

Os ciclos de vida dos dados sob a perspectiva da Curadoria digital

Gabriela de Oliveira Souza^I

<https://orcid.org/0000-0001-7519-6624>

Daniel Martínez-Ávila^{II}

<https://orcid.org/0000-0003-2236-553X>

Maria José Vicentini Jorente^I

<https://orcid.org/0000-0002-0492-0918>

^I Universidade Estadual Paulista, Marília, SP, Brasil

^{II} Universidad de León, Leon, Espanha

Resumo: O objetivo geral do presente estudo é analisar modelos de ciclo de vida dos dados sob a perspectiva da Curadoria Digital. Os objetivos específicos são: construir um referencial teórico sobre Curadoria Digital; realizar um estudo comparativo entre o Modelo do Ciclo de Vida da Curadoria do *Digital Curation Centre*, o *DataONE Lifecycle Model*, o *Data Life Cycle Labs Model* e o *Big Data Life Cycle Model*. O presente estudo é de natureza qualitativa, e de caráter descritivo e exploratório. O método utilizado foi o *Design Thinking*. A partir da análise de cada um dos ciclos de vida dos dados, foi possível observar algumas correspondências com o Ciclo de Vida da Curadoria Digital. Destaca-se que nenhum dos modelos apresentou elementos para a participação da comunidade. Nesse contexto, é possível adaptar o ciclo de vida da Curadoria Digital conforme as necessidades da comunidade a qual se destina. Faz-se necessário considerar a estrutura de outros modelos para curadoria de dados, e compreender suas possíveis aplicações e convergências com a Curadoria Digital, com o intuito de incluir suas práticas ao modelo. Isso possibilitaria adaptações ao modelo original conforme a necessidade da comunidade e a natureza dos dados tratados.

Palavras-chave: curadoria digital; ciclo de vida dos dados; informação e tecnologia

1 Introdução

As transformações das Tecnologias de Informação e Comunicação suscitaram uma preocupação com a preservação dos dados e informações em seus diversos meios de criação, armazenamento e compartilhamento (Souza; Oliveira; D'Ávila; Chaves, 2012). Entretanto, as tecnologias permitem a criação, o armazenamento e o compartilhamento de uma grande quantidade de informações, mas apresentam fragilidades que podem colocar em risco a salvaguarda dos objetos digitais. Isto ocorre porque eles apresentam características distintas dos documentos tradicionais e necessitam de diferentes metodologias para seu tratamento e gestão (Sayão; Sales, 2012).

Nesse cenário, a Curadoria Digital (CD) apresenta práticas e metodologias contínuas para preservação e gestão de objetos digitais a longo prazo, capazes de solucionar tais impasses. Ressalta-se a importância de modelos de referência para a realização dos processos relativos à curadoria, à preservação e ao tratamento de dados e documentos digitais, como o *DCC Curation Lifecycle Model* (Modelo do Ciclo de Vida da Curadoria Digital), o *DataONE Lifecycle Model* (Modelo do Ciclo de Vida DataONE), o *Data Life Cycle Labs (DLCL) Model* (Modelo DLCL) e o *Big Data Life Cycle Model* (Modelo do Ciclo de Vida do Big Data).

Assim, o presente artigo questiona: Quais os elementos presentes no Modelo do Ciclo de Vida da Curadoria Digital do *Digital Curation Centre* (DCC)? Quais as convergências entre os outros ciclos de vida dos dados e o modelo do DCC? Como os outros modelos de ciclo de vida dos dados poderiam se beneficiar da Curadoria Digital? Considera-se que a Curadoria Digital pode ser aplicada em diferentes contextos e é transversal a todos os modelos de ciclo de vida dos dados. Além disso, está presente no processo de gestão de dados como um todo, de modo que os outros modelos apresentados poderiam ser aprimorados, considerando-se as ações da Curadoria Digital apresentadas no modelo do DCC.

Nesse contexto, o objetivo geral do presente estudo é analisar modelos de ciclo de vida dos dados sob a perspectiva da Curadoria Digital. Os objetivos específicos são: construir um referencial teórico sobre Curadoria Digital;

realizar um estudo comparativo entre o Modelo do Ciclo de Vida da Curadoria, do *Digital Curation Centre* e outros três modelos de ciclo de vida dos dados: Modelo do Ciclo de Vida DataONE, Modelo DLCL e Modelo do Ciclo de Vida do *Big Data*.

2 Procedimentos: revisão narrativa e análise comparativa de modelos

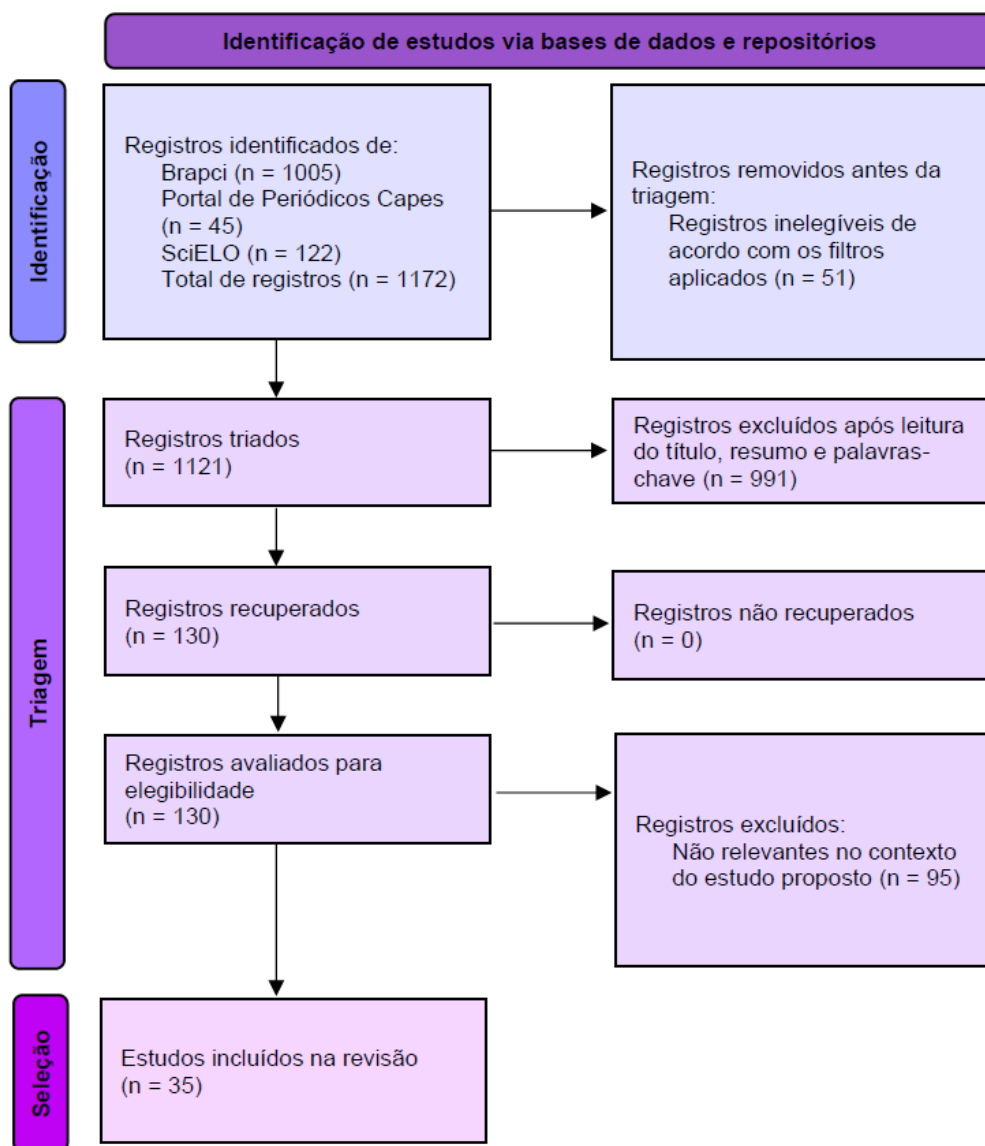
O presente estudo é de natureza qualitativa, e de caráter descritivo e exploratório. O método utilizado foi o *Design Thinking*, que busca encontrar soluções por meio de ideias baseadas no *design* (Nakano; Oliveira; Jorente, 2018). Segundo Nakano, Oliveira e Jorente (2018), o método possui três fases: fase de imersão, fase de ideação e fase de prototipação. A fase de imersão busca identificar problemas e necessidades. Já a fase de ideação consiste no uso de elementos de *design* para encontrar soluções para os problemas identificados anteriormente. A fase de prototipação envolve o desenvolvimento de uma proposta de solução para os problemas identificados nas outras etapas (Nakano; Oliveira; Jorente, 2018).

