

**UNIVERSIDADE FEDERAL DO RIO DO GRANDE DO SUL  
FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

**JAQUELINE INSAURRIAGA SILVEIRA**

**ACESSO ABERTO A DADOS DE PESQUISA EM REPOSITÓRIOS  
UNIVERSITÁRIOS INTERNACIONAIS:**

um estudo sobre políticas de depósito, acesso e uso

Porto Alegre

2021

**JAQUELINE INSAURRIAGA SILVEIRA**

**ACESSO ABERTO A DADOS DE PESQUISA EM REPOSITÓRIOS  
UNIVERSITÁRIOS INTERNACIONAIS:**

um estudo sobre políticas de depósito, acesso e uso

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul como requisito parcial para obtenção do título de Mestre em Ciência da Informação.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Samile Andréa de Souza Vanz

Porto Alegre

2021

## UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Dr. Carlos André Bulhões Mendes

Vice-Reitora: Prof<sup>a</sup>. Dr<sup>a</sup>. Patrícia Helena Lucas Pranke

## FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO

Diretora: Prof<sup>a</sup>. Dr<sup>a</sup>. Ana Maria de Moura

Vice-diretora: Vera Regina Schmitz

## PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Coordenador: Prof. Dr. Thiago Henrique Bragato Barros

Coordenador substituto: Prof. Dr. Moisés Rochemback

---

### CIP - Catalogação na Publicação

Silveira, Jaqueline Insaurreaga  
Acesso aberto a dados de pesquisa em repositórios  
universitários internacionais: um estudo sobre  
políticas de depósito, acesso e uso / Jaqueline  
Insaurreaga Silveira. -- 2021.  
145 f.  
Orientador: Samile Andréa de Souza Vanz.

Dissertação (Mestrado) -- Universidade Federal do  
Rio Grande do Sul, Faculdade de Biblioteconomia e  
Comunicação, Programa de Pós-Graduação em Ciência da  
Informação, Porto Alegre, BR-RS, 2021.

1. Dados abertos de pesquisa. 2. Repositórios de  
dados de pesquisa. 3. Políticas. 4. Repositório  
universitário. I. Vanz, Samile Andréa de Souza,  
orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os  
dados fornecidos pelo(a) autor(a).

---

PPGCIN – UFRGS

Rua Ramiro Barcelos, 2705, Prédio 22201

CEP: 90035-007 Porto Alegre – RS

Telefone: (51) 3308-5067

E-mail: [ppgcin@ufrgs.br](mailto:ppgcin@ufrgs.br)

JAQUELINE INSAURRIAGA SILVEIRA

ACESSO ABERTO A DADOS DE PESQUISA EM REPOSITÓRIOS  
UNIVERSITÁRIOS INTERNACIONAIS:  
um estudo sobre políticas de depósito, acesso e uso

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul como requisito parcial para obtenção do título de Mestre em Ciência da Informação.

Orientadora: Prof<sup>a</sup> Dr<sup>a</sup> Samile Andréa de Souza Vanz

Aprovada em 29 de março de 2021.

Banca examinadora:

---

Prof<sup>a</sup> Dr<sup>a</sup> Caterina Marta Groposo Pavão  
Membro Interno (PPGCIN/UFRGS)

---

Prof. Dr. Rafael Port da Rocha  
Membro Interno (PPGCIN/UFRGS)

---

Prof<sup>a</sup> Dr<sup>a</sup> Sônia Elisa Caregnato  
Membro Externo (PPGCOM/UFRGS)

---

Prof. Dr. Rene Faustino Gabriel Junior  
Suplente (PPGCIN/UFRGS)

## AGRADECIMENTOS

À UFRGS e ao PPGCIN por oportunizar a realização de um sonho.

À minha orientadora Samile, pela disponibilidade em me orientar nessa primeira turma de mestrado do PPGCIN-UFRGS e repassar seus conhecimentos com maestria, respeito e incentivo ao meu desenvolvimento como pesquisadora nesta jornada de eterno aprendizado.

Às professoras Caterina e Sônia, pelas contribuições oferecidas na banca de qualificação, e por aceitarem participar juntamente ao professor Rafael da banca de defesa.

Aos meus filhos Henrique e Thalita, pelo incentivo e por sempre acreditarem na minha capacidade de estudo, a cada novo projeto a que me dedico. E ao meu filho de quatro patas, o Churchill, por todos os momentos de companhia no desenvolvimento desta dissertação e por todos os momentos em que me retirou dos estudos, para nanar o seu soninho de beleza. Amo vocês!

Às amigas Ana, Evelin e Tânia, por serem exemplos em minha jornada, pelo incentivo para cursar o mestrado, amizade e por estarem presentes, mesmo na distância. Às amigas Márcia e Gisa pelo incentivo e confiança na minha aprovação no processo de seleção.

À amiga Iara, parceira de formação na graduação e no mestrado pelas contribuições na leitura desta dissertação.

Aos amigos Diva Maria e Sérgio López, pela presença constante e vibração em cada etapa vencida.

Aos meus colegas de trabalho e amigos da Biblioteca da FCE-UFRGS, pelo incentivo, confiança e apoio para que eu pudesse cursar o mestrado sem interferir nas atividades da Equipe. Ao amigo, colega e parceiro de corridas Vini pelo constante incentivo.

Gratidão à vida por ter me conduzido até aqui.

## RESUMO

Este trabalho aborda as práticas de acesso aberto em repositórios universitários internacionais de dados de pesquisa, por meio da análise de suas políticas, tendo por foco as diretrizes sobre permissões, restrições e licenças para depósito, acesso e uso de dados. Fizeram parte do estudo os repositórios das 18 primeiras universidades elencadas no *ranking Webometrics.info* de 2020. As informações coletadas nos *websites* dos repositórios foram organizadas e categorizadas em quadros elaborados pela autora. A identificação dos repositórios quanto ao seu propósito de atuação revela que, dos 18 repositórios de dados de pesquisa analisados, 17 são institucionais e um é multidisciplinar. Dos 17 repositórios institucionais, nove são para armazenar e preservar toda produção acadêmica e científica das universidades, e os outros oito são exclusivos para dados de pesquisa. O propósito de atuação dos repositórios se reflete nas diretrizes de gestão dos repositórios, sendo que os repositórios não exclusivos para dados apresentam normativas mais generalistas para objetos digitais. Embora os repositórios sejam de acesso aberto, as diretrizes indicam a possibilidade de restrição. As normativas para depósito de dados indicam que os repositórios, em sua maioria, permitem depósito somente para indivíduos com vínculo institucional ou afiliados. Há restrições para depósito de dados e para tipos específicos de dados. Em sua maioria, o acesso é aberto e imediato para uso de dados, e as restrições, quando existem, estão condicionadas às especificadas pelo depositante, envolvem direitos legais, ou são restritos aos usuários com vínculo institucional. A atribuição de licenças para uso de dados indica 77% para as *Creative Commons*, 14% para as *Open Data Commons*, 4% para *AusGOAL* e 5% não especificam. A privacidade no uso de dados sensíveis, em sua maioria, tem por recomendação a estratégia da anonimização. O embargo de dados é identificado nas políticas da maioria dos repositórios, sendo a única exceção o *JHU Data Archive*. O tempo de embargo varia entre seis meses e 10 anos. Todas as políticas analisadas indicam algum tipo de responsabilidade quanto ao uso dos dados e dos serviços dos repositórios. Propõe diretrizes para a elaboração de políticas de depósito, acesso e uso de dados de pesquisas em RDPs.

