

MODELO REFERENCIAL DE CICLO DE VIDA DOS DADOS CIENTÍFICOS: UMA PROPOSTA METODOLÓGICA

LIFE CYCLE REFERENTIAL MODEL OF SCIENTIFIC DATA: A METHODOLOGICAL PROPOSAL

Débora Gomes de Araújo Rodrigues^a
Renata Lemos dos Anjos^b
Guilherme Ataíde Dias^c
Sandra de Albuquerque Siebra^d

RESUMO

Objetivo: Propor um modelo referencial, resultante da interseção do *Digital Curation Centre*, do *Data Observation Network for Earth* e da *Data Documentation Initiative*, para descrição dos dados científicos na Ciência da Informação, com abrangência na grande área das Ciências Sociais. **Metodologia:** Estudo bibliográfico de abordagem qualitativa com finalidade descritiva. Para construção do modelo, realizou-se uma associação dos elementos dos ciclos de vida dos dados supracitados com a literatura acerca dos procedimentos metodológicos das Ciências Sociais. **Resultados:** Observou-se que a literatura aborda de forma evidente etapas do ciclo de vida tais como planejar, coletar e analisar. Porém, outras fases como organizar, preservar e integrar aparecem sutilmente e algumas outras não foram identificadas. Para um maior detalhamento das fases e a fim de evitar a supressão de elementos essenciais, foi desenvolvido um novo modelo envolvendo as seguintes etapas: planejar, coletar, organizar, descrever, assegurar, preservar, encontrar, reutilizar e analisar. **Conclusões:** A pesquisa contribui de forma prática com a gestão dos dados científicos através do modelo elaborado, facilitando os processos de compartilhamento e reúso desses em novas investigações, dando uma maior completude às informações sobre os dados produzidos/coletados.

Descritores: Ciclo de vida dos dados. Ciência da Informação. Ciências Sociais. Dados

^a Doutora em Ciência da Informação pela Universidade Federal da Paraíba (UFPB), João Pessoa, Brasil. E-mail: debora.g.de.araujo@gmail.com.

^b Doutoranda no Programa de Pós-Graduação em Ciência da Informação da Universidade Federal da Paraíba (UFPB), João Pessoa, Brasil. E-mail: renata.anjos@academico.ufpb.br.

^c Doutor pela Universidade Estadual Paulista (UNESP). Docente no Departamento de Ciência da Informação da Universidade Federal da Paraíba (UFPB), João Pessoa, Brasil. E-mail: guilhermeataide@gmail.com.

^d Doutora em Ciências da Computação pela Universidade Federal de Pernambuco (UFPE). Docente do Departamento de Ciência da Informação da Universidade Federal de Pernambuco (UFPE), Recife, Brasil. E-mail: sandra.siebra@ufpe.br.

científicos. Percurso metodológico.

1 INTRODUÇÃO

Na condução de investigações acadêmicas, pesquisadores das diversas áreas do conhecimento lidam constantemente com dados científicos, sobretudo no contexto dos seus procedimentos metodológicos. Nesse sentido, para apoiar a gestão de dados, é necessário que o arcabouço das diversas facetas englobe todas as etapas do Ciclo de Vida dos Dados (CVD) (Sayão; Sales, 2018). Segundo Medeiros (2018), esse processo envolve a fase de coleta até o armazenamento a longo prazo, perpassando por processos de limpeza, curadoria, anotação, indexação e transformação. Assim sendo, o planejamento da gestão dos dados usados e gerados em uma pesquisa integra os procedimentos metodológicos (Rodrigues, 2021).

Diante desse contexto, na literatura identificam-se iniciativas de ciclos de vida que proporcionam uma adequada gestão e curadoria aos dados. Destacam-se o *Digital Curation Centre* (DCC), que propôs o *Curation Lifecycle Model* (DCC, 2020); o *Data Observation Network for Earth* (DataONE) que apresenta o *Data Life Cycle* (DataONE, 2024a); e o *Data Documentation Initiative* (DDI), que oferece o *DDI-Lifecycle* (DDI, 2020).

Para Sanchez, Vidotti e Vechiato (2017), as funções identificadas no modelo do DCC correspondem a uma solução para a preservação digital, com foco em dados, objetos digitais e base de dados.

Segundo Anjos e Dias (2019), uma grande parcela das competências dos profissionais da informação está alinhada com as etapas do CVD do DataONE, possibilitando que os profissionais da área atuem na gestão de dados científicos através deste modelo. Como exemplo, ao elaborarem o Guia “Gestão de dados de pesquisa para bibliotecários e pesquisadores”, Sayão e Sales (2015) empregaram o modelo DataONE.

Em relação à iniciativa do DDI, conforme Vardigan, Heus e Thomas (2008), o modelo possibilita a descrição do CVD e aumenta o número de metadados, visando ampliar a sua contextualização, além de contribuir com a validação da qualidade dos dados. Cabe enfatizar que o *UK Data Service*,

(Serviço de Dados do Reino Unido), emprega o padrão de metadados DDI, especificamente orientado para as necessidades das Ciências Sociais, conforme apontado por Eynden e Corti (2017).

Nesse sentido, o presente estudo é fruto do trabalho de tese, intitulado de “Elementos de Ciclos de Vida dos Dados no Percurso Metodológico das Teses Brasileiras da Área de Ciência da Informação: um estudo diagnóstico”, que buscou o desenvolvimento de um modelo de CVD, com base nas iniciativas supracitadas do DataONE, DCC e DDI, a fim de analisar como os pesquisadores da área da Ciência da Informação (CI), no âmbito das teses, realizaram a gestão dos seus dados científicos. Portanto, o objetivo deste trabalho é propor um modelo referencial, derivado da interseção entre o DataONE, DCC e DDI, para descrever dados científicos na Ciência da Informação, com aplicabilidade abrangente nas Ciências Sociais, oferecendo uma estrutura detalhada para gerenciar dados científicos e aprimorar a metodologia de pesquisa.

2 INICIATIVAS DOS CICLOS DE VIDA DOS DADOS

É válido ressaltar que o CVD segue o caminho dos dados ou as ações necessárias para fazer com que a pesquisa avance para o próximo estágio e que entre os diversos domínios há diferentes versões de um CVD, com suas distintas atribuições (Rice; Southall, 2016; DataONE, 2024a). Para fins do presente trabalho serão evidenciadas as iniciativas do DataONE, DCC e DDI.

2.1 CICLO DE VIDA DOS DADOS – DATAONE

O DataONE (2024b) é um projeto financiado pela *National Science Foundation* (NSF) dos Estados Unidos, com foco na gestão dos dados relacionados às ciências da terra e ambientais. Composto por oito estágios: planejar, coletar, assegurar, descrever, preservar, descobrir, integrar e analisar.

No estágio planejar, é preciso programar e decidir como serão as próximas etapas. Quanto à coleta de dados, Strasser *et al.* (2012) trazem perguntas que precisam ser respondidas, como: Quais dados serão gerados? De que forma serão coletados e analisados? Quais instrumentos serão utilizados

e a infraestrutura pertinente para dar suporte à pesquisa? (Rodrigues, 2021).