Na fase de imersão foi realizado um levantamento bibliográfico e revisão de literatura nas bases de dados Brapci, Portal de Periódicos Capes e SciELO, devido à sua relevância para a Ciência da Informação no âmbito nacional e internacional. Foi utilizada a seguinte expressão de busca: (“Curadoria Digital” OR “*Digital Curation*”) OR (“Ciclo de Vida dos Dados” OR “*Data Lifecycle*”). Foram recuperados artigos publicados em periódicos entre 2010 e 2025, em português, inglês e espanhol e disponíveis em acesso aberto. A seleção dos registros foi feita com base no resumo, no título e nas palavras-chaves, de acordo com a relevância do material para o estudo em questão, de modo que foram selecionados apenas artigos convergentes à temática proposta, ou seja, com foco na Curadoria Digital (CD) e nos ciclos de vida dos dados. Além disso, foi utilizado o material disponibilizado pelo *Digital Curation Centre* (DCC), referência para os estudos sobre Curadoria Digital.

A revisão de literatura foi realizada com base no fluxograma *Preferred Reporting Items for Systematic reviews and Meta-Analyses* (PRISMA), um conjunto de diretrizes para o relato da revisão sistemática de literatura

(PRISMA, 2020). A princípio, o PRISMA foi desenvolvido para revisões sistemáticas de estudos na área da saúde, mas sua aplicabilidade se estende a pesquisas de qualquer área do conhecimento, pesquisas quantitativas ou qualitativas e em conjunto com diversos métodos de pesquisa (Page *et al*, 2022). Nesse sentido, a Figura 1 apresenta o fluxograma PRISMA utilizado para registro e apresentação da revisão de literatura.

Figura 1 - Fluxograma PRISMA



Fonte: Elaborado pelos autores com base em PRISMA (2020).

Conforme apresentado na figura anterior, foram recuperados 1172 artigos: 1105 na BRAPCI, 45 no Portal de Periódicos CAPES e 122 no SciELO.

Após o processo de avaliação e triagem, foram selecionados, ao todo, 35 artigos para o presente estudo.

Na fase de ideação, foi realizada a análise da literatura selecionada a partir do fluxograma PRISMA e a construção do referencial teórico, além da análise dos elementos que compõem os modelos do ciclo de vida dos dados elencados para o presente estudo. Foram selecionados três modelos de ciclo de vida dos dados para análise: Modelo do Ciclo de Vida DataONE, Modelo DLCL e o Modelo do ciclo de vida do *Big Data*, de Line Pouchard.

A última fase consistiu na identificação das convergências entre os modelos apresentados e a CD por meio de uma análise comparativa. A análise foi feita por meio da comparação do Ciclo de Vida da CD do DCC com os três ciclos de vida dos dados selecionados, que não incluem diretamente a CD, mas apresentam elementos convergentes com suas ações e podem ser adaptados conforme a necessidade da instituição que pretende aplicá-los. Foram considerados os seguintes critérios: correspondências com o modelo do Ciclo de Vida do DCC, aderência aos princípios FAIR, aplicabilidade e elementos de participação da comunidade. Destaca-se que, para o presente estudo, considera-se enquanto participação da comunidade a atuação de indivíduos não especialistas, ou seja, do público em geral nos processos curatoriais por meio de atividades participativas.

Referente à análise de dados, a técnica utilizada foi o *visual thinking*, uma técnica própria do *Design Thinking*, que permite a expressão visual das ideias para uma melhor compreensão. Os instrumentos utilizados foram os quadros explicativos (Nakano; Oliveira; Jorente, 2018).

3 Curadoria Digital: principais conceitos e o modelo do DCC

Segundo o *Digital Curation Centre* (DCC), a Curadoria Digital (CD) compreende a gestão e a preservação de objetos e informações digitais a longo prazo, por meio de ações que abarcam manter, preservar e agregar valor aos dados digitais ao longo de todo seu ciclo de vida (Digital Curation Centre, 2004). A CD desenvolveu-se a partir de conceitos presentes na curadoria de museus, na curadoria de dados e na preservação digital, e convergiu tais

conceitos em ações relacionadas à criação, preservação, compartilhamento e acesso à informação na Web (Santos, 2014).

O termo Curadoria foi utilizado de diferentes formas antes de ser utilizado no contexto da Curadoria Digital. Os termos Curadoria e curador apresentam diferentes significados dependendo do contexto em que são utilizados, no entanto, na maioria das vezes eles estão relacionados a ações de preservação e cuidado (Amaral, 2012).

Castilho (2015) ainda afirma que a CD emerge como uma solução para o grande volume de dados e informações presentes na Web. A CD possui metodologias e estratégias que viabilizam o acesso e a confiabilidade dos objetos digitais, além de abrir espaço para a colaboração entre os profissionais da informação e os internautas (Sayão; Sales, 2012).

A implementação da CD apresenta vantagens a curto e a longo prazo. A curto prazo, o uso de padrões, a verificação da autenticidade, e a preservação, ocasionam uma melhora nas formas de acesso, uso e compartilhamento de informações. A longo prazo, a CD apresenta como principais vantagens a preservação e proteção dos objetos digitais, pois evita a perda de informações e a obsolescência tecnológica (Jorente; Silva; Padua, 2021).

As ações, técnicas, estratégias e metodologias da CD contribuem de forma significativa com a qualidade e confiabilidade dos dados (Sayão; Sales, 2012). Assim, o uso de padrões, diretrizes e princípios de modo convergente às práticas da CD pode apresentar benefícios ao processo como um todo. Dentre eles, destaca-se os princípios FAIR (*Findable, Accessible, Interoperable, Reusable*), que orientam as boas práticas para a gestão de objetos informacionais digitais.

Os princípios FAIR não sugerem tecnologias ou apresentam soluções para a sua aplicação, mas indicam boas práticas para que os profissionais que atuam com gestão de dados e objetos informacionais digitais os tornem encontráveis, acessíveis, interoperáveis e reutilizáveis (Campos; Dias; Sousa, 2023). Ao considerar a CD enquanto área emergente e transversal a outras áreas (Sayão; Sales, 2012), como a preservação digital, por exemplo, os princípios

FAIR podem auxiliar na aplicação prática da CD, bem como na avaliação dos resultados do processo.

