



## Contribuições para a gestão de dados científicos: análise comparativa entre modelos de ciclo de vida dos dados

*Contributions to the management of scientific data: comparative analysis among life cycle models*

Débora Gomes de Araújo \*

Marco Antonio Almeida Llarena \*\*

Sandra de Albuquerque Siebra \*\*\*

Guilherme Ataíde Dias \*\*\*\*

### RESUMO

O objetivo foi analisar as intersecções entre os elementos dos modelos de ciclos de vida dos dados das iniciativas do DCC, DataONE e o CVD-CI. Trata-se de uma pesquisa descritiva, qualitativa e bibliográfica. Verificou-se que há correspondências entre etapas (nem sempre de um para um) dos ciclos de vida dos dados analisados. Foi possível constatar que o CVD-CI condensa várias atividades em uma única etapa, o que pode dificultar a sua aplicabilidade. De uma maneira geral, os modelos propostos ainda carecem de maior detalhamento para poderem ser aplicados diretamente por pesquisadores/curadores.

**Palavras-chave:** Ciclo de Vida dos Dados;

### ABSTRACT

The objective was to analyze the intersections among the data life cycle model elements of the DCC, DataONE and CVD-CI initiatives. It is a descriptive, qualitative and bibliographical research. It was verified that there are correspondences between stages (not always one-to-one) of the analyzed data life cycles. It was possible to verify that the CVD-CI condenses several activities in a single step, which can hinder its applicability. In general, the proposed models still need to be further detailed so that they can be directly applied by researchers/curators.

**Keywords:** Data Life Cycle; Digital Curation; Scientific Data; Information Technology.

\* Mestre em Gestão em Organizações Aprendentes pela Universidade Federal da Paraíba -UFPB. Endereço: Programa de Pós-Graduação em Ciência da Informação da Universidade Federal da Paraíba, S/N, Cidade Universitária, CEP 58051-900, João Pessoa, PB. Telefone: (83) 3216-7483. E-mail: debora.g.de.araujo@gmail.com

\*\* Doutorado em Educação pela Universidad del Norte- UNINORTE - PY. Endereço: Programa de Pós-Graduação em Ciência da Informação da Universidade Federal da Paraíba, S/N, Cidade Universitária, CEP 58051-900, João Pessoa, PB. Telefone: (83) 3216-7483. E-mail: llarenaifpb@gmail.com

\*\*\* Doutora em Ciência da Computação pela Universidade Federal de Pernambuco. Professora do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco (UFPE). Endereço: Avenida da Arquitetura, S/N, Cidade Universitária, CEP 50740-550, Recife, PE. Telefone: (81) 2126-8304. E-mail: sandra.siebra@gmail.com

\*\*\*\* Doutor em Ciência da Informação pela Universidade de São Paulo (USP). Professor do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal da Paraíba (UFPB). Endereço: Departamento de Ciência da Informação, S/N, Cidade Universitária, CEP 58051-900, João Pessoa, PB. Telefone: (83) 3216-7264. E-mail: guilhermeataide@ccsa.ufpb.br

## INTRODUÇÃO

Hodiernamente, vivenciamos um cenário em que são produzidos grandes volumes de dados, processados de forma cada vez mais veloz e que possuem fontes e formatos variados. Este contexto caracteriza o fenômeno do *Big Data*<sup>1</sup>, que está relacionado com a heterogeneidade dos dados, com a velocidade da sua geração e processamento, assim como da sua significativa quantidade disponível em ambientes digitais (MAYER-SCHÖNBERGER; CUKIER, 2013).

De acordo com Coneglian, Segundo e Sant'Ana (2017), o fenômeno do *Big Data* vem provocando transformações nas esferas governamentais, empresariais, assim como nos campos científicos. No âmbito científico, foco desse artigo, o desenvolvimento de uma nova geração de experimentos, sensores, instrumentos e *software* de simulação possibilita que a pesquisa científica contemporânea trabalhe com uma quantidade considerável de dados, o que caracteriza o conceito de *e-Science* (SAYÃO; SALES, 2014). Por meio de um conjunto de ferramentas tecnológicas torna-se possível a coleta e análise de dados de pesquisa. Esta realidade permite que novos enfoques, aplicações, inovações e serviços sejam ofertados pela ciência.

O grande volume de dados em meio digital traz a preocupação com a forma adequada de gerir e preservar esses dados, de forma que possam continuar acessíveis e passíveis de serem usados e reusados agora e no futuro, o que remete ao processo de curadoria digital, mais especificamente à curadoria de dados. Esta abarca atividades de gestão requeridas para manter, gerir e preservar dados de pesquisa no longo prazo, de forma que possam estar disponíveis para uso e reuso (SIEBRA; BORBA; MIRANDA, 2016). Favorecendo assim a colaboração entre pesquisadores, o avanço da ciência e a preservação do conhecimento científico (SANTOS, 2014).

Para promover a curadoria de dados é fundamental o desenvolvimento e/ou adoção de metodologias (tecnológicas e gerenciais) que guiem a produção de dados, o desenvolvimento de coleções de dados e seu armazenamento, análise e interpretação (SAYÃO; SALES, 2012). Logo, torna-se primordial a adoção de algum modelo de ciclo de vida dos dados, pois esse tipo de iniciativa apresenta uma estrutura que caracteriza as diversas etapas que necessitam ser empreendidas sobre os dados e seus conjuntos/coleções no decorrer da sua vida útil, permitindo que o uso dos mesmos possa ser otimizado e estendido (SAYÃO; SALES, 2015).

Nesse contexto, o objetivo geral desse artigo é analisar as intersecções entre os elementos dos modelos de ciclos de vida dos dados do Modelo do *Digital Curation Centre - DCC*<sup>2</sup>, do *Data Observation Network for Earth - DataOne*<sup>3</sup> e do Ciclo de Vida dos Dados para a Ciência da Informação (CVD-CI). Sendo as duas primeiras iniciativas de destaque no cenário internacional e, a última, uma das primeiras propostas desenvolvidas no contexto brasileiro por Sant'Ana (2013), no âmbito do Programa de

---

<sup>1</sup> *Big Data* consiste de conjuntos de dados, caracterizados pelo volume, velocidade e variedade, que requerem arquiteturas escaláveis para que sejam eficientes os processos de armazenamento, manipulação (NIST, 2016).

<sup>2</sup> Disponível em: <http://www.dcc.ac.uk/resources/curation-lifecycle-model>

<sup>3</sup> Disponível em: <https://www.dataone.org/data-life-cycle>

Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista (UNESP). Indicamos que não foram encontradas na literatura da área, pesquisas relacionadas com a investigação aqui apresentada.

De forma a minimizar divergências conceituais entre gestão e curadoria de dados de pesquisa, esclarecemos que os processos de curadoria são mais amplos, porque incluem etapas adicionais além da gestão, tais como: arquivamento e preservação. No contexto dos modelos estudados, optamos por utilizar o conceito de gestão como sinônimo de curadoria.

O estudo da intersecção desses modelos traz como contribuição a possibilidade do desenvolvimento de estratégias para a gestão de dados científicos com base em suas etapas, além de apontar o caminho para novas pesquisas que tenham como objetivo adaptar ou propor um ciclo de vida que contemple especificamente os dados de pesquisa com etapas mais refinadas dos três modelos previamente indicados. A presente pesquisa torna possível também à identificação de eventuais fases do CVD que não tenham sido contempladas anteriormente.

Essa pesquisa “desenvolvida no âmbito da Ciência da Informação (CI),” indica a responsabilidade da área com os pesquisadores em geral. Trata-se de um olhar da CI voltado para o crescimento da ciência ao estimular a gestão de dados de pesquisa, o que contribui com novos estudos, com a qualidade da pesquisa e consequentemente com a possibilidade de uma maior credibilidade das investigações científicas.