No que se refere ao armazenamento, as decisões abrangem a escolha de um repositório de dados adequado aos dados que serão gerados e destinados à comunidade que os utilizará. Algumas condições devem ser consideradas, tais como, envio de documentação de dados, uso de padrões de metadados e qualquer restrição de uso que possa existir (como direitos de propriedade intelectual). Strasser *et al.* (2012) indicam alguns aspectos que precisam ser esclarecidos, tais como: Quem será o responsável pela gestão dos dados? Como deve ocorrer o controle de versão? De que forma é realizado o *backup* dos dados e com que frequência? Sobre esses questionamentos, o modelo referencial, especificamente na etapa planejar, contempla o momento de decidir os colaboradores e suas funções na gestão dos dados, o controle de versão que pode ser incluído no plano de gestão de dados, a frequência e a maneira dos *backups*, considerando os direitos autorais e permissões de acesso.

O momento de planejar a descrição dos dados deve considerar: Como será feito o registro de metadados e com qual padrão? Ao compartilhar os dados, é preciso elaborar um plano com a equipe do projeto, com colaboradores e com a comunidade científica. Além de algumas tomadas de decisões como, o período de embargo e de que forma os dados serão liberados (Strasser *et al.*, 2012).

O planejamento da preservação dos dados ocorre por meio da elaboração de um plano de preservação de dados a curto prazo, o qual deve considerar orçamento disponível, pessoal necessário para executar o plano de gestão dos dados, quais recursos computacionais serão utilizados (*hardware* e *software*) e que outras despesas poderão surgir (Strasser *et al.*, 2012). Nesse momento, Oliveira e Silva (2016) orientam o uso da ferramenta de livre acesso *Data Management Plan Tool* (DMPTool) para criação do plano de gestão de dados considerada uma boa prática de pesquisa.

No estágio **coletar**, é preciso garantir que os dados serão usados posteriormente. É fundamental considerar os métodos e a documentação antes de começar a coleta, sendo necessário a criação de um modelo para utilizar no transcorrer dessa fase. Recomenda-se a descrição do conteúdo dos arquivos de dados, como por exemplo, o formato e as unidades utilizadas (Strasser *et al.*,

2012). Essas questões são contempladas nas etapas planejar, coletar e organizar do modelo referencial, em que os dados devem ser coletados e registrados de acordo com o que foi planejado, e organizados na etapa posterior.

O terceiro estágio é o **assegurar**. Busca-se promover a qualidade dos dados durante a sua coleta e análise. Sendo recomendado o registro de qualquer situação ocorrida na coleta que possa influenciar em sua qualidade. É necessário verificar os valores estimados ao analisar os dados inseridos de forma manual e utilizar sinalizadores de nível de qualidade para identificar possíveis problemas. Para que o conjunto de dados seja consistente, sugere-se a identificação dos seus respectivos formatos (Strasser *et al.*, 2012). Complementando, Oliveira e Silva (2016) destacam a importância da adoção de recomendações legais antes que os dados passem para o próximo estágio, no qual podem receber um status (aberto, fechado ou embargado) (Rodrigues, 2021).

O quarto estágio é o **descrever**, para que os dados sejam recuperados e compreendidos no futuro, faz-se necessário uma documentação que contextualiza o arquivo de dados, intitulada de metadados. No qual deve-se considerar a descrição do contexto digital com a intitulação do conjunto de dados; o nome do(s) arquivo(s) e a data da última modificação; o *software* (informando a versão) e a forma de processamento utilizada (Rodrigues, 2021). Além disso, deve-se apresentar uma descrição do pessoal e das partes interessadas; apresentar o contexto científico e/ou ambiental em que os dados foram coletados e quais os instrumentos foram utilizados (Strasser *et al.*, 2012). Dialogando com a etapa descrever do modelo referencial, em que os metadados devem ser preenchidos de acordo com o padrão de metadados adotado no planejar, sendo os aspectos abordados relacionados aos elementos desse padrão, o qual deve ser o mais adequado à natureza da pesquisa.

Para os metadados serem descritos com qualidade, Oliveira e Silva (2016) sugerem a utilização de vocabulários controlados, tesouros e padrões como o W3C-DTF que contribuem com a descrição dos formatos de data e tempo e o *Internet Media Type* (IMT) que direciona a descrição do tipo, formato e extensão para a normatização de diversos suportes e mídias.

O quinto estágio é o **preservar**. Recomenda-se o uso de um *data center*,

que pode disponibilizar algumas orientações sobre como preparar os metadados e preservá-los, o formato de arquivo e prover serviços extras para possíveis usuários que tenham interesse em reutilizá-los. Nesse momento, deve-se arquivar os dados para possíveis futuros usuários (geralmente dados brutos); armazená-los com precisão; e fazer uso de terminologia padrão (ontologias da área). É preciso atentar ainda, para as políticas e questões legais que podem refletir na reutilização dos dados (Strasser *et al.*, 2012).

O sexto, sétimo e oitavo estágio são respectivamente **descobrir**, **integrar** e **analisar**. Esses, facilitados por um conjunto de ferramentas, possibilitam que os dados sejam descobertos, integrados e analisados. Nesse cenário, quando um conjunto de dados se constitui como o ponto de partida para novos conjuntos de dados, é importante documentar e identificar o conjunto de dados derivados. Possibilitando o entendimento da proveniência dos conjuntos e de seus elementos, a citação dos autores dos conjuntos originais e o levantamento de possíveis erros que possam existir no conjunto original (Strasser *et al.*, 2012).

No estágio descobrir, ocorre o encontro dos dados por outros usuários, devido a publicação dos dados, pois se não estiverem armazenados em repositórios adequados, voltados ao compartilhamento, disponíveis para visualização e/ou reúso, é provável que sejam esquecidos, “engavetados” ou até descartados (Silva Segundo; Araújo, 2019). Neste contexto, Sayão e Sales (2018) concordam e alertam sobre a necessidade dos pesquisadores identificarem repositórios apropriados para a publicação dos seus dados. Curty e Aventurier (2017) além de apresentarem os repositórios, acrescentam duas abordagens, as publicações ampliadas e os artigos de dados (Rodrigues, 2021).

No que se refere ao estágio integrar, Silva Segundo e Araújo (2019) mostram ser o momento de reutilizar dados, sendo preciso uma gestão adequada nas etapas anteriores e que os dados estejam acessíveis para serem integrados a outros projetos. Oliveira e Silva (2016) destacam a importância da adesão a protocolos de interoperabilidade para integrar os metadados proporcionando visibilidade, novas análises e pesquisas (Rodrigues, 2021).

Por fim, há o estágio analisar, que para Silva Segundo e Araújo (2019) é quando se verifica o alcance dos objetivos propostos. Para tanto, o curador dos

dados deve analisar com frequência se o plano de gestão possibilitará a segurança e a preservação dos dados. Para o DataONE (2024a), considera-se a compatibilidade dos dados que estão sendo integrados, além de descrever as etapas de processamento para produtos de dados derivados, que devem ser devidamente documentadas possibilitando a avaliação e interpretação dos resultados por terceiros (Rodrigues, 2021).

As etapas do ciclo de vida do DataONE referem-se a um momento da pesquisa e possibilitam a manutenção do valor dos dados funcionando como uma bússola que orienta o pesquisador na busca por conhecimento através da gestão dos dados científicos.