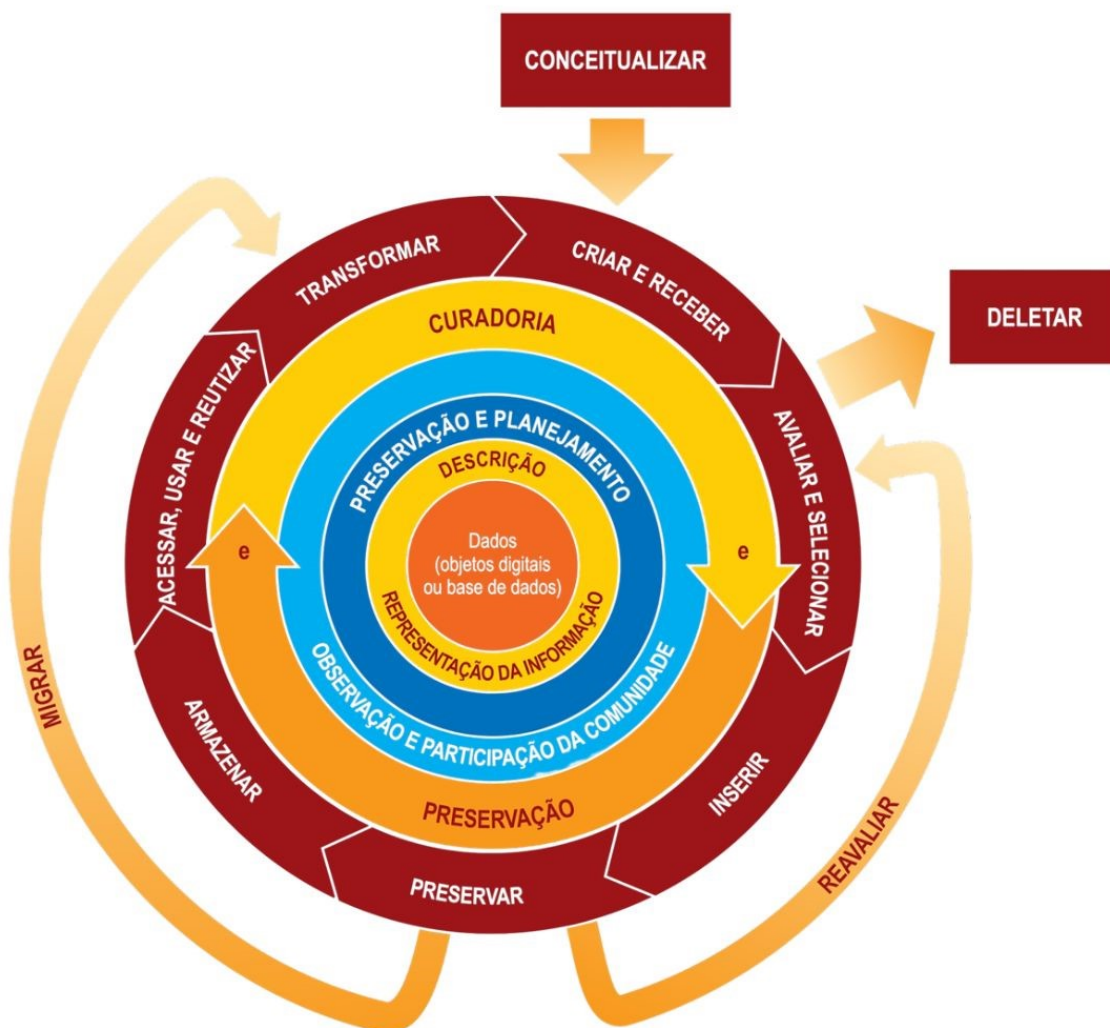
O primeiro princípio, *findable* (encontrável), aponta que dados e metadados devem ser facilmente encontrados, tanto por humanos quanto por máquinas. De acordo com o princípio *accessible* (acessível), os dados e metadados, além de encontráveis, devem se acessíveis, ou seja, o internauta deve conseguir acessar os dados.

Além disso, os dados e metadados devem estar representados e organizados de modo a permitir a interoperabilidade com diferentes aplicativos e softwares, o que representa o terceiro princípio, *interoperable* (interoperável). Por fim, os princípios FAIR buscam otimizar a reutilização de dados, de modo que o último princípio, *reusable* (reutilizável), aponta que dados e metadados devem ser descritos de forma eficiente para que sejam utilizados e reutilizados em diferentes contextos (GOFAIR, 2025; Veiga *et al*, 2025).

Tais princípios podem apoiar as ações e práticas da CD, ao fornecerem uma base consistente para a verificação e validação de todo o processo. Para sua aplicação, são necessários, também, modelos de referência que possam nortear a implementação das ações da CD, como o modelo do Ciclo de Vida da Curadoria Digital do DCC.

Higgins (2008) apresentou o Ciclo de Vida da Curadoria Digital do DCC (*DCC Curation Lifecycle Model*), uma representação gráfica das etapas da CD, compreendida como um processo contínuo e cíclico. A figura a seguir (Figura 2) apresenta uma versão traduzida do modelo.

Figura 2 - Modelo do Ciclo de Vida da Curadoria Digital do DCC



Fonte: Traduzido pelos autores de Higgins (2008).

O Modelo do Ciclo de Vida da Curadoria Digital (Figura 1) é dividido em **ações essenciais** (dados; descrição e representação da informação; preservação e planejamento; participação e observação da comunidade; e curadoria e preservação), **ações sequenciais** (conceitualizar; criar e receber; avaliar e selecionar; inserir; preservar; armazenar; acessar, usar e reutilizar; e transformar) e **ações ocasionais** (deletar, reavaliar e migrar) (Higgins, 2008).

Os **dados** abrangem qualquer informação em formato digital binário (Higgins, 2008). As ações de **descrição e representação da informação** dizem respeito à atribuição de metadados aos objetos digitais (Higgins, 2008).

Preservação e planejamento consistem no planejamento das demais ações do ciclo de vida (Yamaoka, 2012). A etapa de **participação e observação da comunidade** corresponde ao desenvolvimento colaborativo de padrões, recursos e *softwares* pela comunidade de profissionais. **Curadoria e preservação**, envolve o encaminhamento das ações planejadas para promover a curadoria dos objetos informacionais digitais (Higgins, 2008).

Conceitualizar compreende as práticas de concepção e planejamento da criação de dados, bem como dos métodos e técnicas utilizados na coleta e no armazenamento (Yamaoka, 2012). As ações **criar e receber** correspondem à criação de objetos digitais e inclusão de metadados, além da recepção de dados e metadados coletados em outras ações do modelo. **Avaliar e selecionar**, consiste na avaliação dos objetos informacionais digitais e seleção daqueles que necessitam de preservação a longo prazo (Higgins, 2008).

Inserir representa a transferência dos dados para um arquivo, repositório digital, centro de dados ou instituição responsável por sua salvaguarda. **Preservar** compreende as práticas necessárias para assegurar a preservação a longo prazo (Higgins, 2008). A ação seguinte, **armazenar**, consiste no armazenamento dos dados de forma segura em conformidade com os padrões vigentes. **Acessar, usar e reutilizar** garantem que as informações estejam disponíveis e acessíveis aos sujeitos informacionais. **Transformar** corresponde à criação de novos dados a partir do dado original (Higgins, 2008).

Deletar consiste na eliminação dos dados que não foram selecionados para a preservação a longo prazo. Na ação de **reavaliar**, os dados que obtiveram falhas nos procedimentos de validação passam novamente pela avaliação e pela seleção, com o intuito de garantir a confiabilidade do processo de CD (Higgins, 2008). A última ação consiste em **migrar** os dados para um novo formato, a fim de evitar a obsolescência tecnológica e garantir a compatibilidade com o sistema de armazenamento (Sayão; Sales, 2012). As ações do Ciclo de Vida da Curadoria Digital estão sintetizadas no quadro a seguir (Quadro 1):

Quadro 1 - Ações do Modelo do Ciclo de Vida da Curadoria Digital do DCC

Ações essenciais	Ações sequenciais	Ações ocasionais
- Dados - Descrição e representação da informação - Preservação e planejamento - Participação e observação da comunidade - Curadoria e preservação	- Conceitualizar - Criar e receber - Avaliar e selecionar - Inserir - Preservar - Armazenar - Acessar, usar e reutilizar - Transformar	- Deletar - Reavaliar - Migrar

Fonte: Elaborado pelos autores.

Desse modo, o ciclo de vida proposto pelo DCC apresenta diversas ações contínuas necessárias ao desenvolvimento da CD.

4 Modelos de ciclo de vida dos dados

Devido às transformações na produção científica, os fluxos informacionais tornaram-se mais intensos e os pesquisadores passaram a produzir dados mais complexos (Pouchard, 2016). Nesse contexto, os modelos de ciclo de vida dos dados representam graficamente os processos e atividades necessárias para a gestão de dados em um projeto ou em uma instituição.

