

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO
DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO**

ANDRÉ DE OLIVEIRA

**CICLO DE VIDA DO DADO DE PESQUISA: modelo de implementação para o
Centro de Documentação e Acervo Digital da Pesquisa-UFRGS.**

PORTO ALEGRE

2018

ANDRÉ DE OLIVEIRA

**CICLO DE VIDA DO DADO DE PESQUISA: modelo de implementação para o
Centro de Documentação e Acervo Digital da Pesquisa-UFRGS**

Trabalho de Conclusão de Curso apresentado como requisito parcial para a obtenção do título de Bacharel em Biblioteconomia pela Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul.

Orientador: Prof. Dr. Rene Faustino Gabriel Junior

Porto Alegre

2018

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Dr. Rui Vicente Oppermann
Vice-Reitora: Profa. Dra. Jane Fraga Tutikian

FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO

Diretora: Profa. Dra. Karla Maria Müller
Vice-Diretora: Profa. Dra. Ilza Maria Tourinho Girardi

DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO

Chefe: Profa. Dra. Jeniffer Alves Cuty
Chefe Substituta: Profa. Dra. Eliane Lourdes da Silva Moro

COMISSÃO DE GRADUAÇÃO DO CURSO DE BIBLIOTECONOMIA

Coordenadora: Profa. Dra. Rita do Carmo Ferreira Laipelt
Coordenador Substituto: Prof. Dr. Rene Faustino Gabriel Júnior

O48 Oliveira, André de

Ciclo de vida do dado de pesquisa: modelo de implementação para o Centro de Documentação e Acervo Digital Da Pesquisa-UFRGS / André De Oliveira. -- Porto Alegre 2018.

66 f.

Orientador: René Faustino Gabriel Júnior.

Trabalho de conclusão de curso (Graduação) – Universidade Federal do Rio Grande do Sul, Faculdade de Biblioteconomia e Comunicação, Curso de Biblioteconomia, Porto Alegre, RS, 2018.

1. Ciclo de Vida do Dado de Pesquisa. 2. Gestão do Dado de Pesquisa. 3. Dado de Pesquisa. I. Faustino Gabriel Júnior, Rene, orient. II. Título.

ANDRÉ DE OLIVEIRA

**CICLO DE VIDA DO DADO DE PESQUISA: modelo de implementação para o
Centro de Documentação e Acervo Digital da Pesquisa-UFRGS**

Trabalho de Conclusão de Curso apresentado como requisito parcial para a obtenção do título de Bacharel em Biblioteconomia, pelo Departamento de Ciências da Informação, da Faculdade de Biblioteconomia e Comunicação, da Universidade Federal do Rio Grande do Sul.

Aprovado em _____ de _____ de 2018

BANCA EXAMINADORA

Prof. Dr. Rene Faustino Gabriel Junior - UFRGS Orientador

Prof. Dr. Rafael Port da Rocha – UFRGS Examinador

Prof. Dr. Fabiano Couto Corrêa da Silva – UFRGS Examinador

AGRADECIMENTOS

Agradeço primeiramente à minha família pelo apoio, especialmente à minha mãe, uma mulher forte e guerreira que mesmo não sendo alfabetizada sempre soube da importância dos estudos e sempre me incentivou.

À minha namorada Fabrina, companheira e grande amor da minha vida, que esteve ao meu lado nos últimos onze anos e sem a qual não existiria esse trabalho.

Ao Glauber, outro “culpado” por esse trabalho, agradeço pela longa amizade e pelo auxílio e leitura paciente dos meus textos.

Ao professor Rene, que me orientou com sabedoria e paciência, e mostrou os caminhos da pesquisa acadêmica.

Ao professor Rafael, que me indicou o tema e a quem peço desculpas por não ter concluído o trabalho sob sua orientação.

Aos amigos da FMP, do passado e do presente: Kátia, Jaqueline, Émerson, Juliano, Fernanda, Robson, Bruno, Marta, Cristini, Nathália e Bruna. Agradeço também e especialmente à minha chefe Patricia, pelo carinho, incentivo e ensinamentos.

Por fim, agradeço a todos os meus amigos e amigas que estiveram presentes nos últimos anos.

“Dados bons vão para o céu”

Luís Fernando Sayão

RESUMO

Este estudo analisa diretrizes para a elaboração de Ciclo de Vida de Dados de Pesquisa no contexto da estrutura de repositórios institucionais. Apresenta um panorama teórico e conceitual sobre o assunto, descrevendo seus marcos teóricos na produção científica no campo da Biblioteconomia e Ciência da Informação. Exibe um quadro referencial a partir do qual propõe observações pertinentes para a elaboração de um CVDP para o repositório digital do Centro de Documentação e Acervo Digital da Pesquisa - CEDAP, órgão auxiliar da Faculdade de Biblioteconomia e Comunicação. Reflete sobre a importância da definição de CVDP para a preservação do conhecimento gerado em instituições de ensino e pesquisa. Demonstra que, ainda que na área da Biblioteconomia e Ciência da Informação as pesquisas sobre o tema estejam em fase de desenvolvimento, elas já apontam alguns caminhos metodológicos para a aplicação prática para a formação de repositórios institucionais. Constata que cada um dos modelos analisados foram elaborados para cumprir com objetivos institucionais bastante pontuais, o que sugere que cada CVDP possui uma variável estrutural definida em termos dos objetivos da instituição. A pesquisa apresentou limitações e dificuldades, como a complexidade na categorização dos modelos de CVDP. Este trabalho pretende contribuir para o incentivo a futuras pesquisas nessa área e auxiliar na elaboração de políticas de gestão de dados de pesquisa.

Palavras-chave: Dados de Pesquisa. Ciclo de Vida do Dado de Pesquisa. Gestão de Dados de Pesquisa. CEDAP – UFRGS.

RESUMEN

Esta investigación analiza las directrices para elaboración de Ciclos de Vida de Datos de Pesquisa en el contexto de la estructura de repositorios institucionales. Presenta un panorama teórico y conceptual acerca del asunto, describiendo sus marcos teóricos en la producción científica en el campo de la Biblioteconomía y Ciencias de la Información. Exhibe un cuadro referencial a fin de proponer pertinentes observaciones para la elaboración de un CVDP para el repositorio digital del Centro de Documentação e Acervo Digital da Pesquisa – CEDAP, órgano auxiliar de la Faculdade de Biblioteconomia e Comunicação. Reflexiona sobre la importancia de la definición de CVDP para la preservación del conocimiento generado en instituciones de enseñanza y pesquisa. Demuestra que, mismo que el área de la Biblioteconomía y Ciencia de la Información las investigaciones acerca de la temática estén en fase de desarrollo, ya señalan algunos caminos metodológicos para la aplicación en la construcción de repositorios institucionales. Constata que cada uno de los modelos analizados fueron elaborados para cumplir con objetivos institucionales puntuales, lo que sugiere que cada CVDP posee una variable estructural definida no que se refiere a los objetivos de la institución. La investigación presentó limitaciones y dificultades, como la complejidad en la categorización de los modelos de CVDP. Esta pesquisa busca contribuir incentivando las futuras investigaciones en el área y auxiliando en la elaboración de políticas de gestión de datos de investigación.

Palabras clave: Datos de investigación. Ciclo de Vida de Datos de Pesquisa. CEDAP – UFRGS.

LISTA DE QUADROS

Quadro 1 - Comparativo dos conceitos aplicados ao ciclo de vida de dados de pesquisa	32
Quadro 2 - Comparativo entre os modelos de CVDP	56

LISTA DE FIGURAS

Figura 1 - O ciclo de vida dos dados DataONE.....	39
Figura 2 - Ciclo de vida dos dados GFBio	42
Figura 3 - O ciclo de vida dos dados Spiron.....	47
Figura 4 - O ciclo de vida dos dados USGS Science Data Lifecycle Model (SDLM).....	49
Figura 5 - O ciclo de vida dos dados UK Data Service.....	52
Figura 6 - O ciclo de vida dos dados DCC	54

SUMÁRIO

1 INTRODUÇÃO	10
1.1 Problema de pesquisa	13
1.2 Objetivo geral.....	13
1.3 Objetivos específicos	13
2 METODOLOGIA	14
2.1 Abordagem e natureza da pesquisa	16
2.2 Objetivos e procedimentos da pesquisa	16
2.3 Coleta E ANÁLISE dOS dados	16
3 CONTEXTO DO ESTUDO	18
4 REFERENCIAL TEÓRICO	20
4.1 DADOS DE PESQUISA.....	20
4.2 CICLO DE VIDA DOS DADOS DE PESQUISA (CVDP).....	23
4.3 GESTÃO DE DADOS DE PESQUISA.....	25
5 ANÁLISE DOS RESULTADOS DA PESQUISA	29
5.2 apresentação dos MODELOS DE CVDP.....	38
5.3 Comparativo entre os modelos de CVDP	56
5.4 desenvolvimento de uma CVDP para o CEDAP	57
6 CONSIDERAÇÕES FINAIS	59
REFERÊNCIAS	62

1 INTRODUÇÃO

O desenvolvimento da pesquisa científica, aliado à evolução tecnológica, tem lançado novos desafios para a comunidade científica. Em todo o mundo, a divulgação do conhecimento científico está intimamente ligada à necessidade de elaboração de instrumentos de coleta e análise de dados, a fim de que a pesquisa moderna consiga agregar novas aplicações e serviços para a ciência. Para que isso seja uma realidade, é necessária uma efetiva gestão do volume cada vez maior de dados produzidos e gerados a partir da atividade científica em todo o mundo.

No contexto da e-Science¹, a pesquisa científica tem gerado e aplicado uma quantidade de dados que cresce significativamente. Na atualidade, o desafio que se impõe à gestão de dados é como preservar e garantir a transmissão de conhecimento ao longo do tempo. Assim, a capacidade de produção de conhecimento da pesquisa científica depende da gestão desses dados no contexto de explosão informacional dos últimos anos.

Ao definir a disseminação e preservação de dados como um dos maiores desafios da ciência moderna é preciso reconhecer a importância da noção de Ciclo de Vida de Dados de Pesquisa (CVDP), conceito que se refere ao valor dos dados gerados por instituições de pesquisa ao longo do tempo. Em tese, as coleções e dados de pesquisa tendem a manifestar um tempo de vida, que define a capacidade das informações terem usabilidade estendida e aproveitada.

Existem vários modelos de CVDP, e cada qual segue um determinado conceito de estrutura dos procedimentos realizados no registro de dados ao longo do seu desenvolvimento. Cada campo do conhecimento ou entidade exige uma metodologia para estabelecimento do ciclo de vida dos dados gerados na pesquisa científica. Dessa forma, o CVDP é um instrumento norteador das políticas de gestão documental que

¹ A e-Science (GRAY, 2009) caracteriza-se por explorar, no próprio fazer científico, ferramentas computacionais avançadas, que possibilitam amplo acesso a recursos geograficamente dispersos, incluindo coleções de dados, instrumentos científicos e mecanismos de visualização de alto desempenho. Como movimento internacional de crescente interesse e debate, a meta essencial da e-Science é utilizar-se das Tecnologias da Informação e da Comunicação (TICs) como fundamento 'e-infraestrutura' para uma transformação profunda no método científico, de modo a extrair o maior proveito possível dos resultados de pesquisa.

estabelece etapas específicas pelas quais os dados vão se desenvolver ao longo do tempo e os processos necessários para garantir a conservação desses dados em cada etapa.

Nesse sentido, centros de documentação que trabalham com repositórios digitais precisam elaborar seus CVDP. Essa necessidade tem contribuído para o desenvolvimento de novos campos de atuação para bibliotecários e gestores de dados. Pesquisadores, por sua vez, estão envolvidos em todos os processos do ciclo de vida dos dados, como criadores e também como usuários dos dados.

Na fase de delineamento da pesquisa, o planejamento, que é a etapa mais importante da gestão de dados, leva em conta como os dados são organizados durante o desenvolvimento do projeto e de que maneira serão posteriormente compartilhados. Essa decisão é formalizada no Plano de Gestão de Dados (PGD), que define como os dados serão coletados ou gerados, processados, arquivados e disseminados em um projeto de pesquisa. Trata-se de estabelecer um compromisso em relação à forma como esses dados serão tratados ao longo da pesquisa e após a sua conclusão.

Na prática, como se trata de um campo relativamente recente entre bibliotecários, muitos estudos estão surgindo em uma tentativa de definição de conceitos e metodologias necessárias para dar conta das novas demandas. Uma das dificuldades está na elaboração, por parte dos pesquisadores e gestores de informação, de uma conceitualização consistente a respeito do CVDP. Para o grupo de trabalho citado, a pesquisa científica e o ensino da ciência podem se beneficiar do tratamento, preservação e gerenciamento adequado dos dados de pesquisa. O guia contribui de forma decisiva na atuação de bibliotecários, visto que estabelece que a Gestão de Dados de Pesquisa (GDP) se aplica aos diversos aspectos de manipulação, organização, documentação e agregação de valor aos dados. Da mesma forma, propõe que a gestão tenha como objetivo final ser uma ferramenta facilitadora no processo de compartilhamento de dados, na garantia de sustentabilidade e acessibilidade aos dados a longo prazo.

Os bibliotecários, ao atuar em colaboração com os pesquisadores, encontram-se bem posicionados, pois as bibliotecas (digitais e físicas) são as entidades mais indicadas para a custódia documental dos dados gerados a partir da pesquisa científica.

Para que a gestão dos dados de pesquisa seja efetiva ao longo do seu ciclo de vida, é importante que os bibliotecários agreguem um sólido repertório conceitual sobre o assunto.

Um dos conceitos mais elementares que envolve a gestão de dados de pesquisa é do próprio dado de pesquisa. A definição dada pela Organização de Cooperação para o Desenvolvimento Econômico (OCDE) pode sugerir uma complexidade de formas pelas quais esses tipos de dados podem se manifestar. Em relatório publicado em 2007, a OCDE define dados de pesquisa como registros factuais usados como fonte primária para pesquisa científica e que são comumente aceitos pelos pesquisadores como necessários para validar os resultados do trabalho científico. Porém, essa noção pode variar entre pesquisadores e áreas do conhecimento, influenciando a forma pela qual os dados podem se manifestar: texto, vídeo, áudio, modelos, equações, entre outros.

Nesse contexto, o Centro de Documentação e Acervo Digital da Pesquisa - CEDAP - UFRGS busca implementar um modelo de CVDP, e este trabalho objetiva contribuir com ferramentas conceituais e informações metodológicas, propondo um modelo de CVDP que atenda às suas necessidades.