**Palavras-chave:** Dados abertos de pesquisa. Repositório de dados de pesquisa. Políticas. Repositório universitário.

## ABSTRACT

This study addresses open access practices in international university repositories of research data, through the analysis of their policies, focusing on guidelines on permissions, restrictions, and licenses for submission, access and use of data. The study included the repositories of the first 18 universities listed in the Webometrics.info 2020 ranking. The information collected on the websites of the repositories were organized and categorized into tables prepared by the author. The identification of the repositories as to their purpose of operation reveals that, of the 18 research data repositories analyzed, 17 are institutional and one is multidisciplinary. Of the 17 institutional repositories, nine are for storing and preserving all academic and scientific production of universities, and the other eight are exclusive to research data. The purpose of the repositories is reflected in the data management guidelines, and the repositories are not exclusive to data that present more generalist regulations for digital objects. Although repositories are open access, the guidelines indicate the possibility of restriction. The rules for data submission indicate that repositories mostly allow deposit only for individuals with institutional or affiliated ties. There are restrictions on data submission and for specific types of data. Most access is open and immediate for data use, and restrictions when they exist, are conditional on those specified by the depositor, or involve legal rights, or are restricted to users with institutional ties. The allocation of licenses for data use indicates 77% for Creative Commons, 14% for Open Data Commons, 4% for AusGOAL and 5% do not specify. Privacy in the use of sensitive data is mostly recommended by the anonymization strategy. The data embargo is identified in the policies of most repositories, the only exception being the JHU Data Archive. The embargo time varies between six months and 10 years. All policies analyzed indicate some kind of responsibility for the use of data and services of repositories. It proposes guidelines for the development of policies for the submission, access and use of research data in RDPs.

**Keywords:** Open research data. Research data repository. Policies. Academic repository.

## LISTA DE ILUSTRAÇÕES

Figura 1 — Modelo de ciclo de vida da curadoria de dados formulado pela DCC ...	34
Figura 2 — Modelo de ciclo de vida formulado pela DataONE .....	35
Figura 3 — Modelo de ciclo de vida dos dados formulado pela DDI .....	36
Figura 4 — Modelo de ciclo de vida dos dados de pesquisa formulado pela UK Data Archive Data .....	37
Gráfico 1 — Licenças de uso indicadas nas políticas dos repositórios .....	116

## LISTA DE QUADROS

Quadro 1 — Taxonomia de dados de pesquisa por Sayão e Sales (2019) .....	27
Quadro 2 — Universidades selecionadas para pesquisa, nome do repositório e <i>website</i> .....	54
Quadro 3 — Controle da coleta de dados .....	58
Quadro 4 — Propósito de atuação dos repositórios e data de sua fundação .....	88
Quadro 5 — Políticas identificadas nos <i>websites</i> dos Repositórios analisados .....	90
Quadro 6 — Proveniência dos dados para depósito .....	93
Quadro 7 — Tipos de acesso aos dados .....	95
Quadro 8 — Restrições de acesso aos dados .....	98
Quadro 9 — Permissão de depósito por vínculo institucional .....	100
Quadro 10 — Permissão de depósito sem vínculo institucional .....	101
Quadro 11 — Restrições para depósito de dados .....	103
Quadro 12 — Restrições quanto ao tipo de dados para depósito .....	108
Quadro 13 — Permissões e restrições para uso de dados baixados .....	111
Quadro 14 — Atribuição de licenças de uso dos dados de pesquisa .....	114
Quadro 15 — Diretrizes sobre privacidade no uso de dados pessoais .....	118
Quadro 16 — Comparativo entre restrições ao tipo de dados para depósito X diretrizes sobre privacidade no uso de dados pessoais .....	119
Quadro 17 — Diretrizes sobre embargo de dados .....	121
Quadro 18 — Tempo de embargo identificado nas políticas .....	123
Quadro 19 — Responsabilidade no uso dos Repositórios .....	124
Quadro 20 — Comparativo das responsabilidades identificadas nas políticas .....	126

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>14</b>
1.1 OBJETIVOS .....	18
1.1.1 Objetivo geral .....	18
1.1.2 Objetivos específicos.....	18
1.2 JUSTIFICATIVA .....	19
<b>2 REFERENCIAL TEÓRICO</b> .....	<b>23</b>
2.1 DADOS DE PESQUISA .....	23
2.2 GESTÃO DE DADOS DE PESQUISA.....	30
2.2.1 Ciclo de vida e curadoria digital de dados de pesquisa.....	31
2.2.2 Acesso, licenciamento, embargo e privacidade de dados.....	38
2.3 REPOSITÓRIOS DE DADOS DE PESQUISA (RDPs) .....	47
<b>3 PROCEDIMENTOS METODOLÓGICOS</b> .....	<b>52</b>
3.1 DELINEAMENTO DA PESQUISA.....	52
3.2 PROCEDIMENTO DE IDENTIFICAÇÃO DOS REPOSITÓRIOS.....	53
3.3 DESCRIÇÃO DA ETAPA DE COLETA E ANÁLISE DE DADOS .....	55
<b>4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS</b> .....	<b>60</b>
4.1 APRESENTAÇÃO DAS UNIVERSIDADES E SEUS REPOSITÓRIOS .....	60
4.1.1 Harvard Dataverse .....	60
4.1.2 Stanford Digital Repository (SDR).....	62
4.1.3 DSpace@MIT.....	63
4.1.4 Data Repository [Berkeley].....	64
4.1.5 Research Works Archive .....	65
4.1.6 Deep Blue Data (DBD) .....	67
4.1.7 ORA-Data.....	68
4.1.8 JHU Data Archive.....	69
4.1.9 Columbia Academic Commons .....	71
4.1.10 eCommons .....	72

4.1.11 Scholarly Commons .....	74
4.1.12 Apollo .....	75
4.1.13 Merritt .....	77
4.1.14 ISPS Data Archive.....	79
4.1.15 MINDS@UW .....	80
4.1.16 UC San Diego Library Digital Collections .....	81
4.1.17 Data Repository for the University of Minnesota (DRUM) .....	83
4.1.18 UCL Research Data Repository (RDR) .....	85
<b>4.2 CONSIDERAÇÕES SOBRE OS REPOSITÓRIOS .....</b>	<b>87</b>
<b>4.3 ANÁLISE DAS POLÍTICAS .....</b>	<b>92</b>
4.3.1 Análise quanto à proveniência dos dados para depósito .....	92
4.3.2 Análise dos tipos de acesso aos dados nos repositórios .....	95
4.3.3 Análise quanto a restrições de acesso aos dados nos repositórios .....	98
4.3.4 Análise quanto as permissões para depósito de dados nos repositórios .....	100
4.3.5 Análise das restrições para depósito de dados nos repositórios.....	103
4.3.6 Análise das restrições quanto aos tipos de dados para depósito.....	108
4.3.7 Análise das permissões e restrições para uso de dados.....	110
4.3.8 Análise das permissões de uso de dados por atribuição de licenças.....	114
4.3.9 Análise quanto à privacidade no uso de dados pessoais.....	118
4.3.10 Análise quanto ao embargo de dados.....	121
4.3.11 Análise quanto as responsabilidades no uso do Repositório .....	124
<b>4.4 DIRETRIZES PARA A ELABORAÇÃO DE POLÍTICAS DE DEPÓSITO, ACESSO E USO DE DADOS DE PESQUISAS EM RDPS.....</b>	<b>129</b>
<b>5 CONSIDERAÇÕES FINAIS .....</b>	<b>134</b>
<b><u>REFERÊNCIAS.....</u></b>	<b>137</b>

## 1 INTRODUÇÃO

O desenvolvimento científico e tecnológico ocorrido ao longo dos tempos proporcionou, a partir das novas descobertas, diversos avanços em prol da sociedade. Contudo, foi no pós-Grande Guerra e, posteriormente com a expansão da Internet e das Tecnologias da Informação e Comunicação (TICs), que esses avanços se intensificaram, transformando o fazer científico, possibilitando novas alternativas de comunicação entre os pesquisadores e, conseqüentemente, provocando alterações nos paradigmas da ciência (KURAMOTO, 2008; MEDEIROS; CAREGNATO, 2012; SANTOS, 2014).

Segundo proposta elaborada por Jim Gray (HEY; TANSLEY; TOLLE, 2009), existem quatro paradigmas da ciência: o **paradigma da ciência empírica**, tendo por base os experimentos científicos; o **paradigma da ciência teórica**, que formula modelos para explicar os fenômenos naturais; o **paradigma da ciência computacional e das grandes gerações de dados**; e, por fim, o **quarto paradigma, o e-Science**, entendido como sendo a infraestrutura que possibilita que cientistas e pesquisadores tenham “acesso a dados científicos primários distribuídos, utilizando o acesso remoto a esses conteúdos”. Indo além das questões de informática, o quarto paradigma possibilita e capacita o acesso, movimento, manipulação e extração de dados, que é o ponto central das ciências colaborativas (MEDEIROS; CAREGNATO, 2012, p. 312).