Os modelos também são uma forma de comunicar os processos de gestão de dados ao público-alvo, que pode ser composto por pesquisadores, cientistas de dados, curadores, profissionais da tecnologia, bibliotecários e demais profissionais da informação responsáveis ou interessados no processo. Desse modo, os modelos podem ser utilizados para mapear as etapas do ciclo de vida dos dados para a identificação de lacunas e tomada de decisões sobre o processo de gestão de dados (Pouchard, 2016).

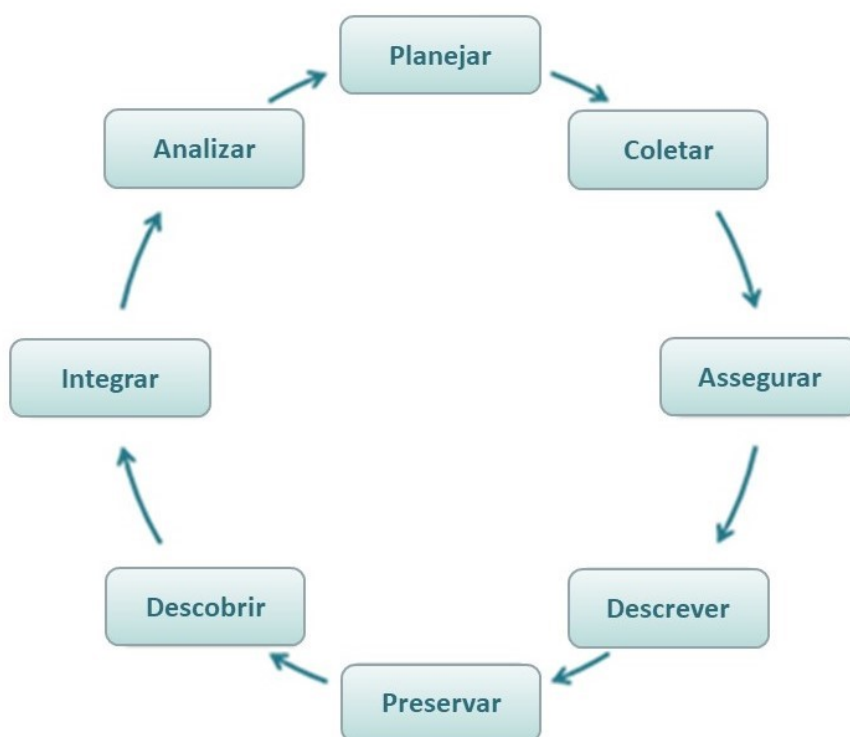
No entanto, os modelos geralmente representam os processos relacionados aos dados de maneira ordenada e linear, o que não reflete a realidade, e podem ignorar a diversidade de atividades práticas presentes em todo o processo (Pouchard, 2016). Tais aspectos podem ser solucionados se consideradas as principais características e necessidades da comunidade a qual

os dados se destinam. Assim, no contexto da Curadoria Digital, considera-se relevante a análise de outros modelos de ciclo de vida dos dados, além do modelo proposto pelo DCC, com o intuito de verificar sua aplicabilidade.

4.1 Modelo do Ciclo de Vida DataONE

O Ciclo de Vida DataONE é um modelo proposto pela iniciativa *Data Observation Network for Earth* (DataONE), para sistematizar e organizar as práticas de Curadoria e Preservação de dados. A DataONE reúne centros de dados e arquivos que propõe a curadoria, a preservação e o acesso à dados sobre as Ciências da Terra para uso a longo prazo, além de fornecer orientações acerca de boas práticas sobre curadoria de dados (DataONE, 2019; Plale; Kouper, 2017; Pouchard, 2016), conforme pode ser observado na figura a seguir (Figura 3).

Figura 3 - Modelo do Ciclo de Vida DataONE



Fonte: Traduzido pelos autores de Plale e Kouper (2017).

O Modelo do Ciclo de Vida DataONE é composto por oito etapas: planejar, coletar, assegurar, descrever, preservar, descobrir, integrar e analisar (DataONE, 2019; Anjos; Dias, 2019; Plale; Kouper, 2017; Pouchard, 2016).

A primeira etapa do ciclo, **planejar**, consiste no planejamento de todo o processo de gestão de dados, como por exemplo: métodos para coleta de dados, escolha do repositório, estratégias para organização dos dados, informações para gestão dos dados, escolha dos padrões de metadados, requisitos para o compartilhamento dos dados, definição do plano de preservação e do orçamento (Anjos; Dias, 2019).

A etapa seguinte, **coletar**, corresponde à coleta dos dados primários. **Assegurar** corresponde à implementação de critérios e atividades que garantam a qualidade, a validação e a autenticidade dos dados a longo prazo. **Descrever** consiste na descrição dos dados por meio de padrões de metadados, além de permitir a interoperabilidade dos dados. Na quinta etapa, **preservar**, são elaborados os planos de preservação a curto e longo prazo (Anjos; Dias, 2019).

Na etapa **descobrir** são identificados novos conjuntos de dados que complementem a pesquisa, além da divulgação dos dados. **Integrar** representa a integração de diversos conjuntos de dados para proporcionar novos estudos. A última etapa, **analisar**, compreende a análise e apresentação dos dados por meio de publicações e divulgação científica (Anjos; Dias, 2019). Nesse sentido, as etapas do Ciclo de Vida dos Dados DataONE estão sintetizadas no quadro a seguir (Quadro 2):

Quadro 2 - Etapas do Modelo do Ciclo de Vida DataONE

Etapas
Planejar
Coletar
Assegurar
Descrever
Preservar
Descobrir
Integrar
Analisar

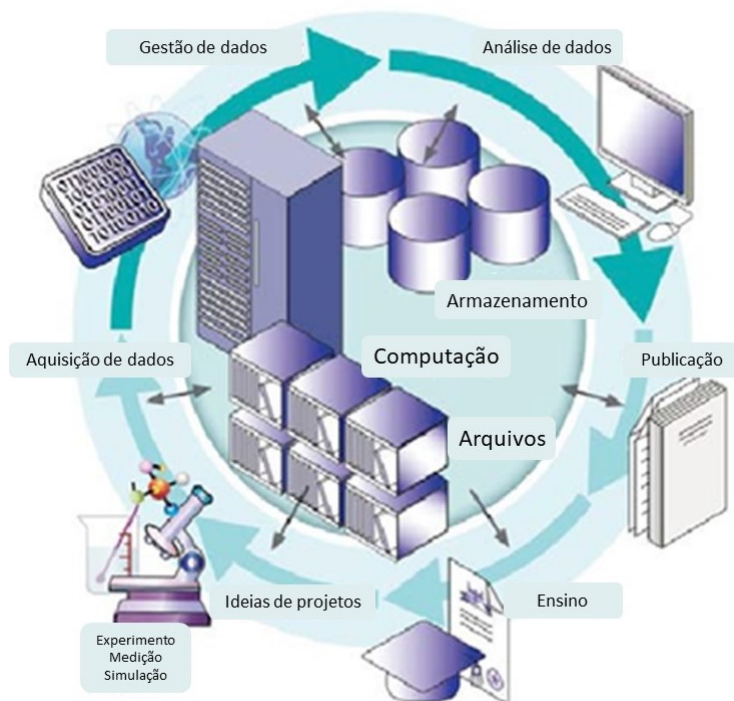
Fonte: Elaborado pelos autores.

Destaca-se que as etapas do Modelo do Ciclo de Vida DataONE são independentes, e podem ser realizadas conforme a necessidade. Ainda que a realização de todas as etapas não seja uma exigência para a aplicação do modelo, atividades relativas à preservação, descrição e autenticidade são essenciais para qualquer projeto (Anjos; Dias, 2019; Plale; Kouper, 2017), bem como princípios que norteiam tais atividades.