Dessa forma, no primeiro capítulo deste estudo apresenta-se a metodologia de pesquisa, descrevendo a abordagem teórico-metodológica utilizada - quantitativa e qualitativa - e o caráter exploratório da investigação. Ademais, descreve-se a coleta de dados, realizada nas bases Brapci e Scielo.

No segundo capítulo, discorre-se acerca do contexto do estudo, ou seja, sobre o Centro de Documentação e Acervo Digital da Pesquisa - CEDAP - UFRGS, sua história, os objetivos de sua implantação, além dos serviços que serão oferecidos à comunidade acadêmica.

No terceiro capítulo, apresenta-se o referencial teórico desta investigação, abordando-se os conceitos de Dados de Pesquisa, Ciclo de Vida dos Dados de Pesquisa (CVDP) e Gestão de Dados de Pesquisa segundo os principais estudiosos da área.

Por sua vez, o quarto capítulo é dedicado à análise dos resultados da pesquisa, apresentando os principais modelos de CVDP selecionados a partir de uma busca no

Google pelo termo “data life cycle”. Após, traz-se o comparativo entre os modelos de CVDP descritos anteriormente. Por fim, sugere-se um modelo para utilização no CEDAP, escolhido levando-se em consideração as necessidades, características e particularidades do órgão.

No último capítulo apresentam-se as considerações finais desta investigação, explanando os objetivos alcançados, as dificuldades encontradas e as sugestões e reflexões para novos estudos relacionados ao tema dos modelos de CVDP.

1.1 PROBLEMA DE PESQUISA

Em alguns países e em determinadas áreas do conhecimento, observa-se um grau de desenvolvimento conceitual sobre repositórios digitais e CVDP bastante reconhecido e consistente. Dessa forma, os bibliotecários precisam agregar um sólido conhecimento sobre o CVDP. Na literatura é possível localizar diversos modelos de CVDP, parte-se da inquietação de qual o melhor modelo aplicável ao CEDAP, e busca-se verificar o que esses modelos podem oferecer, ou seja, de que maneira os repositórios digitais mais destacados podem contribuir com ferramentas conceituais e informações metodológicas para a criação do CVDP do CEDAP?

1.2 OBJETIVO GERAL

Recomendar um modelo de Ciclo de Vida de Dados de Pesquisa (CVDP) para o Centro de Documentação e Acervo Digital da Pesquisa - UFRGS.

1.3 OBJETIVOS ESPECÍFICOS

- a) identificar os conceitos de Dados de Pesquisa, CVDP e Gestão de Dado da Pesquisa;
- b) comparar as metodologias de elaboração de CVDP;
- c) recomendar subsídios para o desenvolvimento de um CVDP para o CEDAP.

2 METODOLOGIA

No campo da metodologia de pesquisa, vários autores têm trabalhado na definição do termo pesquisa e tentado distinguir os vários tipos de pesquisas realizadas na produção de conhecimento científico. A demarcação do conceito de pesquisa é fundamental para entendermos os vários objetivos ligados à produção de conhecimento na sociedade.

Para construir e reconstruir conhecimento, é necessária a realização de procedimentos sistemáticos que controlam os fenômenos e os objetos de pesquisa. Demo (2008, p. 20), partindo desse princípio, define pesquisa como “procedimento de fabricação do conhecimento”, e, ao mesmo tempo como “procedimento da aprendizagem (princípio científico e educativo), sendo parte integrante de todo o processo reconstutivo de conhecimento”. Nesse sentido, a pesquisa é entendida principalmente como um meio pelo qual se alcançam os resultados da produção de conhecimento.

Ainda na tentativa de aprofundar o conceito, o autor propõe a existência de quatro tipos de pesquisa (DEMO, 2008): a pesquisa teórica, “dedicada a reconstruir teorias, conceitos, ideias, ideologias, polêmicas, tendo em vista, em termos imediatos, aprimorar fundamentos teóricos e, em termos imediatos, aprimorar práticas”; a pesquisa metodológica, que trata de “inquirir métodos e procedimentos a serviço da cientificidade, polêmicas e paradigmas metodológicos, usos e abusos, tanto em âmbito mais epistemológico, quanto de controle empírico”; a pesquisa empírica, que investiga a “face empírica e factual da realidade, de preferência mensurável; produz e analisa dados, procedendo sempre pela via do controle empírico e fatural”; e a pesquisa prática, que é ligada à práxis, ou seja, à prática histórica em termos de usar conhecimento científico para fins explícitos de intervenção”.

Demo (2011), ainda explorando a conceituação sobre pesquisa, expõe que, no sentido prático, a pesquisa está ligada a algumas funções específicas na construção do conhecimento. Para o autor, a pesquisa é entendida também como um dispositivo de regulação e avaliação do quadro de conhecimentos consolidados. Nesse contexto, o autor refere-se especificamente à pesquisa que está associada ao conhecimento

científico, que é dinâmico e mutável, alcançando uma complexidade cada vez maior com o passar do tempo. Nessa perspectiva, a pesquisa coloca o desafio de “desfazer a aparência visível, observável, para surpreender a realidade por trás disso” (DEMO, 2011).

Outra função da pesquisa destacada por Gil (2010) também está afinada com a noção da pesquisa como um processo regulador do próprio quadro de conhecimentos. Para esse autor, a pesquisa é requerida quando não se dispõe de informação suficiente para responder ao problema, ou então quando a informação disponível se encontra em tal estado de desordem que não possa ser adequadamente relacionada ao problema (Gil, 2010).

As razões que motivam a realização da pesquisa, no entanto, são determinadas pela natureza incompleta dos diversos campos do conhecimento. O que é comum em todos os tipos de pesquisa é que elas envolvem certos tipos de procedimentos específicos que são fixados e limitados por alguns fatores especiais.

Para Gil (2010), os procedimentos técnicos da pesquisa são influenciados por alguns elementos, como conhecimento do assunto a ser pesquisado, curiosidade, integridade intelectual, sensibilidade social, etc. Ainda assim, a pesquisa necessita de recursos específicos para ser realizada, como recursos humanos, materiais e financeiros.

Nesse contexto, outro consenso sobre a pesquisa é que ela precisa ser planejada para que seus resultados cumpram determinados objetivos. Portanto, em uma investigação é necessário que se adotem procedimentos técnicos e teóricos controlados, com a finalidade de construir conhecimento de vários tipos e dentro de várias categorias de competências científicas.

Dessa forma, pode-se afirmar que a metodologia é o estudo da organização e dos caminhos a serem percorridos a fim de se realizar uma pesquisa e que ela visa orientar a produção da ciência, norteando passo a passo sua trajetória. Com o objetivo de destacar os caminhos e instrumentos utilizados, a seguir será apresentada a metodologia de pesquisa que guiou esta investigação.

2.1 ABORDAGEM E NATUREZA DA PESQUISA

Quanto à abordagem, esta pesquisa pode ser classificada como qualitativa. Já em relação à natureza, este estudo possui caráter aplicado, visto que objetiva gerar conhecimentos para aplicação prática, dirigidos à solução de problemas específicos, no caso, relacionados à escolha e uso de um modelo de CVDP pelo CEDAP-UFRGS.

2.2 OBJETIVOS E PROCEDIMENTOS DA PESQUISA

Segundo seu objetivo mais geral, a pesquisa é do tipo exploratória, na medida em que busca familiaridade com o objeto CVDP, objetivando propor uma aplicação desse objeto em uma realidade específica. Conforme seus procedimentos, esta pesquisa classifica-se como bibliográfica, pois é feita a partir do levantamento de referências teóricas relacionadas aos conceitos de Dados de Pesquisa, Ciclo de Vida dos Dados de Pesquisa (CVDP) e Gestão de Dados de Pesquisa. Fazendo uma análise do que está sendo produzido nas pesquisas científicas de âmbito nacional.

2.3 COLETA E ANÁLISE DOS DADOS

Para a concretização dos objetivos deste estudo, buscou-se, na literatura, os conceitos de Dados de Pesquisa, Ciclo de Vida dos Dados de Pesquisa (CVDP) e Gestão de Dados de Pesquisa em bases de dados na Brapci e Scielo. A delimitação foi para artigos publicados em todo período de abrangência das bases. A fim de organizar os conceitos coletados, criou-se uma tabela, possibilitando a comparação e análise dos conceitos de diferentes autores.

Posteriormente, realizou-se uma pesquisa de modelos de CVDP no Google. Essa pesquisa foi realizada a partir do termo "Data Life Cycle", resultando em inúmeros modelos de CVDP. A partir desses resultados brutos foram selecionados apenas os modelos que continham uma definição de CVDP, resultando em sete modelos.

Após a análise e comparação entre os modelos, foi escolhido o modelo a ser sugerido como o mais adequado para utilização pelo CEDAP-UFRGS.

3 CONTEXTO DO ESTUDO

O CEDAP é um órgão auxiliar da UFRGS, instituído em 2012 e vinculado à FABICO, conforme consta no art. 96 do estatuto UFRGS. Desde sua concepção, o centro tem por objetivo dar suporte à pesquisa científica, tecnológica, artística e cultural realizada na UFRGS, e também fora dela, através da curadoria de dados de pesquisa, natos ou provenientes da digitalização, a fim de permitir o seu reuso, validar resultados, preservar memória e aplicar essas experiências no ensino e para o bem público.

Oferece assessoria em gestão e curadoria de dados da pesquisa, apoiando pesquisadores durante os estágios ativos do ciclo de vida dos dados (planos de gestão de dados, coleta e organização dos dados), assim como oferecendo recursos para disseminação e para armazenamento a longo prazo (repositório de dados da pesquisa).

Dispõem de serviços para conversão de documentos para o formato digital, por meio de uma abordagem interdisciplinar das áreas da ciência da informação, comunicação e computação, desenvolvendo projetos de investigação, participando de redes com temáticas afins, estimulando a produção científica e promovendo o compartilhamento e a disseminação de novos conhecimentos.

O CEDAP tem como filosofia o processo de inovação aberta, visto que explora tanto ideias internas como externas, trata a Pesquisa e Desenvolvimento (P&D) como um sistema aberto e assume que o conhecimento útil deva ser amplamente distribuído. Para tal desenvolveu um ambiente fundamentado em trabalho colaborativo, em que o conhecimento e as práticas em dados da pesquisa são adquiridos e socializados por meio de um sistema wiki, baseado em corpos de conhecimento, processos de inovação e modelos conceituais.

O Centro está em fase de estudos e operacionalização de seu repositório de dados em DSpace, empregando uma estratégia que passa a considerar o repositório de dados como parte final de uma cadeia de ações a ser explorada. Como decorrência, estuda-se todo o ciclo de vida do dado da pesquisa, que envolve, desde o apoio aos pesquisadores na elaboração de seus planos de gestão de dados de pesquisa, na coleta e organização dos dados, assim como no armazenamento a curto e longo prazo desses dados em um repositório confiável.

A constituição do CEDAP reforçou o pioneirismo da UFRGS entre as universidades brasileiras no compartilhamento de dados de pesquisa, e acompanha uma tendência mundial encontrada em universidades que estão bem posicionadas no cenário internacional. Ela também representou mais uma parceria da FABICO com o CPD da UFRGS.

4 REFERENCIAL TEÓRICO

Como suporte conceitual do trabalho, descreve-se, no presente capítulo, os marcos teóricos sobre o assunto principal e aqueles que são fundamentais para o entendimento do contexto da pesquisa. Em nosso idioma, as pesquisas sobre os assuntos do trabalho têm crescido nos últimos anos devido à relevância dos repositórios digitais na atualidade. Verifica-se uma certa consistência na definição dos termos ao mesmo tempo em que alguns poucos pesquisadores no país tem sido referenciados em quase todas as pesquisas posteriores.

Dentro do quadro referencial trazido para esta pesquisa, ainda predomina as investigações que partem do campo prático da elaboração de repositórios digitais para o desenvolvimento de teorias e conceitos sobre o assunto.

4.1 DADOS DE PESQUISA

Os estudos e o desenvolvimento teórico sobre dados de pesquisa foram impulsionados pela publicação do relatório da Organização para a Cooperação e desenvolvimento Econômico (2007). Desde então, o termo tem se desenvolvido tanto quanto a sua aplicação, visto sua importância para o progresso do conhecimento científico.

A descrição de suas características básicas pode apresentar algumas diferenças conforme o campo da ciência onde ele é enunciado, porém o termo tem sido amplamente reconhecido, pois sua divulgação é importante no processo de validação dos resultados obtidos na pesquisa, proporcionando seu reuso e impulsionando novas pesquisas.

Os dados de pesquisa podem se manifestar em vários formatos e ser utilizados para várias finalidades no contexto da pesquisa científica. Segundo Sayão e Sales (2016), dado de pesquisa pode ser o resultado de um experimento realizado em um ambiente controlado de laboratório, um estudo empírico na área das ciências sociais ou a observação de um fenômeno cultural ou da erupção de um vulcão num determinado momento e lugar.

Isso significa que o dado de pesquisa pode assumir várias aparências e formatos (imagens, fórmulas, figuras, tabelas, entre outros), em diferentes formatos e suportes.

Da mesma forma, pode assumir diferentes níveis de processamento e agregação. A categorização dos dados de pesquisa, ajuda no entendimento da sua natureza e finalidade e, ao mesmo tempo, oferece um quadro contextualizado que pode sustentar a gestão de dados de pesquisa. O relatório da National Science Board, segundo Sayão e Sales (2016) categorizam dados de pesquisa em dados experimentais, resultados de estudos em ambientes controlados de laboratórios; dados computacionais, que são produtos da execução de modelos computacionais que simulam uma dada realidade; e dados observacionais, que são resultados de observações de fenômenos que se desenrolam em lugares e tempos específicos.

Cada uma dessas categorias exige um nível específico de gestão, relacionado à forma de tratamento do dado de pesquisa e sua utilização ao longo do tempo. O planejamento adequado de repositórios e coleções de dados de pesquisa se justifica a partir da necessidade de controle e formas de armazenamento necessárias para cada uma das categorias desse tipo de dado.