O movimento Ciência Aberta insere-se nesse ambiente da *e-Science* como um novo fazer científico que envolve práticas e iniciativas de abertura da ciência em diferentes âmbitos. Entendida por Oliveira e Silva (2016, p.6) como “[...] um fio condutor de investigações científicas apoiadas por uma ciberinfraestrutura tecnológica e metodológica”, a ciência aberta “[...] permite o uso, reuso e reprodutibilidade de dados de pesquisa”.

Impulsionada pelo movimento Acesso Aberto (AA), a ciência aberta, com todas as suas ramificações, promove novas práticas de vislumbrar e compartilhar ciência. Dentre essas práticas, está a de abertura, compartilhamento e reuso de dados de pesquisa, como forma de minimizar tempo e recursos de pesquisa. Em seus pressupostos, a ciência aberta enuncia uma ciência mais colaborativa, permitindo que o conhecimento científico esteja livre para ser usado, reutilizado e distribuído sem embargos legais, tecnológicos ou sociais (SAYÃO; SALES, 2016b).

Motivada pelo movimento AA, a ciência aberta tornou-se um movimento maior, com uma taxonomia abrangente, envolvendo outros tantos movimentos de abertura da ciência, “[...] como os dados científicos abertos, as ferramentas científicas abertas (*software* e *hardware*), os cadernos abertos de laboratório, a educação aberta e a ciência cidadã.” (ALBAGLI, 2017, p. 659).

Por sua vez, o AA, ou *Open Access*, movimento de abertura à literatura científica, tem por marco a iniciativa de Budapeste em 2002 (*Budapest Open Access Initiative* - BOAI), a qual propôs o conceito e as estratégias para o acesso aberto por intermédio da via dourada (*Gold Road*) e da via verde (*Green Road*). A via dourada se refere às revistas de acesso aberto com revisão por pares, e a via verde ao autoarquivamento de artigos pelos autores em repositórios institucionais de livre acesso (SANTOS, 2017). Os repositórios institucionais são uma realidade alicerçada mundialmente, permitindo o acesso à produção científica gerada no âmbito das instituições acadêmicas e de pesquisa. Contudo, no Brasil, o autoarquivamento de artigos pelos autores, representado pela via verde, ainda é uma prática a ser consolidada, bem como o armazenamento de dados de pesquisa em repositórios brasileiros.

Em busca realizada por repositórios de dados no diretório **re3data.org**<sup>1</sup>, identificou-se o registro de 2005 repositórios disciplinares, 626 repositórios institucionais, e 285 repositórios classificados como outros, sendo que, dentre esse montante recuperado, apenas nove repositórios são provenientes do Brasil. Em estudo realizado há três anos por Pavão *et al.* (2018b), ao ampliar a busca para além do **re3data.org**, os autores identificaram a presença de 15 repositórios brasileiros que disponibilizam dados de pesquisa. Dentre esses repositórios, quatro são de abrangência internacional (envolvendo várias instituições), sete são de abrangência nacional, e cinco de abrangência multi-institucional. Mesmo com a ampliação da busca pelos autores, constata-se que as iniciativas de instituições brasileiras para o compartilhamento de dados de pesquisa ainda é muito incipiente e tem um longo caminho a trilhar nas diferentes dimensões de uma ciência mais colaborativa.

Pressupõe-se que, diante desse novo fazer científico de uma ciência mais colaborativa, novas regras e práticas sejam oportunizadas pela comunidade científica, pelas instituições de fomento, empresas e pelo governo, demandando que se pense

---

<sup>1</sup> Pesquisa realizada em re3data.org em 29/06/2020 às 19h47min.

amplamente nas dimensões que sustentam a efetivação da ciência aberta. Nas diferentes dimensões da ciência aberta propostas por Oliveira e Silva (2016, p. 29), deve-se ter em mente que:

- a) na **dimensão epistemológica**, a ciência voltada aos dados não substitui os paradigmas anteriores, mas muda o *modus operandi* das investigações científicas;
- b) na **dimensão política**, ainda há um longo caminho a trilhar na conscientização dos atores envolvidos, bem como a necessidade de implementação de normativas institucionais e governamentais que amparem as boas práticas desse novo fazer científico;
- c) na **dimensão ético-legal-cultural**, é importante elucidar as dúvidas dos pesquisadores quanto às questões de direitos de propriedade intelectual diante da abertura dos dados, uma vez que, culturalmente, a pesquisa científica apoia-se em sistemas de propriedade intelectual de natureza privada, em que o reconhecimento científico se dá pela transmissão de titularidade de direitos para sua exploração comercial;
- d) na **dimensão morfológica**, é necessário mostrar a dinâmica da *e-Science* em relação às mudanças no *modus operandi* da pesquisa científica, por meio da ciberinfraestrutura da comunicação científica; e
- e) na **dimensão tecnológica**, é importante adotar modelos e tecnologias abertas, soluções tecnológicas que “[...] devem prover a sustentabilidade de uma ciência orientada aos dados e proporcionar a coleta, gestão, curadoria, preservação para o acesso, uso e reuso em longo prazo.”

A ciência contemporânea, ao produzir e consumir uma grande variedade de dados, desloca-os do papel de mero subproduto da pesquisa científica ao patamar de recursos informacionais de valor, como protagonistas nesse novo paradigma científico pautado na evolução das práticas científicas e na preocupação com a abertura, compartilhamento e reuso desses dados (MONTEIRO; LUCAS, 2019; SAYÃO; SALES, 2016).

A instituição de políticas de gestão de repositórios e a disponibilização destas de forma clara e acessível aos usuários garante confiabilidade ao repositório, minimiza dúvidas e questões de judicialização, bem como traz segurança aos usuários em relação às práticas adotadas pelo repositório referente a permissões, restrições e licenciamentos para o depósito, acesso e uso de dados de pesquisa. Segundo

Carvalho (2012), os repositórios podem acabar sendo subutilizados em decorrência das inseguranças geradas pela falta de adequação e clareza de suas políticas.

Assim sendo, para que os dados de pesquisa atendam às premissas de acesso, compartilhamento e reuso, estes necessitam de uma infraestrutura tecnológica e gerencial que permita que sejam cuidados ao longo do seu ciclo de vida (SAYÃO; SALES, 2016), que se relacionam, ainda segundo Sayão e Sales (2015), ao depósito desses dados em ambientes que garantam sua preservação ativa por longo prazo, mantendo suas características de autenticidade, integridade e proveniência, e garantindo, assim, que estejam sempre disponíveis e prontos para o uso.

Em relação à proveniência dos dados para depósito nos repositórios, entende-se que sua identificação se faça necessária, uma vez que, segundo Freud, Sembay e Macedo (2019, p. 807), “a proveniência de dados pode ser entendida como o registro que descreve pessoas, instituições, entidades e atividades envolvidas na produção, influência ou entrega de um dado ou objeto.” No caso do presente estudo, a proveniência mostrará a origem dos dados em relação aos seus pesquisadores e instituições.

A gestão de dados de pesquisa está também condicionada ao atendimento das boas práticas de gestão e partilha de dados, práticas resumidas e descritas pelos princípios FAIR (*Findable, Accessible, Interoperable, Reusable*), os quais preconizam que, para ser FAIR, os dados devem ser localizáveis, acessíveis, interoperáveis e reutilizáveis (RODRIGUES *et al.*, 2019). Assim, requerem ações por parte das instituições quanto às diretrizes de armazenamento desses dados.

A premissa de que o acesso, compartilhamento e reuso de dados de pesquisa através de seu armazenamento seja feita em repositórios institucionais ou em repositórios de domínios específicos traz uma nova demanda às instituições de pesquisa, dentre elas as universidades, as quais são as principais produtoras de ciência no Brasil. Contudo, as incipientes iniciativas brasileiras de implementação de Repositórios de Dados de Pesquisa (RDPs) em universidades acarretam uma falta de instrumentos comparativos em relação às políticas de gestão dos repositórios ao que se refere ao depósito, acesso e uso desses dados. Como demonstrado no estudo de Pavão *et al.* (2018), somente nove dentre as 15 iniciativas brasileiras de implantação de RDPs disponibilizam suas políticas e regras institucionais, enquanto as outras seis não disponibilizam nenhuma informação quanto às políticas adotadas.