4.2 Modelo *Data Life Cycle Labs*

Data Life Cycle Labs (DLCL) são iniciativas do projeto *Large Scale Data Management and Analysis* (LSDMA) da Associação Helmholtz de centros de investigação na Alemanha. A comunidade considerada pelo modelo envolve pesquisadores que trabalham com temáticas relacionadas a energia, terra e meio ambiente, saúde, tecnologias e estrutura da matéria (Pouchard, 2016). O Modelo DLCL pode ser visualizado na figura a seguir (Figura 4):

Figura 4 - Modelo DLCL



Fonte: Traduzido pelos autores de Pouchard (2016).

O Modelo DLCL é composto por seis fases na parte externa do ciclo e três processos na parte interna, que perpassam todo o processo. Os processos

internos são: arquivos, computação e armazenamento. As fases, por sua vez, são: ideias de projetos, aquisição de dados, gestão de dados, análise de dados, publicação e ensino (Pouchard, 2016).

A fase inicial do modelo em questão corresponde às **ideias de projetos**, que por sua vez, se subdividem em três itens: **experimento**, **medição**, **simulação**. Os dados são gerados por meio de medições, observações, experimentos e simulações, e precisam ser preservados. Assim, a **aquisição de dados** é realizada pelos pesquisadores durante suas pesquisas (Pouchard, 2016).

A fase de **gestão de dados** inclui processos como: armazenamento dos dados, acesso, uso de metadados e ontologias para identificação e recuperação dos dados, padronização de formatos e segurança de dados. A **análise de dados** corresponde à próxima fase do processo, seguida de **publicação e ensino**, que contemplam a divulgação científica (Pouchard, 2016).

O modelo DLCL foi criado para aplicação no contexto das pesquisas financiadas e desenvolvidas por pesquisadores da Associação Helmholtz (Pouchard, 2016). No modelo, há a presença de arquivos, computação e armazenamento enquanto elementos centrais, que perpassam todas as etapas do ciclo, no entanto, tais elementos apenas são mencionados. O Quadro 3 apresenta as etapas do Modelo DLCL:

Quadro 3 - Fases do Modelo DLCL

Processos gerais	Fases
• Arquivos • Computação • Armazenamento	• Ideias de projetos • Aquisição de dados • Gerenciamento de dados • Análise de dados • Publicação • Ensino

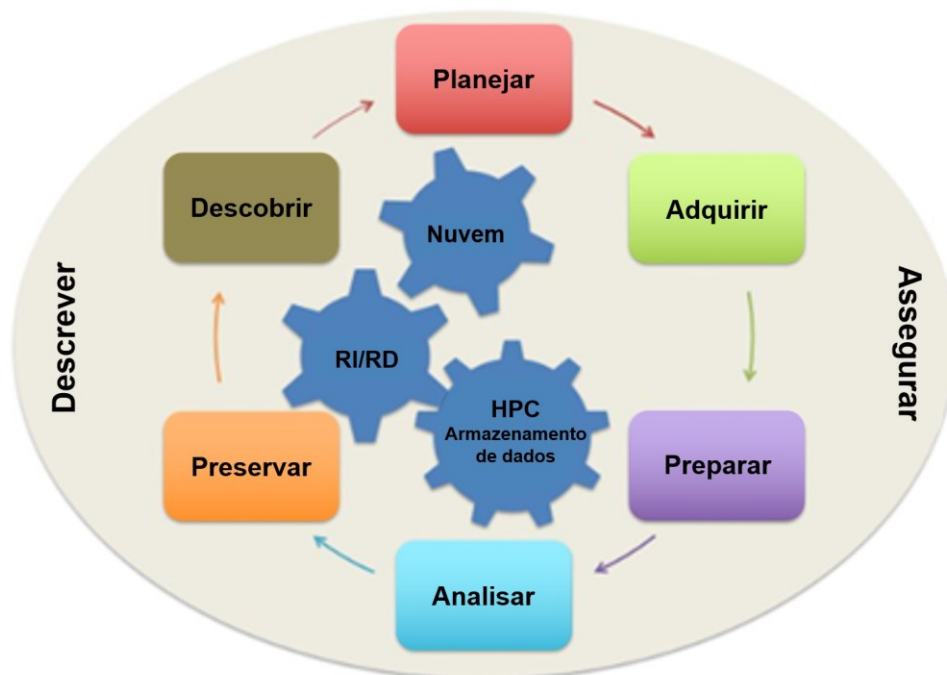
Fonte: Elaborado pelos autores.

As iniciativas do DLCL contemplam recursos para análise e gestão de dados, de acordo com o ciclo de vida dos dados e com as necessidades da comunidade a qual se destina (Pouchard, 2016).

4.3 Modelo do ciclo de vida do *Big Data*

O Modelo do ciclo de vida do *Big Data* foi desenvolvido por Line Pouchard a partir da análise dos dois modelos anteriores, e sob a perspectiva do *Big Data*. O modelo considera a perspectiva da curadoria de dados, e apresenta as etapas necessárias para a gestão e curadoria do *Big Data* (Pouchard, 2016), conforme pode ser analisado na Figura 5.

Figura 5 - Modelo do ciclo de vida do *Big Data*



Fonte: Traduzido pelos autores de Pouchard (2016).

O modelo do ciclo de vida do *Big Data* é composto por três tipos de elementos: atividades presentes em todo o ciclo, atividades e infraestrutura computacional. As atividades presentes em todo o ciclo são: descrever e assegurar. As demais atividades são: planejar, adquirir, preparar, analisar, preservar e descobrir. A infraestrutura computacional é composta por nuvem (infraestrutura de armazenamento em nuvem), RI/RD (repositório institucional - IR / repositório disciplinar - DR) e HPC armazenamento de dados (armazenamento de dados em um centro de computação de alto desempenho - HPC).

O modelo pode ser iniciado a partir de diferentes etapas, e os conjuntos de dados de um projeto podem estar alinhados à diferentes atividades do ciclo, de modo que nem todas as atividades serão aplicáveis a todos os tipos de dados. Nesse sentido, a atividade **descrever** corresponde à descrição dos dados e procedimentos aplicados a cada etapa. **Assegurar** compreende a garantia de qualidade dos dados (Pouchard, 2016).

A atividade **planejar** compreende o planejamento dos procedimentos de seleção de dados para preservação. **Adquirir** se refere aos processos de produção dos dados ao longo do tempo. **Preparar** consiste na preparação dos conjuntos de dados para posterior análise. Ao tratar o *Big Data*, a **análise** representa uma etapa complexa do ciclo, uma vez que demanda métodos e técnicas distintas e depende do tipo de pesquisa que o cientista está desenvolvendo. **Preservar** diz respeito aos processos necessários para a preservação dos dados a longo prazo. Por fim, **descobrir** envolve os processos necessários para garantir que os conjuntos de dados possam ser encontrados pela comunidade de modo geral (Pouchard, 2016).

Os elementos centrais do ciclo, representados por engrenagens, indicam a infraestrutura computacional necessária ao *Big Data*. **Nuvem** corresponde à infraestrutura necessária para suporte dos dados armazenados em nuvem. **RI/RD** representa os repositórios institucionais (RI) e os repositórios disciplinares (RD). Por fim, **HPC armazenamento de dados** representa o armazenamento de dados em um centro de computação de alto desempenho (HPC). O quadro a seguir (Quadro 4) representa as atividades do modelo em questão:

Quadro 4 - Atividades do Modelo do ciclo de vida do Big Data

Atividades presentes em todo o ciclo	Atividades	Infraestrutura computacional
• Descrever • Assegurar	• Planejar • Adquirir • Preparar • Analisar • Preservar • Descobrir	• Armazenamento em nuvem • Repositórios – RI/RD • Armazenamento de dados em HPC

Fonte: Elaborado pelos autores.

Segundo Pouchard (2016), a curadoria de *Big Data* se assemelha a outros tipos de curadoria, entretanto, os processos apresentam desafios diferentes, e a perspectiva da curadoria não foi abordada nos outros modelos de modo que pudesse ser aplicada ao *Big Data*, por esse motivo, foi identificada a necessidade de criação de um novo ciclo.

