digitais da instituição. Neste sentido, gostaríamos de contar com a sua valiosa colaboração, respondendo a estas perguntas. Garantimos que as informações aqui fornecidas serão de uso exclusivo para o desenvolvimento desta pesquisa.

Desde já agradecemos sua atenção.

Atenciosamente,

Morgana Ramos Albuquerque

### **COLETA**

→ fase de obtenção dos dados. Consiste na seleção de avaliação do que será inserido no repositório. Identifica as fontes dos dados que faz parte do escopo do repositório

Perguntas:

1. Como é feito o recebimento de documentos digitais que compõe o repositório?
2. As formas de recebimento dos documento garantem a privacidade de informações?
3. Fica claro a procedência e integridade dos documentos?

4. ex. documento em formato aberto (word) podem ser alterados.
5. Obs. neste caso a integração garantirá a qualidade e o direito autoral.
6. Os documentos recebidos garantem os dados necessários para a descrição?

## ARMAZENAMENTO

→ envolve ações para persistência dos dados focando em seu uso posterior (reutilização). Envolve especificações físicas e lógicas para o registro dos dados.  
Catalogação/ descrição/ tratamento

Perguntas:

1. Que tipo de metadados são descritos?
2. Quem pode acessar os documentos inseridos no SGBD/repositório?
3. Qual formato os objetos são tratados? Há transformação de suporte?
4. Quais as formas de acesso aos documentos? ex. SGBD
5. Considerando a parte lógica, onde os arquivos são armazenados? ex. Servidor da instituição
6. Para armazenamento dos documentos é concedido permissão pelo autor? ex. Termo de cessão
7. Quais são as formas de indexação e categorização dos documentos? Utiliza vocabulário controlado?
8. Que ações que promove a preservação dos documentos são feitas? ex. Migração para formatos abertos

## RECUPERAÇÃO

→ Torna os dados disponíveis para acesso e uso. Preocupa-se com meios que ampliem os níveis de utilização destes dados. Exemplo: cópia/download e interface de visualização dos dados.

Perguntas:

1. Existem documentos com dados sensíveis? Como são apresentados esses dados no sistema?
2. Na interface de apresentação há relacionamento entre os documentos? ex. Categorias por curso e documentos do mesmo tema
3. Na interface, como é a forma de visualização dos documentos? Há disponibilidade de download? Há recursos de acessibilidade?

4. Na interface do sistema fica claro quem tem permissão de acesso e uso do documento? ex. Termo de compromisso.
5. Que informações são apresentadas quando há dados disponíveis mas restritos? ex. Informação no site.
6. Há controle sobre o mecanismo de busca (solr) para manter o padrão dos resultados?

## DESCARTE

→ Dados que não necessários ou que devem ser excluídos da base, ou ainda que estão acima da capacidade de tratamento da base. Consiste na Limpeza ou desativação. Pode ser parcial (vertical) ou total (horizontal).

1. Há exclusão de documentos? Se sim por quais motivos?
2. Há exclusão de campos de catalogação? Neste caso é avaliado se a eliminação dos metadados não prejudica a integridade dos documentos? Essa exclusão interferiu nos dados de autoria?
3. A biblioteca tem autonomia para excluir os documentos ou metadados?
4. Já foi solicitada a retirada de algum documento em virtude do direito de esquecimento? ex. Mudança de nome.
5. Os processos de descarte são registrados? Se sim, como?