2.2 CICLO DE VIDA DOS DADOS – DIGITAL CURATION CENTRE

O *Digital Curation Centre* (DCC), fundado em 2004 no Reino Unido, é um líder global em curadoria digital, que na busca por uma curadoria de sucesso, desenvolveu o *Curation Lifecycle Model* (DCC, 2020).

Nesse contexto, Siebra *et al.* (2013) apresentam o objeto digital como a unidade principal de gestão do ciclo, sendo qualquer informação em formato digital. Assim, em sua parte central estão os objetos digitais simples, complexos e as bases de dados. Ao redor do objeto digital (como um dado de pesquisa) encontram-se as ações indicadas no decorrer da curadoria.

As ações para todo o CVD recebem essa classificação por corresponderem a atividades que o permeiam, agindo como anéis concêntricos, rodeando os objetos de dados que estão no centro do modelo (Sayão, Sales, 2012). As ações estão expressas no Quadro 1.

Quadro 1 – Ações para todo o ciclo de vida dos dados do DCC

Ações para todo o ciclo	Descrição
Descrição e Representação Informação	Definir os metadados administrativos, descritivos, técnicos, estruturais e de preservação, através de padrões apropriados, para assegurar descrição e controle a longo prazo. Definir a forma de atribuição de informações de representação apropriada para compreender e contextualizar o material digital.

Planejamento da Preservação	Planejar e aplicar as estratégias de preservação durante todo o ciclo da curadoria do material digital.
Acompanhamento e participação da comunidade	Motivar e envolver a comunidade interna (produtores de conteúdo/dados, desenvolvedores, pesquisadores) e externa (comunidade em geral, usuários da informação) no ciclo de curadoria para cooperar na representação da informação, definir a melhor maneira de acessar os dados, contribuir na definição de padrões, ferramentas e produtos de <i>software</i> pertinentes, entre outros.
Curar e Preservar	Realizar ações administrativas e gerenciais para colaborar com a curadoria e a preservação ao longo de todo o ciclo da curadoria. Considerar aspectos de direitos autorais e permissões de acesso.

Fonte: Baseado em DCC (2020) e Rodrigues (2021).

As ações sequenciais integram etapas que precisam ser realizadas repetitivamente de modo a assegurar que o dado se mantenha em processo contínuo de curadoria. A sequência deve ser cumprida de forma constante durante todo tempo em que o dado estiver sob curadoria, de maneira cíclica, não necessariamente na ordem apresentada (Sayão; Sales, 2012; Rodrigues, 2021). É válido destacar que é factível não cumprir alguma ação do ciclo e que ele pode ser personalizado segundo as necessidades do pesquisador. As ações sequenciais estão dispostas no Quadro 2.

Quadro 2 – Ações sequenciais do ciclo de vida dos dados do DCC

Ações sequenciais	Descrição
Conceituar	Conceber e planejar a criação de dados com o método de captura e opções de armazenamento.
Criar ou receber	Criar dados, acoplado a eles metadados administrativos, descritivos, estruturais, técnicos e, se necessário, de preservação. Receber dados, segundo políticas de coleta documentadas, de criadores de dados, outros arquivos, repositórios ou centros de dados e, se preciso, atribuir metadados pertinentes.
Avaliar e Selecionar	Avaliar os dados e selecionar para curadoria de longo prazo e preservação, com base em orientações documentadas, políticas ou requisitos legais.
Inserir	Transferir dados de um arquivo, repositório, centro de dados ou outra entidade custodiante, sob orientações documentadas, políticas ou requisitos legais.
Ações de preservação	Realizar ações para garantir a preservação a longo prazo e a retenção da natureza autorizativa dos dados. As ações de preservação devem assegurar que os dados permaneçam autênticos, confiáveis, utilizáveis e íntegros.

Armazenar	Armazenar os dados com segurança, sob a orientação de padrões relevantes. Além de seguir rotinas de <i>backup</i> .
Acessar, usar e reusar	Garantir que os dados estejam acessíveis aos usuários designados no cotidiano, adequadamente ao público-alvo.
Transformar	Criar dados com base no original, por exemplo: por meio da migração para um formato diferente.

Fonte: Baseado em DCC (2020) e Rodrigues (2021).

Por fim, de acordo com Sayão e Sales (2012), as ações que são de natureza ocasional ocorrem de forma esporádica e têm o potencial de interromper ou reordenar a sequência planejada de atividades, dependendo de uma decisão. Essas ações estão elencadas no Quadro 3.

Quadro 3 – Ações ocasionais do ciclo de vida dos dados do DCC

Ações ocasionais	Descrição
Descartar	Eliminar dados que não foram designados para curadoria e preservação de longo prazo, conforme as políticas documentadas, orientação ou requisitos legais ou que já atingiram seu período máximo de curadoria, conforme tabelas de temporalidades adotadas.
Reavaliar	Retornar dados que passaram por procedimentos de validação falhos, para uma avaliação e seleção adicional. Reavaliar todas as atividades executadas no ciclo para verificar a necessidade de ajustes/melhorias.
Migrar	Migrar dados para um formato diferente. Podendo ser realizado de acordo com o ambiente de armazenamento ou para assegurar a preservação, evitando a obsolescência de <i>hardware</i> ou <i>software</i> .

Fonte: Baseado em DCC (2020) e Rodrigues (2021).

Silva Segundo e Araújo (2019) enfatizam a necessidade de estruturas tecnológicas robustas e colaboradores habilidosos para lidar com dados na curadoria digital. É preciso ainda a existência de políticas públicas que considerem a dimensão legal dos dados, assim como dos recursos financeiros necessários para financiar as estruturas, tecnologias e pessoas.

2.3 CICLO DE VIDA DOS DADOS – DATA DOCUMENTATION INITIATIVE

A *Data Documentation Initiative* (DDI) teve início em 1995, em Quebec no Canadá, com um pequeno grupo internacional focado em metadados de Ciências Sociais (Vardigan, 2013). O DDI – *Lifecycle* consiste em um dos produtos dessa iniciativa, descrevendo além dos resultados da coleta de dados, os metadados de todo o processo, a partir da conceituação até o arquivamento

dos dados resultantes. Ball (2012) apresenta de forma detalhada as características das fases desse modelo, conforme o Quadro 4 (Rodrigues, 2021).

Quadro 4 – Descrição das fases do ciclo de vida dos dados do DDI versão 3.0

Fases	Descrição
Conceito	Momento de planejar a pesquisa. Envolve a escolha da questão de pesquisa e a metodologia para coletar os dados necessários, planos para a maneira como os dados serão processados, analisados e usados para responder à questão de pesquisa e a forma que essa resposta terá. Definir as relações que acontecerão entre os produtos de dados da pesquisa.
Coleta	Métodos e fontes de coleta de dados incluem pesquisas, censos, registros de votação ou de saúde, estatísticas de comércio ou coleções baseadas na <i>web</i> . Fontes de dados primárias e secundárias precisam ser distinguidas.
Processamento	Dados de entrada são processados e analisados gerando dados de saída (a exemplo, uma estatística ou conjunto dela), trazendo resposta à problemática da pesquisa. Podem ser registrados em uma forma legível por máquina ou de orientação humana, como um relatório técnico.
Arquivamento	Dados devem ser transferidos para um arquivo, em vez de serem apenas mantidos por pesquisadores. Preservando dados (e metadados) e agregando valor no decorrer do tempo.
Distribuição	Dados são distribuídos aos usuários diretamente ou através de uma biblioteca ou arquivo de dados.
Descoberta	Dados podem ser disseminados através de uma variedade de canais, incluindo livros, artigos em periódicos científicos, páginas da <i>web</i> ou demais serviços disponíveis na internet.
Análise	Dados podem ser usados por outros dentro dos limites da conceituação original (ao optar por estatísticas importantes para um relatório de pesquisa).
Reaproveitamento	Dados podem ser usados dentro de uma estrutura conceitual distinta; envolvendo amostragem ou reestruturação, combinação com outros conjuntos similares, ou produção de materiais pedagógicos.