5 Resultados

A partir da análise de cada um dos ciclos de vida dos dados, foi possível observar algumas correspondências com o Ciclo de Vida da Curadoria Digital. O quadro a seguir (Quadro 5) apresenta a comparação entre o Modelo do Ciclo de Vida da CD e os demais modelos analisados. Destaca-se que os elementos dos outros ciclos foram ordenados no quadro conforme as ações da CD presentes no modelo do DCC, e que alguns se repetem por corresponderem a mais de uma ação da CD.

Quadro 5 - Correspondências entre o Ciclo de Vida da Curadoria Digital do DCC e os ciclos de vida dos dados analisados

TIPOS DE AÇÕES DA CD	CICLOS DE VIDA DOS DADOS			
	Ciclo de Vida da Curadoria Digital	Modelo do Ciclo de Vida DataONE	Modelo DLCL	Modelo do ciclo de vida do Big Data
Ações essenciais	Dados	Coletar	Arquivos Aquisição de dados	Adquirir
	Descrição e representação da informação	Descrever	Gerenciamento de dados Análise de dados	Descrever
	Preservação e planejamento	Planejar Preservar	Armazenamento	Planejar
	Participação e observação da comunidade			
	Curadoria e preservação	Preservar	Gerenciamento de dados	
Ações sequenciais	Conceitualizar		Gerenciamento de dados	Analisar
	Criar e receber		Computação	Preparar
	Avaliar e selecionar	Assegurar	Gerenciamento de dados	Assegurar
	Inserir		Armazenamento	
	Preservar	Preservar	Gerenciamento de dados	Preservar
	Armazenar		Armazenamento	
	Acessar, usar e reutilizar	Descobrir	Ideias de projetos	Descobrir
	Transformar	Integrar Analisar	Publicação Ensino	Descobrir
Ações ocasionais	Deletar			
	Reavaliar			
	Migrar			

Fonte: Elaborado pelos autores.

O Quadro 5 apresenta as correspondências entre os ciclos de vida dos dados analisados anteriormente. As ações da CD que não apresentaram

correspondências com os outros modelos foram representadas na cor cinza. Todos os ciclos analisados apresentam elementos que correspondem a algumas ações essenciais e sequenciais da CD, mas não apresentam ações que correspondem às ações ocasionais. As correspondências foram analisadas com base nos processos realizados em cada elemento dos ciclos de vida, considerando-se seu funcionamento de modo geral, em comparação com as ações presentes no Ciclo de Vida da Curadoria Digital do DCC.

Além disso, considera-se que o Ciclo de Vida da Curadoria Digital contempla práticas convergentes aos princípios FAIR, de modo que permite a verificação e a validação de todas as etapas necessárias ao processo de CD. As ações essenciais representam o princípio *findable*. As ações sequenciais representam os princípios *accessible* e *interoperable*. Por fim, a ação sequencial transformar e as ações ocasionais representam o último princípio, *reusable*.

O Modelo do Ciclo de Vida DataONE não apresenta etapas correspondentes às seguintes ações do Ciclo de Vida da Curadoria Digital: participação e observação da comunidade; conceitualizar; criar e receber; inserir; armazenar; deletar; reavaliar; migrar. O modelo é flexível e pode ser adaptado às necessidades da comunidade, ainda que não apresente ações específicas para a participação da comunidade. Possui uma estrutura cíclica, assim como o apresentado pelo DCC, e suas etapas correspondem às ações da CD, ainda que de modo mais simplificado. Nesse sentido, o modelo DataONE poderia ser aplicado em atividades de CD, de acordo com o contexto e as necessidades da comunidade.

Referente aos princípios FAIR, as etapas planejar, coletar, assegurar e descrever representam o princípio *findable*. Preservar corresponde ao princípio *accessible*. As etapas descrever e integrar representam o princípio *interoperable*. Por fim, a etapa analisar representa o último princípio, *reusable*. Nesse sentido, considera-se que o Ciclo de Vida da Curadoria Digital contempla práticas convergentes aos princípios FAIR, de modo que permite a verificação e a validação de todas as etapas necessárias ao processo de CD.

O Modelo DLCL, por sua vez, não apresentou correspondências com: participação e observação da comunidade; deletar; reavaliar; migrar. Ainda que

apresente mais correspondências com o modelo do DCC, a descrição do modelo DLCL não é detalhada, de modo que alguns elementos constam no modelo, mas sua função e os procedimentos que devem ser realizados não estão claros. O modelo em questão não trata especificamente da curadoria de dados (Pouchard, 2016), ainda que sejam identificados elementos que correspondem às ações da CD, de forma mais simplificada, uma vez que o modelo considera a perspectiva do pesquisador.

Os princípios FAIR são identificáveis nas ações de gerenciamento de dados, análise de dados e publicação. Tais fases representam ações práticas de tratamento e compartilhamento de dados, o que corresponde, de modo convergente, aos quatro princípios FAIR.

O Modelo do ciclo de vida do *Big Data* não apresenta correspondências com as seguintes ações da CD: participação e observação da comunidade; curadoria e preservação; inserir; armazenar; deletar; reavaliar; migrar. A partir da análise do modelo em questão, pode-se perceber que o modelo do ciclo de vida da CD poderia ser aplicado ao *Big Data*. O modelo proposto por Line Pouchard é mais detalhado do que os dois modelos anteriores, e apresenta três tipos de atividades, assim como o Ciclo de Vida da Curadoria Digital do DCC apresenta três tipos de ações.

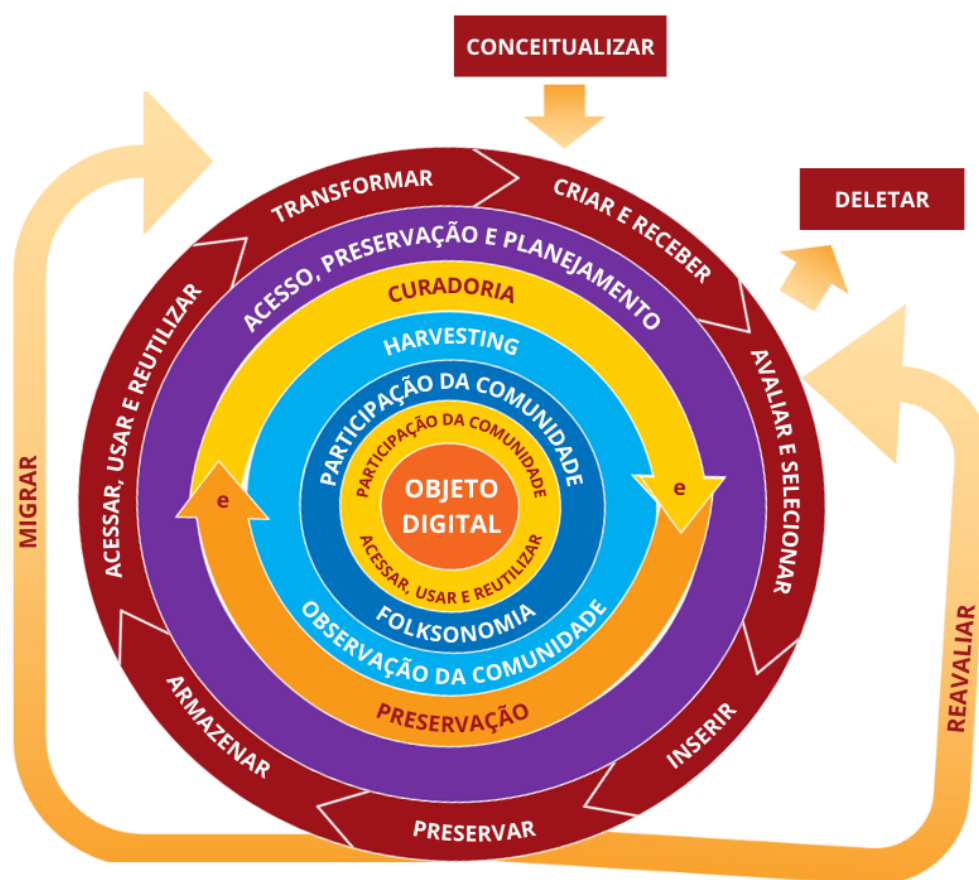
Referente aos princípios FAIR, as atividades descrever e assegurar correspondem ao princípio *findable*. As atividades planejar, adquirir, preparar, analisar e preservar representam os princípios *accessible* e *interoperable*. Por fim, a ação descobrir e a infraestrutura computacional correspondem ao princípio *reusable*.