## DADOS

Pomerantz (2015) considera os dados como objetos potencialmente informativos, que podem ser uma representação da informação (SETZER, 2001). Sayão e Sales (2015) ao tratarem dos dados de pesquisa, revelam que eles podem ser expressos de diferentes formas, sendo que a sua concepção pode variar de modo considerável entre os pesquisadores e entre as áreas do conhecimento. Uma vez que eles são gerados para finalidades distintas, através de diferentes processos, por e para diversas comunidades acadêmicas e científicas.

Os dados podem ser fatos, números, letras e símbolos (NATIONAL RESEARCH COUNCIL, 1999). Segundo Borgman (2015), os dados podem ser de origem observacional, sendo considerados os mais importantes para preservar por serem menos replicáveis. A autora revela que os dados também podem ser de origem computacional. Estes necessitam para a reutilização futura de um modelo computacional, de uma documentação extensiva de *hardware*, *software*, dados de entrada e as etapas intermediárias. E, quanto a origem dos dados experimentais, caso um experimento seja projetado para ser replicável, esses dados podem ser mais fáceis de replicar, do que preservar. Desta forma, a origem dos dados pode influenciar nas decisões sobre a preservação dos mesmos.

Este foco maior nos dados, especialmente em meio digital, é relativamente recente e ainda não é possível ter uma visão completa sobre todos os aspectos envolvidos. Porém, neste processo de desenvolvimento é crucial uma compreensão comum de pontos chave a serem trabalhados para geri-los e preservá-los, independente de tecnologias, linguagens ou sistemas específicos (MÖLLER, 2013). Por isso mesmo, diante da obsolescência tecnológica e dos riscos inerentes ao meio digital (como, por exemplo, problemas de *hardware*), Sales e Cavalcanti (2015) alertam sobre a importância de adotar práticas de preservação dos dados digitais. Além de enfatizarem que a curadoria digital de dados de pesquisa se apresenta como uma

solução a ser adotada para se atingir o objetivo de ter segurança na preservação e compartilhamento dos dados, sendo o desenvolvimento de repositórios de dados uma alternativa para tal preservação e exercício da curadoria digital.

É importante destacar que a gestão e a curadoria de dados de pesquisa não se limitam aos grandes volumes de dados em meio digital. Sendo elas igualmente relevante em ambientes com grandes e pequenos volumes de dados, sendo estes digitais e/ou não digitais. E, para sistematizar a aplicação das atividades vinculadas à curadoria, modelos de ciclos de vida surgiram e vem sendo usados (SIEBRA; BORBA; MIRANDA, 2016), como será discutido na próxima seção.

## CICLO DE VIDA DOS DADOS (CVD)

Um ciclo de vida dos dados delinea a atividade de um pesquisador durante um projeto de pesquisa, de forma temporária, ele apresenta o caminho dos dados ou as ações pertinentes que possibilitam o avanço da pesquisa para o próximo estágio (RICE; SOUTHALL, 2016).

Sayão e Sales (2015) revelam que existem vários tipos de modelos de ciclo de vida dos dados de pesquisa, que possuem os seus próprios objetivos e particularidades, os quais frequentemente são direcionados para áreas específicas do conhecimento. Sant’Ana (2013) identificou dezoito (18) iniciativas de modelos de CVD, de 1997 a 2013 e destacou suas características e usos. No Quadro 1 são apresentadas as iniciativas de modelos encontrados em Sant’Ana (2013), foram acrescentados o seu modelo voltado para CI e dois modelos destacados por Rice e Southall (2016).

**Quadro 1 – Iniciativas de Modelos de Ciclos de Vida dos Dados (1997 - 2013)**

INICIATIVAS	CARACTERÍSTICAS	FONTES
Entropia de Dados	Denuncia desordem e degradação dos dados e metadados em função do tempo e da obsolescência.	Michener et al. (1997)
CVD – Componentes e recursos estratégicos	Mostra insumos e resultados do processo como produtos bem delineados e definidos; característica de aplicabilidade.	Ferderer (2001)
Ciclo de Vida de Metadados - CVMD	Sinaliza padrões adequados para assegurar a descrição e o controle ao longo do tempo.	Chen et al. (2003)
Ciclo de Vida dos Dados – DDI	Composto por oito fases identificadas em sequência; base para outros modelos.	Data Documentation Initiative – DDI (2004)
CVD e a Pesquisa Translacional	Caracterizado como “ <i>Translational Research</i> ”, o diagrama é composto por quatro ciclos que apresentam a percepção do acesso aos dados.	VUMC (2005)
CVD de Conhecimento de Pesquisa	Ilustra as principais considerações para o arquivamento em cada etapa do processo de criação de dados.	Humphrey (2006)
CVD – Pennock	Identifica fase de acesso que o registro é utilizado; fase de acesso e reuso que pode gerar a criação de novos registros.	Pennock (2007)

CVD e Publicações	Citado na CI: descreve fases envolvidas no processo de publicação de produções científicas.	Gold (2007)
Modelo de Ciclo de Vida de Dados Digitais	Extrapolando o fluxo informacional em si e incorpora elementos contextuais como a própria organização, perfis profissionais necessários e políticas de informação.	IWGDD (2009)
Ciclo de Vida de Dados Abertos Vinculados	Colabora no processo de compreensão das diferenças nos ciclos de vida de dados ordinários e de dados ligados.	Villazón-Terrazas et al. (2009)
Tarefas e atores envolvidos no gerenciamento da base de dados da SHARE	Mostra sequência circular e linear de fases, acrescentando ainda as tarefas e atores envolvidos em cada uma.	Hunkler et al. (2011)
CVD de Pesquisa – UCL*	Inclui as principais tarefas envolvidas em cada fase e enfoque em dados coletados por meio de entrevistas.	UCL (2012)
CVD – DAMA	Figura da operação de descarte dos dados: “ <i>retention and destruction</i> ”.	DAMA (2012)
O compartilhamento de dados para pesquisa demográfica	Cita fases que congregam mais de uma tarefa e apresentam no diagrama estes objetivos, facilitando o entendimento.	DSDR (2012)
CVD - comparado com o projeto que o cria	Proporciona um cenário de análise para o acompanhamento de cada uma das fases.	UKDA (2013)
Ciclo de Vida dos dados de Materiais	Identifica práticas de gestão, análise e uso de informações sobre materiais, vinculando suas iniciativas a aplicação.	MDMC (2013)
Ciclo de Vida dos Dados	Atua com estruturas distribuídas e sustentáveis que atendam às necessidades da ciência e da sociedade de acesso aberto, destacando a pesquisa na área ambiental.	DataOne (2013)
Ciclo de Vida de Curadoria	Destaca gestão e ações administrativas planejadas para promover a curadoria e preservação ao longo do ciclo de vida da curadoria.	DCC (2004)
CVD – CI	Propõe um modelo CVD que seja compartilhado pela CI, fortalecendo a compreensão dos esforços e resultados das pesquisas realizadas.	Sant’Ana (2013)
Ciclo de Vida de Dados de Pesquisa	Mostra que os serviços podem ser mapeados em um fluxo de trabalho centrado no pesquisador, com diagrama circular sinalizando que os dados continuam a viver após a pesquisa.	Jisc (2016)
Fluxo de trabalho do RDM da Biblioteca do Imperial College	Fornece um exemplo de serviços que podem ser mapeados em modelo de ciclo de vida de dados do <i>Imperial College London</i> .	Barnes (2016)

London		
--------	--	--

Fonte: Adaptado de Sant’Ana (2013) e Rice e Southall (2016).