No ambiente de pesquisa, todas as informações geradas tanto no processo quanto nos resultados podem ser encaradas como dados de pesquisa, pois caracterizam-se como informações que oferecem a narrativa sequencial dos procedimentos da pesquisa e podem ser utilizadas no desenvolvimento de outros procedimentos científicos e institucionais. Para que sejam aproveitadas essas informações, esses dados precisam incorporar metadados e documentações que descrevam seu contexto e as ferramentas utilizadas para criá-los. Para Sayão (2015), esse tipo de dado pode ser caracterizado a partir de vários pontos de vista. Em sua categorização, os dados de pesquisa podem ser encarados conforme sua natureza, sua origem e seu status de fluxo de trabalho de pesquisa.

Sayão (2015) estabelece, a partir do *Policy-making for research data in repositories: a guide*, três origens para os dados de pesquisa: dados observacionais, dados computacionais e dados experimentais. Os dados observacionais são obtidos por meio de observações diretas, que podem ser associadas a lugares e tempos específicos, como por exemplo, a erupção de determinado vulcão numa data

específica, a fotografia de uma supernova, o levantamento das atitudes de uma comunidade (SAYÃO, 2015). Portanto, esses dados são registros históricos que não podem ser coletados uma segunda vez, por isso devem receber um tratamento de preservação permanente.

Diferentemente dos dados observacionais, os dados computacionais resultam da execução de modelos computacionais ou de simulação, por exemplo, no domínio da física ou para a criação de ambientes virtuais culturais ou educacionais (SAYÃO, 2015). Esta categoria de dados pode ser replicada ao longo do tempo, por isso a sua preservação permanente pode não ser necessária, pois os dados de pesquisa são reutilizáveis e seu uso é interminável.

Outra categoria de dados surge a partir de situações controladas em laboratório: são os dados experimentais. Esse tipo de dados pode ou não ser objeto preservação permanente, pois nem sempre é fácil conduzir as mesmas condições de experimentos realizados em laboratório.

Cabe destacar, portanto, que o conhecimento das categorias descritas anteriormente pode ajudar na aplicação de uma estratégia de arquivamento e preservação dos dados de pesquisa.

A categoria natureza, por sua vez, define tanto os diversos suportes quanto às formas nas quais os dados de pesquisa podem se manifestar: número, imagem, vídeo, equações, entre outras. Os dados de pesquisa podem ainda pertencer a três fases distintas da pesquisa, segundo classificação de Sayão (2015):

- a) dados brutos, crus ou preliminares (*raw data* em inglês): são dados que vêm diretamente dos instrumentos científicos;
- b) dados derivados: são resultados do processamento ou combinação de dados brutos ou de outros dados;
- c) dados canônico ou dados referenciais: são coleções de dados consolidados e arquivados geralmente em grandes centros de dados, por exemplo, sequência genética, estrutura química, etc.

Cabe destacar que no debate sobre o acesso à informação em diferentes níveis, tem se criado novos modelos de disseminação dos dados de pesquisa. Tais modelos

promovem tratamentos alinhados à divulgação e ao compartilhamento do conhecimento acumulado nos vários campos de pesquisa. Esse movimento de ampliação do acesso é justificado, segundo Sayão e Sales (2016), porque, na maioria das vezes, trata-se de pesquisas financiadas com recursos do poder público. Portanto, é fundamental, para o progresso científico, que os dados de pesquisa sejam compartilhados, a fim de agregar autenticidade e veracidade ao corpo de conhecimentos gerados nas pesquisas.

4.2 CICLO DE VIDA DOS DADOS DE PESQUISA (CVDP)

Para que os dados gerados no processo de pesquisa sejam otimizados quanto ao seu uso e reuso, inclusive depois de finalizada a pesquisa, é necessária uma análise e a delimitação das diversas fases relacionadas ao acesso e uso dos dados. Tornar evidentes os fatores e os diferentes momentos desse processo é a função do Ciclo de Vida dos Dados de Pesquisa (CVDP).

No geral, a vida útil dos dados de pesquisa e de suas coleções ultrapassam o projeto de pesquisa, pois, pesquisadores, professores e outros profissionais podem continuar trabalhando nesses dados mesmo quando a pesquisa em si tenha sido concluída. Da mesma forma, novos projetos de pesquisa podem surgir a partir e utilizando-se desses dados.

A elaboração do CVDP está subordinada ao campo do conhecimento onde ela é aplicada. Por isso, este ciclo está orientado por modelos específicos que se adequam à determinados contextos. Segundo Sayão (2015).

Há uma série de concepções de modelos de ciclo de vida de dados de pesquisa, cada um com particularidades e objetivos determinados, muitas vezes orientados para domínios de conhecimentos específicos. A importância desses modelos é que eles oferecem uma estrutura que representa as muitas operações que precisarão ser realizadas sobre os registros de dados durante a sua vida, garantindo que eles possam ter sua usabilidade otimizada e estendida.

Corrêa (2016, tradução nossa) especifica que os modelos de ciclo de vida dos dados de pesquisa concluem sua vida útil após serem criados, analisados e preparados para serem depositados em um repositório.

Dentre os modelos usados por profissionais da informação, pesquisadores e gestores de informação estão: o Digital Curation Center (DCC) Curation Lifecycle Model, o DataOne Data Lifecycle e o UK Data Archive Data Lifecycle.

Dentro deste ciclo, os pesquisadores podem estar envolvidos em todas as etapas, tanto como criadores, como usuários dos dados. No entanto, alguns projetos de pesquisa podem partir e se desenvolver dentro de uma etapa específica, ou mesmo não seguir de forma linear estas etapas. Esta flexibilidade pode se manifestar mais expressivamente em determinados tipos de pesquisa, conforme objetivos de seu projeto.

Partindo da perspectiva da Ciência da Informação, Sant'Ana (2016) enfatiza a importância da interação entre profissionais da informação e da computação para garantir algumas qualidades importantes aos dados gerados no processo de pesquisa.

Nesse sentido, o conhecimento advindo da Ciência da Informação e da Computação interagem com o usuário para assegurar integridade, qualidade e preservação dos dados dentro de seu ciclo de vida. Segundo este autor, essas características são fundamentais no contexto de elaboração de um repositório de dados.

A importância de se garantir a integridade dos dados em cada fase da pesquisa, somado ao acesso facilitado às ferramentas tecnológicas de gestão de informação faz com que, no processo de planejamento de uma base de dados, se tenha a preocupação com a definição de suportes de informação. Para Sant'Ana, essa preocupação gera as seguintes perguntas quais são os dados disponíveis? Quais destes dados serão armazenados? Qual estrutura (física e lógica) será utilizada para seu armazenamento? Como garantir a permanência dos dados complementares sobre a coleta para que se tenha garantido o contexto de sua obtenção? Estes dados podem representar um risco a privacidade dos indivíduos ou instituições neles referenciados de alguma forma? Como as partes de sua estrutura lógica serão interligadas e como serão mantidas as interligações com outros conjuntos de dados? Como garantir que os elementos que sustentam a sua qualidade sejam mantidos?

Tem-se o direito de armazenar estes dados? Todos os aspectos que podem contribuir para sua encontrabilidade estão sendo armazenados? Todos os fatores para sua utilização ao longo do tempo estão sendo mantidos? (SANT'ANA, 2016).

Outras questões que devem guiar o processo de elaboração de CVDP estão relacionadas à privacidade, à qualidade, aos direitos autorais, integração, disseminação e preservação. No contexto de um projeto de repositório de dados, estas características devem estar presentes em todas as fases do CVDP, pois são elas que atribuirão consistência e eficiência ao repositório.

4.3 GESTÃO DE DADOS DE PESQUISA

Em todos os campos científicos, a importância da Gestão de Dados de Pesquisa (GDP) tem sido reconhecida e discutida com grande interesse pelos pesquisadores. Isso ocorre porque a gestão de dados de pesquisa tem apoiado decisivamente a produção de conhecimento científico.

Em um cenário onde a relevância da utilização dos dados de pesquisa cresce, é natural e justificável que esses dados sejam tratados de forma a permitir o acesso e o uso mais eficiente possível. Desse fato se pode apreender que, para que sejam verdadeiramente úteis, os dados de pesquisa devem possuir estrutura e organização. É nesse sentido que se configura a necessidade de aprimoramento dos mecanismos e das práticas de gestão de dados de pesquisa em universidades e organizações em P & D (VISOLI; DRUCKER, 2017).

O cenário brasileiro de pesquisa e produção de conhecimento começa a apresentar, na última década, algumas iniciativas que representam a preocupação com a gestão de dados de pesquisa. No entanto, é na Europa que vêm surgindo esforços mais determinados no sentido de atribuir planejamento e gestão aos sistemas de armazenamento e acesso aos produtos da e-Science. Segundo Vaz (2011), o Reino Unido criou, em 2011, um programa que visava prover infraestrutura e facilidades necessárias para a pesquisa colaborativa, acelerar a emergência da próxima geração de padrões de plataforma aberta para serviços globais de informação e resolver os

principais desafios em processamento, comunicação e armazenamento de grandes volumes de dados.

Foi na Europa que também surgiu o projeto Horizonte 2020 (CAVALCANTI; SALES, 2017), maior programa europeu de apoio à pesquisa e à inovação, tem novas oportunidades de apoio a pesquisas colaborativas entre cientistas. Esta política incentiva aos participantes do Horizonte 2020 a tornarem os seus dados de investigação localizáveis, acessíveis, interoperáveis e reutilizáveis.

O projeto tem um documento descreve como preparar um plano para submissão a projetos dentro do programa Horizon2020. Enumera perguntas que um pesquisador se deve fazer no preparo de planos. Com isso, auxilia na criação do plano de gestão de dados de pesquisa.

Entende-se que o processo de gestão de dados de pesquisa deve, para que seja efetivo, envolver políticas concretas e investimento em diversos tipos de recursos. Entre os recursos fundamentais para que a gestão de dados de pesquisa seja uma realidade está a infraestrutura tecnológica, traduzida em desenvolvimento de banco de dados que dinamizam o processo de armazenamento e acesso aos dados de pesquisa. Os Estados Unidos foi pioneiros nesse tipo de ação quando, em 2007, criou o programa Sustainable Digital Data Preservation and Access Network Partner (DataNet) a fim de desenvolver um conjunto de organizações exemplares de infraestrutura nacionais e globais de dados de pesquisa (apelidado DataNet Partners) que oferecem oportunidades únicas para as comunidades de pesquisadores para avançar a ciência e/ou engenharia, pesquisa e aprendizagem (CHOUDHURY, 2010)

No caso dos EUA, é considerável o interesse das grandes corporações pelo tema. A Microsoft, por exemplo, mantém profissionais em sua área de pesquisa que são especializados no estudo da e-Science e de repositórios digitais. Nesse sentido, segundo Costa e Cunha (2014), Tony Hey, gestor da área de pesquisa da Microsoft, é um dos mais ativos colaboradores da e-Science. Além de ter sido o diretor do e-Science Core Programme do Reino Unido, ele é um dos nomes mais recuperados no Google Acadêmico quando o assunto é e-Science.

Outra dimensão da gestão de dados de pesquisa envolve a fixação de procedimentos operacionais específicos de tratamento desses dados. Conceito

examinado por Sayão e Sales (2015), o Plano de Gestão de Dados (PGD) é uma ferramenta que controla o tratamento dos dados de pesquisa, no sentido de garantir as várias características que asseguram a consistência e o acesso às informações geradas no âmbito da pesquisa. Trata-se do planejamento que se inicia quando a pesquisa ainda está sendo delineada e deve considerar como os dados serão gerenciados durante o desenvolvimento do projeto e como eles serão compartilhados depois. Dessa forma, é necessário formalizar as ações e compromissos que serão estabelecidos em relação aos dados desde os seus primeiros estágios (SAYÃO; SALES, 2015).

O PGD propõe o ciclo de vida da gestão de todos os dados que serão trabalhados durante o processo de pesquisa, formalizado em um documento que descreve, entre outros, a metodologia que será aplicada em todo o processo.

A necessidade de um PGD se justifica também em função do pré-requisito para participação de projetos de auxílio, pois cada vez mais as agências financiadoras exigem que este plano faça parte dos projetos de pesquisa.

Existe também o aspecto ético da pesquisa, o qual envolve a necessidade de elaboração de Plano de Gestão de Dados. Nesse sentido, na comunidade científica há a preocupação acerca de como os dados de pesquisa serão preservados e compartilhados de forma mais ampla. Para Sayão e Sales (2015), as principais razões para o estabelecimento de um PGD são:

- a) Ajustar o seu projeto de pesquisa às políticas mandatórias da sua instituição e/ou dos órgãos de fomento à pesquisa.
- b) Assegurar a integridade da pesquisa e o seu potencial de replicação.
- c) Assegurar que os dados e demais registros de pesquisa sejam acurados, completos, autênticos e confiáveis.
- d) Aumentar a sua eficiência como pesquisador – um plano que organize os dados e seu armazenamento permite que você foque na sua pesquisa. Você estará mais capacitado a localizar e usar os seus dados e compartilhá-los com os seus colaboradores.

- e) Permitir que os seus dados sejam compreensíveis agora e no futuro – se os dados são bem documentados antes e durante a formação da coleção de dados, eles serão mais facilmente entendidos e reutilizados.
- f) Economizar tempo e recursos a longo prazo.
- g) Aumentar a segurança dos dados e minimizar os riscos de perda.
- h) Evitar a duplicação de esforços na coleta ou regeneração dos dados, possibilitando que outros pesquisadores se beneficiem dos seus dados e os interprete em outros contextos e com novas visões.
- i) Aumentar a visibilidade da pesquisa – se os seus dados foram planejados para estarem organizados e corretamente arquivados, eles poderão ser identificados, recuperados e citados, aumentando a visibilidade da sua pesquisa e o seu prestígio como pesquisador.
- j) Tornar mais fácil a preservação e o arquivamento – ajustando antecipadamente a geração de dados tomando como referência as práticas, formatos e demais padrões mais adequados ao arquivamento e à preservação de longo prazo, torna a gestão de dados mais fácil e menos custosa; além do mais, tornam os dados mais aderentes aos requisitos dos repositórios e centros de dados.

A Biblioteconomia tem, aos poucos, avançado em direção a práticas de gestão de dados de pesquisa. A importância da atuação de profissionais bibliotecários no gerenciamento desse tipo de dados tem se firmado ao passo que as pesquisas vão esclarecendo e, de certa forma, introduzindo noções e conceitos sobre esse campo de trabalho que tem crescido ao mesmo tempo em que as práticas científicas se tornam mais complexas, demandando novas necessidades da prática da organização da informação.