Fonte: Ball (2012) e Rodrigues (2021).

Para Gupta e Müller-Birn (2018), o modelo do DDI abrange de forma efetiva vários aspectos da pesquisa, com algumas limitações, tais como: o foco na coleta de dados primários, a falta de foco na publicação de dados que foram analisados e a agilidade limitada. Além dessas limitações, não foi possível identificar no *site* do DDI a descrição das fases do CVD, o que prejudica a compreensão de cada uma delas. Em contrapartida, o modelo destaca-se na pesquisa científica ao priorizar o uso eficiente de metadados, especialmente nas Ciências Sociais, servindo como referência para outros modelos.

3 O CAMINHO DOS DADOS NOS PERCURSOS METODOLÓGICOS DAS CIÊNCIAS SOCIAIS

O percurso metodológico faz parte dos elementos fundamentais da investigação científica (Richardson, 2017). Nesse sentido, segundo Deslandes (2009), este momento inclui a descrição da fase de exploração de campo, as etapas do trabalho de campo bem como os procedimentos para análise. Para facilitar o entendimento das etapas metodológicas, as seções subsequentes foram estruturadas de forma a esclarecer o processo de obtenção, coleta e análise dos dados que emergem dessa investigação.

3.1 DESCRIÇÃO DA FASE DE PLANEJAMENTO DA OBTENÇÃO DOS DADOS

A fase planejar a obtenção dos dados é o momento de definir procedimentos para a coleta e análise dos dados, o tipo da pesquisa, o campo de estudo (sujeitos, locais, amostragem) e a seleção de métodos e técnicas. Michel (2015) sugere o planejamento e organização da coleta dos dados, seguindo um cronograma, incluindo fase de registro e análise dos dados, ou seja, o desenvolvimento de um plano tático de ação para a pesquisa (Rodrigues, 2021).

No processo da coleta de dados, é fundamental decidir sobre a identificação e escolha dos sujeitos que trarão respostas às problemáticas levantadas ou que terão seus comportamentos observados. Nesse contexto, o pesquisador necessita refletir sobre quais questionamentos serão feitos, onde e quando será realizado o estudo, quem fará parte dele, de que forma vai obter permissão e colaboração dos participantes, quais os dados serão escolhidos e como serão reunidos (Richardson, 2017).

De acordo com Borgman (2015), geralmente devem ser feitas escolhas acerca dos dados que podem ou devem ser coletados, como serão obtidos e o porquê, bem como serão relatados e divulgados. Por se tratar de pesquisas que podem envolver humanos, é necessário considerar questões éticas, atentando ao fato de que dados pessoais são submetidos a regras e regulamentos sobre como podem ser coletados e em que circunstâncias, como a Lei Geral de

Proteção aos Dados Pessoais (LGPD) (Brasil, 2018; Rodrigues, 2021).

Segundo Gil (2019), normalmente as pesquisas sociais podem envolver um universo de elementos, sendo complexa a análise da sua totalidade. Em razão disso, é comum a análise de uma parcela desses elementos, ou seja, a amostra. Na literatura há diferentes tipos de amostragem, que podem ser selecionados pelo pesquisador de acordo com a sua investigação.

A seleção dos métodos e técnicas a serem utilizados na pesquisa científica pode ser baseada na proposição do problema, na formulação das hipóteses e na limitação do universo e da amostra, denotando uma relação com o problema a ser pesquisado (Marconi; Lakatos, 2019; Rodrigues, 2021).

As técnicas geralmente utilizadas na coleta de dados na pesquisa social são: a análise de documentos da empresa ou grupo social, a observação pessoal, as diversas modalidades de entrevista e o questionário. Cada uma delas tem o seu objetivo, um público específico e, à sua maneira, agrega valor ao processo de análise (Michel, 2015). Quanto aos métodos, os pesquisadores adotam mais de um ou até uma combinação de dois ou mais. A organização do instrumento da pesquisa demanda tempo, mas é um momento importante a ser planejado (Marconi; Lakatos, 2019; Rodrigues, 2021).

Observa-se que esse momento faz menção a etapa planejar, que direciona decisões necessárias tanto para o rumo da pesquisa como para a obtenção e análise dos dados. A partir dessa previsão, que pode ser flexível, considera as decisões nos recursos disponíveis como tempo e pessoal. Portanto, com base em uma preparação preliminar estabelece os requisitos e as ações necessárias para as decisões relacionadas aos dados científicos.

3.2 DESCRIÇÃO DA FASE DE COLETA DOS DADOS

De acordo com Gil (2019), no processo de coleta de dados é possível a utilização de três procedimentos: pesquisa bibliográfica, documental e contatos diretos ou de campo. No primeiro, os dados são obtidos através da leitura de livros, artigos de periódicos, anais de eventos e impressos diversos. O segundo, pode acontecer por meio da exploração de arquivos, análise de documentos ou de artefatos físicos. O terceiro, os contatos diretos são estabelecidos com

pessoas que podem disponibilizar dados ou apontar possíveis fontes de informações relevantes, podendo a pesquisa bibliográfica pode ser realizada junto com a de campo (Marconi; Lakatos, 2019; Rodrigues, 2021).

Neste contexto, Gil (2019) evidencia que a coleta de dados é realizada *in situ*, captando os fenômenos em sua manifestação natural por meio de técnicas como observação, aplicação de questionários e condução de entrevistas.

Diante dos apontamentos dos autores (Gil, 2019; Marconi; Lakatos, 2019), os dados podem ser obtidos através de naturezas ou procedimentos diferentes, ou seja, a partir de fontes diferentes que podem ser combinadas, a critério do pesquisador, de forma que possam contribuir para alcançar os objetivos definidos para a pesquisa.

Nesse sentido, a fase da coleta refere-se à execução da pesquisa. Conforme Michel (2015), ela deve estar alinhada com o cronograma pré-estabelecido, conectada ao plano de ação. É fundamental que os dados sejam armazenados à medida que forem obtidos, mesmo que provisoriamente e sem formato deliberado, visando a preservação. É importante ressaltar que, para possibilitar o futuro reuso, a qualidade dos dados deve ser assegurada.

Marconi e Lakatos (2019) revelam que a coleta compreende o momento da pesquisa em que há a aplicação dos instrumentos elaborados e das técnicas selecionadas. Esta situação demanda tempo e preparo do pesquisador, que precisa ser atencioso com o registro dos dados. Portanto, é preciso realizar o trabalho em tempo adequado, não ir além do orçamento previsto e preparar o pessoal. Assim, quanto mais planejamento, melhor. A realidade exposta evidencia que a fase de obtenção dos dados está diretamente ligada ao planejamento (Rodrigues, 2021).