Entre os modelos, três foram destacados no Quadro 1 e serão utilizados para fins do presente estudo, o modelo do CVD-CI, por estar voltado para a CI e ser uma proposta brasileira, cuja elaboração foi baseada em diversos modelos de ciclo de vida dos dados (SANT’ANA, 2013). Trata-se de um modelo mais recente, em que foi encontrado apenas um trabalho na literatura que fez uso prático do mesmo no contexto de um repositório institucional (ALBUQUERQUE, 2018).

O modelo do DCC por ter como foco em seu centro a curadoria de dados, objetos digitais (simples e complexos) e base de dados e ser um dos modelos mais utilizados em vários contextos, desde a gestão de dados de pesquisa, assim como para a curadoria de acervos memoriais, históricos e culturais (SIEBRA; BORBA, 2018).

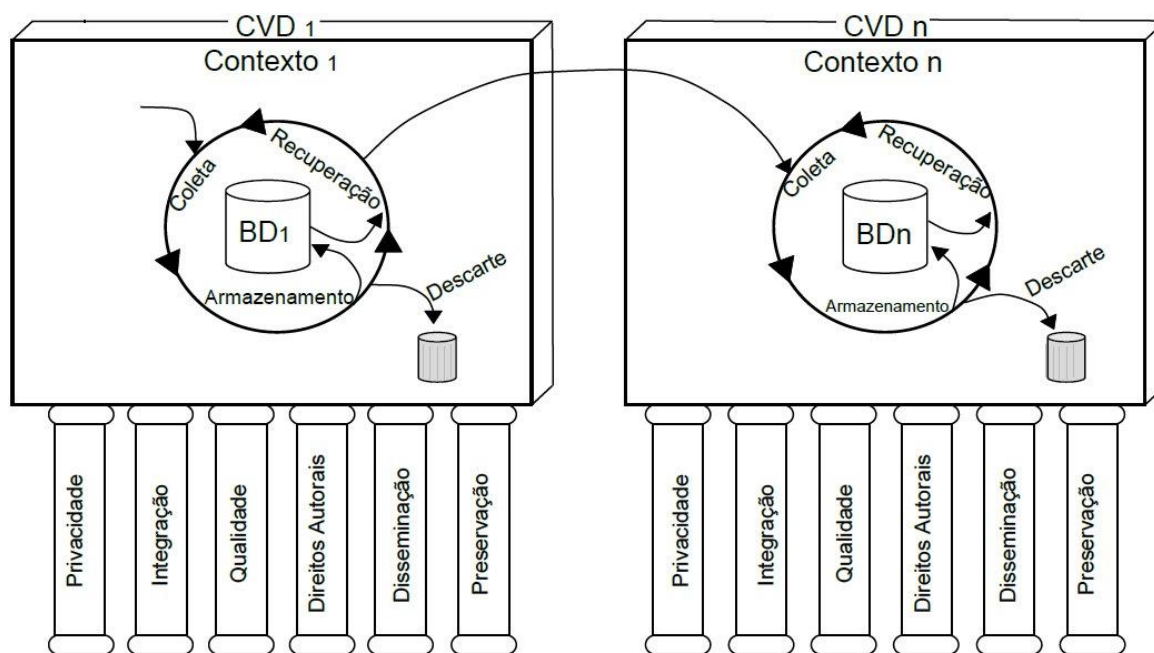
Finalmente, o modelo do *DataONE* que foi fundamentado na sua relevância no âmbito da preservação dos dados para uso e reuso, propiciando boas práticas de pesquisa. Este modelo vem sendo mais utilizado para a curadoria de dados de pesquisa (DATAONE, 2018) e, apesar de ser um modelo norte-americano voltado para as ciências da terra e ambientais, não se limita a tal área específica, pois por tratar de dados de pesquisas pode ser utilizado por diversas áreas do conhecimento. Desta forma, mesmo que os modelos supracitados possam ser usados para dados de qualquer natureza, eles foram inicialmente criados com o pensamento voltado para o contexto dos dados de pesquisa.

### **Ciclo de vida dos dados para a Ciência da Informação – CVD-CI**

O modelo de Ciclo de Vida dos Dados (Figura 1), proposto por Sant’Ana (2013), leva em conta características e especificidades da Ciência da Informação (CVD-CI), partindo da pesquisa realizada sobre modelos de ciclo de vida de áreas correlatas. Considera o estudo do comportamento da informação, de seu fluxo e dos meios para o seu acesso como objetivos da Ciência da Informação. Para efeito de compreensão do processo de acesso a dados, o autor propôs quatro fases: coleta, armazenamento, recuperação e descarte, as quais são permeadas pelos fatores: privacidade, qualidade, direitos autorais, integração, disseminação e preservação.



Figura 1 - Ciclo de Vida dos Dados para Ciência da Informação – (CVD-CI).



Fonte: Sant'Ana (2016)

A **coleta** dos dados é a primeira fase. É preciso obter os dados que podem ser utilizados para contemplar uma demanda pontual ou uma necessidade prevista de informações sobre um determinado panorama. Por conseguinte, são necessárias ações coordenadas para que se possa estabelecer um plano de ação, análise da viabilidade, bem como a execução da coleta dos dados. De acordo com Sant'Ana (2016, p. 18), nessa fase é preciso fazer “a definição inicial dos dados a serem utilizados, seja na elaboração do planejamento de como serão obtidos, filtrados e organizados, identificando-se a estrutura, formato e meios de descrição que será utilizado”. O autor ainda complementa afirmando que nessa fase, os dados também devem ser descritos em metadados, avaliados e selecionados. Para isso, são necessárias competências específicas, ainda que não totalmente dependentes de um conhecimento profundo sobre as tecnologias digitais, mas muito próximo da necessidade informacional que motiva a coleta.

Nesta fase de coleta, ao tratar dos fatores, Sant'Ana (2016) revela a importância de levar em consideração com base nos dados coletados, a exposição da **privacidade** das pessoas e das instituições. O fator **integração** deve identificar e validar os aspectos que terão a responsabilidade para levantar de forma única cada registro, assim como os correspondentes de outras entidades. O **direito autoral** tem o foco no responsável pelas fontes de dados. A **disseminação** é proposta nesse momento para fomentar a encontrabilidade e uso dos dados e a **preservação** de dados que pode requerer a outros adicionais.

Na fase de **armazenamento**, os esforços, segundo Sant'Ana (2016, p. 18) focam no “processamento, transformação, inserção, migração transmissão e toda e qualquer ação que vise a persistência dos dados em suporte digital.” Portanto, nesta fase, se faz necessário um conjunto de planejamentos e ações que requerem um conhecimento mais profundo da Ciência da Computação, mas que ainda apresentam forte potencial de participação para a Ciência da Informação (CI). A participação do usuário nessa fase é muito reduzida, limitando-se à validação dos modelos de estruturas definidos para os dados.

No que se refere aos fatores, à **privacidade** está relacionada a quem poderá ter acesso aos dados. A **integração** deve buscar responder a questão de como os dados serão acessados. A **qualidade** refere-se às definições acerca do armazenamento. Os **direitos autorais** estão voltados para a fonte dos dados coletados. A **disseminação** ao acesso futuro e a **preservação** com a interpretação dos dados no futuro (SANT'ANA, 2016).