5 ANÁLISE DOS RESULTADOS DA PESQUISA

De forma a atender o primeiro objetivo específico de identificar conceitos sobre Gestão de Dado da Pesquisa e CVDP, foi realizada uma busca na Brapci utilizando os termos definidos na metodologia, como resultado da busca, foram localizados vinte e um trabalhos, dos quais foram realizadas leituras nos os títulos, resumos e palavras-chave para verificar sua pertinência ao tema. Resultando em seis trabalhos representativos.

A mesma busca foi realizada na base do Scielo, onde foram identificados doze artigos, dos quais três relevantes, entretanto dois já haviam sido identificados na Brapci. Totalizando o *corpus* de análise em sete artigos.

O primeiro artigo identificado é de autoria de Sales e Sayão (2012) com o título “O impacto da curadoria digital dos dados de pesquisa na comunicação científica”, publicado na revista Encontros Bibli. O trabalho apresentou o conceito de curadoria digital, bem como a tendência que começava a surgir pela demanda de acesso ao livre aos dados de pesquisa, a criação de repositórios digitais, o gerenciamento e a guarda desses dados.

O segundo artigo identificado, publicado na revista Tendências da Pesquisa Brasileira em Ciência da Informação, também autoria de Sales e Sayão (2013), denominado de “Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país”, este trabalho tem o intuito de expor o desafio da gestão de dados de pesquisa, seu uso e reuso, apresentando elementos necessários para a definição de um modelo de curadoria digital, onde, que para isso foi usado como principal recurso metodológico o exame de três relatórios considerados fundamentais no endereçamento de questões de curadoria de dados de pesquisa: os relatórios da National Science Foundation (NSF), do Digital Data Curation (DCC) e da Organização para a Cooperação e Desenvolvimento Econômico (OCDE). Por fim apresenta os seguintes pontos a serem apreciados para definição desse modelo: aspectos políticos, infraestrutura organizacional e tecnológica, pesquisa em curadoria digital, desenvolvimento de coleções, formação de especialistas, sustentabilidade econômica, implicações sociais, éticas e legais e oferecimento de serviços.

O artigo de Sant'ana (2013), terceiro identificado, sob o título “Ciclo de vida dos dados e o papel da Ciência da Informação”, foi publicado nos Anais do 10 ° Encontro Nacional de Pesquisa em Ciência da Informação, o autor oferece um novo olhar sobre o ciclo de vida dos dados apresentando as fases que compõem um ciclo: coleta, armazenamento, recuperação e descarte, permeadas por fatores transversais e presentes em todas as fases: privacidade, integração, qualidade, direito autoral, disseminação e preservação. Mostra como a Ciência da Informação pode oferecer um novo enfoque centrado nos dados e fazendo a ponte entre os usuários e os dados que eles necessitam.

Com o título de “Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação” igualmente de autoria de Sant'ana (2016), este quarto artigo encontrado foi publicado na revista Informação & Informação e demonstra uma nova visão para o CVDP, que pressupõe, como elemento central, os próprios dados, amparando-se nos conceitos e contribuições que a Ciência da Informação pode oferecer. Um novo enfoque, agora centrado nos dados, e contribuir para a otimização do Ciclo de Vida dos Dados como um todo, ampliando as pontes entre os usuários e os dados que necessitam, sem abrir mão da reflexão sobre o papel de outras áreas chave como a Ciência da Computação.

Na revista Ponto de Acesso foi publicado o quinto artigo dos autores Bertin; Visoli; Drucker (2017), nomeado de “A gestão de dados de pesquisa no contexto da e-Science: benefícios, desafios e oportunidades para organizações de P&D”, esse trabalho situa a gestão de dados de pesquisa como elemento propulsor do avanço científico e tecnológico no novo paradigma da e-Science, e alerta quanto aos desafios que se impõem às universidades e organizações atuantes no setor de Pesquisa e Desenvolvimento. Discute os princípios e os fundamentos da gestão de dados de pesquisa e propõe um roteiro para a implantação de um programa institucional de gestão dados de pesquisa.

O artigo de Cavalcanti; Sales (2017), sob o título de “Gestão de dados de pesquisa: um panorama da atuação da União Europeia”, publicado na revista BIBLOS, é o sexto localizado, e apresenta um panorama atual da gestão de dados de pesquisa,

dos repositórios existentes e um panorama geral das políticas implementadas e em desenvolvimento na União Europeia.

Por fim publicado na revista Pesquisa Brasileira em Ciência da Informação e Biblioteconomia o artigo intitulado “A contribuição da curadoria digital em repositórios digitais” dos autores Sanchez, Vidotti e Vechiato (2018), que explana sobre Curadoria Digital em Repositórios Digitais, a partir do modelo de ciclo de vida. Verifica que as aplicações de atributos presentes no modelo do ciclo de vida da Curadoria Digital usado para essa pesquisa transformam ambientes informacionais digitais em espaços colaborativos de qualidade em relação a padrões e normas, organização, recuperação, acesso, disseminação e preservação da informação.

De forma a sistematizar os conhecimentos aplicados nos CVDP, buscou-se concentrar em um quadro os principais conceitos identificados e os autores dos artigos recuperados. O resultado dos conceitos identificados na literatura é apresentado no Quadro 1.

Quadro 1 - Comparativo dos conceitos aplicados ao ciclo de vida de dados de pesquisa

Conceito	Autores	Descrição
Ciclo de vida do dado de pesquisa (CVDP)	SAYÃO; SALES, 2012.	O CVDP permite uma visão coletiva sobre o conjunto de funções necessárias à curadoria e à preservação de dados de pesquisa. Além de definir papéis, responsabilidades e conceitos, ele explicita a infraestrutura de padronização e as tecnologias que devem ser implementadas.
	SANT'ANA, 2013	CVDP define as fases de Coleta, Armazenamento, Recuperação e Descarte e os fatores que estão presentes em cada uma destas fases: Privacidade, Integração, Qualidade, Direitos Autorais, Disseminação e, Preservação. Para explicitar cada uma destas fases e fatores do CVDP, passamos a analisar cada uma delas em relação ao contexto de um projeto de repositório de dados.
	SAYÃO; SALES, 2016	CVDP é um ciclo de geração de conhecimento que necessita de infraestruturas informacionais formalizadas para que os dados se tornem visíveis para as devidas comunidades acadêmicas, para as instituições de pesquisa e agências de fomento e para a sociedade como um todo. Trata-se dos dados digitais de pesquisa que muito rapidamente deixam de ser meros subprodutos das atividades de pesquisa e se tornam um foco de grande interesse para todo o mundo científico.
Dados	NATIONAL RESEARCH COUNCIL, 1999.	Dados são fatos, números, letras e símbolos que descrevem um objeto, uma ideia, condição, situação ou outros fatores.
Dados de pesquisa	NATIONAL SCIENCE BOARD, LONG-LIVED DIGITAL DATA COLLECTIONS, 2005	Qualquer informação que possa ser armazenada em formato digital, incluindo texto, números, imagens, vídeos ou filmes, áudio, software, algoritmos, equações, animações, modelos, simulações, etc.
	OECD, 2007.	Dados científicos ou dados de pesquisas, conforme definição da OCDE, podem ser definidos como "registros de fatos usados como fontes primárias na investigação científica e que geralmente são aceitos na comunidade científica como necessários para a validação dos resultados da pesquisa."
	NATIONAL SCIENCE FOUNDATION (NSF), 2010.	Apresenta uma categorização das origens dos dados: dados de observação (registros históricos, como medidas de precipitação, que não podem ser reproduzidos e necessitam de preservação permanente), dados computacionais (resultantes de simulações, que teoricamente podem ser reproduzidos se for preservada informação sobre o modelo e a sua execução) e dados experimentais (que também não

		são facilmente reproduzíveis).
	SAYÃO E SALES, 2016.	Dados de pesquisa podem se manifestar em vários formatos a serem utilizados para várias finalidades no contexto da pesquisa científica.
	SAYÃO E SALES, 2016.	Dado de pesquisa tem uma amplitude de significados que vão se transformando de acordo com domínios científicos específicos, objetos de pesquisas, metodologias de geração e coleta de dados e muitas outras variáveis.
	CAVALCANTI; SALES, 2017.	Dados de pesquisa podem ser entendidos como um conjunto de informações brutas, primárias, produzido durante o desenvolvimento da pesquisa e que precisa ser analisado e interpretado para a conclusão do trabalho.
Gestão de Dados de Pesquisa (GDP)	RIBEIRO, Cristina; et al, (2010)	Na gestão de dados científicos incluem-se as ações de representação dos dados e seu armazenamento, a associação de metadados que os descrevem, que ajudam a interpretá-los e documentam o seu uso, a organização dos dados em coleções, a indexação dos dados para pesquisa e todas as formas de apresentação dos dados.
	SAYÃO; SALES, 2013.	GDP abre a possibilidade de verificação confiável dos resultados e permite pesquisas transversais e inovadoras desenvolvidas sobre informações já existentes, encurtando o ciclo clássico de comunicação científica e abrindo novas formas de interlocução e de otimização de recursos financeiros.
	SAYÃO; SALES, 2015.	GDP tenta assegurar que os dados possam ser compreendidos e interpretados por outros pesquisadores ao longo do tempo. Para isso é essencial uma descrição clara e detalhada dos dados, anotações adicionais e informações que contextualizam os dados e possibilitem que transmitam informação e conhecimento no tempo e no espaço.
	SANT'ANA, 2016.	GDP requer, por parte de seus detentores, planejamento e ações concretas que tragam eficiência não só para coleta e armazenamento como também e, principalmente, para fase de recuperação desses dados ampliando sua visibilidade e potencial uso.
	SAYÃO; SALES, 2016.	GDP são ações que coletivamente permeiam o ciclo DVDP e se consolidam, de maneira ideal, em consonância com padrões de ampla aceitação, condicionantes de domínios disciplinares, requisitos estabelecidos pelos pesquisadores e, quando existentes, políticas institucionais e diretrizes de alcance nacional e internacional.

	SESARTIC; TOWE, 2016.	A gestão de dados científicos vem ganhando atenção crescente nos últimos anos, devido, em grande parte, à conscientização sobre o valor contido nos dados científicos e dos riscos de perder esses dados ao longo do tempo.
	ARELLANO, 2018.	Um sistema de gestão de dados de pesquisa garante que esses dados, sua codificação e fluxo poderão estar disponíveis e serem citados. Sua documentação é um dos princípios que destaca a reprodutividade dos dados.
	ARELLANO, 2018.	As práticas de gerenciamento de dados cobrem todo o ciclo de vida dos dados, desde o planejamento da investigação até a realização, bem como de fazer backup de dados, uma vez que eles são utilizados para a preservação de dados a longo prazo entregues após a conclusão da investigação.
	ARELLANO, 2018.	O objetivo de um plano de gestão de dados é considerar os muitos aspectos do gerenciamento de dados, geração de metadados, preservação de dados e análise antes do início do projeto, garantindo que os dados sejam bem gerenciados no presente e preparados para preservação no futuro.
	ARELLANO, 2018.	A gestão de dados científicos cobre todo o chamado “ciclo de vida” dos dados, ou seja, desde a sua coleta até o armazenamento de longo prazo, passando por uma série de processamentos de limpeza, curadoria, anotação, indexação e transformação. Grande parte da pesquisa científica de hoje exige algum tipo de análise e processamento de dados. Com isto, o planejamento da gestão dos dados utilizados e gerados em uma pesquisa passou a fazer parte integral da metodologia científica, sendo, inclusive, considerado como um dos itens necessários de boas práticas de pesquisa.

Fonte: Dados da pesquisa (2018).

Em relação ao conceito de CVDP, cabe destacar que, nos artigos estudados, os autores não chegam a um consenso na definição de CVDP. Sayão; Sales (2012) exemplificam que o modelo de CVDP é capaz de demonstrar o conjunto de funções imprescindíveis à organização e preservação dos dados de pesquisa. Nesse contexto, esse modelo permitiria que sua estruturação fosse realizada a partir de conceitos, funções e responsabilidades bem delimitadas, com auxílio de certa infraestrutura e de tecnologias padronizadas. Para esses autores, é necessário que a produção de conhecimento científico possa contar com infraestruturas informacionais formalizadas, de forma a ganhar maior visibilidade na comunidade acadêmica. Dessa forma, os dados produzidos deixam de ser descartáveis e passam a gerar um novo interesse no mundo científico. (SAYÃO; SALES, 2016).

Sant'Ana (2013), por sua vez, demonstra as etapas e fatores necessários no modelo CVDP. Ao coletar, armazenar, recuperar e descartar os dados de pesquisa, devemos observar fatores como privacidade, integração, qualidade, direitos autorais, disseminação e preservação.

A partir do exposto, observa-se uma variedade de concepções em relação ao CVDP, sem haver uma definição. Há, no entanto, um consenso entre os autores de que certas etapas são imprescindíveis ao se propor um CVDP. Acredita-se que a inexistência de uma definição de CVDP se deva ao fato de que os estudos sobre esse tema ainda são recentes.

Como parte imprescindível do CVDP, cabe abordar as definições de dados de pesquisa. Para a National Research Council (1999), um dado pode ser um fato, número, letra ou símbolo que descreve um objeto, uma ideia, condição, situação, entre outros. Essa definição foi aprofundada ao considerar a especificidade dos dados de pesquisa. Nesse sentido, novamente os autores divergem quanto à definição desse termo. Para a *National Science Board* (1999), os dados de pesquisa referem-se a "... QUALQUER informação que possa ser armazenada em formato digital" (destaque nosso). A *National Science Foundation* (NSF, 2010) categoriza os dados de pesquisa em dados observacionais, que nem sempre podem ser reproduzidos em outras coletas, dados computacionais, que podem ser reproduzidos, além de dados experimentais, que não são facilmente reproduzíveis. Delimitando esse conceito, a Organization for

Economic Co-operation and Development - OECD (2007) determina que esses dados são fontes primárias da investigação, ou seja, dados coletados pelo pesquisador para sua pesquisa.

Cavalcanti e Sales (2017) afirmam que os dados de pesquisa são um conjunto de informações brutas e primárias, fruto do desenvolvimento da pesquisa, os quais devem ser analisados e interpretados para validar o trabalho. Sayão e Sales (2016) corroboram afirmando que o termo dados de pesquisa possui uma amplitude de significados, os quais podem variar de acordo com a área, objeto de pesquisa, metodologias de geração e coleta, entre outros fatores. Ou seja, os dados de pesquisa poderão revelar-se em vários formatos e serem utilizados com diferentes finalidades na variedade de áreas da pesquisa científica. (SAYÃO E SALES, 2016).