Segundo Michel (2015), para que os dados brutos coletados se transformem em resultados pertinentes para a pesquisa, é necessário adotar procedimentos que possibilitem ao pesquisador a sistematização, categorização e análise dos dados através das técnicas, o que será enfatizado no próximo tópico. Desta forma, compreende-se que para facilitar o processo de análise dos dados, é necessário que eles sejam devidamente organizados, constituindo um momento fundamental que envolve o tratamento dos dados.

3.3 PROCEDIMENTOS DE ANÁLISE DOS DADOS

A obtenção do significado dos dados coletados/gerados pela pesquisa, envolve a necessidade de inclusão do processo de análise e interpretação. Desta forma, é preciso que sejam tabulados, resumidos, organizados e apresentados em tabelas, gráficos ou diagramas (Gil, 2019). Richardson (2017) acrescenta que, os dados obtidos apresentam as respostas para o questionamento da pesquisa e conseqüentemente, os resultados da investigação.

Gil (2019) chama atenção para a possibilidade de confrontar dados provenientes da pesquisa original com outros dados oriundos de arquivos ou de pesquisas realizadas. Sendo possível verificar a necessidade de disponibilizar os dados de uma investigação científica para servir de referência para outras pesquisas, assim como para comprovar a sua veracidade.

Os dados levantados devem ser interpretados e analisados à luz de alguma teoria. Possibilitando uma interpretação significativa, uma vez que, uma das funções da teoria é certificar uma maior relevância aos dados (GIL, 2019).

Nesse contexto, Richardson (2017) alerta que, se for realizada uma análise quantitativa, é necessário indicar como os dados serão tratados: tabelas, gráficos e testes estatísticos. Porém se a pesquisa tiver um enfoque qualitativo, explicitar as técnicas adotadas: tipo de análise (documentária, de conteúdo ou histórica). Na literatura é possível verificar diversos métodos de pesquisa usados para a análise e interpretação de textos, como, a análise qualitativa do conteúdo, a etnografia, a teoria fundamentada, a fenomenologia e a pesquisa histórica. A análise do discurso constitui uma ferramenta de interpretação primordial nas Ciências Sociais (Richardson, 2017; Rodrigues, 2021).

Desta forma, verifica-se que a análise dos dados produz os resultados da pesquisa estando intimamente relacionada com o planejamento e a coleta de dados. Diante do exposto, observa-se que essas três fases relacionadas aos dados são mais evidenciadas no discurso dos autores das Ciências Sociais (Deslandes, 2009; Gil, 2019; Marconi; Lakatos, 2019; Michel, 2015; Richardson, 2017). Verifica-se, portanto, que antes da análise os dados necessitam ser sistematizados e/ou categorizados, ou seja, organizados e armazenados à

medida que são coletados, relacionando-se a etapa preservar. Por fim, foi visto que se pode fazer a reutilização de dados de outros estudos para confrontar com dados da pesquisa original.

4 PROCEDIMENTOS METODOLÓGICOS

A presente investigação se configura como estudo bibliográfico de abordagem qualitativa com finalidade descritiva (Gil, 2019; Marconi; Lakatos, 2019).

Para construção do modelo de referência estudou-se os modelos DCC, DataONE e DDI, suas etapas/fases e elementos constituintes. A partir deste estudo, buscou-se usar no modelo referencial a interseção destes três modelos. Na sequência, procurou-se identificar a associação dos elementos do modelo criado com as ações referentes aos dados encontradas na literatura que aborda os procedimentos metodológicos das Ciências Sociais, identificando nos referidos procedimentos o tratamento dos dados das pesquisas.

Como resultado, foi desenvolvido um modelo de CVD e um quadro com as características de cada fase para que os pesquisadores da área de CI possam compreender as ações necessárias na aplicação do modelo.

5 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Para criação do modelo de referência, buscou-se identificar etapas em comum nos ciclos de vida do DCC, DataONE e DDI.

No tocante à etapa sugerida de “**planejar**”, a sua relação com o DCC está atrelada às decisões sobre os metadados que serão descritos, as estratégias de preservação, a equipe de trabalho e a identificação dos potenciais usuários que poderão fazer o (re)uso dos dados, considerando direitos autorais e permissões de acesso. A correlação com o DataONE acontece no planejamento do processo de coleta, de armazenamento, do período de embargo e a forma como os dados serão descritos. A ligação com o DDI também se refere às decisões sobre a coleta, além do processamento e análise de dados (Rodrigues, 2021).

A etapa proposta “**coletar**” está correlacionada com o DCC com a

execução da coleta propriamente dita. No que tange ao DataONE, está sintonizada com a forma de obtenção de dados. Já em relação ao DDI, apresenta alinhamento quanto aos métodos e fontes de dados propostos na etapa coleção de dados. Verificou-se que os três ciclos, de modo geral, incluem ações de forma direta ou indireta inerentes à organização dos dados, ou seja, a forma de agrupar os dados (Rodrigues, 2021). Desta forma, para efeito deste estudo, foi criada uma fase, intitulada de “**organizar**”, a fim de evitar resumir atividades essenciais em uma única etapa ou que alguma atividade importante fique subentendida, de forma a não prejudicar a aplicação do ciclo. Além disso, segundo Gil (2019) é necessário que os dados estejam organizados para posterior análise.

A etapa sugerida “**descrever**” apresenta correlação com o DCC no tocante à utilização de metadados identificados na ação de criação e recebimento. Ocorrendo o mesmo com o DataONE que atrela à adoção de um padrão de metadados para a descrição dos dados. Alinhando-se também ao DDI, uma vez que essa descrição deve possibilitar que os dados sejam legíveis por humanos e por máquinas (Rodrigues, 2021).

A etapa proposta de “**assegurar a qualidade**” está em sintonia com o DCC no acompanhamento e participação da comunidade, quando as pessoas ligadas ao projeto devem garantir que os dados são autênticos, confiáveis e (re)utilizáveis. Também se relaciona com às ações de avaliar e selecionar, pela adoção de ações para que os dados sejam preservados, verificar a qualidade dos dados coletados, sua autenticidade e integridade. Desta forma, alguns dados poderão ser descartados, não passando para a próxima etapa de preservar. Apresenta correlação com o DataONE no tocante a assegurar uma coleta adequada e com o DDI, no momento das checagens para verificar a confiabilidade dos dados. Cabe destacar que a etapa assegurar a qualidade, que no DataONE está posicionada antes da etapa descrever, foi realocada posteriormente, pois entende-se ser o momento de assegurar que o dado foi coletado e a correspondência descritiva dele está adequada (Rodrigues, 2021).

A sugestão da etapa “**preservar**” está alinhada com o DCC, por ser um momento de tomar como parâmetro documentos e requisitos legais para depositar os dados em um arquivo com acesso a longo prazo, conforme

orientação do DataONE, de forma que os dados sejam armazenados com segurança, de acordo com a indicação do DDI (Rodrigues, 2021).