A **recuperação** dos dados constitui a terceira fase. Gera preocupações e esforços voltados para que os dados possam ser encontrados, acessados e interpretados com facilidade. Nesta fase, fica explícita a necessidade de conhecimentos advindos da CI, instrumentalizados pela Ciência da Computação e, preferencialmente, contando com aqueles que detenham conhecimento sobre o público alvo ou em potencial, bem como das necessidades e usos previstos, por mais amplas que sejam essas definições. No que tange aos fatores que permeiam esta fase, a **privacidade** está voltada aos envolvidos com os conteúdos disponibilizados, a **integração** deve permitir que entidades distintas sejam capazes de analisar os dados armazenados. A **qualidade** tem o seu foco na interação do usuário. Os **direitos autorais** devem evidenciar quem pode usar os dados e como eles podem ser utilizados. A **disseminação** enriquece os dados com elementos que possibilite a encontrabilidade dos que vão fazer uso dos mesmos e a **preservação** deve primar pela interpretação dos dados (SANT'ANA, 2016).

O momento em que se decide que dados já não são mais necessários ou que não devem ser mantidos, sinaliza o **descarte**, entrando na quarta e última fase. Neste momento ocorre então, a eliminação de parte dos dados, o que pode ocorrer em bloco, horizontalmente ou verticalmente (SANTOS; SANT'ANA, 2013). Para esta fase, segundo Sant'Ana (2013) também são necessários conhecimentos específicos para os esforços de planejamento e execução.

A **privacidade** na fase de descarte deve ser levada em consideração, uma vez que pode levar a complicações caso uma pessoa tenha a necessidade de ter os seus dados retirados de uma base de dados. A **integração** também é fundamental, pois a exclusão de dados em uma base pode prejudicar o relacionamento entre diferentes bases. A **qualidade** se relaciona com a integração. O **direito autoral** consiste na manutenção da autoria das informações. Quanto a **disseminação**, os dados não estão mais disponíveis, porém deve-se persistir na **preservação**, mesmo quando os dados não têm mais utilidade.

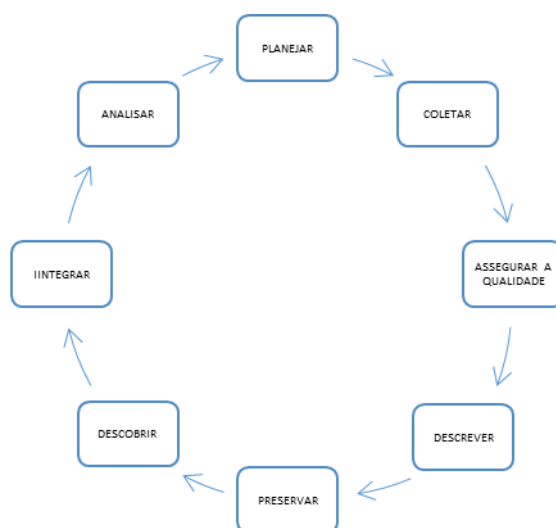
### **Ciclo de vida dos dados – DataONE**

A *Data Observation Network for Earth (DataONE)*, financiado pela *National Science Foundation (NSF)* dos Estados Unidos, trata-se de um projeto norte-americano, que por meio de um estrutura distribuída e infraestrutura cibernética sustentável busca atender as necessidades da ciência e da sociedade de acesso aberto aos dados provenientes das observações da terra. Se propõe a isso de modo persistente, robusto e seguro, constituindo a base da nova ciência ambiental inovadora, a qual foi constituída por meio de um acordo cooperativo, envolvendo a colaboração de muitas organizações parceiras. Isso porque a área ambiental tem se expandido de modo acentuado, representando uma fronteira de pesquisa relacionada ao acesso a dados, especialmente diante do vasto volume de dados gerados, com o barateamento e disseminação de sensores e outros instrumentos (SANT'ANA, 2013).



O ciclo de vida dos dados *DataONE* (Figura 2) apresenta uma visão abrangente dos estágios relacionados com a gestão bem-sucedida e com a preservação de dados para que possam ser usados e reutilizados. Este ciclo foi criado pela Equipe de Liderança *DataONE*, em parceria com a comunidade *DataONE* e foi construído sobre o modelo de ciclo de vida proposto pela *National Science Foundation (NSF)* (DATAONE, 2018). Trata-se de um modelo acessível, que pode abranger várias dinâmicas institucionais e diversas tipologias de dados (OLIVEIRA; SILVA, 2016).

**Figura 2: Ciclo de Vida dos Dados – DataONE**



Fonte: Adaptado de DataONE (2018).

O *DataONE* possui oito etapas (DATAONE, 2018), conforme ilustrado na Figura 2. Na etapa **planejar** selecionam-se os dados a serem gerenciados, são definidas as pessoas e suas responsabilidades, bem como a definição de quais dados serão curados, como serão descritos e a forma como serão gerenciados e mantidos. Na etapa seguinte, **coletar**, os dados podem ser obtidos manualmente, através de sensores, por outros instrumentos ou por meio de digitalização. Ressalta-se que é importante registrar todas as informações relacionadas ao dado coletado. Na etapa de **garantir**, através de verificações e inspeções a qualidade dos dados é assegurada. A seguir, na etapa de **descrever**, os dados são descritos de forma precisa e detalhada através do uso de metadados pertinentes, segundo um padrão pré-definido. Posteriormente, vem o momento de **preservar**, em que os dados são encaminhados para dispositivos de armazenamento apropriados (por exemplo, *data centers*), de forma a garantir a sua recuperação e acesso a longo prazo. Na fase de **descobrir**, os dados potencialmente úteis são encontrados e obtidos juntamente com as informações relevantes a respeito dos dados (metadados), o que possibilita a criação de novos dados e a geração de conhecimento. Na fase de **integrar**, os dados provenientes de fontes distintas são combinados para gerar um conjunto homogêneo de dados, ocorre à fase **analisar**, que é o momento em que os dados são analisados com o intuito de fornecerem informações relevantes ou não para futuras pesquisas (STRASSER et al., 2012).

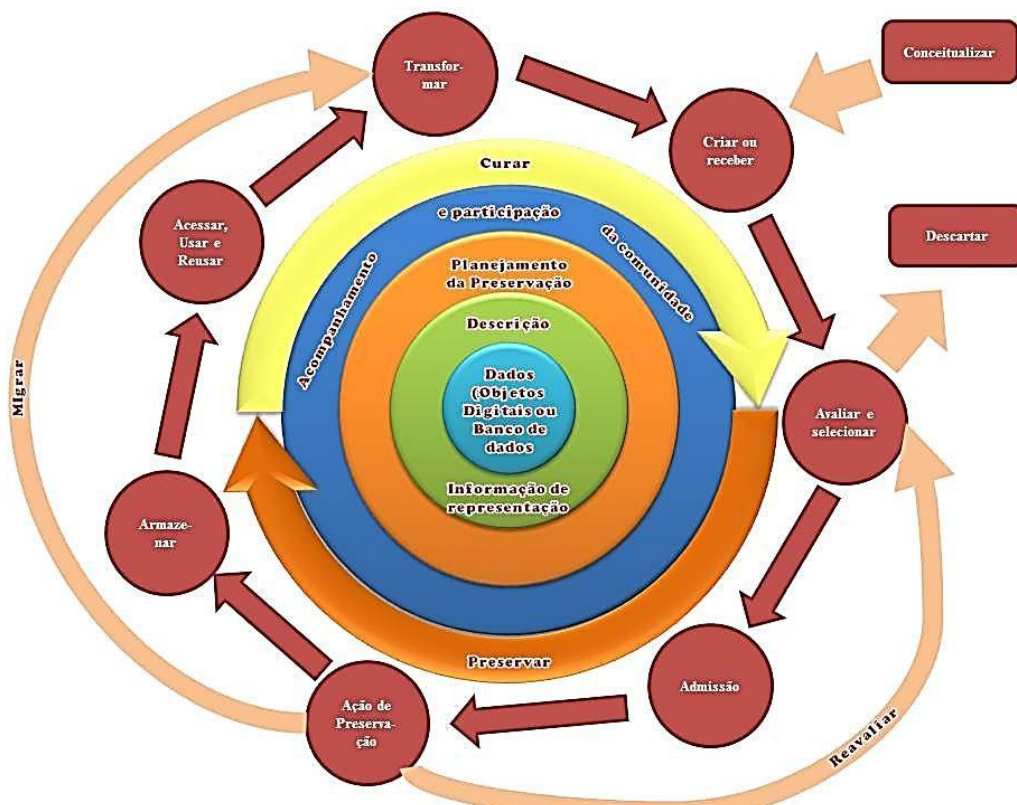
Um pesquisador ou uma equipe de pesquisadores geralmente está envolvida em todos os aspectos do ciclo de vida dos dados, na função de criador(es), assim como de usuário(s) dos dados. Os autores exemplificam que algumas equipes de cientistas, aquelas ligadas à modelagem e sínteses podem criar novos dados no processo de