Considerando-se as definições expostas, observa-se que os dados de pesquisa caracterizam-se principalmente por sua abundância, que deve ser lapidada e interpretada conforme a área e a necessidade do pesquisador, assim como os objetivos da pesquisa. Assim, os dados de pesquisa devem ser preservados e contar com uma melhor descrição de seus metadados, para futuro reaproveitamento em novas investigações.

Para um bom encaminhamento e organização dos dados de pesquisa, é necessário um planejamento de como serão utilizados, tratados, organizados e armazenados esses dados. Para isso, cabe desenvolver um Plano de Gestão de Dados de Pesquisa.

Conforme Sesartic e Towe (2016), o risco de perda dos dados de pesquisa ao longo do tempo demonstrou a importância de uma conscientização sobre o valor desses dados e da necessidade de gestão dos mesmos. A gestão de dados de pesquisa pode ser definida como ações coletivas que permeiam o CVDP. Essas ações devem estar alinhadas com padrões amplamente aceitos em cada área, requisitos estabelecidos pelos investigadores e, ademais, políticas institucionais e diretrizes nacionais e internacionais. (SAYÃO; SALES, 2016).

Sayão e Sales (2015) destacam que o principal objetivo da gestão de dados científicos é assegurar a reutilização dos dados por outros investigadores, ao longo do tempo. Para Arellano (2018), o objetivo do plano de gestão de dados é considerar os

muitos aspectos do gerenciamento de dados, a geração de metadados, a preservação de dados e a análise antes do início do projeto.

Segundo Arellano (2018), para que a gestão de dados de pesquisa seja realizada de forma adequada, algumas etapas devem ser percorridas, começando pela coleta e armazenamento de longo prazo, passando por processos de limpeza, curadoria, anotação, indexação e transformação. Para Ribeiro e outros autores (2010), a gestão é composta de ações de registro, armazenamento, associação a metadados, organização em coleções, indexação para pesquisa e todas as formas de descrição dos dados.

Por fim, a gestão de dados de pesquisa possibilita a verificação confiável dos resultados de uma investigação. Além disso, permite o desenvolvimento de novas pesquisas a partir de informações já existentes, otimizando a comunicação científica e os recursos financeiros. (SAYÃO; SALES, 2013). Arellano (2018) também destaca a possibilidade de reutilização dos dados e enfatiza a documentação como um dos princípios essenciais para que esse reaproveitamento ocorra.

Sant'Ana (2016) afirma que a gestão de dados demanda planejamento e ações concretas que resultem em eficiência não só nas etapas de coleta e armazenamento, mas principalmente na fase de recuperação desses dados, aumentando sua visibilidade e potencial uso. Portanto, o planejamento da gestão de dados passou a fazer parte integral da metodologia científica, sendo considerado como um dos itens necessários para as boas práticas de pesquisa. (Arellano, 2018).

Assim sendo, o planejamento da gestão dos dados de pesquisa se torna um item essencial do CVDP, na medida em que o pesquisador, desde o início de sua investigação, planeja as etapas a serem seguidas, facilitando a recuperação e o reaproveitamento dos dados de pesquisa.

5.2 APRESENTAÇÃO DOS MODELOS DE CVDP

Nesta seção serão analisados os modelos de CVDP identificados por meio de uma pesquisa no site Google, os quais são utilizados em diferentes áreas do conhecimento. Essa pesquisa foi realizada a partir do termo “Data Life Cycle”, resultando em inúmeros modelos de CVDP. A partir desses resultados brutos foram selecionados apenas os modelos que continham uma definição de CVDP, resultando em sete modelos.

A partir desses dados, identificaram-se as etapas que compõem cada modelo de CVDP, mesmo alguns sendo utilizados apenas para arquivamento de documentos. Cabe salientar que o entendimento sobre os modelos adotados por diferentes repositórios digitais pode oferecer um quadro referencial para elaboração de um novo CVDP. Dessa forma, apresenta-se a seguir os CVDP encontrados e quais suas principais características.

5.2.1 Data Observation Network for Earth (DataONE)

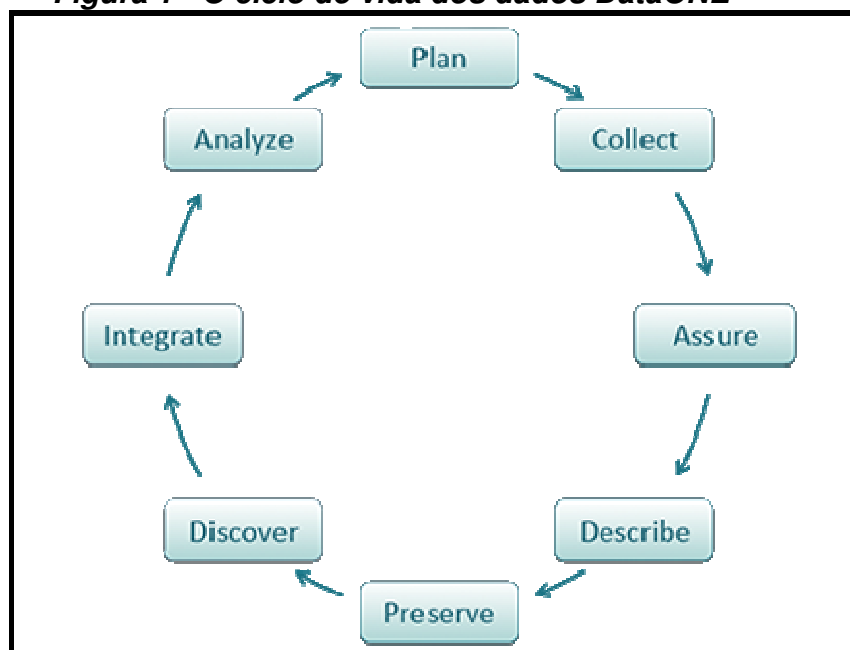
Data Observation Network for Earth (DataONE) é uma colaboração entre muitas organizações parceiras e é financiado pela *National Science Foundation (NSF)* dos EUA sob um Acordo Cooperativo. É uma base de ciência ambiental inovadora, que por meio de uma estrutura distribuída e infra-estrutura cibernética sustentável que atende às necessidades da ciência e da sociedade para acesso aberto, persistente, robusto e seguro a observações observacionais de dados da Terra.

Atualmente, o *DataONE* hospeda três *data centers* que fornecem serviços de suporte, indexação e replicação em toda a rede para aprimorar a interoperabilidade dos centros dos membros. Esses centros fornecem um catálogo que replica as participações de centros de membros, facilitam para os cientistas, descobrirem os dados onde quer que residam, permitindo também que os repositórios de dados disponibilizem seus dados e serviços de forma mais ampla para a comunidade internacional. O *DataONE* tem centros localizados na Universidade do Novo México, na

Universidade da Califórnia em Santa Bárbara e na Universidade do Tennessee (em colaboração com o Oak Ridge National Laboratory).

O modelo DataOne propõe um CVDP em oito etapas que são planejar, coletar, assegurar, descrever, preservar, descobrir, integrar, analisar.

Figura 1 - O ciclo de vida dos dados DataONE



Fonte: DataOne (2018).

Esse modelo se desenrola de forma cíclica, dando início a um novo ciclo, estabelecido nas seguintes etapas e suas definições:

- a) **Planejar:** descrição dos dados que serão compilados e como eles serão gerenciados e disponibilizados para acesso durante o seu tempo de vida.
- b) **Coletar:** as observações são realizadas manualmente ou por sensores ou outros instrumentos, e os dados são colocados em formas digitais.

- c) **Assegurar:** a qualidade dos dados é assegurada por meio de controles e inspeção.
- d) **Descrever:** dados são descritos de forma acurada usando padrões de metadados apropriados.
- e) **Preservar:** dados são submetidos a um arquivo apropriado (centro de dados) para preservação de longo prazo.
- f) **Descobrir:** dados potencialmente úteis são descobertos e acessados juntamente com informações relevantes sobre os dados (metadados).
- g) **Integrar:** dados de diversas fontes são combinados para formar um conjunto de dados homogêneo que pode ser prontamente analisado.
- h) **Analisar:** os dados são analisados.

Segundo o próprio DataONE (2018) não é necessário seguir todas as etapas para a utilização dos dados de pesquisa. Seu modelo permite que o pesquisador utilize apenas algumas etapas do CVDP, podendo criar novos dados em qualquer parte do mesmo.

5.2.2 TechTarget

A TechTarget é uma empresa americana que oferece serviços de marketing orientados por dados para fornecedores de tecnologia business-to-business. Ele usa dados de intenção de compra coletados dos leitores de seus sites focados na tecnologia para ajudar os fornecedores de tecnologia a alcançar compradores que pesquisam ativamente produtos e serviços de TI relevantes. Fundada em 1999 por Greg Strakosch e Don Hawk como um spin-off do United Communications Group (UCG), atualmente é controlada por um conglomerado de empresas. Rouse define o ciclo de vida dos dados da empresa como uma sequência de estágios pelos quais uma determinada unidade de dados passa desde sua geração ou captura inicial até seu arquivamento e, ou, exclusão no final de sua vida útil. Embora possam variar normalmente eles utilizam seis etapas no ciclo de vida dos dados.

- a) **Geração ou captura:** nesta fase, os dados entram em uma organização, geralmente por meio de entrada de dados, aquisição de uma fonte externa ou recepção de sinal, como dados de sensores transmitidos.
- b) **Manutenção:** Nesta fase, os dados são processados antes do seu uso. Os dados podem estar sujeitos a processos como integração, depuração e extração-transformação-carga.
- c) **Uso ativo:** nesta fase, os dados são usados para apoiar os objetivos e operações da organização.
- d) **Publicação:** nesta fase, os dados não são necessariamente disponibilizados para o público em geral, mas são enviados apenas para fora da organização. A publicação pode ou não fazer parte do ciclo de vida de uma determinada unidade de dados.
- e) **Arquivamento:** nesta fase, os dados são removidos de todos os ambientes de produção ativos. Ele não é mais processado, usado ou publicado, mas é armazenado caso seja necessário novamente no futuro.
- f) **Limpeza (*Purging*):** Nesta fase, todas as cópias de dados são excluídas. Normalmente, isso é executado em dados que já estão arquivados.

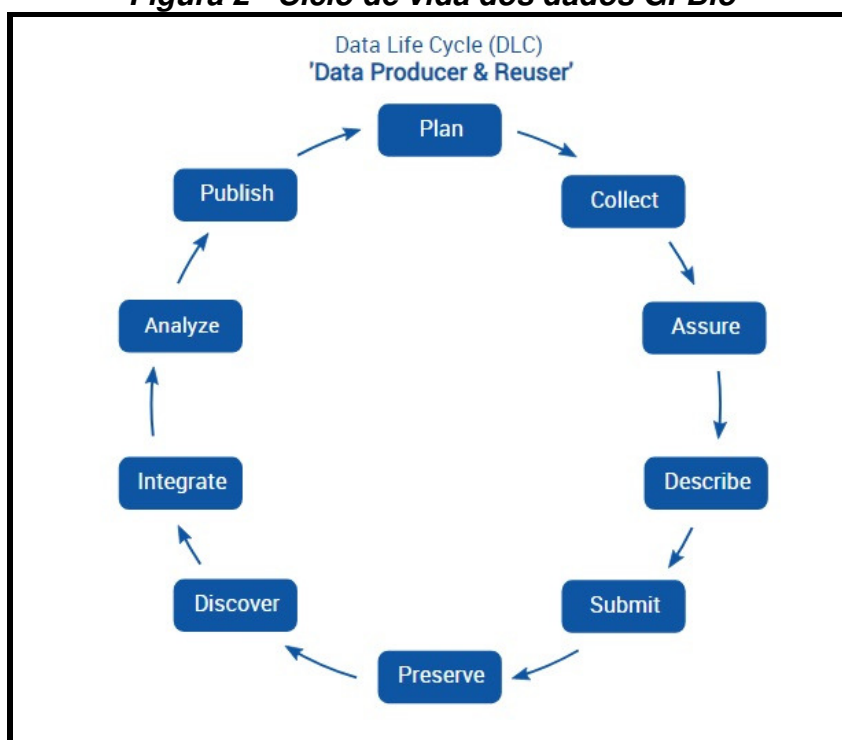
A TechTarget (2018) define o ciclo de vida de dado como uma abordagem para gerenciar os dados de uma organização, envolvendo procedimentos e práticas específicas. Entre seus clientes estão a IBM, Microsoft, Intel e Dell, entre outros.

5.2.3 GFBio

O German Federation for Biological Data (GFBio) é uma infraestrutura de dados nacional sustentável, orientada a serviços, que facilita o compartilhamento de dados para pesquisas biológicas e ambientais. Criada e administrada por um consórcio multidisciplinar é representado por especialistas de diversas instituições, desde coleções de história natural, bibliotecas, bioinformática a arquivos de dados ambientais pelo sendo o contato oficial e autorizado para questões relativas à gestão e padronização de dados de pesquisas biológicas e ambientais durante todo o CVDP (da aquisição ao arquivamento e publicação de dados).

O CVDP da GFBio é uma ferramenta conceitual que ajuda a entender os diferentes passos que os dados seguem, desde a geração de dados até a criação do conhecimento. O compartilhamento e reutilização de dados começam com uma boa gestão de dados em todas as etapas do CVDP. Compreende os dados como matéria prima para publicações acadêmicas buscando serem recompensados com produtos valiosos ao final do processo. Promove um CVDP composto por 10 etapas: planejamento, coleta e organização de dados, garantia e controle de qualidade, criação de metadados, submissão, preservação, descoberta de dados, integração, análise e visualização e publicação de dados.

Figura 2 - Ciclo de vida dos dados GFBio



Fonte: GFBio (2018)

O CVDP do GFBio apresenta as seguintes definições de suas etapas:

- a) **Planejamento:** Na proposta de pesquisa já se deve pensar em gerenciamento, compartilhamento e reutilização dos dados da pesquisa, criando um plano de

gerenciamento de dados para economizar tempo depois. Considerando quais dados serão coletados, os métodos, como se dará o *backup*, a preservação e compartilhamento.