Sendo assim, o presente trabalho aborda práticas de Acesso Aberto a Dados de Pesquisa (AADP) adotadas em repositórios universitários, em âmbito internacional, por meio de reconhecimento e análise de suas políticas, referentes ao depósito, acesso e uso de dados de pesquisa. Nas políticas analisadas, observou-se os seguintes aspectos: proveniência dos dados para depósito nos repositórios; permissões e restrições para depósito, acesso e uso de dados; licenças de uso dos dados; uso de dados sensíveis; utilização do embargo como instrumento limitante ao acesso aos dados e; responsabilidade no uso dos dados e serviços oferecidos pelos repositórios. Esses aspectos analisados serviram de subsídio para diretrizes para a elaboração de políticas de depósito, acesso e uso de dados de pesquisa, direcionadas a repositórios universitários brasileiros.

Dessa forma, para observar tais políticas, formulou-se a seguinte questão de pesquisa: *quais práticas de acesso aberto a dados de pesquisas estão sendo orientadas nas políticas de gestão de repositórios de dados de universidades em âmbito internacional referente às permissões, restrições e licenciamento para depósito, acesso e uso de dados de pesquisa?*

A partir da questão de pesquisa acima formulada para o presente estudo, passa-se a apresentar, a seguir, os objetivos propostos para respondê-la, separando-os em objetivos geral e específicos.

## 1.1 OBJETIVOS

Os objetivos desta pesquisa estão divididos em geral e específicos, conforme os itens abaixo.

### 1.1.1 Objetivo geral

Investigar práticas de acesso aberto a dados de pesquisa instituídas em políticas de gestão de repositórios de dados de universidades internacionais, referente às permissões, restrições e licenciamento para depósito, acesso e uso de dados de pesquisa.

### 1.1.2 Objetivos específicos

Para atender ao objetivo geral de pesquisa, têm-se os seguintes objetivos específicos:

- a) elencar as políticas de Acesso Aberto a Dados de Pesquisa em repositórios institucionais de dados de pesquisa de universidades internacionais;
- b) analisar e comparar as políticas de gestão dos repositórios quanto aos seguintes aspectos:
  - 1) **proveniência dos dados para depósito;**
  - 2) **tipos de acesso aos dados nos repositórios;**
  - 3) **restrições de acesso aos dados nos repositórios;**
  - 4) **permissão para depósito de dados;**
  - 5) **restrições para depósito de dados;**
  - 6) **restrições aos tipos de dados para depósito;**
  - 7) **permissões e restrições para uso de dados baixados;**
  - 8) **licença de uso dos dados;**
  - 9) **privacidade no uso de dados pessoais;**
  - 10) **embargo no uso de dados;**
  - 11) **responsabilidade no uso dos repositórios e seus conteúdos.**
- c) Propor diretrizes para a elaboração de políticas para RDPs de universidades brasileiras referente a permissões, restrições e licenças para depósito, acesso e uso de dados de pesquisa.

As seções seguintes apresentam a justificativa para o presente estudo, o referencial teórico que serviu de embasamento para as temáticas abordadas e a metodologia utilizada. Na sequência, são mostrados os repositórios selecionados para a pesquisa com uma breve contextualização das universidades a qual estão vinculados, as análises de suas políticas referente aos aspectos especificados nos objetivos deste estudo e, as diretrizes propostas para a elaboração de políticas de depósito, acesso e uso de dados de pesquisa em RDPs. Por fim, são apresentadas as considerações finais.

## 1.2 JUSTIFICATIVA

A abertura, compartilhamento e reuso de dados de pesquisa, tendo por premissa uma ciência colaborativa que minimize custos, otimize tempo de pesquisa, e promova o progresso e a inovação científica, requer a prescrição de planos de

gestão e curadoria, que estabeleçam como esses dados serão acessados, compartilhados e reutilizados ao longo do seu ciclo de vida. Da mesma forma, requer a implantação de sistemas com infraestruturas que deem suporte ao armazenamento, preservação e publicação desses dados.

A intensificação de estudos que aprofundem discussões sobre dados de pesquisa é importante, visto que, ao abrangerem diferenças disciplinares na produção/obtenção, compartilhamento e reuso de dados, esses estudos, além de contribuírem para os processos de tomada de decisão em instituições de pesquisa, fornecem subsídios para a criação ou aperfeiçoamento de sistemas de informação caracterizados, sobretudo, por lidarem com dados de pesquisa de diferentes disciplinas (CARVALHO, 2018).

Em diversos países, estão em andamento mecanismos de manutenção e gestão de RDPs (MORENO, 2018). A gestão e representação de dados ou conjuntos de dados armazenados em repositórios é um desafio para as instituições de pesquisa por envolver altos custos de investimento financeiro, tecnológico e de recursos humanos, além de abranger questões legais, envolvendo o acesso, uso e compartilhamento desses dados.

A literatura brasileira apresenta diferentes estudos sobre curadoria e gestão de dados de pesquisa, contudo ainda se mostra incipiente referente às temáticas envolvendo normativas para depósito, acesso e uso de dados em RDPs, fator que motiva o presente estudo.

A implantação de RDPs é uma demanda das universidades e institutos de pesquisa, em consonância ao movimento de abertura dos dados de pesquisa. Por meio de repositórios de dados, as universidades propiciam a disseminação dos conjuntos de dados produzidos por suas comunidades, potencializando, dessa forma, a divulgação e acesso aos dados, bem como a visibilidade e reconhecimento de seus pesquisadores (MONTEIRO, 2017).

Um exemplo de repositório institucional é o Lume, repositório da Universidade Federal do Rio Grande do Sul (UFRGS), uma instituição octogenária de ensino, pesquisa e extensão, avaliada como a segunda melhor universidade federal do Brasil, segundo o *ranking* The Center for World University Ranking (CWUR) 2020-2021<sup>2</sup>. A instituição ocupa, ainda, segundo o CWUR, a quinta posição no *ranking* geral de

---

<sup>2</sup> CWUR. World University Rankings 2020-21. Disponível em: <https://cwur.org/2020-21.php>. Acesso em: 08 jun. 2020.

universidades do país e a 476ª posição a nível mundial. O LUME é responsável por reunir, preservar, divulgar e permitir o acesso confiável e permanente aos documentos acadêmicos, científicos, artísticos e administrativos gerados na Universidade.

Dentre os repositórios de instituições de ensino e de pesquisa brasileiras e latino-americanas, o Lume ocupa a primeira posição, segundo a 8ª edição do *Transparent Ranking: Institutional Repositories by Google Scholar*, e destaca-se ainda como terceiro colocado no mundo dentre repositórios institucionais e oitavo dentre o total de 3.084 repositórios mundiais (UFRGS, 2020). Seu acervo conta atualmente com 234.761<sup>3</sup> documentos digitais, distribuídos em distintas comunidades/coleções.

Embora a UFRGS possua expertise no gerenciamento de repositórios digitais, visto ter implementado o LUME em janeiro de 2008, e este, ao longo dos anos, ter crescido em proporção de acervo e em qualidade, a implantação de um repositório de dados na Universidade requer estudos, análises e adequações a sua realidade. A observância de questões referentes à gestão e curadoria dos diferentes tipos de dados gerados nas diferentes áreas do conhecimento no âmbito da Universidade se faz necessária, bem como de questões de infraestrutura referente ao armazenamento desses dados, se serão agregados ao LUME ou não. Ainda, é fundamental a observação de questões referentes ao treinamento de recursos humanos, bem como ao estabelecimento de políticas de gestão do repositório.

Tais políticas devem se atentar aos aspectos referentes a permissões e restrições para depósito, acesso e uso de dados de pesquisa, observando também sua privacidade e embargo. Além disso, deve atentar-se ao licenciamento no uso e compartilhamento desses dados e às responsabilidades no uso dos mesmos e do repositório. Acredita-se que o estabelecimento e publicação dessas políticas venha garantir confiabilidade e segurança aos pesquisadores e usuários no uso do repositório, além de proporcionar segurança aos gestores do repositório em relação a isenção de responsabilidades diante da sua utilização por terceiros.

A abrangência das diferentes áreas do conhecimento na produção científica das universidades, com peculiaridades distintas em cada área, acarreta um fator desafiador para a gestão de um repositório de dados, exigindo que se busque por experiências consolidadas em repositórios universitários internacionais, haja vista a incipiência de iniciativas nacionais. A identificação, e posterior análise das políticas de

---

<sup>3</sup> Dados coletados no LUME em 31/01/2021 às 15:47.

gestão de repositórios, estabelecidas por estas instituições internacionais, no que tange as normativas referentes a depósito, acesso e uso de dados de pesquisa, possibilitarão o estabelecimento de diretrizes para as políticas direcionadas aos RDPs de universidades brasileiras.