A etapa proposta “**encontrar**” apresenta correlação com o DCC em relação aos dados estarem acessíveis aos usuários potenciais; ao DataONE, no aspecto desses serem descobertos por meio dos seus metadados e ao DDI, quanto à visibilidade dos dados através de publicações.

A etapa recomendada de “**reutilizar**” apresenta ligação com os três ciclos, no que se refere a referências qualificadas a dados de outros estudos com a finalidade de enriquecer o conjunto de dados já existentes, para realizar comparações que poderão aperfeiçoar o estudo, ou analisados sob uma nova perspectiva. O que será enfatizado na próxima etapa proposta “**analisar**”, que é o momento ligado ao DCC de reavaliar e analisar do DataONE no que tange à verificação dos ajustes que poderão contribuir com os dados, assim como a aplicação de métodos de pesquisa para realizar a análise (Rodrigues, 2021).

Diante do exposto, propõe-se um modelo referencial com etapas adaptadas à gestão dos dados científico em pesquisas da área de CI, que tomou como base os modelos do DCC, DataONE e DDI. Ressalta-se que, embora este novo ciclo tenha sido desenvolvido com foco na CI, ele pode ser utilizado pelos pesquisadores da grande área das Ciências Sociais Aplicadas, uma vez que as características do modelo estão baseadas na literatura discutida acerca dos procedimentos metodológicos deste domínio. Além disso, foi utilizado o CVD do DDI cujo foco está direcionado às Ciências Sociais. O CVD referência com suas nove etapas está ilustrado na Figura 1.

Figura 1 – Ciclo de vida dos dados científicos



Fonte: Adaptado de DataONE (2024a), DCC (2020), DDI (2020) e Rodrigues (2021).

É importante destacar que a etapa “planejar” é o momento de refletir sobre as ações necessárias em todo o ciclo, de modo que as demais etapas orbitam em torno dela e influenciam no seu aprimoramento para aplicações subsequentes. Além disso, o modelo tem em seu centro os dados científicos, foco do estudo. Cada etapa deste ciclo está detalhada no Quadro 5, com as respectivas ações sugestivas que devem ser aplicadas em cada uma delas. As características de cada etapa envolvem o debate dos autores que discutem os procedimentos metodológicos no âmbito das Ciências Sociais (Deslandes, 2009; Gil, 2019; Marconi; Lakatos, 2019; Michel, 2015; Richardson, 2017), além da literatura que aborda os ciclos de vida supracitados.

Quadro 5 – Detalhamento das etapas do ciclo de vida dos dados científicos

ETAPAS	CARACTERÍSTICAS
Planejar	<p>Devem ser tomadas decisões referentes a cada etapa do ciclo de vida, por meio do plano de gestão de dados e um cronograma para que ocorra um adequado gerenciamento dos dados. É fundamental a adoção de algumas práticas, a saber:</p> <ul style="list-style-type: none"> ✓ Decidir sobre quais dados serão obtidos e o que é necessário realizar para alcançá-los. Deve ser identificado: sujeitos que trarão respostas à questão da pesquisa, local onde se realizará o estudo, a definição da amostragem, bem como a escolha do instrumento de pesquisa.

	<ul style="list-style-type: none">✓ Definir o período da coleta alinhado com o cronograma, levando em consideração o prazo para concluir o trabalho;✓ Levar em consideração as questões éticas quanto ao consentimento para uso e divulgação dos dados. Quando a pesquisa incluir seres humanos, incluir a aprovação do comitê de ética.✓ Considerar questões legais relacionadas aos dados, como a privacidade de dados, e as legislações pertinentes para o tipo de dado sendo trabalhado.✓ Planejar a descrição dos dados, no tocante aos tipos de metadados e o padrão que será adotado.✓ Definir estratégias para preservação quanto à escolha de um arquivo de dados, garantindo o valor dos dados a longo prazo, esclarecer a frequência de <i>backup</i>, e de que maneira acontecerá. Refletindo sobre questões de direito autoral e permissões de acesso.✓ Definir os colaboradores e suas responsabilidades no âmbito da equipe de trabalho e identificar os possíveis usuários dos dados;✓ Esclarecer para a equipe de trabalho o que é essencial, sob pena de comprometer a qualidade dos dados;✓ Decidir como os dados serão analisados considerando as técnicas;✓ Definir metas, o que se espera após a execução de cada etapa e;✓ Levantar os recursos financeiros e tecnológicos necessários e disponíveis para o funcionamento das etapas do ciclo.
Coletar	<p>Geração dos dados com base no que foi planejado. Por meio de instrumentos ocorre a aquisição, criação ou transformação dos dados que necessitam ser registrados para não serem perdidos.</p> <p>Nas Ciências Sociais, por exemplo, a coleta dos dados poderá acontecer por meio de uma pesquisa bibliográfica, documental e/ou de campo. Onde podem ser aplicadas técnicas como observação direta ou indireta, entrevistas, questionários, entre outros.</p>
Organizar	<p>Agrupamento, codificação ou categorização dos dados.</p>
Descrever	<p>Preencher os metadados dos dados que foram obtidos, com base no padrão de metadados que foi previamente planejado para ser adotado, a fim de que os dados no futuro possam ser recuperados e compreendidos, de forma a serem legíveis por máquinas e humanos.</p>
Assegurar a qualidade	<p>Assegurar que os dados sejam autênticos, íntegros, confiáveis e utilizáveis, considerando uma relação com a equipe de trabalho que têm as informações sobre a infraestrutura, os instrumentos utilizados e o contexto de geração de dados. Avaliar e selecionar os dados que serão preservados e decidir os dados que serão descartados ou mantidos.</p>
Preservar	<p>Preservar dados em um arquivo, repositório ou bases de dados, com vista à longevidade dos dados. Considerar a questão legal dos dados no tocante a aspectos de direitos autorais, privacidade de dados e o período de embargo, conseqüentemente ficará claro o momento que os dados podem estar acessíveis para os potenciais usuários.</p>
Encontrar	<p>Descoberta de dados acessíveis para serem usados e reutilizados, contribuindo com a visibilidade dos dados, podendo ser encontrados em periódicos de dados, publicações ampliadas e repositórios de dados.</p>
Reutilizar	<p>Agregar um conjunto de dados à pesquisa. É necessário considerar a documentação dos dados, questões de interoperabilidade dos sistemas, além de requerer o <i>feedback</i> dos usuários dos dados para a identificação de necessidades visando o reuso da melhor maneira.</p>

Analisar	Descrever os métodos e as técnicas utilizadas na análise dos dados, averiguando se os dados necessitam ser ajustados. Confrontar dados da pesquisa com dados de outras pesquisas já realizadas para verificar a veracidade dos dados e assim analisar a necessidade de repensar o processo de curadoria como um todo.
-----------------	---

Fonte: Rodrigues (2021).

A execução das etapas deve ser adaptada ao projeto de cada pesquisador. Nos casos em que os dados sejam provenientes da própria pesquisa, é recomendável cumprir cada etapa do ciclo e levar em consideração a interação entre elas. Alguns pontos merecem destaque, tais como: a relação da etapa planejar com o restante do ciclo; a importância da realização de pré-teste de questionário ou entrevistas antes da coleta, visando a assegurar a qualidade dos dados, de forma que eles sejam coletados de forma autêntica, íntegra e confiável; a importância da etapa organizar para facilitar a análise dos dados; a ligação da preservação com a descoberta, já que para os dados serem encontrados, precisam ser disponibilizados e preservados. Desta forma, as ações adotadas em um momento podem afetar o ciclo como um todo, por isso a importância da interação entre as partes do ciclo e a necessidade de uma revisão constante das definições feitas no planejamento.