descobrir, integrar, analisar e sintetizar dados existentes. Ressalta-se que, nem todos os projetos utilizam todos os componentes, alguns podem utilizar apenas parte do ciclo e não necessariamente seguir o caminho linear apresentado em formato de ciclo (STRASSER et al., 2012).

### Ciclo de vida dos dados - DCC

O *Digital Curation Centre* (DCC) foi criado em 2004 no Reino Unido, visando acompanhar os desafios voltados para a curadoria digital. No contexto de desenvolvimento da iniciativa, a comunidade de pesquisa do ensino superior do Reino Unido recebeu apoio para a curadoria dos seus dados de pesquisa (DCC, 2018). O modelo de ciclo de vida dos dados do DCC compreende a manutenção, preservação e agregação de valor aos dados de pesquisa digital, em um processo que envolve todo o seu ciclo de vida (DCC, 2018). Neste sentido, **de acordo com Siebra et al. (2013)**, a unidade principal de gestão do ciclo, destacada em seu centro, é o dado, que corresponde a qualquer objeto digital (simples ou complexo) ou base de dados, codificado em formato binário. Em volta do dado digital estão as ações que necessitam ser executadas durante o processo de curadoria. Essas ações são classificadas, de acordo com o modelo DCC (2018) em três tipos: ações para todo o ciclo de vida; ações sequenciais e ações ocasionais, ilustrado na Figura 3.

Figura 3 - Ciclo de vida de dados do *Digital Curation Centre*



Fonte: Yamaoka (2012)

Sayão e Sales (2012) afirmam que as ações para o ciclo de vida completo são atividades que envolvem todo o ciclo de vida da curadoria digital. Para traduzir essa ideia de presença constante, essas ações estão simbolizadas no gráfico como anéis

concêntricos permeando os dados que estão no centro do modelo. As ações para todo o ciclo de vida dos dados possuem as seguintes etapas, segundo o DCC (2018):

- **Descrição Informação de Representação – Definir os** metadados administrativos, descritivos, técnicos, estruturais e de preservação, utilizando padrões apropriados, para garantir descrição e controle a longo prazo. Definir a forma de atribuição de informações de representação pertinente para compreender e contextualizar o material digital;
- **Planejamento da Preservação** - Planejar as estratégias de preservação a serem aplicadas no decorrer de todo o ciclo de vida da curadoria de material digital.
- **Acompanhamento e participação da comunidade** – motivar e envolver a comunidade interna (produtores do conteúdo/dados, desenvolvedores, pessoas da equipe) e externa (comunidade em geral, usuários da informação) no ciclo de curadoria para ajudar na representação da informação, definir a melhor forma de acesso aos dados, colaborar na definição de padrões, ferramentas e produtos de *software* apropriados, entre outros;
- **Curar e Preservar** - Estar consciente e realizar ações administrativas e gerenciais planejadas e intencionais para contribuir com a curadoria e a preservação no decorrer de todo o ciclo de vida da curadoria. Tratar questões de direitos autorais e permissões de acesso.

As ações sequenciais, conforme Sayão e Sales (2012), constituem etapas que precisam ser realizadas repetidamente para garantir que o dado se mantenha em contínuo processo de curadoria. Tal sequência não deve ser cumprida apenas uma vez do começo ao fim, mas continuamente durante todo tempo que o dado estiver sob curadoria, de forma cíclica, não necessariamente na ordem apresentada. Essas etapas são, de acordo com o DCC (2018):

- **Conceitualizar** - Conceber e planejar a criação de dados, contendo o método de captura e opções de armazenamento;
- **Criar ou receber** - Criar dados, acoplado a eles metadados administrativos, descritivos, estruturais e técnicos. Metadados de preservação também podem ser incluídos no momento da criação. Receber dados, de acordo com políticas de coleta documentadas, de criadores de dados, outros arquivos, repositórios ou centros de dados e, se preciso, atribuir metadados apropriados;
- **Avaliar e Selecionar** - Avaliar os dados e selecionar para curadoria de longo prazo e preservação com base em orientações documentadas, políticas ou requisitos legais;
- **Admissão** - Transferir dados de um arquivo, repositório, centro de dados ou outra entidade custodiante, de acordo com orientações documentadas, políticas ou requisitos legais;
- **Ação de preservação** - Executar ações para assegurar a preservação a longo prazo e a retenção da natureza autorizativa dos dados. As ações de preservação precisam garantir que os dados permaneçam autênticos, confiáveis e usáveis, mantendo sua integridade;
- **Armazenar** – Realizar o armazenamento dos dados com segurança, de acordo com padrões relevantes. Além de seguir rotinas de *backup*.

- **Acessar, usar e reusar** - Garantir que os dados estejam acessíveis aos usuários designados no cotidiano, de forma adequada ao público ao qual se destinam;
- **Transformar** - Criar novos dados a partir do original, por exemplo: através da migração para um formato diferente.

Por fim, Sayão e Sales (2012) falam das ações ocasionais, que são realizados eventualmente, podendo interromper ou reordenar as ações sequenciais, mediante uma decisão. Segundo o DCC (2018), essas ações são:

- **Descartar - Eliminar** os dados que não foram escolhidos para curadoria e preservação de longo prazo, segundo as políticas documentadas, orientação ou requisitos legais ou aqueles que já alcançaram seu período máximo de curadoria, de acordo com tabelas de temporalidades adotadas. Ressalta-se que o descarte, muitas vezes, é uma atividade necessária, mas que deve ser realizado seguindo procedimentos bem definidos, como, por exemplo, os recomendados pelo CONARQ (2014);
- **Reavaliar** - Retornar os dados que passaram por procedimentos de validação falhos, para uma avaliação e seleção adicional. Reavaliar todas as atividades realizadas no ciclo para verificar a necessidade de ajustes/melhorias;
- **Migrar** – Fazer a migração dos dados para um formato diferente. Isso pode ser feito de acordo com o ambiente de armazenamento ou para garantir a preservação dos dados evitando a obsolescência de *hardware* ou *software*.

O modelo permite uma visão coletiva sobre o conjunto de funções necessárias à curadoria e à preservação de dados de pesquisa (SAYÃO; SALES, 2012). Importante frisar que responsabilidades, infraestrutura de padronização, tecnologias e papéis devem ser definidos de modo adequado para evitar surpresas futuras durante a execução do ciclo.

## PROCEDIMENTOS METODOLÓGICOS

A pesquisa configura-se como descritiva e apresenta uma abordagem qualitativa, por seu foco estar voltado aos aspectos da realidade que não podem ser calculados (SILVEIRA; CÓRDOVA, 2009), preocupando-se com o entendimento e explicação das relações existentes entre os três modelos de ciclo de vida dos dados escolhidos: CVD-CI, *DataONE* e *DCC*.