Isso posto, o presente estudo contribui socialmente ao identificar políticas de gestão de repositórios que possam servir de diretrizes aos RDPs brasileiros, evitando, assim, futuras questões de judicialização entre usuários e instituições mantenedoras dos repositórios de dados. Contribui também para a gestão da pesquisa, uma vez que norteia o pesquisador em relação a essas políticas para que o Plano de Gestão de Dados (PGD) atenda às necessidades do repositório, bem como orienta em relação a aspectos técnico/tecnológicos para que o repositório responda as exigências das políticas financiadoras no que diz respeito a preservação, acesso e uso de dados de pesquisa.

Para a UFRGS contribui com diretrizes sobre a gestão de RDPs referentes às permissões, restrições e licenças para depósito, acesso, e uso de dados de pesquisa, no âmbito do seu repositório de dados em fase de planejamento. Também agrega valor ao Sistema de Bibliotecas da Universidade e ao seu corpo técnico de bibliotecários, do qual a autora é integrante como responsável pelo processamento técnico da produção intelectual da comunidade acadêmica da Faculdade de Ciências Econômicas.

Para a Ciência da Informação, a presente pesquisa contribui como um aporte teórico, ainda incipiente na literatura brasileira, sobre políticas de depósito, acesso e uso de dados de pesquisa em RDPs de universidades. Ao Programa de Pós-Graduação em Ciência da Informação da UFRGS (PPGCIN/UFRGS), contribui como produção acadêmica, haja vista ser um programa de pós-graduação recente, iniciado no ano de 2019, do qual a autora faz parte, integrando a primeira turma de mestrandos. Por fim, contribui com o projeto de pesquisa da Rede de Dados de Pesquisa Brasileiros (RDP Brasil), desenvolvido na Faculdade de Biblioteconomia e Comunicação da UFRGS (FABICO/UFRGS), ao se inserir no âmbito da temática estudada pelo Grupo.

A seção a seguir apresenta a contextualização teórica que amparou o presente estudo.

## 2 REFERENCIAL TEÓRICO

Esta seção apresenta a fundamentação teórica para as temáticas tratadas neste estudo. Os temas abordados são: dados de pesquisa segundo sua diversidade e heterogeneidade na perspectiva da ciência aberta; gestão de dados de pesquisa; ciclo de vida dos dados de pesquisa e a curadoria digital como aspecto imprescindível para o processamento e preservação dos dados; acesso, licenciamento, embargo e privacidade de dados; e repositórios de dados de pesquisa segundo suas características e propósito de atuação.

### 2.1 DADOS DE PESQUISA

A ciência, no seu fazer, produz, ao longo dos tempos, diferentes tipos de dados que sustentam suas atividades de pesquisa e validam seus resultados. Contudo, foi a partir dos eventos *Big Data* e *e-Science*, impulsionados pelo desenvolvimento tecnológico e pelo surgimento e expansão da Internet e das TICs (KURAMOTO, 2008) que se desencadeou, segundo Jorge (2018), o processo de “dataficação” da ciência, o qual envolve a produção massiva de dados de pesquisa, trazendo mudanças à ciência e à forma de ver e lidar com os dados gerados nas atividades científicas.

Os dados de pesquisa adquirem valor a partir do movimento de abertura da ciência. Anteriormente restritos aos membros de uma equipe de pesquisa, ou simplesmente esquecidos em anotações engavetadas, os dados perdiam-se ao longo do tempo, impossibilitando seu uso por outros pesquisadores. Com a ciência aberta, os dados de pesquisa passam a ser vistos como elementos importantes, que agregam valor aos projetos de pesquisa e, conseqüentemente, à ciência. Por meio do seu compartilhamento e reuso, permitem que outros cientistas confirmem ou refutem os resultados obtidos em estudos realizados, bem como possibilitam o seu reuso em novas pesquisas e por diferentes pesquisadores.

À vista disso, os dados de pesquisa são considerados “um dos principais produtos das pesquisas acadêmicas nas universidades e centros de pesquisas” (SANTA ANNA; DIAS; MACULAN, 2019, p. 7), grandes responsáveis pela geração dos dados cauda longa, os quais representam pequenas quantidades de dados distribuídos num grande número de iniciativas de pesquisas científicas.

O valor que os dados atribuem à pesquisa científica, segundo Borgman (2010, p. 3), está relacionado à “categoria a qual pertencem, sendo que alguns tipos de dados têm valor imediato e duradouro, outros ganham valor ao longo do tempo, alguns têm valor transitório e outros são mais fáceis de recriar do que selecionar.”

Diferentes dados são gerados nas pesquisas, e o seu compartilhamento impulsiona uma nova ciência, melhorando sua reprodutibilidade e robustez. Contudo, ao não se conseguir prever o valor futuro de um conjunto de dados, perde-se a noção de quanto o não compartilhamento desses dados poderá vir a ser um impedimento para os cientistas futuramente (MELIS, 2018).

O movimento de abertura dos dados de pesquisa, assim como o movimento de abertura da ciência ao qual faz parte, é um processo em construção, envolvendo diferentes interesses, pontos de vistas e interpretações (ALBAGLI, 2015). Esse contexto, além de envolver tensões e disputas, também envolve sentimento de insegurança e desconfiança por parte dos pesquisadores no que tange a questões sobre propriedade intelectual, prioridade e reconhecimento nas descobertas científicas, segurança, privacidade e confidencialidade de dados.

Embora os dados de pesquisa sejam objetos complexos, com propósitos e métodos de produção variados nos diferentes campos científicos, inclusive nos seus critérios de compartilhamento, Borgman (2010, p. 2, *tradução nossa*<sup>4</sup>) afirma que, “para o enfrentamento desses desafios, é necessário examinar os argumentos para o compartilhamento de dados e como esses argumentos correspondem às motivações e interesses da comunidade científica e do público.”

Além das questões de insegurança anteriormente expostas, a compreensão do que seja um dado de pesquisa pode gerar controvérsias e mais incertezas. Faz-se, assim, necessário um entendimento epistemológico do termo “dado”, uma vez que este pode assumir uma grande diversidade de formas, bem como uma variação terminológica quanto a sua definição junto aos pesquisadores e às diferentes áreas do conhecimento. Essa preocupação está presente em Oliveira e Silva (2016), ao enfatizarem a importância de um entendimento terminológico para o termo “dado” pelo

---

<sup>4</sup> Tradução livre do trecho: “To address these challenges, it is necessary to examine the arguments for sharing data and how those arguments match the motivations and interests of the scientific community and the public.”

fato de encontrarem na literatura diferentes expressões, como “dados de pesquisa”, “dados científicos” e “dados abertos”.

Costa (2107), igualmente, sinaliza uma variedade tipológica de dados encontrada na literatura, bem como a dificuldade terminológica para definir o termo “dado”. Dessa forma, a autora busca amparo nas considerações de Sayão e Sales (2013), que argumentam que essa variação terminológica ocorre devido aos dados serem gerados para diferentes finalidades e por distintas comunidades acadêmicas e científicas, por meio de diversos processos.

Outra questão importante relativa ao entendimento sobre o termo “dado” refere-se a sua produção, distinguindo-se os dados produzidos, coletados e processados pela pesquisa, comumente chamados de ‘dados de pesquisa’, dos dados “disponibilizados por organizações para serem usados em pesquisas, comumente chamados ‘dados para pesquisa’.” (ROCHA *et al.*, 2018, p. 8).

Terminologicamente, para fins do presente estudo, adotar-se-á o conceito apresentado no relatório da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) publicado em 2007, o qual define dados de pesquisa como “registros factuais usados como fonte primária para a pesquisa científica e que são comumente aceitos pelos pesquisadores como necessários para validar os resultados do trabalho científico” (OCDE, 2007<sup>5</sup>, p. 13, *tradução nossa*). Excluem-se desse conceito “cadernos de laboratório, análises preliminares, rascunhos de artigos científicos, planos para futuras pesquisas, avaliação por pares ou comunicações pessoais, objetos físicos como amostras de laboratório [...]” (OCDE, 2007<sup>6</sup>, p. 14, *tradução nossa*).