Nos casos em que o pesquisador estiver fazendo o reuso de dados, algumas etapas podem ser opcionais. De todo modo, é fundamental a descrição dos dados, com a finalidade de assegurar a sua qualidade, bem como é preciso que sejam preservados para que possam ser descobertos e reutilizados. Neste contexto, visando o armazenamento e descoberta de dados, Eynden e Corti (2017) indicam o uso de repositório de dados específico voltado às Ciências Sociais, Reshare¹. No contexto desta pesquisa recomenda-se consultar o diretório global de repositórios de acesso aberto, OpenDoar², onde é possível encontrar repositórios que publicam conjuntos de dados de diferentes tipos. Outras opções identificadas na literatura dizem respeito ao uso de publicações ampliadas (Mucheroni; Silva; Paletta, 2015) e periódicos de dados (Curty; Aventurier, 2017). No Brasil, pode ser citada como exemplo a Revista Encontros

¹ Disponível em: UK Data Service (c2024).

² Disponível em: OpenDOAR ([2024?]).

Bibli³ que passou a publicar artigos de dados (Rodrigues, 2021).

6 CONSIDERAÇÕES FINAIS

Por meio do estudo realizado, foi possível atingir o objetivo de propor um modelo referencial de CVD para a descrição dos dados científicos voltado à CI, podendo ser aplicado pela grande área das Ciências Sociais, tomando como base a análise das iniciativas do DataONE, DCC e DDI.

Identificou-se que o CVD do DataONE foi construído com o foco nas ciências da terra; o do DCC não se limita a curadoria dos dados científicos, podendo ser utilizado com outros tipos de objetos digitais e o da DDI é especificamente voltado às Ciências Sociais. Assim sendo, considerou-se as particularidades de cada ciclo que, apesar das suas diferenças, foram essenciais para a criação do ciclo adaptado.

A literatura utilizada no presente estudo mostrou que algumas etapas como planejar, coletar e analisar estão mais presentes no discurso dos autores que tratam sobre os procedimentos metodológicos das Ciências Sociais. Já algumas outras etapas apareceram de forma sutil como organizar, preservar e integrar. O que mostra uma associação da metodologia dos trabalhos científicos às etapas do CVD, de forma direta ou indireta.

A pesquisa também apontou que a execução de cada etapa, primeiramente passa pelo momento de planejar. De modo geral, é relevante a integração entre as partes do ciclo, visto que uma ação executada pode influenciar no desempenho de todo o ciclo, por isso as etapas foram detalhadas para uma melhor aplicabilidade por parte dos pesquisadores.

Espera-se que a pesquisa contribua de forma prática com a gestão dos dados científicos produzidos nas pesquisas da área de CI, uma vez que mostra que a própria definição e aplicação dos procedimentos metodológicos das pesquisas tem relação direta com o ciclo de vida dos dados que estão sendo coletados ou produzidos nas pesquisas. Logo, a aplicação do modelo referencial proposto pode contribuir para aprimorar a forma como as definições são

³ Disponível em: Encontros BIBLI ([2024?]).

realizadas, dando uma maior completude às informações sobre os dados produzidos/coletados, de forma a favorecer que os pesquisadores possam disponibilizá-los para serem reutilizados em novas investigações.

REFERÊNCIAS

ANJOS, R. L.; DIAS, G. A. 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: <http://dx.doi.org/10.5433/1981-8920.2019v24n1p80>. Acesso em: 6 abr. 2020.

BALL, A. **Review of data management lifecycle models**. Bath: University of Bath, 2012. Disponível em: <https://purehost.bath.ac.uk/ws/portalfiles/portal/206543/redm1rep120110ab10.pdf>. Acesso em: 12 set. 2020.

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

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da União**: seção 1, Brasília, DF, n. 157, p. 59, 15 ago. 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 16 nov. 2021.

CURTY, R. G.; AVENTURIER, P. O paradigma da publicação de dados e suas diferentes abordagens. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 18., Marília, 2017. **Anais** [...]. Marília: UNESP, 2017. p. 1-21. Disponível em: <https://www.brapci.inf.br/index.php/res/v/105144>. Acesso em: 8 jan. 2020.

DATA DOCUMENTATION INITIATIVE (DDI). 2020. **DDI Lifecycle 3.3**, [S. l.], 2020. Disponível em: <https://ddialliance.org/Specification/DDI-Lifecycle/3.3/>. Acesso em: 23 jul. 2020.

DATAONE. **Best Practices of Data Management**. [S. l.], [2024a?]. Disponível em: <https://dataoneorg.github.io/Education/bestpractices/>. Acesso em: 05 jul. 2024.

DATAONE. **About DataONE**. [S. l.], [2024b?]. Disponível em: <https://www.dataone.org/>. Acesso em: 05 jul. 2024.

DESLANDES, S. F. O projeto de pesquisa como exercício científico e artesanato intelectual. *In*: MINAYO, M. C. S. (org.). **Pesquisa Social: teoria, método e criatividade**. Petrópolis: Vozes, 2009. p. 31-60.

DIGITAL CURATION CENTRE (DCC). **Curation lifecycle model**. Edinburgh: DCC, 2020. Disponível em: <http://www.dcc.ac.uk/resources/curation-lifecycle-model>. Acesso em: 13 ago. 2019.

ENCONTROS BIBLI. Florianópolis: UFSC, [2024?]. ISSN 1518-2924. Disponível em: <https://periodicos.ufsc.br/index.php/eb>. Acesso em: 13 maio 2024.

EYNDEN, V. V. D.; CORTI, L. Advancing research data publishing practices for the social sciences: from archive activity to empowering researchers. **International Journal on Digital Libraries**, New York, v. 18, p. 113–121, 2017. Disponível em: <https://doi.org/10.1007/s00799-016-0177-3>. Acesso em: 25 out. 2020.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 7. ed. São Paulo: Atlas, 2019.

GUPTA, S.; MÜLLER-BIRN, C. A study of e-Research and its relation with research data life cycle: a literature perspective. **Benchmarking: an international journal**, [England], v. 25, n. 6, p.1656-1680, 2018. Disponível em: <https://doi.org/10.1108/BIJ-02-2017-0030>. Acesso em: 15 out. 2020.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 8. ed. São Paulo: Atlas, 2019.

MEDEIROS, C. B. Gestão de dados científicos – da coleta à preservação. **SciELO em Perspectiva**, São Paulo, n. 39, 2018. Disponível em: <https://blog.scielo.org/blog/2018/06/22/gestao-de-dados-cientificos-da-coleta-a-preservacao/>. Acesso em: 10 jun. 2020.

MICHEL, M. H. **Metodologia e pesquisa científica em ciências sociais**: um guia prático para acompanhamento da disciplina e elaboração de trabalhos monográficos. 3. ed. São Paulo: Atlas, 2015.