O estudo é bibliográfico (GIL, 2017) e fez uso de material impresso e digital, como dissertações, livros, revistas e anais de eventos científicos, além de materiais disponibilizados nos sites do *DataONE* e do *DCC*, onde foi possível obter informações sobre os seus respectivos modelos de ciclo de vida dos dados.

## COMPARANDO E DISCUTINDO OS MODELOS

Diante dos estudos realizados a partir dos modelos do CVD-CI (SANT'ANA, 2013, 2016); *DCC* (2018) e *DataONE* (2018) foram identificadas intersecções e relações entre as fases desses modelos, expressas nos Quadros 2, 3.

**Quadro 2: Intersecções/Relações entre o modelo DataONE e os modelos DCC e CVD-CI**

DataONE	DCC	CVD-CI
<b>Planejar</b>	Descrição informação de representação Planejamento da preservação Curar e preservar Avaliar e selecionar Acompanhamento e participação da comunidade	Coleta
<b>Coletar</b>	Conceitualizar Criar e receber Transformar	Coleta
<b>Assegurar a qualidade</b>	Ação de preservação Acompanhamento e participação da comunidade Reavaliar	Coleta Armazenamento Recuperação.
<b>Descrever</b>	Criar ou Receber	Coleta
<b>Preservar</b>	Avaliar e Selecionar Admissão Armazenar Descartar Ação de preservação	Armazenamento Descarte
<b>Descobrir</b>	Acessar, usar e reusar.	Recuperação
<b>Integrar</b>	Acessar, usar e reusar	Recuperação
<b>Analisar</b>	Acessar, usar e reusar Reavaliar.	Recuperação

Fonte: Os Autores, 2019.

De acordo com os elementos apresentados no Quadro 2, o modelo *DataONE* se correlaciona com os modelos *DCC* e *CVD-CI* da seguinte forma: a etapa **planejar** combina com as ações do *DCC* de **descrição informação de representação**, pois é o momento de se planejar a descrição dos dados visando garantir que os arquivos possam ser futuramente localizados e que o seu conteúdo possa ser compreendido; está relacionada com as ações de **planejamento da preservação e curar e preservar**, pois estratégias nesse sentido devem ser planejadas e podem afetar várias outras etapas; além de se planejar como e quais dados serão gerenciados, relacionando assim com as etapas de avaliar e selecionar do *DCC*. Também está relacionada com a ação de **acompanhamento e participação da comunidade**, mas de forma parcial, porque no *DataONE* é mencionada a necessidade de se definir pessoas e suas responsabilidades dentro da equipe de trabalho, mas não há menção explícita a participação do usuário da informação, como no *DCC*.

A etapa **planejar** do *DataONE* se relaciona com o *CVD-CI* com a fase de **coleta**, que consiste em elaborar um plano de ação para proporcionar um tratamento adequado para os dados. Porém, quando se compara a fase de **coleta** do *CVD-CI* com os outros modelos (Quadro 3), verifica-se que ela acaba abrangendo várias etapas deles, o que acaba misturando as ações de planejamento com a execução em si do ato de coletar e de descrever os dados. Isso pode tornar confuso para os curadores que forem

trabalhar com o modelo o que especificamente deve ser realizado, pela existência de muitas ações a serem realizadas em uma só etapa. Como consequência, o planejamento em si, etapa tão relevante para o sucesso do processo de curadoria como um todo, pode acabar não sendo realizado ou sendo feito de maneira breve, incompleta e não-abrangente.

**Quadro 3: Intersecções/Relações entre o modelo CVD-CI e os modelos DataONE e DCC**

CVD-CI	DataONE	DCC
<b>Coleta</b>	Planejar; Coletar; Assegurar a qualidade Descrever.	Curar e preservar Descrição informação de representação Acompanhamento e participação da comunidade Planejamento da preservação; Conceitualizar Criar ou receber Avaliar e selecionar Descartar
<b>Armazenar</b>	Preservar Assegurar a qualidade	Ação de preservação Admissão Armazenar Transformar Migrar
<b>Recuperação</b>	Integrar; Descobrir; Analisar Assegurar a qualidade	Acessar, usar e reusar Acompanhamento e participação da comunidade
<b>Descarte</b>	Preservar	Descartar

Fonte: Os Autores, 2019.

A etapa de **coletar** do *DataONE* (Quadro 2) relaciona-se com as ações de **conceitualizar** do DCC, pois é o momento de conceber o método de captura dos dados com base no que foi planejado; engloba parcialmente a ação de **criar ou receber**, por nessa etapa ser feita a aquisição/criação de dados; e, também, com a etapa **transformar**, pois como explicitado no modelo *DataONE*, algumas vezes durante a coleta, se fazem necessárias algumas transformações. A etapa **coletar** do *DataONE* se relaciona com o CVD-CI apenas com sua etapa de **coleta**, com ênfase na obtenção de dados, uma vez que esta etapa no CVD-CI é muito mais ampla.

A etapa de **assegurar a qualidade** do *DataONE* (Quadro 2) está ligada com a **ação de preservação** do DCC, em que as práticas devem estar voltadas para garantir autenticidade, confiabilidade, usabilidade, ou seja, a qualidade dos dados. Como, em alguns casos a verificação da qualidade dos dados depende daqueles que detém conhecimento sobre o negócio contexto dos dados ou das necessidades e usos que podem ser feitos desses dados, acaba por também haver uma relação com o **acompanhamento e participação da comunidade** do DCC. Adicionalmente, há uma relação com a etapa de **reavaliar**, visto que os dados que não passarem pela validação de qualidade será reavaliado nessa etapa. No CVD-CI, a qualidade é um



fator que permeia todas as etapas do modelo. Assim, a etapa **assegurar a qualidade** do *DataONE* acaba por ter relação com as etapas de **Coleta, Armazenamento e Recuperação** do CVD-CI, com ênfase no fator qualidade.

A etapa **descrever** do *DataONE* (Quadro 2) se correlaciona com a etapa de **criar ou receber** do DCC, onde os dados são criados/recebidos e seus metadados preenchidos, de acordo com o definido na etapa de descrição informação de representação. Já a conexão com o CVC-CI acontece em sua fase de **coleta**, pois ela engloba a descrição dos dados.

A etapa de **preservar** do *DataONE* (Quadro 2) é um momento de decidir os dados que serão preservados/encaminhados para um arquivo apropriado, para armazená-los de forma apropriada, o que está em sintonia com as seguintes ações do DCC: **avaliar e selecionar**, que faz a avaliação e seleção dos dados para a curadoria de longo prazo; ação de **admissão e armazenar**, pois os dados são encaminhados para um arquivo apropriado e devidamente armazenados. Apesar de não explicitado no desenho do modelo *DataONE* ou em sua descrição, subentende-se que essa etapa de **preservar** acaba também por ter relação com a ação de **descartar** do DCC, ao se considerar que os dados que não foram escolhidos para preservação poderão vir a ser descartados. Adicionalmente, também há uma relação indireta com a etapa de **ação de preservação** do DCC, visto que é especificado no *DataONE* que o arquivo/repositório/base para armazenamento dos dados deve ser apropriado para armazenamento a longo prazo, precisando, assim, ser submetido a estratégias e ações de preservação. A etapa de **preservar** do *DataONE* se conecta com a etapa de **armazenar** do CVD-CI, que visa a persistência dos dados a longo prazo e com a etapa de **descarte**, visto que os dados que não forem selecionados para preservação podem ser descartados.