Destaca-se que, de modo geral, os modelos analisados apresentam pouca abertura para a participação da comunidade, ou seja, há poucos elementos que integrem a participação de sujeitos não especialistas nos processos curatoriais. Tal ação consta no modelo do DCC, mas se refere a uma comunidade especializada. Nesse contexto, é possível adaptar o ciclo de vida da CD, conforme as necessidades da comunidade a qual se destina, como por exemplo, o ciclo pode ser adaptado para incluir ações colaborativas. Brayner (2018) destacou a ausência de discussões voltadas à participação dos sujeitos

informacionais nas etapas da CD. A participação da comunidade é mencionada, no entanto, o termo comunidade é, na maioria das vezes, relacionado a uma comunidade de profissionais da informação ou a equipes da tecnologia da informação (Brayner, 2018).

Assim, apresenta-se a seguir tal exemplo de adaptação do Ciclo de Vida da Curadoria Digital (Figura 6), inspirado no ciclo de vida da CD do DCC, voltado às ações que envolvem a participação da comunidade de modo geral.

Figura 6 - Modelo do Ciclo de Vida da Curadoria Digital voltado à participação da comunidade



Fonte: Elaborado Souza (2023) com base no DCC *Curation Lifecycle Model*, de Higgins (2008).

O modelo (Figura 6) representa graficamente um Modelo do Ciclo de Vida da Curadoria com ações mais colaborativas, e foi elaborado com base no modelo original, apresentado por Higgins (2008). Além das ações anteriormente apontadas, a adaptação do modelo do ciclo de vida da CD apresenta ações colaborativas em outros estágios do ciclo de vida, o que salienta a necessidade

da participação dos sujeitos informacionais em grande parte dos processos curatoriais (Souza, 2023).

No modelo em questão, a participação da comunidade foi incluída em duas ações: participação da comunidade – acessar, usar e reutilizar; e participação da comunidade – Folksonomia. Na primeira - acessar, usar e reutilizar -, os objetos digitais são compartilhados para que os internautas colaborem no processo de preservação. Na segunda ação - participação da comunidade – Folksonomia -, os internautas podem colaborar com o processamento técnico dos objetos informacionais, ao inserir etiquetas a eles.

A ação *Harvesting* – observação da comunidade corresponde à coleta dos termos inseridos pelos internautas por meio da Folksonomia (*harvesting*). Após a coleta, deve ser realizada a análise e a organização dos termos para a construção de instrumentos como vocabulários controlados e tesouros. Também foi acrescentada a ação acesso, preservação e planejamento, que consiste na aplicação do produto da Folksonomia para o processamento técnico dos objetos digitais, para garantir a preservação. Também compreende, por meio da análise das ações anteriores, o planejamento das próximas ações da CD.

A adaptação do modelo do ciclo de vida da CD não exclui o modelo original. Ele apresenta outra perspectiva das ações da CD e, portanto, complementa o outro, e demonstra que ele pode ser adaptado.

6 Considerações finais

Com base na análise dos ciclos de vida dos dados, bem como da literatura selecionada, pode-se concluir que os modelos são complementares, uma vez que cada modelo é eficiente para o contexto para o qual foi criado, o que representa o caráter complexo da curadoria de dados. No entanto, o presente artigo apresenta uma análise comparativa dos modelos de ciclo de vida dos dados sob a perspectiva da Curadoria Digital, que tradicionalmente é representada pelo modelo do DCC. Além disso, destaca-se que todos os modelos analisados apresentam elementos que convergem com os princípios FAIR, o que representa a preocupação com as boas práticas de gestão e compartilhamento de dados.

O modelo do DCC é eficiente, no sentido que atende aos requisitos necessários ao processo de Curadoria Digital. É possível ainda adaptá-lo, conforme explicitado por meio da customização, para atividades colaborativas. Assim, dentre os modelos analisados, considera-se que o modelo do DCC seja o mais complexo e completo, devido à sua estrutura, bem como às ações que ele apresenta.

Por outro lado, embora o modelo proposto por Line Pouchard seja o mais adequado para o tratamento do *Big Data*, poderia ser ainda aperfeiçoado com algumas ações da Curadoria Digital, como a participação da comunidade, por exemplo.

Os demais modelos também apresentam limitações no que se refere à participação da comunidade, que não é mencionada em nenhum deles. Também não dispõem elementos que correspondem às ações ocasionais da CD: deletar, reavaliar e migrar. Desse modo, tais modelos podem ser ampliados para contemplar os processos de descarte, reavaliação e migração de dados.

Na adaptação do modelo do DCC por nós sugerida, acrescentam-se camadas periféricas de acesso. Como resultado, por meio de tais acréscimos, apresenta-se outra perspectiva ao modelo tradicional do DCC: o acesso passa a ser elemento chave para a preservação, quando inicialmente falava-se em preservação para o acesso.

A ação de inverter a perspectiva com ênfase no acesso pode ser praticada também nos outros modelos apresentados, de modo que mais elementos relacionados às colaborações idealmente provindas do maior número de acessos podem ser otimizados.

Considera-se, finalmente, que muitos desses modelos foram concebidos ainda no contexto da *Web 1.0*, e não consideram a participação do internauta nos processos de gestão, tratamento e preservação dos dados. No contexto da *Web 2.0*, que apresenta a colaboração e o compartilhamento como principais características, é possível acrescentar novas camadas de participação e interação aos processos técnicos tradicionais. Com a participação dos internautas, a complexidade de tais processos aumenta, fazendo emergir a inversão do processo: acesso enquanto elemento de preservação.

Tendo em vista que as plataformas da Web estão em constante transformação, a dinâmica de sua complexidade é contínua e exige novas percepções das necessidades dos objetos informacionais digitais ali simulados, bem como dos outros elementos da complexidade, tais quais os sujeitos informacionais e as comunidades de interesse envolvidas no acesso aos dados compartilhados nos ambientes e plataformas. Neste sentido, estudar outros modelos de curadoria de dados e compreender os diálogos possíveis entre eles e a CD possibilitará conversações *ad hoc* e constantes adaptações aos modelos conforme as necessidades e a natureza dos dados.

Financiamento

O presente artigo foi elaborado a partir de pesquisa financiada pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) – Processo n. 2023/07820-5.

Agradecimentos

Agradecemos ao Professor Dietmar Wolfram por seus apontamentos e colaborações nas discussões sobre o presente artigo, e aos pareceristas pela cuidadosa leitura e apontamentos. Todas as contribuições foram essenciais para o desenvolvimento e publicação do artigo.

Referências

AMARAL, Adriana. Curadoria de informação e conteúdo na web: uma abordagem cultural: uma abordagem cultural. *In*: CORREA, Elizabeth Nicolau Saad (org.). **Curadoria digital e o campo da comunicação**. São Paulo: ECA - USP, 2012. p. 40-50.

ANJOS, Renata Lemos; DIAS, Guilherme Ataíde. Atuação dos profissionais da informação no Ciclo de Vida dos Dados - DATAONE: um estudo comparado. **Informação & Informação**, Londrina, v. 24, n. 1, p. 80-101, 2019. Disponível em: <https://doi.org/10.5433/1981-8920.2019v24n1p80>. Acesso em: 21 nov. 2024.