MUCHERONI, M. L.; SILVA, J. F. M.; PALETTA, C. F. Entre a publicação ampliada e a multimodalidade. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 16., João Pessoa, 2015. **Anais [...]**. João Pessoa: UFPB, 2015. p. 1-19. Disponível em: <https://www.eca.usp.br/acervo/producao-academica/002735273.pdf>. Acesso em: 26 jun. 2024.

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, 2016. Disponível em: <http://dx.doi.org/10.5433/1981-8920.2016v21n2p5>. Acesso em: 20 abr. 2020.

OPENDOAR. **Welcome to OpenDOAR**. [S. l.]: OpenDOAR, [2024?]. Disponível em: <https://v2.sherpa.ac.uk/opensoar/>. Acesso em: 13 maio 2024.

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

RICHARDSON, R. J. **Pesquisa social: métodos e técnicas**. 4. ed. São Paulo: Atlas, 2017.

RODRIGUES, D. G. A. **Elementos de ciclos de vida dos dados no percurso metodológico das teses brasileiras da área de ciência da informação: um estudo diagnóstico**. 2021. 183 f. Tese (Doutorado em Ciência da Informação) – Universidade Federal da Paraíba, João Pessoa, 2021. Disponível em: <https://repositorio.ufpb.br/jspui/handle/123456789/21993>. Acesso em: 19 abr. 2024.

SANCHEZ, F. A.; VIDOTTI, S. A. B. G.; VECHIATO, F. L. A contribuição da curadoria digital em repositórios digitais. **Revista Informação na Sociedade Contemporânea**, Lagoa Nova, v. 1, p. 1-17, 2017. Disponível em: <https://doi.org/10.21680/2447-0198.2017v1n0ID12280>. Acesso em: 15 mar. 2020.

SAYÃO, L. F.; SALES, L. F. Subsídios para a construção de um modelo de avaliação de sistemas de gestão de dados de pesquisa. **Ponto de Acesso**, Salvador, v. 12, n. 3, p. 80-108, 2018. Disponível em: <http://dx.doi.org/10.9771/rpa.v12i3.28965>. Acesso em: 16 jun. 2020.

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para a preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**, João Pessoa, v. 22, n. 3, p. 179-191, 2012. Disponível em: <https://www.marilia.unesp.br/Home/Instituicao/Docentes/EdbertoFerneda/curadoria-digital---sayao.pdf>. Acesso em: 20 mar. 2020.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: Comissão Nacional de Energia Nuclear, 2015. Disponível em: http://carpedien.ien.gov.br/bitstream/ien/1624/1/GUIA_DE_DADOS_DE_PESQUISA.pdf. Acesso em: 10 out. 2017.

SIEBRA, S. A.; BORBA, V. R.; GALINDO, M.; MIRANDA, M. K. F. O.; TAVARES, L. L. L.; OLIVEIRA, J. N. N. Curadoria digital: além da questão da preservação digital. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 14., Florianópolis, 2013. **Anais** [...]. Florianópolis: UFSC, 2013. p. 1-20. 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-0>. Acesso em 20 jun. 2018.

SILVA SEGUNDO, S. J.; ARAÚJO, W. J. Curadoria e ciclo de vida de dados. *In*: DIAS, G. A.; OLIVEIRA, B. M. J. F. (org.). **Dados científicos: perspectivas e desafios**. João Pessoa: UFPB, 2019. p. 113-152.

STRASSER, C.; COOK, R.; MICHENER, W.; BUDDEN, A. **Primer on data management**: what you always wanted to know. California: CDL, 2012. Disponível em: <https://doi.org/doi:10.5060/D2251G48>. Acesso em: 13 ago. 2019.

UK DATA SERVICE. **ReShare repository**. [S. l.]: UK Data Service, c2024. Disponível em: <https://reshare.ukdataservice.ac.uk/>. Acesso em: 13 maio 2024.

VARDIGAN, M. The DDI Matures: 1997 to the Present. **IASSIST Quarterly**, Chapel Hill, v. 37, n. 1-4, p. 45-50, 2013. Disponível em: <https://doi.org/10.29173/iq501>. Acesso em: 20 jan. 2021.

VARDIGAN, M.; HEUS, P.; THOMAS, W. Data Documentation Initiative: toward a standard for the social sciences. **International Journal of Digital Curation**, New York, v. 3, n. 1, p. 107-113, 2008. Disponível em: <http://dx.doi.org/10.2218/ijdc.v3i1.45>. Acesso em: 21 out. 2020.

LIFE CYCLE REFERENTIAL MODEL OF SCIENTIFIC DATA: A METHODOLOGICAL PROPOSAL

ABSTRACT

Objective: To propose a reference model, resulting from the intersection of the Digital Curation Centre, the Data Observation Network for Earth, and the Data Documentation Initiative, for describing scientific data in Information Science, with a focus on the broad area of Social Sciences. **Methodology:** Qualitative bibliographic study with a descriptive purpose. To construct this model, an association was made between the elements of the data life cycles and the literature on methodological procedures in the Social Sciences. **Results:** It was observed that the literature clearly addresses stages of the life cycle such as planning, collecting, and analyzing. However, other phases of the cycle such as organizing, preserving, and integrating appear subtly, and some others were not identified. For a more detailed understanding of the phases and to avoid the suppression of essential elements, a new model was developed involving the following stages: planning, collecting, organizing, describing, ensuring, preserving, finding, reusing, and analyzing. **Conclusions:** The research practically contributes to the management of scientific data through the developed model, facilitating the processes of sharing and reusing them in new investigations, providing greater completeness to the information about the produced/collected data.

Descriptors: Data lifecycle. Information Science. Social Sciences. Scientific data. Methodological path.

CICLO DE VIDA MODELO REFERENCIAL DE DATOS CIENTÍFICOS: UMA PROPUESTA METODOLÓGICA

RESUMEN

Objetivo: Proponer un modelo de referencia, resultado de la intersección del Centro de Curación Digital, la Red de Observación de Datos para la Tierra y la Iniciativa de Documentación de Datos, para describir datos científicos en la Ciencia de la Información, con un enfoque en el amplio campo de las Ciencias Sociales. **Metodología:** Estudio bibliográfico cualitativo con un propósito descriptivo. Para construir este modelo, se realizó una asociación entre los elementos de los ciclos de vida de datos mencionados y la literatura sobre procedimientos metodológicos en las Ciencias Sociales. **Resultados:** Se observó que la literatura aborda claramente etapas del ciclo de vida como planificar, recopilar y analizar. Sin embargo, otras fases del ciclo como organizar, preservar e integrar aparecen sutilmente, y algunas otras no fueron identificadas. Para una comprensión más detallada de las fases y evitar la supresión de elementos esenciales, se desarrolló un nuevo modelo que involucra las siguientes etapas: planificar, recopilar, organizar, describir, asegurar, preservar, encontrar, reutilizar y analizar. **Conclusiones:** La investigación contribuye de manera práctica a la gestión de datos científicos a través del modelo desarrollado, facilitando los procesos de compartir y reutilizarlos en nuevas investigaciones, proporcionando una mayor completitud a la información sobre los datos producidos/recopilados.

Descriptores: Ciclo de vida de los datos. Ciencia de la Información. Ciencias Sociales. Datos científicos. Trayectoria metodológica.

Recebido em: 23.04.2022

Aceito em: 04.05.2024