Tomando a etapa de **armazenar** do CVD-CI como referência (Quadro 3), como visto, ela se relaciona com a etapa **preservar** do *DataONE* e, também, com a etapa de **assegurar a qualidade**, uma vez que a verificação da qualidade é um fator que permeia as etapas do CVD-CI. Com o DCC, a etapa de armazenar do CVD-CI engloba diversas etapas, tais como: **admissão** (pois os dados podem ser transmitidos para uma entidade custodiante, **armazenar** (como o nome já afirma, é onde os dados serão armazenados), **transformar** e **migrar** podem ser aplicadas visando manter a preservação dos dados no meio de armazenamento, assim como podem ser aplicadas outras **ações de preservação** que tem relação direta com os fatores do CVD-CI, com ênfase para os **fatores preservação, privacidade, direitos autorais e integração**.

As etapas **descobrir, integrar e analisar** do *DataONE* (Quadro 2) se relacionam com a etapa **acessar, usar e reusar** do DCC, pois nela procura-se garantir que os dados estejam acessíveis e possam ser usados e reusados agora e no futuro. Ressalta-se que a etapa **analisar** do *DataONE*, acaba por ter relação, também, com a etapa de **reavaliar** do DCC, visto que, a partir da análise dos dados, eles podem levar a ajustes nos mesmos, assim como no repensar do processo de curadoria como um todo. Essas mesmas etapas do *DataONE* se relacionam com a etapa de **recuperação** do CVD-CI, pois nela os dados são encontrados, acessados, estruturados, tratados, refinados e se interage com eles (SANT'ANA, 2013).

Tomando como base a etapa de **recuperação** do CVD-CI (Quadro 3), como já mencionado, ela se relaciona com as etapas **integrar, descobrir, analisar e assegurar a qualidade** do *DataONE*. Porém, chama-se atenção para sua relação com o DCC por meio das etapas **acessar, usar e reusar** e também **acompanhamento e participação da comunidade**. Visto que, nesta etapa, é recomendável a participação e/ou *feedback*

dos usuários dos dados ou daqueles que detenham conhecimento do público alvo ou em potencial, bem como das necessidades e usos previstos para um melhor aproveitamento, uso e reuso dos dados armazenados.

Por fim, a etapa descarte do CVD-CI (Quadro 3), relaciona-se com as ações de **preservar** do *DataONE*, pois o descarte está implícito dentro desta etapa e com a etapa de **descartar** do *DCC*, momento em que é decidido quais os dados que poderão ser eliminados do processo de curadoria.

Diante do exposto, pode-se verificar que os modelos *DataONE*, *DCC* e *CVD-CI* têm etapas que se relacionam mais ou menos fortemente. Percebe-se que é preciso mais cuidado na comparação com o modelo *CVD-CI* visto que os fatores impactam fortemente cada etapa do modelo, não sendo suficiente analisar a descrição fornecida da etapa.

A análise feita acaba por ser um pouco discrepante do comparativo de fases dos modelos de ciclo de vida apresentado por Sant'ana (2013) em seu artigo. As principais discrepâncias são: o autor não atribuiu correlação da etapa armazenar do *CVD-CI* com qualquer etapa do *DataONE*, porém verificou-se que há relação sim com as etapas de assegurar a qualidade e de preservar. No que se refere a fase de recuperação, concorda-se parcialmente com Sant'Ana (2013), pois além das etapas de descobrir e analisar propostas pelo autor na comparação dos modelos, verifica-se que há relação também com as etapas de assegurar a qualidade e integrar. Pois, na própria definição de Sant'Ana (2016) destaca-se que uma questão a ser respondida no momento da recuperação é de que forma explicitar e operacionalizar a integração envolvendo as várias estruturas de dados com outros conjuntos de dados, além de se possuir o fator integração permeando tal fase. E por fim, na fase de descarte do *CVD-CI* para a qual não foi identificada relação pelo autor com qualquer fase do *DataONE*, verifica-se uma relação implícita com a etapa de preservar deste modelo, uma vez que ao decidir os dados que serão mantidos no processo de curadoria, naturalmente, escolhe-se aqueles dados que deverão ser eliminados.

Algumas ações do *DCC* estão presentes em mais de uma etapa do *DataONE* como acompanhamento e participação da comunidade, criar ou receber, acessar, usar e reusar. Verifica-se que no *DataONE* algumas ações acabam por ficar implícitas na descrição das etapas, como o descarte e o acompanhamento e participação das comunidades. Também fica implícito na descrição do *CVD-CI* como se dá o acompanhamento e participação da comunidade, pois a descrição do modelo foca em equipes de Ciência da Informação e de Ciência da Computação, mas há outros perfis de comunidade e o próprio público alvo que podem ser envolvidos para garantir o sucesso do processo de curadoria digital como um todo.

Ressalta-se que tanto o *DataONE*, quanto o *DCC* não precisam ter todas as suas etapas seguidas, podendo ser o modelo redimensionado dependendo da necessidade do projeto que se tem em mãos. Inclusive, no site do *DataONE*<sup>4</sup> foram especificados alguns *cases* baseados em *personas* onde o modelo é aplicado de diferentes formas, dependendo da necessidade dos usuários. Frisa-se, novamente, que o *CVD-CI* resume as atividades essenciais de um ciclo de vida e agrega mais de uma etapa encontrada em outros modelos, porém o que na teoria pode parecer uma otimização/simplificação, na prática, pode trazer dificuldade no momento de aplicar o modelo, especialmente porque algumas etapas explicitadas em outros modelos

---

<sup>4</sup> <https://www.dataone.org/user-personas>

tornaram-se fatores no CVD-CI que permeiam suas etapas, deixando ainda mais abstratas como as atividades devem ser executadas.

De uma maneira geral, os modelos analisados ainda carecem de maior detalhamento para poderem ser aplicados diretamente por pesquisadores/curadores. Pois a descrição do que precisa ser feito, por quem deve ser feito e o que não se deve deixar de fora da execução (o que é essencial), assim como o resultado efetivo esperado após a execução de cada etapa (de forma que possa ser medido se a etapa foi alcançada com sucesso ou não) não são bem definidos ou o são de maneira muito superficial. Percebe-se que alguns trabalhos vêm sendo desenvolvidos em busca de sanar esse problema, como, por exemplo, o DCC está aprimorando seu site<sup>5</sup>, vem produzindo capítulos que detalham etapa por etapa do ciclo. Porém, esse trabalho ainda está em estágio inicial de desenvolvimento.

## CONSIDERAÇÕES FINAIS

Por meio da pesquisa realizada foi possível caracterizar os elementos dos ciclos de vida dos dados das iniciativas DCC, DataONE e CVD-CI e elaborar quadros comparativos entre os modelos, alcançando assim o objetivo inicial do trabalho.

O CVD-CI, cujo desenvolvimento é, a priori, voltado para a CI no cenário brasileiro, pode tornar tal área uma referência no gerenciamento de dados, porém algumas de suas etapas simplificam o processo, as quais poderiam estar separadas e não reunir vários estágios em uma mesma etapa.

Todos os modelos podem colaborar com boas práticas de pesquisa ao proporcionar uma gestão efetiva dos dados científicos, contribuindo com uma preservação adequada e desta forma tornando possível que os dados possam ser encontrados, acessados, usados e reutilizados, o que conseqüentemente proporciona o desenvolvimento científico.

Artigo recebido em 03/05/2019 e aprovado em 15/10/2019.

## REFERÊNCIAS

ALBUQUERQUE, M. R. *A curadoria em um repositório institucional: uma análise sob a ótica do ciclo de vida dos dados (CVD-CI) de Sant'Ana*. 2018. 87f. - Dissertação (Mestrado em Ciência da Informação) - Programa de Pós-graduação em Ciência da Informação, Universidade Federal do Ceará, Fortaleza, 2018.