BRAYNER, Aquiles Alencar. Curadoria digital: novos modelos de participação pública na descrição de conteúdos em instituições culturais. **RICI -Revista Ibero-Americana de Ciência da Informação**, Brasília, v. 12, n. 1, p. 53-65,

2018. Disponível em: <https://doi.org/10.26512/rici.v12.n1.2019.10521>. Acesso em: 21 nov. 2024.

CAMPOS, Arthur Ferreira; DIAS, Guilherme Ataíde; SOUSA, Marckson Roberto Ferreira. Princípios FAIR, encontrabilidade e dados: por um diálogo teórico e uma aplicação prática. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 28, p. 1-21, 2023. Disponível em: <https://doi.org/10.5007/1518-2924.2023.e88060>. Acesso em: 21 nov. 2024.

CASTILHO, Carlos Albano Volkmer. **O papel da curadoria na promoção do fluxo de notícias em espaços informativos voltados para a produção e conhecimento**. 2015. Tese (Doutorado em Engenharia e Gestão do Conhecimento) - Universidade Federal de Santa Catarina, Florianópolis, 2015.

DATAONE. **About DataONE**. [S. l.: s. n.], 2019.

DIGITAL CURATION CENTRE. **What is digital curation?** [S. l.: s. n.], 2004.

GOFAIR. **FAIR Principles**. [S. l.: s. n.], 2025.

HIGGINS, Sarah. The DCC curation lifecycle model. **The International Journal of Digital Curation**, New York, v. 3, n. 1, p. 134-140, 2008. Disponível em: <https://doi.org/10.2218/ijdc.v3i1.48>. Acesso em: 10 mar. 2025.

JORENTE, Maria José Vicentini; SILVA, Stephanie Cerqueira; PADUA, Mariana Cantisani. Digital curation and information design in digital environments: women's museums panorama. **Transinformação**, Campinas, v. 33, p. 1-14, 2021. Disponível em: <https://doi.org/10.1590/2318-0889202133e210013>. Acesso em: 10 mar. 2025.

NAKANO, Natalia; OLIVEIRA, João Augusto Dias Barreira; JORENTE, Maria José Vicentini. Design thinking as a dynamic methodology for Information Science. **Information and Learning Science**, Leeds, v. 119, n. 12, p. 743-757, 2018. Disponível em: <https://doi.org/10.1108/ILS-07-2018-0061>. Acesso em: 11 set. 2024.

PAGE, Matthew J. *et al.* A declaração PRISMA 2020: diretriz atualizada para relatar revisões sistemáticas. **Revista Panamericana de Salud Pública**, São Paulo, v. 31, n. 2, p. 1-12, 2022. Disponível em: <https://doi.org/10.26633/RPSP.2022.112>. Acesso em: 16 out. 2025.

PLALE, Beth; KOUPER, Inna. The centrality of data: data lifecycle and data pipelines. In: CHOWDHURY, Mashrur; APON, Amy; DEY, Kakan (ed.), **Data analytics for intelligent transportation systems**. Cambridge: Elsevier, 2017.

POUCHARD, Line. Revisiting the data lifecycle with big data curation. **International Journal of Digital Curation**, New York, v. 10, n. 2, p. 176-192, 2016. Disponível em: <https://doi.org/10.2218/ijdc.v10i2.342>. Acesso em: 20 out. 2024.

PRISMA. **PRISMA 2020**. [S. l.: s. n.], 2020.

SANTOS, Thayse Natália Cantanhede. **Curadoria digital: o conceito no período de 2000 a 2013**. 2014. Dissertação (Mestrado em Ciência da Informação) - Universidade de Brasília, Brasília, 2014.

SAYÃO, Luis Fernando; SALES, Luana Farias. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**, João Pessoa, v. 3, n. 22, p. 179-191, 2012.

SOUZA, Gabriela Oliveira. **A folksonomia como elemento de preservação na curadoria digital: um estudo de caso do Museu da Pessoa**. 2023. Dissertação (Mestrado) - Universidade Estadual Paulista, São Paulo, 2023.

SOUZA, Arthur Heleno Lima Rodrigues; OLIVEIRA, Alexandre Faria; D'AVILA, Raquel Tavares; CHAVES, Erinalva Pereira Silva Sales. O modelo de referência OAIS e a preservação digital distribuída. **Ciência da Informação**, Brasília, v. 41, n. 1, p. 65-73, 2012. Disponível em: <https://doi.org/10.18225/ci.inf.v41i1.1352>. Acesso em: 28 jun. 2023.

VEIGA, Viviane; HENNING, Patricia; PRÍNCIPE, Pedro; SANTOS, Luiz Olavo Bonino Silva. Um panorama dos princípios de dados FAIR: teoria, práticas e serviços. **Informação & Informação**, Londrina, v. 29, n. 4, p. 267-282, 2025. Disponível em: <https://doi.org/10.5433/1981-8920.2024v29n4p267>. Acesso em: 24 out. 2025.

YAMAOKA, Eloi Juniti. Ontologia para mapeamento da dependência tecnológica de objetos digitais no contexto da curadoria e preservação digital. **Atoz: Novas Práticas em Informação e Conhecimento**, Curitiba, v. 1, n. 2, p.65-78, 2012. Disponível em: <https://doi.org/10.5380/atoz.v1i2.41313>. Acesso em: 24 out. 2025.

Data life cycles from the perspective of Digital Curation

Abstract: The general objective of the present study is to analyze data life cycle models from the perspective of Digital Curation. The specific objectives are: to build a theoretical framework on Digital Curation; carry out a comparative study between the Digital Curation Center Curation Lifecycle Model, the DataONE Lifecycle Model, the Data Life Cycle Labs Model and the Big Data Life Cycle Model. The present study is qualitative in nature, and descriptive and exploratory in nature. The method used was Design Thinking. From the analysis of each of the data life cycles, it was possible to observe some correspondences with the Digital Curation Life Cycle. It is noteworthy that none of the models presented elements for community participation. In this context, it is possible to adapt the Digital Curation life cycle according to the needs of the community it is intended for. It is necessary to consider the structure of other models for data curation, and understand their possible applications and convergences with Digital Curation, with the aim of including their practices in the model. This would allow adaptations to the original model according to the needs of the community and the nature of the data processed.

Keywords: digital curation; data lifecycle; information and technology

Declaração de autoria

Concepção e elaboração do estudo: Gabriela de Oliveira Souza, Daniel Martínez-Ávila e Maria José Vicentini Jorente.

Coleta de dados: Gabriela de Oliveira Souza.

Análise e interpretação de dados: Gabriela de Oliveira Souza, Daniel Martínez-Ávila e Maria José Vicentini Jorente.

Redação: Gabriela de Oliveira Souza, Daniel Martínez-Ávila e Maria José Vicentini Jorente.

Revisão crítica do manuscrito: Daniel Martínez-Ávila e Maria José Vicentini Jorente.

Declaração de disponibilidade de dados

Não se aplica.

Autoria para correspondência

Gabriela de Oliveira Souza
gabriela.oliveira@unesp.br

Editor-chefe

Thiago Henrique Bragato Barros

Como citar

SOUZA, Gabriela de Oliveira; MARTÍNEZ-ÁVILA, Daniel; JORENTE, Maria José Vicentini. Os ciclos de vida dos dados sob a perspectiva da Curadoria digital. **Em Questão**, Porto Alegre, v. 32, e-146953, 2026. DOI: <https://doi.org/10.1590/1808-5245.32.146953>

Parecer(es) aberto(s):

<https://doi.org/10.1590/1808-5245.32.146953A>

<https://doi.org/10.1590/1808-5245.32.146953B>

Recebido: 17/04/2025

Aceito: 08/10/2025

Copyright (c) 2026 Gabriela de Oliveira Souza, Daniel Martínez-Ávila, Maria José Vicentini Jorente. Este trabalho está licenciado sob uma licença [Creative Commons Atribuição 4.0 Internacional](#). Autores mantêm os direitos autorais e concedem à revista o direito de primeira publicação, com o trabalho licenciado sob a Licença Creative Commons Attribution (CC BY 4.0), que permite o compartilhamento do trabalho com reconhecimento da autoria. Os artigos são de acesso aberto e uso gratuito.