- b) **Coleta e organização de dados:** os dados podem ser coletados manualmente ou automaticamente no laboratório ou campo. Outra forma é por meio de pesquisa bibliográfica em de repositórios de dados de pesquisa. Os dados emergem por meio de observação, entrevistas, contagem, medição com instrumentos ou simulação por algoritmos. Os pesquisadores usam diferentes tecnologias, como gravadores de fita ou vídeo, computadores, instrumentos (portáteis), folhas de dados manuscritas, questionários, ferramentas, bancadas de trabalho, imagens aéreas ou de satélite.
- c) **Garantia e controle de qualidade:** assegurar e controlar a qualidade dos dados pode evitar erros durante a transferência de dados e supostamente para eliminar erros existentes em um conjunto de dados. A garantia é importante para monitorar e manter a qualidade dos dados durante todo o ciclo de vida dos dados. A qualidade dos dados é um dos principais desafios ao promover a reutilização de dados, portanto a qualidade deve ser controlada em todas as etapas do tratamento de dados - quando os dados são coletados, inseridos e analisados. Assim que os dados são "adequados para uso", eles devem ser acessíveis, precisos, completos, consistentes com outras fontes, relevantes, abrangentes, legíveis e interpretáveis.
- d) **Criação de metadados:** uma descrição detalhada dos dados pode garantir sua qualidade e que outros usuários possam facilmente encontrar esses dados. A descrição dos dados gera informações estruturadas, os chamados metadados, que devem responder às seguintes perguntas: Por que os dados foram gerados? QUEM criou os dados? ONDE e QUANDO os dados foram coletados? QUAL é o conteúdo dos dados? Como os dados foram avaliados?
- e) **Submissão:** após assegurar qualidade e a descrição dos metadados, os dados de pesquisa particular devem ser enviados para um domínio de pesquisa compartilhado. O GFBio oferece suporte para um processo de envio fácil e semi automatizado para transferir dados para um data center especializado e

responder as perguntas sobre padronização e formatos de metadados, bem como termos de reutilização e citação. No âmbito da boa prática científica, espera-se que os dados sejam submetidos a um *data center* de longo prazo para serem preservados e disponíveis para descoberta. O GFBio oferece armazenamento a longo prazo de seus dados em seus data centers colaboradores.

- f) **Preservação:** preservar é mais que um backup. É um conjunto de ações que garantem o arquivamento e a integridade dos dados por um longo período de tempo, mantendo a acessibilidade (os dados podem ser recuperados, exibidos e usados), a autenticidade (os dados não foram manipulados, substituídos ou falsificados) e a longevidade (os dados são reutilizáveis para longo prazo, independentemente das mudanças nos softwares e hardwares). Objetos digitais são frágeis, sendo suscetíveis ao “apodrecimento de dados”, o que pode influenciar sua acessibilidade e autenticidade. A preservação visa garantir os dados estejam na melhor forma para serem armazenados, descobertos, acessados e reutilizados. Existem diferentes métodos de preservação digital, como migração, emulação, arqueologia digital e preservação de tecnologia.
- g) **Descoberta de dados:** nesse estágio de descoberta deve-se procurar por dados disponíveis que possam ser valiosos e úteis para um novo estudo ou projeto de pesquisa. Para descobrir dados e seus metadados, eles precisam estar visíveis e acessíveis. Isso significa que os dados foram enviados para uma infraestrutura de dados (arquivo, data center, repositório de dados).
- h) **Integração:** a integração é a mesclarem de vários conjuntos de dados de diferentes origens, como os dados coletados recentemente com antigos dados de outros proprietários, resultando em um novo conjunto de dados maior. Quando os dados de outros autores são reutilizados, é fundamental fornecer crédito aos criadores de dados por meio de uma prática robusta de citação de dados, que funciona melhor quando os dados são equipados com um identificador persistente.
- i) **Análise e visualização:** fazendo o upload dos seus dados da pesquisa no GFBio será disponibilizado as ferramentas de análise deles. Em seguida, eles

podem ser explorados estatisticamente por meio de visualização, sobreposição, transformação, análise estatística ou modelagem. Esse conjunto de ações deve fornecer uma nova visão sobre questões de pesquisa, levando a conclusões que podem ajudar os tomadores de decisão a agir. O software e o hardware usados para análise de dados, bem como a análise de dados, variam entre disciplinas. Apesar das enormes diferenças nas características, quantidade e formato, os dados de pesquisa de vários campos científicos podem ser categorizados em três tipos básicos de dados: Dados brutos: dados iniciais não processados. Exemplos são dados medidos sem correções, fotos ou amostras. Dados primários: dados processados. Eles podem resultar da conversão ou correção de medições. Outros exemplos são compilações de dados ou anotações específicas. Dados secundários: os dados são dados secundários quando são reutilizados. Esses dados foram originalmente coletados para um propósito diferente e usados em um novo contexto.

- j) **Publicação de dados:** a prática mais amplamente aplicada, para publicar dados, é acompanhada de uma publicação de pesquisa. O número de periódicos que colaboram com os *data centers* está aumentando (por exemplo, Nature, The American Naturalist, Ecologia Molecular, PNAS, PLOS), e a publicação de dados associados está se tornando um requisito comum. Isso porque os dados são reconhecidos como parte fundamental do processo de pesquisa e tão importantes quanto as discussões e conclusões deles derivadas. Os dados tornam-se publicamente disponíveis, citáveis e identificáveis de forma única por meio de um identificador persistente (PID). Além disso, a publicação de conjuntos de dados promove a transparência no ciclo de vida da pesquisa, facilita a verificação e a reprodutibilidade dos resultados e, muito provavelmente, aumenta sua taxa de citações. Portanto, é importante fomentar uma prática, como os geneticistas já estabelecidos com o "Genbank" (desde 1982), que permite o compartilhamento de dados. Outra possibilidade de publicação de dados é escrever um "documento de dados" descrevendo seus dados e seu uso potencial em detalhes, e publicá-los em um diário de dados designado (exemplos de periódicos de dados: Biodiversity Data Journal, Genomics Data, Earth System

Science Data). Se você pretende publicar dados que não estejam conectados a uma publicação de pesquisa ou de um diário de dados, poderá publicá-los diretamente por meio de um data center.

No entanto o GFBio (2018) se concentrou em dois grupos principais, produtor e reutilizador de dados. Fazendo com que CVDP seja implementado de formas diferentes, dependendo da relação entre o pesquisador, o projeto de pesquisa e os dados.

5.2.4 Spirion

A Spirion, com sede em São Petersburgo, Flórida, é a principal fornecedora de software de gerenciamento de dados corporativos, auxiliando as empresas a reduzir o volume de dados confidenciais e minimizar proativamente os riscos, custos e danos à reputação no caso de ataques cibernéticos. Ajuda as organizações a evitar violações dispendiosas de dados, descobrindo, classificando, monitorando e protegendo informações pessoais, registros médicos, números de cartões de crédito e propriedade intelectual armazenados em toda a empresa, dentro de e-mail e na nuvem.

É especializada na pesquisa de alta precisão e na classificação automatizada de dados não estruturados, usando a precisão inigualável de seu mecanismo AnyFind ao analisar textos e imagens geradas por pessoas. A Spirion tem milhares de clientes entre as principais empresas dos setores de saúde, setor público, varejo, educação, serviços financeiros, energia.

A Spirion utiliza a Data Lifecycle Management (DLM) como um processo que ajuda as organizações a gerenciar o fluxo de dados em todo o seu ciclo de vida, da criação, uso, compartilhamento, arquivamento e exclusão. Rastrea dados com precisão durante todo o ciclo de vida das informações como base de uma estratégia de proteção de dados confidenciais e para ajudar a determinar onde aplicar os controles de segurança.

Figura 2 - O ciclo de vida dos dados Spiron



Fonte: Spiron (2018)

Apresenta um CVDP com seis etapas mostradas abaixo:

- a) **Etapa 1 - Criação de Dados:** durante a primeira fase, são criados dados estruturados ou não estruturados. Pode ser na forma de um documento do MS Office 365, um arquivo PDF, e-mail, informações em um banco de dados ou até mesmo imagens. Frequentemente, nessa fase, a classificação de segurança ocorre com base em uma política de segurança de dados corporativos.
- b) **Etapa 2 - Gerenciamento de armazenamento:** depois que um arquivo é criado, ele é armazenado em algum lugar. Neste ponto, deve-se garantir que os dados armazenados estejam protegidos e que os controles de segurança de dados necessários sejam implementados.
- c) **Etapa 3 - Segurança baseada em função e uso de dados:** depois de criado o arquivo, ele é armazenado e usado. Durante este estágio, os dados críticos são visualizados, processados, modificados e salvos. Controles de segurança são

aplicados aos dados no ponto de uso. Aqui é preciso monitorar a atividade do usuário e aplicar controles de segurança baseados em função para garantir a prevenção de vazamento de dados.

- d) **Etapa 4 - Compartilhamento de dados:** os dados são constantemente compartilhados entre funcionários, clientes e parceiros, exigindo uma estratégia que monitore continuamente os armazenamentos de dados para novas instâncias de dados confidenciais. Os dados se movem entre uma variedade de locais de armazenamento públicos e privados, aplicativos e ambientes operacionais e são acessados por vários proprietários de dados de diferentes dispositivos e plataformas de armazenamento. Isso pode acontecer em qualquer estágio do ciclo de vida da segurança de dados, e é por isso que é importante aplicar os controles de segurança certos no momento certo.

- e) **Etapa 5 - Soluções de arquivamento:** em algum momento, os dados podem deixar o uso ativo e serem arquivados. Depois de arquivado, é importante que os dados estejam disponíveis e protegidos. Cada setor tem estipulações específicas de retenção de dados para permanecer em conformidade. Após período de retenção deve-se decidir o que fazer com esses dados.

- f) **Etapa 6 - Destruir permanentemente os dados:** o volume de dados arquivados inevitavelmente aumenta e fica inviável armazenar todos os dados para sempre. O custo de armazenamento e os problemas de conformidade exercem pressão para destruir dados que não são mais utilizados.

A Spirion é utilizada por mais de 10.000 de organizações que confiaram seus dados sigilosos a ela.

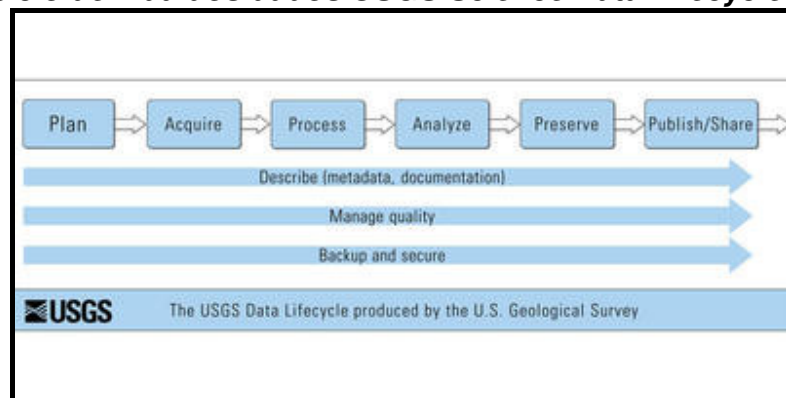
5.2.5 United States Geological Survey (USGS)

A United States Geological Survey (USGS) foi criada por um ato do Congresso em 1879, são agência de mapeamento civil, ciências biológicas e geológicas da terra, a USGS coleta, monitora, analisa e fornece ciência sobre condições, problemas e problemas dos recursos naturais. Realiza investigações multidisciplinares em grande escala e fornecer informações científicas imparciais aos gerentes de recursos, planejadores e outros clientes. Foi criada e é mantida do pelo governo dos Estados Unidos da América, através do Departamento do Interior.

A USGS (2018) definiu o CVDP como dados ativos corporativos com valor além da necessidade imediata de uso e que devem ser gerenciados durante todo o ciclo de vida do dado. Questão de documentação, armazenamento, garantia de qualidade e propriedades precisam ser respondidas em cada etapa do ciclo de vida.

Criaram um modelo chamado de USGS Science Data Lifecycle Model (SDLM) que ilustra os estágios do gerenciamento de dados e descreve como os dados fluem através de um projeto de pesquisa do início ao fim. Seu ciclo consiste em seis etapas, sendo que inclui mais três elementos de corte transversal.

Figura 3 - O ciclo de vida dos dados USGS Science Data Lifecycle Model (SDLM)



Fonte: USGS (2018)

As etapas são apresentadas em seguida:

- a) **Plano:** criar um plano de gerenciamento de dados e aprender sobre atividades de planejamento importantes. Antes de iniciar um projeto, é importante planejar como os dados serão gerenciados durante todo o ciclo de vida.

- b) **Adquirir:** adquirir dados para um projeto envolve coletar ou gerar novos dados ou obter dados existentes.
- c) **Processo:** o processamento de dados envolve várias atividades associadas à preparação de entradas de dados novas ou coletadas anteriormente, incluindo: validação de dados, resumo de dados, transformação de dados, integrando dados, subconjunto de dados, derivando dados.
- d) **Analisar:** a análise de dados envolve várias atividades associadas à exploração e interpretação de dados processados. As atividades de análise incluem: análise estatística, visualização, análise espacial, análise de imagem, modelagem, interpretação.
- e) **Preservar:** preservação envolve ações e procedimentos usados para garantir a viabilidade em longo prazo e a acessibilidade dos dados. Esta etapa discute os seguintes tópicos: arquivamento de dados, disposição de dados, repositórios para dados.
- f) **Publicar / compartilhar:** a publicação e o compartilhamento de dados é uma etapa importante e necessária no processo de pesquisa, assim como a publicação de artigos de periódicos revisados por especialistas.

Os elementos transversais descrevem atividades que devem ser executadas continuamente em todos os estágios do ciclo de vida para ajudar no suporte ao gerenciamento de dados efetivo. São eles:

Descrever (Metadados, Documentação): durante todo o ciclo de vida dos dados, a documentação deve ser criada e atualizada para refletir as ações tomadas nos dados. Esta seção envolve os seguintes tópicos: criação de metadados, a importância dos dicionários de dados e usando palavras-chave para descoberta de dados.

Gerenciar qualidade: o gerenciamento de qualidade de dados é um processo em que protocolos e métodos são empregados para garantir que os dados sejam coletados, manipulados, processados, usados e mantidos adequadamente em todos os estágios do ciclo de vida dos dados científicos. Abrange os seguintes tópicos: Planos de garantia de qualidade, garantia da qualidade, controle da qualidade e documentando a qualidade dos dados.