ANDRADE, M. M. *Introdução à metodologia do trabalho científico*. 10 ed. São Paulo: Atlas, 2010.

BORGMAN, C. L. *Big data, little data, no data: scholarship in the networked world*. Massachusetts London, England: The MIT Press Cambridge, 2015.

BORKO, H. Information science: what is it?. *American Documentation*, v. 19, n. 1, p. 3-5, Jan. 1968.

---

<sup>5</sup> <http://www.dcc.ac.uk/resources/curation-reference-manual/chapters-production>

CONEGLIAN, C. S.; SEGUNDO, J. E. S.; SANT'ANA, R. C. S. G. A. Big data: fatores potencialmente discriminatórios em análise de dados. *Em Questão*, Porto Alegre, v. 23, n. 1, p. 62-86, jan/abr. 2017. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/62122>. Acesso em: 10 out. 2017.

CONSELHO NACIONAL DE ARQUIVOS. Resolução n. 40, de de 9 de dezembro de 2014. Dispõe sobre os procedimentos para a eliminação de documentos no âmbito dos órgãos e entidades integrantes do Sistema Nacional de Arquivos - SINAR. Disponível em: <http://www.conarq.arquivonacional.gov.br/cgi/cgilua.exe/sys/start.htm?inoid=1017&sid=46>. Acesso em: 17 dez. 2014.

DATAONE. *Primer on data management: what you always wanted know*. Disponível em: <https://www.dataone.org/sites/all/documents/DataONEBPPrimero20212.pdf>. Acesso em: 20 jun. 2018.

DATAONE. *What is DataONE? Online*. Disponível em: <https://www.dataone.org/what-dataone>. Acesso em: 20 jun. 2018.

DIGITAL CURATION CENTRE. *Curation lifecycle model*. Disponível em: <http://www.dcc.ac.uk/resources/curation-lifecycle-model>. Acesso em: 20 jun. 2018.

GIL, A. C. *Como elaborar projetos de pesquisa*. 6. ed. São Paulo: Atlas, 2017.

MAYER-SCHÖNBERGER, V.; CUKIER, K. *Big data: a revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt, 2013.

MÖLLER, K. Lifecycle models of data-centric systems and domains: the abstract data lifecycle model. *Semantic Web Journal*, v. 4, n. 1, 2013. DOI:10.3233/SW-2012-0060. Disponível em: <http://www.semantic-web-journal.net>. Acesso em: 15 jul. 2013.

NATIONAL RESEARCH COUNCIL. *A question of balance: private rights and the public interest in scientific and technical databases*. Washington, DC: National Academies Press, 1999.

NIST. *Big data definitions*. Disponível em: <http://dx.doi.org/10.6028/NIST.SP.1500-1>. Acesso em: 03 set. 2016.

OLIVEIRA, A. C. S.; SILVA, E. M. Ciência aberta: dimensões para um novo fazer científico. *Informação & Informação*, Londrina, v. 21, n. 2, p. 5-39, maio/ago. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27666>. Acesso em: 10 out. 2017.

POMERANTZ, J. *Metadata*. Cambridge, Massachusetts; London, England: The MIT Press, 2015.

RICE, R.; SOUTHALL, J. *The data librarian's handbook*. London: Facet Publishing, 2016.

SALES, L. F.; CAVALCANTI, M. T. Seleção e avaliação de coleções de dados digitais de pesquisa: uma possível abordagem metodológica. *Informação & Tecnologia (ITEC)*, v. 2, n. 2, p. 88-105, jul./dez. 2015. Disponível em: <http://periodicos.ufpb.br/index.php/itec/article/view/34134/17537>. Acesso em: 10 out. 2017.

SANT'ANA, R. C. G. 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* [...]. Florianópolis, 2013. Disponível em:

<http://www.brapci.inf.br/index.php/article/view/0000013772/fac7d2fecfd2c83f1e9208204b2fefed10>. Acesso em: 20 jun. 2018.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. *Informação & Informação*, Londrina, v. 21, n. 2, p. 116–142, maio/ago. 2016. DOI: <http://dx.doi.org/10.5433/1981-8920.2016v21n2p116>. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940/20124>. Acesso em: 20 jun. 2018.

SANTOS, T. N. C. *Curadoria digital: o conceito no período de 2000 a 2013*. 2014. 165 f. Dissertação (Mestrado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, Universidade de Brasília, Brasília, 2014.

SAYÃO, L. F.; SALES, L. F. Curadoria geral: um novo patamar para a preservação de dados digitais de pesquisa. *Informação & Sociedade*, v. 22, n. 3. p. 179-191, set./dez. 2012. Disponível em: <http://www.ies.ufpb.br/ojs/index.php/ies/article/viewFile/12224/8586>. Acesso em: 10 out. 2017.

SAYÃO, L. F.; SALES, L. F. *Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores*. Rio de Janeiro: CNEN/IEN, 2015. Disponível em: <http://carpedien.ien.gov.br/bitstream/ien/1624/1/GUIADEDADOSDEPESQUISA.pdf>. Acesso em: 10 out. 2017.

SAYÃO, L. F.; SALES, L. F. Dados abertos de pesquisa: ampliando o conceito de acesso livre. *Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*, Rio de Janeiro, v. 8, n. 2, p. 76-92. 2014. Disponível em: <http://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611/1252>. Acesso em: 05 jul. 2017.

SETZER, V. W. Dado, informação, conhecimento e competência. In: SETZER, V.W (org.). *Os meios eletrônicos e a educação: uma visão alternativa*. São Paulo: Escrituras, Coleção “Ensaio Transversais”, 2001. V. 10. Disponível em: <https://www.ime.usp.br/~vwsetzer/dado-info.html>. Acesso em: 20 jun. 2018.

SIEBRA, S. de A. et al. Curadoria digital: além da questão da preservação digital. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. *Anais [...]*. Florianópolis: UFSC, 2013. Disponível em: <http://www.egov.ufsc.br/portal/conteudo/curadoria-digital-al%C3%A9m-da-quest%C3%A3o-da-preserva%C3%A7%C3%A3o-digital-o>. Acesso em 20 jun. 2018.

SIEBRA, S. A.; BORBA, V. R.; MIRANDA, M. J. K. F. O. Curadoria digital: um termo interdisciplinar. *Informação & Tecnologia (ITEC)*, v. 3, n. 2, p. 21-38, jul./dez. 2016. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/itec/article/view/38408/20163>. Acesso em 5 jul. 2018.

SIEBRA, S. A.; BORBA, V. R. Análise, tendências e perspectivas da produção científica em curadoria digital. In: OLIVEIRA, H. P. C.; VIDOTTI, S. A. B. G. (org.). *Informação e tecnologias: desenhando fronteiras científicas*. João Pessoa: Editora Ufpb, 2018. p. 97-112.

SILVEIRA, D. T.; CORDOVA, F. P. A pesquisa científica. In: GERHARDT, T. E.; SILVEIRA, D. T. (org.). *Métodos de pesquisa*. Porto Alegre: Editora da UFRGS, 2009.

STRASSER, C. et al. *Primer on data management: what you always wanted to know*. California: CDL, 2012. Disponível em: <http://escholarship.org/uc/item/7tf5q7n3#page-1>. Acesso em: 13 jun. 2018.

YAMAOKA, E. J. Ontologia para mapeamento da dependência tecnológica de objetos digitais no contexto da curadoria e preservação digital. *AtoZ*, v. 1, n. 2, p. 65-78, jan./dez. 2012. Disponível em: <https://revistas.ufpr.br/atoz/article/view/41313/25240>. Acesso em 5 jul. 2018.