Buscando uma categorização dos dados de pesquisa, Sales (2014) apresenta as diversas classificações dos dados de pesquisa. A proposta da National Science Foundation - NSF (2005) considera que os dados se apresentam de diferentes formas, as quais incluem sua natureza, sua reprodutibilidade e o nível de processamento ao qual foram submetidos. Quanto a sua natureza, os dados podem ser diversificados, incluindo entre outros, números, imagens, vídeo ou áudio, informações sobre

---

<sup>5</sup> Tradução livre do trecho: “Research data” are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings.”

<sup>6</sup> This term does not cover the following: laboratory notebooks, preliminary analyses, and drafts of scientific papers, plans for future research, peer reviews, or personal communications with colleagues or physical objects (e.g. laboratory samples, strains of bacteria and test animals such as mice).

*softwares*, algoritmos, equações, animações ou modelos/simulações. De acordo com sua origem, os dados podem ser categorizados em: observacionais, computacionais e experimentais. Segundo NSF (2005), tal distinção é crucial para as escolhas feitas para arquivamento e preservação. Ademais, há uma categoria de dados que são os dados classificados de acordo com o nível de processamento ao qual foram submetidos, podendo, além dos dados brutos, gerar dados derivados, os quais, segundo a NSF (2005), podem ser mais facilmente reutilizados.

A proposta de Borgman (2010) categoriza os dados igualmente em observacionais, computacionais e experimentais, mas acrescenta entre eles os “dados de registros”, que consistem em dados provenientes do governo, de empresas públicas e privadas e são úteis para pesquisas científicas, sociais e humanísticas. De acordo com Borgman (2010), embora essa categorização seja útil como uma estruturação dos dados de pesquisa, tal divisão tende a obscurecer os diversos tipos de dados gerados numa pesquisa acadêmica.

Já Harvey (2010<sup>7</sup> *apud* Sales 2014) tem outra classificação para os dados de pesquisa, dividindo-os em coleções, a saber: coleção de pesquisa; coleção de comunidade; e coleção de referência, incluindo, ainda, em sua classificação, os “metadados” como outra categoria de tipos de dados.

Por fim, Lyon (2007) observa que, embora os dados possam ser categorizados genericamente, é importante reconhecer sua complexidade e heterogeneidade. Para classificá-los, apoia-se na categorização do Natural Environment Research Council (NERC), em que os dados são divididos em: dados canônicos (dados com variação mínima) e dados episódicos (aqueles que podem mudar, como uma informação climática). Na classificação do International Union of Crystallography<sup>21</sup> (IUCr), por sua vez, os dados são classificados em: dados crus (imagem de arquivos), dados primários (fatores de estrutura) e dados derivados (seis dimensões do modelo estrutural).

Souza (2012, p. 4), porém, ressalta que, de acordo com duas premissas fundamentais no que diz respeito às classificações, “os mesmos objetos e ideias podem ser organizados e representados de formas diferentes e toda classificação está relacionada a um propósito definido de construção e uso de informação”. Sendo

---

<sup>7</sup> HARVEY, Douglas Ross. *Digital curation: a how-to-do-it manual*. London: Facet, 2010.

assim, todas essas classificações “refletem o propósito para o qual foram construídas em um contexto específico” (SALES, 2014, p. 52).

A partir do entendimento de que dados de pesquisa são objetos complexos, diversificados e heterogêneos, Sayão e Sales (2015) propõem a seguinte classificação, categorizando-os:

- a) **quanto a sua origem** (dados observacionais, experimentais e computacionais);
- b) **quanto a sua natureza** (números, imagens, vídeos ou áudio, *software*, algoritmos, equações, animações ou modelos e simulações);
- c) **quanto à fase de pesquisa** (dados brutos ou dados primários; dados derivados; dados canônicos ou dados referenciais).

No entanto, os autores, após revisitarem o estudo de Sales (2014) e considerarem que as classificações expostas pela autora (por meio de um levantamento na literatura) apresentavam uma visão única sobre os dados, que parecia, por vezes, sobrepor-se em alguns conceitos e divergir em outros (SALES; SAYÃO, 2019), apontam uma proposta de taxonomia para dados de pesquisa. Sendo assim, após um estudo conceitual sobre a tipologia dos dados de pesquisa e a partir das visões encontradas na literatura e dos tipos de dados indicados em entrevista realizada com pesquisadores da área de ciência nucleares, os autores sistematizam os dados de pesquisa, passando a incluir abordagens quanto a sua **materialidade, perenidade, nível de sensibilidade e abertura**.

Dessa forma, Sales e Sayão (2019) chegaram na sistematização apresentada no Quadro 1 a seguir.

Quadro 1 – Taxonomia de dados de pesquisa por Sayão e Sales (2019)

<b>DADOS DE PESQUISA</b>	<b>Quanto à origem</b>	<b>Dados de pesquisa</b>	Observacionais Experimentais Computacionais	
		<b>Dados para a pesquisa</b>	Governamentais Registros Dados transacionais	
	<b>Quanto ao grau de processamento</b>	<b>Dados brutos/primários/crus</b>		
		<b>Dados intermediários/pré-processados/ secundários</b>	Dados limpos Dados analisados Dados anonimizados Dados derivados	
		<b>Dados finais/processados/terciários</b>	Dados publicados/arquivados Dados de alta densidade	
	<b>Quanto à abordagem</b>	Dados qualitativos Dados quantitativos		

<b>DADOS DE PESQUISA</b>	<b>Quanto à natureza</b>	<b>Número</b>	Medidas Resultado de levantamento Fórmula Equação Algoritmo
		<b>Multimídia</b>	Imagem Vídeo Áudio Animação Filme Fotografia
		<b>Software</b>	Base de dados Simulação Códigos
		<b>Visualização</b>	Tabelas Gráficos Diagramas Modelo em 3D Modelo reduzido Desenho
		<b>Textual</b>	Metadado Questionário Entrevista Anotação Certificado Caderno de laboratório Transcrição Correspondência Diário Caderno de campo
		<b>Artefato</b>	Espécime Amostra Maquete Planton/Manequim
		<b>Processo</b>	Procedimentos operacionais padronizados <i>Workflow</i> Protocolo Teste
	<b>Quanto ao nível de sensibilidade</b>	<b>Dados não sensíveis/ostensivos</b>	
		<b>Dados confidenciais</b>	
		<b>Dados pessoais</b>	Dados pessoais anonimizados Dados pessoais sensíveis
	<b>Quanto à materialidade</b>	<b>Dados digitais</b>	
		<b>Dados físicos</b>	Amostra Espécime Artefato
	<b>Quanto à perenidade</b>	<b>Dados canônicos/referenciais</b>	
		<b>Dados episódicos</b>	Coleções institucionais * Coleções comunitárias Coleções individuais
	<b>Quanto à abertura</b>	<b>Grau de abertura</b>	Fechado Aberto Parcialmente aberto Embargado
		<b>Tipo de licença para reuso</b>	Creative Commons Outras

**Fonte:** elaborado pela autora com referência na proposta de Sales e Sayão (2019).

\*Nota: Informação fornecida pela Luana Sales via contato no LinkedIn, em 27/09/2020.

Nessa proposta de taxonomia elaborada pelos autores, os tipos de dados de pesquisa são divididos a partir de oito características, referentes: à origem dos dados;

ao grau de processamento; à abordagem da pesquisa; à natureza dos dados; ao nível de sensibilidade dos dados; à materialidade; à perenidade; e, por fim, à abertura de dados (SALES; SAYÃO, 2019).

Ainda de acordo com Sales e Sayão (2019), as características adotadas por eles na proposta de taxonomia de dados de pesquisa ajudam a distinguir as abordagens a serem consideradas no mapeamento e curadoria dos dados, possibilitando a identificação de metadados específicos no tratamento de cada tipo de dado a ser curado.

Outro aspecto importante a considerar, de acordo como Monteiro e Lucas (2019), é o tempo de vida dos dados de pesquisa, haja vista ser um ponto primordial para a questão do reuso dos dados. Torna-se imprescindível, portanto, a gestão dos dados, planejando sua coleta, descrição e preservação, de forma a garantir sua qualidade para reuso por parte de outros pesquisadores e a possibilitar a descoberta de novos dados, a integração e a análise de dados já existentes (MONTEIRO; LUCAS, 2019).