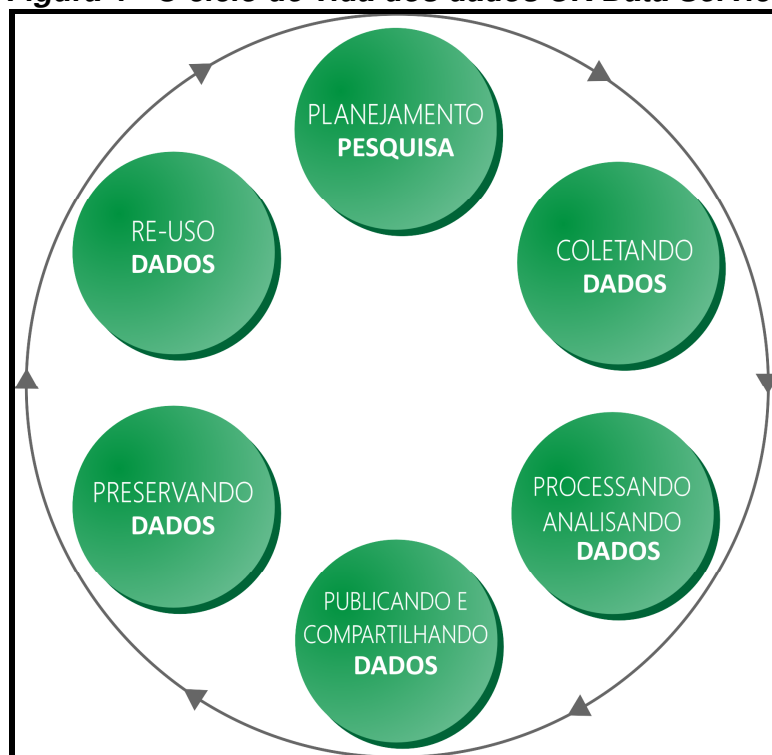
Backup e proteção: fazer o backup e proteger os dados envolve proteger os dados contra perda acidental de dados, corrupção e acesso não autorizado.

5.2.6 UK Data Service

O UK Data Service foi criado em 1967, na Universidade de Essex, com o apoio do então Social Science Research Council, com o objetivo de curar dados de pesquisa de alta qualidade para análise e reutilização. Em 2005 foi designado como Local de Depósito pelo National Archives, permitindo a organização de registros públicos.

O UK Data Service oferece acesso a dados sociais e econômicos regionais, nacionais e internacionais, apoio a pesquisa e orientação relevantes para políticas e treinamento para o desenvolvimento de habilidades no uso de dados. Tem o UK Data Archive como principal serviço, que fornece acesso unificado à maior coleção de dados sociais, econômicos e populacionais do Reino Unido. Financiado pelo Economic and Social Research Council (ESRC), O UK Data Archive é um centro de conhecimento reconhecido internacionalmente para adquirir, curar e fornecer acesso a dados de ciências sociais e humanidades.

Figura 4 - O ciclo de vida dos dados UK Data Service



Fonte: UK Data Service (2018)

Seu CVDP é descrito em seis etapas:

- a) **Planejamento** – inclui o desenho da pesquisa, o planejamento e gerenciamento dos dados, a obtenção do consentimento para compartilhamento (dados de humanos ou copyrights), planejamento da coleta de dados, dos protocolos e modelos de processamento e explora as fontes de dados existentes.
- b) **Coletando dados** - coleta os dados, captura dados com metadados e reaproveita dados existentes de terceiros.
- c) **Processando e analisando dados** – digitalizar, transcrever, insere e traduz dado. Verifica e valida dados. Deriva, descreve e documentar

dados, gerenciando e armazenado. Analisa e interpreta dados e produz resultados de pesquisa e cita as fontes de dados.

- d) **Publicando e compartilhando dados** - Estabelecer direitos autorais e de uso. Criar documentação do usuário. Criar metadados de descoberta. Selecionar o acesso apropriado aos dados. Publicar, compartilhar dados e promover dados.
- e) **Preservando dados** - Migrar dados para o melhor formato e mídia. Armazenar e fazer backup de dados. Criar documentação de preservação. Preservar e curar dados.
- f) **Reutilizando dados** - Realizar análise secundária. Realizar pesquisas de acompanhamento. Realizar análises de pesquisa. Examinar as descobertas. Usar dados para ensino e pesquisa

Os dados geralmente têm uma vida útil mais longa do que o projeto de pesquisa que os cria, que sempre tem um início e um fim, enquanto os dados podem continuar a serem utilizados, ou adicionados novos dados em projetos de acompanhamento, mesmo após o término do financiamento, e ainda esses dados podem ser reutilizados por outros pesquisadores.

Dados bem organizados, bem documentados, preservados e compartilhados são inestimáveis para avançar na investigação científica e aumentar as oportunidades de aprendizado e inovação. (UK Data Service)

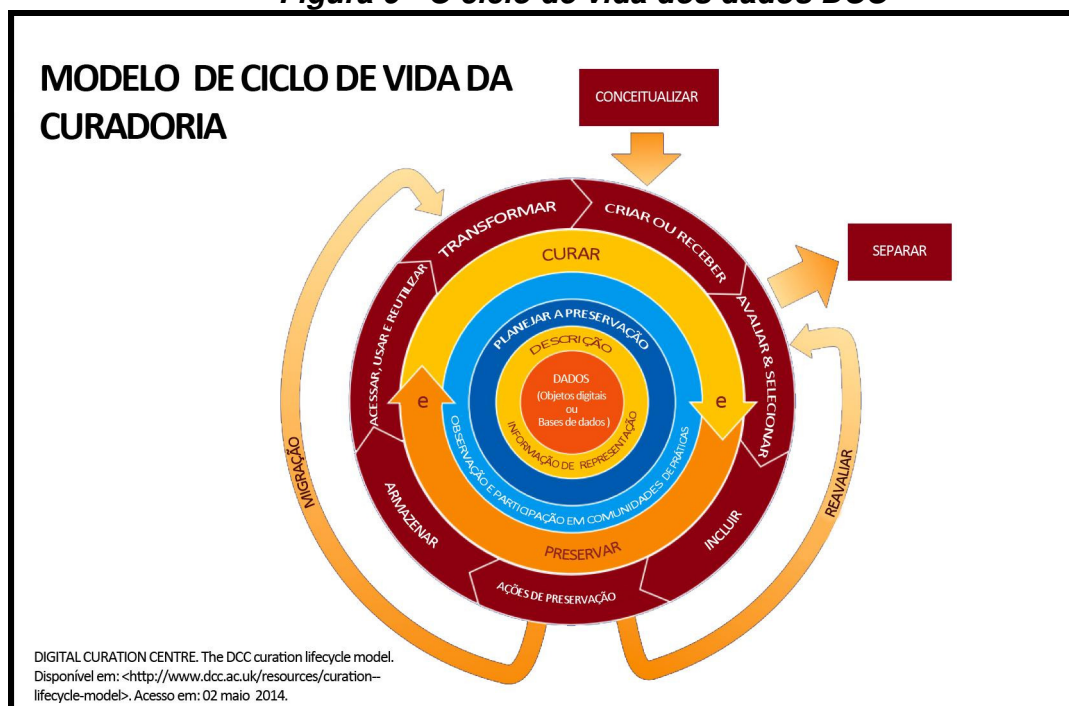
5.2.7 Digital Curation Centre (DCC)

O Digital Curation Centre (DCC) foi lançado em 1 de março de 2004, após uma resposta bem-sucedida à Circular JISC 6/038 por um consórcio formado pelas Universidades de Edimburgo e Glasgow (que juntos abrigaram o Centro Nacional de e-Science), UKOLN na Universidade de Bath, e STFC, que gerenciou os Laboratórios Rutherford Appleton e Daresbury.

Desenvolvido para fornecer aconselhamento especializado e ajuda prática a qualquer pessoa no ensino superior e pesquisa no Reino Unido que deseja armazenar, gerenciar, proteger e compartilhar dados de pesquisa digital.

Seu modelo de ciclo de vida do dado de pesquisa está entre os modelos mais usados por bibliotecários, pesquisadores e gestores de informação.

Figura 6 - O ciclo de vida dos dados DCC



Fonte: DCC (2018)

Utiliza sete etapas que são descritas abaixo:

- Conceitualizar:** conceber e planejar a criação de objetos digitais, incluindo métodos de captura de dados e opções de armazenamento.
- Criar ou receber dados:** além do conteúdo, poderão ser incluídos metadados administrativos, descritivos, estruturais e técnicos. Metadados de preservação também poderão ser incluídos neste momento.
- Avaliar e selecionar dados:** avaliar e selecionar dados para curadoria e preservação. Devem estar em conformidade com a documentação, políticas e requisitos legais estabelecidos.

- d) **Incluir dados:** transferir dados para um arquivo, repositório, data center ou outro tipo de custódia de conteúdo. Deve estar em conformidade com a documentação, políticas e requisitos legais estabelecidos.
- e) **Ação de preservação:** limpeza dos dados, validação, atribuição de metadados de preservação, atribuição de informações de representação. Deve-se garantir a estrutura dos dados e formatos de arquivos;
- f) **Armazenar:** armazenar os dados de forma segura garantindo padrões relevantes de armazenamento, segurança e preservação.
- g) **Acessar, usar e reutilizar:** garantir que os dados estarão acessíveis para uso e reuso ao longo do tempo. Controles robustos de acesso e autenticação podem ser utilizados.
- h) **Transformar:** criar novos dados baseados em originais (migrando para formato diferente ou criando outro estudo baseado em dados de pesquisa já existentes).

5.3 COMPARATIVO ENTRE OS MODELOS DE CVDP

Com a identificação dos conceitos e dos modelos de CVDP, busca-se realizar uma comparação entre os modelos, a qual é apresentada no Quadro 2.

Quadro 2 - Comparativo entre os modelos de CVDP

	DataONE	TechTarget	GFBio	Spirion	USGS	UK Data Service	DCC
Área do conhecimento	Ciência Ambiental	Ciências da Computação	Ciências Biológicas	Ciências da Computação	Ciências Biológicas/Geologia	Ciências Sociais e Econômicas	Não especificado
Utilização livre	Sim	Não	Sim	Não	Sim	Sim	Sim
Criador/Gestor	DataONE	TechTarget	GFBio	Spirion	Governo dos Estados Unidos da América	Universidade de Essex / Social Science Research Council	Universidades de Edimburgo e Glasgow
Financiamento público	Sim	Não	Sim	Não	Sim	Sim	Sim
Aplicação	Dados de Pesquisa	Documentos empresariais	Dados de Pesquisa	Documentos empresariais	Dados de Pesquisa	Dados de Pesquisa	Documentos
Permite customização	Sim	Não especificado	Não especificado	Não especificado	Não especificado	Sim	Sim
Curva de aprendizado	Razoável	Longo	Rápido	Razoável	Razoável	Rápido	Complexa e Demorada

Fonte: Dados da pesquisa (2018).

Para uma melhor visualização dos modelos de CVDP e suas características, organizou-se o quadro acima considerando os seguintes pontos: a área do conhecimento, se a utilização é livre, criador/gestor do CVDP, se utiliza financiamento público, sua aplicação, se permite customização e de que maneira se dá sua curva de aprendizado. Cabe salientar que o termo curva de aprendizado pode ser definido como o tempo que se leva para aprender e entender o funcionamento de uma nova atividade ou ferramenta.

A primeira característica apresentada no quadro refere-se à área do conhecimento. Observa-se que todos os modelos de CVDP foram desenvolvidos para atender a uma área de conhecimento específica, a exceção do modelo DCC.

Em relação à utilização do modelo, TechTarget e Spirion caracterizam-se por uma utilização restrita, ou seja, seus modelos de CVDP destinam-se apenas a clientes dessas empresas. DataONE, GFBio, USGS, UK Data Service e DCC, por sua vez, configuram-se como modelos de livre utilização.

Cabe destacar ainda que os modelos TechTarget e Spirion não se utilizam de financiamento público, por serem empresas privadas. DataONE, GFBio, USGS, UK Data Service e DCC, ao contrário, se valem desse tipo de financiamento.

No que se refere à aplicação dos modelos de CVDP, TechTarget e Spirion são utilizados para documentos empresariais, enquanto DCC apenas preserva e faz a curadoria de documentos. DataOne, GFBio, USGS e UK Data Service ocupam-se verdadeiramente de dados de pesquisa.

TechTarget, GFBio, Spirion e USGS não especificam se há a possibilidade de customização de seus modelos, enquanto DataONE, UK Data Service e DCC permitem customização em seus modelos de CVDP.

Por fim, observa-se que os modelos DataONE, GFBio, Spirion, USGS e UK Data Service apresentam modelos simples, de fácil compreensão e que requerem um tempo menor para entendimento e utilização. Por sua vez, TechTarget não cria um modelo visual de seu ciclo, fato que o torna mais complexo e longo, sendo o DCC o modelo que apresenta um grau maior de complexidade em seu entendimento e utilização.

O modelo da DCC está ligado à curadoria e preservação de documentos, sendo que sua curva de aprendizado é complexa e demorada e, mesmo oferecendo alguns

materiais para o entendimento de seu complexo CVDP, sua utilização ainda apresenta certa dificuldade.

Já os modelos GFBio e United States Geological Survey (USGS) são repositórios especializados da área das Ciências Naturais (Biologia e Geologia, respectivamente). Acredita-se que essa especificidade pode dificultar o processo de reaproveitamento de seu CVDP. Suas curvas de aprendizagem são de razoável compreensão, mas adaptados do modelo DataOne.

5.4 DESENVOLVIMENTO DE UMA CVDP PARA O CEDAP

O objetivo da pesquisa estava na proposta de desenvolvimento de um CVDP para o CEDAP, entretanto após análise das definições dos conceitos de dados de pesquisa, ciclo de vida de dados de pesquisa e gestão de dados de pesquisa, bem como os diversos modelos de CVDP encontrado na literatura, optou-se em adaptar um ciclo já existente.

Pelo CEDAP trabalhar com uma diversidade de dados de pesquisa, cauda longa, estabeleceu-se algumas ponderações sobre o melhor CVDP para uso. Considerando este perfil heterogêneo, restaram os modelos DataONE e UK Data Service, os quais são tomados como referências para pesquisadores, arquivistas e bibliotecários, segundo alguns autores, juntamente com o modelo da DCC, que já foi anteriormente descartado.