O compartilhamento e reuso de dados de pesquisa é um comportamento recente na prática científica nacional e, sendo assim, pode ainda gerar equívocos e dúvidas entre os pesquisadores. Em estudo realizado sobre as práticas e percepções dos usuários nacionais de serviços de acesso aberto a dados de pesquisa, Vanz *et al.* (2018, p. 3) identificaram que, dentre os indivíduos pesquisados, embora haja grande interesse pela temática, ainda “[...] há equívocos em relação ao que significa compartilhamento e reuso de dados de pesquisa.” O estudo também mostra a incipiência quanto à prática de compartilhamento e reuso de dados entre os pesquisadores brasileiros investigados, bem como a indisponibilidade de repositórios institucionais para compartilhamento de dados de pesquisa.

Uma diversidade de dados de pesquisa é gerada constantemente nas diferentes áreas do conhecimento, cada qual com suas especificidades e distintos ciclos de vida. A partir do entendimento de que esses dados só podem ser considerados FAIR se forem tratados e preservados em ambientes seguros ao longo do seu ciclo de vida, aborda-se, na seção a seguir, a gestão de dados de pesquisa.

## 2.2 GESTÃO DE DADOS DE PESQUISA

Em decorrência da grande quantidade e variedade de dados produzidos nas atividades científicas, muitos desses dados não são analisados ou utilizados em uma única pesquisa (JORGE, 2018), fato considerado um elemento incentivador para “[...] práticas de abertura de dados de pesquisa para colaboração entre pesquisadores e equipes de pesquisa, reuso em novas pesquisas, reprodutibilidade de experimentos, e verificação dos resultados científicos entre os pares” (JORGE, 2018, p. 19).

No entanto, para que haja uma efetiva colaboração entre pesquisadores e equipes de pesquisa por meio da abertura dos dados, é necessário seguir princípios e padrões que garantam o compartilhamento e reuso dos dados, bem como seu acesso por todo seu ciclo de vida. Diante das muitas questões que envolvem a abertura de dados de pesquisa, é importante lembrar que este ainda é um processo em construção, envolvendo distintos atores, com suas expectativas, dúvidas e inseguranças.

As instituições de pesquisa e agências de fomento, cada vez mais, têm exigido dos pesquisadores a formalização de um Plano de Gestão de Dados (PGD), o qual corresponde a um documento formal que contém os procedimentos adotados na gestão de dados e considera seu ciclo de vida (MONTEIRO; SANT’ANA, 2018). Indo além de sua aplicabilidade como instrumento norteador das atividades de gerenciamento dos dados, Michener (2015) considera que o PGD pode ser também utilizado como um instrumento de avaliação do mérito de projetos financiados por instituições de pesquisa.

Além de apresentar-se como um desafio devido a sua fragilidade e vulnerabilidade, diante de falhas de armazenamento e obsolescência tecnológica, o armazenamento e o processamento de dados de pesquisa, também pode apresentar desafios perante dados que contenham informações pessoais e necessitem de gerenciamento com medidas de segurança apropriadas (COX; PINFIELD, 2014). Nesse contexto, os repositórios surgem como um sistema de informação com infraestruturas tecnológica e gerencial apropriadas ao armazenamento, gestão e preservação dos dados de pesquisa.

Um dos princípios da gestão de dados de pesquisa está no planejamento, presente desde a concepção do projeto, observando questões referentes: aos tipos, formatos e padrões de dados existentes; aos métodos presentes para coleta de

dados; às políticas de acesso; às questões legais, éticas e relacionadas à propriedade intelectual; aos níveis de acesso aos dados; às formas de compartilhamento e reutilização dos dados; à gestão, curadoria e armazenamento de curto prazo; e ao arquivamento e preservação por longo tempo (BERTIN; VISOLI; DRUCKER, 2017). Embora os autores façam referência à gestão de dados por pesquisadores, o princípio do planejamento para as questões acima citadas deve ser igualmente considerado nas políticas dos repositórios, por meio do estabelecimento de diretrizes que norteiem as ações que envolvem o acesso e o uso dos serviços disponibilizados.

É importante lembrar que os repositórios institucionais e os multidisciplinares se constituem por dados das diferentes áreas do conhecimento, consistindo em distintas abordagens quanto à tipologia e curadoria desses dados. Dessa forma, é imprescindível considerar as particularidades de cada área no processo de gestão dos repositórios, assim como, segundo Monteiro *et al.* (2017), é imperativo tratar das questões de privacidade de dados.

O presente estudo aborda as políticas de depósito, acesso e uso de dados de pesquisa em RDPs por meio de normativas disponibilizadas em seus sites com foco nas permissões, restrições e licenças aplicadas a estes. Estas normativas atribuem confiabilidade aos repositórios e segurança aos seus usuários por explicitarem as ações empenhadas para garantir que os dados depositados estarão adequadamente disponíveis e acessíveis para reutilização ao longo do seu ciclo de vida, tópico este que será abordado a seguir.

### 2.2.1 Ciclo de vida e curadoria digital de dados de pesquisa

O acesso, compartilhamento e reuso de dados de pesquisa dependem de ações que permitam que esses dados sejam FAIR, ou seja, localizáveis, acessíveis, interoperáveis e reutilizáveis. O conhecimento das especificidades dos diferentes tipos de dados produzidos nas várias áreas do conhecimento possibilitam o estabelecimento do ciclo de vida para cada tipologia de dados de pesquisa, assim permitindo ações efetivas para que os dados se tornem FAIR. Dentro desse contexto, encontra-se a curadoria digital.

A curadoria digital envolve operações de arquivamento, autenticação, manutenção, preservação e processamento dos dados de pesquisa, permitindo a descoberta e a recuperação desses dados ao longo do seu ciclo de vida, reduzindo

ameaças ao seu valor de pesquisa a longo prazo e diminuindo o risco de obsolescência digital (DIGITAL CURATION CENTER, 2008). De acordo com o Digital Curation Center (2008), a curadoria digital e a preservação de dados são processos contínuos, os quais, juntos, garantem que os dados permaneçam autênticos, confiáveis e utilizáveis, mantendo sempre sua integridade. Esses processos exigem investimento de tempo e de recursos, tanto financeiros como humanos.

O ciclo de vida dos dados de pesquisa vai além do tempo de vida do projeto que lhe deu origem, visto que os dados podem ser utilizados em continuidade à pesquisa original por outros pesquisadores em novas pesquisas. Isso quer dizer que eles têm um ciclo de vida que corresponde a sua geração, processamento, análise e preservação, podendo ser novamente reutilizados (BERTIN; VISOLI; DRUCKER, 2017).

Segundo Aventurier (2016, *on-line*), o ciclo de vida dos dados de pesquisa tem várias representações e envolve as etapas de:

- a) criação dos dados;
- b) tratamento dos dados;
- c) análise dos dados;
- d) preservação dos dados;
- e) acesso aos dados;
- f) reutilização dos dados.

Silva e Siebra (2017), em pesquisa realizada na literatura, identificaram e analisaram sete modelos de ciclo de vida para a preservação de objetos digitais, sendo eles: o **Joint Information Systems Committee Model (JISC)**; o **Digital Curation Center Lifecycle Model (DCC)**; o **DCC&U – Extended Digital Curation Lifecycle Model**; o **Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval (CASPAR)**; o **UK Data Archive Data Lifecycle**; o modelo da **DigitalNZ**; e o **Dataone Data Lifecycle**. Para as autoras não há clareza de quais ações devem ser realizadas no contexto de cada etapa nem quais documentos devem ser gerados em sua execução, fato que pode gerar dúvidas na aplicabilidade desses modelos por parte dos curadores. Salientam também que, na literatura, faltam exemplos de aplicações desses modelos que possam servir de embasamento às ações de curadoria. Por fim, as autoras enfatizam que a implementação desses modelos deve variar de acordo com a dinâmica e com a rotina das organizações envolvidas na curadoria dos objetos digitais. Vale ressaltar que nem todos os modelos

identificados por Silva e Siebra (2017) são modelos de ciclo de vida de dados de pesquisa, sendo alguns modelos de ciclo de vida de dados em geral e outros modelos de curadoria de dados. No entanto, podem ser aplicados a dados de pesquisa.