O DataOne e UK Data Service têm seu CVDP construídos de forma coerente, conforme os referenciais teóricos estudados, preocupam-se com o planejamento, coleta, gerenciamento, análise dos dados, preservação e reutilização. No entanto o modelo da UK Data Service se mostra mais completo para as pretensões do CEDAP, por se adaptar às suas necessidades.

Em relação à área do conhecimento, o modelo do DataOne não foi escolhido por ser desenvolvido para atender às necessidades específicas da Ciência Ambiental. Já o modelo UK Data Service atende às Ciências Sociais e Econômicas, o que se adéqua ao plano de implantação do CEDAP.

O modelo que mais se adapta ao CEDAP, pelas afinidades com os dados, é o modelo da UK Data Service, por ser um repositório de cauda longa e possuir um modelo que se adapta a vários tipos de dados, incluindo uma etapa de “reuso”, fator que será essencial na implantação do CVDP do CEDAP. Este modelo permite que certas etapas possam ser omitidas durante seu ciclo, mostrando que não é necessário seguir todas as etapas para a utilização dos dados de pesquisa. Isso permitira que certas atividades da pesquisa possam usar apenas parte do CVDP. Com isso, projetos de pesquisa, podem não seguir um caminho linear, podendo sofrer alterações durante o ciclo. Portanto, alguns cientistas ou equipes podem criar novos dados no processo de descobrir, integrar, analisar e sintetizar os dados existentes.

Além disso, dispõe de ferramentas simples de customização, o que permitirá ao CEDAP escolher as melhores formas de preservação e compartilhamento dos dados de pesquisa, bem como ajudará na evolução do processo de reaproveitamento deles. Por último, o modelo é de fácil entendimento, necessitando de um menor tempo despendido pra ser aprendido e aplicado.

6 CONSIDERAÇÕES FINAIS

Neste trabalho abordou-se a importância da preservação dos dados de pesquisa, considerando que há um volume cada vez maior de dados produzidos e gerados a partir da atividade científica em todo o mundo. Dessa forma, sua preservação e reuso estão tomando uma importância maior na sociedade e, considerando-se que 80% da pesquisa centra-se na captação de dados brutos para análise, o reaproveitamento dos mesmos pode encurtar e facilitar novas investigações.

Nesse contexto, mostra-se evidente a importância do desenvolvimento de modelos de Ciclo de Vida do Dado de Pesquisa (CVDP), os quais devem contar com etapas bem definidas e bem documentadas, para que a gestão dos dados de pesquisa seja realizada mais facilmente pelos pesquisadores, gestores de informação, bibliotecários, arquivistas e afins.

Considerando essas informações, buscou-se responder ao seguinte problema de pesquisa: De que maneira os repositórios digitais mais destacados podem contribuir com ferramentas conceituais e informações metodológicas para a criação do CVDP do CEDAP?

Assim, realizou-se a busca dos conceitos de Dados de Pesquisa, Ciclo de Vida dos Dados de Pesquisa (CVDP) e Gestão de Dados de Pesquisa na literatura, em bases de dados na Brapci e Scielo. Essa busca resultou em um corpus textual de sete artigos com relevância para o tema, nos quais foi possível constatar que, por ser um campo relativamente recente, há ainda um número pequeno de publicações e alguns conceitos apresentaram, entre os autores, divergência em suas definições. Entre esses conceitos, cabe destacar as diferenças encontradas na definição do termo Ciclo de Vida do Dado de Pesquisa, na qual não houve consenso entre os autores, demonstrando que ele é composto por um conjunto de etapas.

Posteriormente, realizou-se uma pesquisa de modelos de CVDP no Google. Essa pesquisa foi realizada a partir do termo "Data Life Cycle", resultando em inúmeros modelos de CVDP. A partir desses resultados brutos, foram selecionados apenas os modelos que continham uma definição de CVDP, resultando em sete modelos, quais sejam: DataONE, TechTarget, GFBio, Spirion, USGS, UK Data Service e DCC.

Constatou-se que os modelos TechTarget e Spirion foram concebidos exclusivamente com a finalidade de preservar e arquivar documentos empresariais. Já o modelo DCC ocupa-se da curadoria de vários tipos de documentos, sendo o modelo mais utilizado pelos arquivistas. Os modelos GFBio, USGS, por sua vez, são repositórios de dados de pesquisa que possuem modelos de CVDP com etapas bem definidas e de fácil entendimento, porém dedicam-se exclusivamente à organização e preservação de dados de áreas específicas (Ciências Biológicas e Geológicas).

O modelo DataONE é um dos mais reconhecidos e utilizados por pesquisadores, arquivistas e bibliotecários, segundo alguns autores. No entanto, foi desenvolvido e ocupa-se de dados das Ciências Ambientais. Por fim, observou-se que o modelo UK Data Service mostrou-se mais completo por adaptar-se a vários tipos de dados, incluindo uma etapa de “reuso”. Além disso, dispõe de ferramentas simples de customização e, por último, é de fácil entendimento, necessitando de um menor tempo despendido pra ser aprendido e aplicado. A partir dessa análise, conclui-se que o UK Data Service é o modelo que melhor se adapta às necessidades do CEDAP-UFRGS.

Entre as limitações desta investigação, pode-se destacar a escassa literatura relacionada ao tema, a falta de ferramentas conceituais bem definidas pelos autores e a complexidade de utilização e aprendizado de alguns modelos de CVDP.

Para futuras pesquisas, deve-se desenvolver formas de incentivar o compartilhamento dos dados de pesquisa entre os pesquisadores, buscando a capacitação dos bibliotecários para disseminar a prática de gestão de dados de pesquisa. Uma das formas de incentivar esse compartilhamento seria a obrigatoriedade de um plano de gestão de dados para pesquisadores nas instituições, como condição para o financiamento da pesquisa.

Conclui-se igualmente que, pesquisas realizadas em torno de objetos do universo dos documentos digitais, oportunizam a exploração de um campo que, por sua natureza multidisciplinar, amplia a visão do bibliotecário, principalmente sobre as diversas oportunidades de atuação que vem se abrindo nas instituições. Grande parte do conhecimento em relação ao trabalho bibliotecário com repositórios digitais pode emergir das pesquisas em trabalho de conclusão de curso tanto quanto na formação de disciplinas e súmulas direcionadas a formação teórica e metodológica. O presente

trabalho, mesmo que como objetivo consequente, contribui também nesse sentido, de dar visibilidade ao assunto no contexto da formação básica do bibliotecário.

Por fim, cabe destacar a importância dos Bibliotecários em trabalhar mais de perto com os pesquisadores em cada etapa da pesquisa para desenvolver a cultura do compartilhamento de dados.

REFERÊNCIAS

ARELLANO, Miguel Angel Márdero. **Preservação de dados de pesquisa**. São Paulo: Unifesp, 2018. Disponível em: <<https://unifesp.br/reitoria/bibliotecas/images/bibliotecas/MiguelArellano.pdf>>. Acesso em: 15 abr. 2018.

BERTIN, Patrícia Rocha Bello; VISOLI, Marcos Cezar; DRUCKER, Debora Pignatari. A gestão de dados de pesquisa no contexto da e-Science: benefícios, desafios e oportunidades para organizações de P&D. **Ponto de Acesso**, v. 11, n. 2, p. 34-48, ago. 2017.

CAVALCANTI, Marcia Teixeira; SALES, Luana Farias. Gestão de dados de pesquisa: um panorama da atuação da união europeia. **Biblos**, v. 31, n. 1, 2017.

CAVALCANTI, Marcia Teixeira; SALES, Luana Farias; SAYÃO, Luis Fernando. Critérios para avaliação e seleção de dados de pesquisa no contexto da curadoria digital. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17, 2016, Salvador. **Anais...** Salvador: Universidade Federal da Bahia, 2016. Disponível em: <<http://www.ufpb.br/evento/index.php/enancib2016/enancib2016/paper/view/4064>>. Acesso em: 17 abr. 2018.

CHOUDHURY, Sayeed. *Data curation: an ecological perspective*. *College and Research Library News*, Chicago, v. 71, p. 194-196, Apr. 2010. Disponível em: <<https://crln.acrl.org/index.php/crlnews/article/view/8354/8496>>. Acesso em: 11 out. 2018.

CORRÊA, Fabiano Couto. **Gestión de datos de investigación**. Barcelona: UOC, 2016.

COSTA, Maria Murrieta; CUNHA, Murilo Bastos da. O bibliotecário no tratamento de dados oriundos da e-Science: considerações iniciais. **Perspectivas em Ciência da Informação**, v. 19, n. 3, p. 189-206, jul./set., 2014

DATA Life Cycle. *Data Observation Network for Earth*. Disponível em: <<https://www.dataone.org/data-life-cycle>>. Acesso em: 17 out. 2018.

DATA Life Cycle. *German Federation for Biological Data (GFBio)*., Disponível em: <<https://www.gfbio.org/pt/training/materials/data-lifecycle>>. Acesso em: 17 out. 2018.

DIGITAL Curation Centre. *DCC Curation Lifecycle Model*. Disponível em: <<http://www.dcc.ac.uk/resources/curation-lifecycle-model>>. Acesso em: 17 out. 2018.

DEMO, Pedro. **Metodologia do conhecimento científico**. São Paulo: Atlas, 2000.

DEMO, Pedro. **Pesquisa**: princípio científico e educativo. 14. ed. São Paulo: Cortez, 2011.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010.

GRAY, James Jim. On eScience: A transformed scientific method. In: HEY, Toni; TANSLEY, Stewart.; TOLLE, Kristin (Eds.). **The fourth paradigm**: Data-intensive scientific discovery. Redmond: Microsoft Research, 2009. Disponível em: <<https://www.immagic.com/eLibrary/ARCHIVES/EBOOKS/M091000H.pdf>>. Acesso em: 23 nov. 2018.

MEDEIROS, Claudia Bauzer. Gestão de dados científicos: da coleta à preservação. *SciELO em Perspectiva*, 2018. Disponível em: <<https://blog.scielo.org/blog/2018/06/22/gestao-de-dados-cientificos-da-coleta-a-preservacao/>>. Acesso em: 17 set. 2018.

MONTEIRO, Elizabete Cristina de Souza de Aguiar; SANT´ANA, Ricardo César Gonçalves. Plano de gerenciamento de dados em repositórios de dados de universidades. **Encontros Bibli**, v. 23, n. 53, p. 160-173, set./dez., 2018.

NATIONAL Science Board. Long-Lived Digital Data Collections: enabling Research and Education in the 21st Century. Disponível em: <<https://www.nsf.gov/geo/geo-data-policies/nsb-0540-1.pdf>>. Acesso em: 17 Set. 2018.

NATIONAL Science Foundation. Alexandria. 2010. Disponível em: <<https://nsf.gov/>>. Acesso em: 17 set. 2018.

ORGANIZATION for Economic Co-operation and Development. OECD principles and guidelines for access to research data from public funding. Disponível em: <<https://www.oecd.org/sti/sci-tech/38500813.pdf>>. Acesso em: 14 set. 2018.

RIBEIRO, Cristina; et al. Os repositórios de dados científicos: estado da arte. Lisboa: Repositório Científico de Acesso Aberto de Portugal. <https://www.researchgate.net/publication/235754139_Os_repositorios_de_dados_cientificos_estado_da_arte>. Acesso em: 10 set. 2018.

ROUSE, Margaret. *Data life cycle*. Newton, Massachusetts. Disponível em: <<https://whatis.techtarget.com/definition/data-life-cycle>>. Acesso em: 14 out. 2018.

SALES, Luana Farias; SAYÃO, Luís Fernando. O impacto da curadoria digital dos dados de pesquisa na comunicação científica. **Encontros Bibli**, v. 17, n. esp.2, p. 118-135, 2012. Disponível em: <<https://www.brapci.inf.br/v/a/12634>>. Acesso em: 10 Set. 2018.

SANCHEZ, Fernanda Alves; VIDOTTI, Silvana Aparecida Borsetti Gregório; VECHIATO, Fernando Luiz. A contribuição da curadoria digital em repositórios digitais. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 13, n. 1, 2018.

SANT´ANA, Ricardo César Gonçalves. Ciclo de vida dos dados e o papel da ciência da informação. In: Encontro Nacional de Pesquisa em Ciência da Informação, 14, 2013, Florianópolis. **Anais...** Florianópolis: Universidade Federal de Santa Catarina, 2013. Disponível em: <<http://www.brapci.inf.br/index.php/article/view/0000013772/fac7d2fecfd2c83f1e9208204b2fefdf10>>. Acesso em: 04 abr. 2018.

SANT´ANA, Ricardo César Gonçalves. Ciclo de vida dos dados: uma perspectiva a partir da Ciência da informação. **Informação & Informação**, v. 21, n. 2, p. 116-142, maio/ago. 2016.

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

SAYÃO, Luis Fernando; SALES, Luana Farias. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 6, n. 1, 2013.

SAYÃO, Luis Fernando; SALES, Luana Farias. Guia de Gestão de Dados de Pesquisa: para bibliotecários e pesquisadores. Disponível em: <http://carpedien.ien.gov.br/bitstream/ien/1624/1/GUIA_DE_DADOS_DE_PESQUISA.pdf> Acesso em: 04 abr. 2018.

SAYÃO, Luis Fernando; SALES, Luana Farias. Há futuro para as bibliotecas de pesquisa no ambiente de *e-Science*? **Informação & Tecnologia**, v. 2, n. 1, p. 30-52, jan./jul. 2015.

SESARTIC, Ana; TÖWE, Matthias. Research data services at ETH-Bibliothek. **Journal IFLA, Munich**, v. 42, n. 4, p. 284-291, Nov. 2016. Disponível em: <<http://journals.sagepub.com/doi/abs/10.1177/0340035216674971>>. Acesso em: 5 set. 2018.

SPIRION. *What is Data Lifecycle Management?* Disponível em: <<https://www.spirion.com/data-lifecycle-management/>>. Acesso em: 17 out. 2018.

UNITED States Geological Survey (USGS). *Data Lifecycle*. Disponível em: <<https://www.usgs.gov/products/data-and-tools/data-management/data-lifecycle>>. Acesso em: 17 out. 2018.

UK Data Service. Research data lifecycle. Disponível em:
<<https://www.ukdataservice.ac.uk/manage-data/lifecycle>>. Acesso em: 17 out. 2018.

VAZ, Glauber José. **E-Science na Embrapa.** Campinas: Embrapa Informática Agropecuária, 2011. 58p. (Documentos, 117). Disponível em: <
<https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/921022/1/Livrodoc1171.pdf>>. Acesso em: 20 out. 2018.