Embora existam várias representações de ciclo de vida para dados de pesquisa, cada um com suas especificidades e objetivos, Sayão e Sales (2015) apresentam os modelos identificados por Ball (2012) para o Projeto REDm-MED<sup>8</sup>, descrevendo-os como sendo os modelos referência para pesquisadores, bibliotecários e gestores. Os modelos identificados são: o **DCC Lifecycle Model**; o **Dataone Data Lifecycle**; o **DDI Combined Lifecycle Model**; e o **UK Data Archive Data Lifecycle**, descritos a seguir.

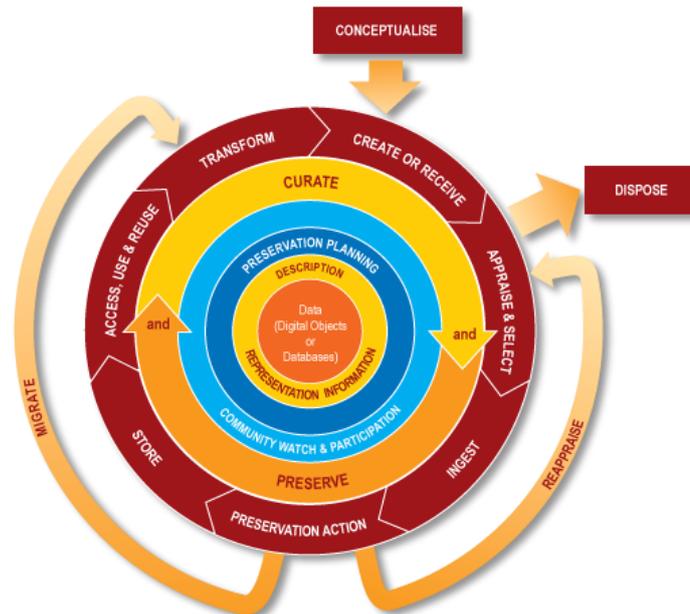
O modelo desenvolvido pelo Digital Curation Centre (DCC), o **Curation Lifecycle Model**, é um modelo em evolução que emprega uma abordagem que, segundo Silva e Siebra (2017, p. 5), “[...] busca assegurar que todas as ações necessárias sejam identificadas, planejadas e implementadas na sequência correta.” As autoras também salientam que esse é um modelo de natureza genérica, servindo como um indicativo, e que nem todas as etapas estarão presentes para todas as instituições. Sendo assim, deve ser adequado às necessidades e realidade destas.

De acordo com Higgins (2008), o modelo desenvolvido pelo DCC (Figura 1) identifica três níveis de ações: o primeiro nível envolve ações de curadoria que são identificadas em todo o ciclo de vida digital; o segundo nível compreende as ações que devem ser executadas continuamente sempre que um objeto digital precise ser curado; e o terceiro abrange as ações ocasionais, que serão executadas conforme seja necessário.

---

<sup>8</sup> Projeto de Gestão de Dados de Pesquisa para Departamentos de Engenharia Mecânica (REDm-MED) da Universidade de Bath.  
Disponível em: <https://researchportal.bath.ac.uk/files/206543/redm1rep120110ab10.pdf>

Figura 1 - Modelo de ciclo de vida da curadoria de dados formulado pela DCC



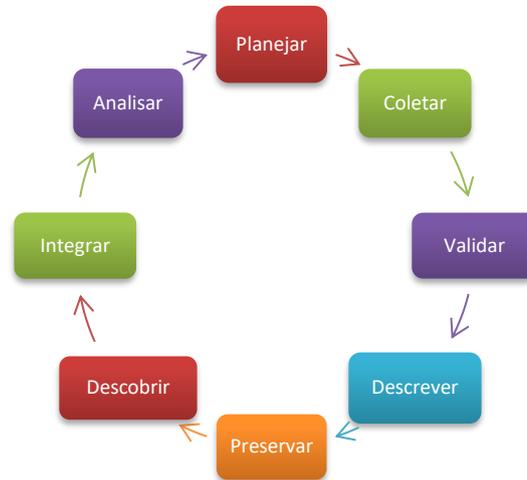
Fonte: HIGGINS (2008).

Observando a Figura 1, as seguintes atividades são identificadas dentro dos três níveis de ações:

- a) **ações de curadoria:** ações localizadas no centro no círculo, envolvendo atividades de descrição e representação da informação; o planejamento da preservação; o acompanhamento e participação da comunidade; e a curadoria e preservação;
- b) **ações sequenciais:** ações que envolvem as atividades de conceitualização; criação e (ou) recebimento; avaliação e seleção; arquivamento; ações de preservação; armazenamento; acesso, uso e reuso; e transformação;
- c) **ações ocasionais:** ações que envolvem as atividades de reapreciação, migração e descarte. A reapreciação consiste na análise de objetos digitais que não foram validados anteriormente. A migração/transferência consiste em transferir objetos para um arquivo morto ou repositório digital confiável, seguindo novamente as orientações, políticas e requisitos legais documentados. O descarte, por sua vez, consiste em liberar os sistemas de objetos digitais não selecionados para curadoria e preservação a longo prazo, observando questões legais de destruição segura desses objetos.

O modelo idealizado pela Data Observation Network for Earth, o **DataONE**, destacado na Figura 2, foi desenvolvido buscando disseminar boas práticas de gestão e preservação dos dados para o seu efetivo uso e reutilização.

Figura 2 – Modelo de ciclo de vida formulado pela DataONE



**Fonte:** Traduzido e adaptado de DataONE.

Analisado por Sayão e Sales (2015); Bertin, Visoli e Drucker (2017), o modelo consiste em oito etapas de ações, apresentadas a seguir:

Etapa 1 – **Planejamento**: consiste na descrição de como os dados serão compilados, gerenciados e disponibilizados para acesso ao longo do seu ciclo de vida;

Etapa 2 – **Coleta**: corresponde à fase em que se define a estratégia de coleta e arquivamento dos dados, para que sejam legíveis e interoperáveis;

Etapa 3 – **Validação**: consiste na definição dos processos que garantirão a qualidade dos dados através de verificações e inspeções;

Etapa 4 – **Descrição**: corresponde à fase em que os dados serão descritos de forma precisa e completa, usando os padrões de metadados apropriados;

Etapa 5 – **Preservação**: consiste na definição de como os dados serão preservados em curto e longo prazo, bem como o local de preservação apropriado (*data center*);

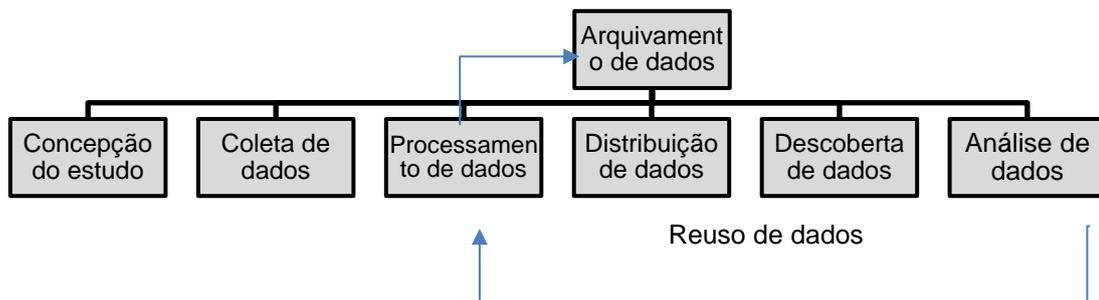
Etapa 6 – **Descoberta**: corresponde à etapa em que dados potencialmente úteis são localizados e obtidos juntamente às informações relevantes sobre os dados (metadados);

Etapa 7 – **Integração**: equivale à fase em que dados de fontes diferentes são combinados para formar um conjunto homogêneo de dados que pode ser facilmente analisado;

Etapa 8 – **Análise**: corresponde ao momento em que os dados reunidos são analisados de acordo com distintos objetivos e por diferentes pesquisadores, instigando novos questionamentos e levando à formulação de um novo ciclo de planejamento e etapas.

Outro modelo tido como referência é o DDI Combined Lifecycle Model, desenvolvido pela Data Documentation Initiative (DDI) especialmente para a área das Ciências Sociais. Na maioria das vezes, é um modelo de ciclo de vida linear, mas pode apresentar caminhos alternativos ao longo do ciclo, oportunizando espaço para *feedback*. O modelo apresentado na Figura 3 é descrito a partir da análise de Ball (2012) em oito etapas:

Figura 3 – Modelo de ciclo de vida dos dados formulado pela DDI



Fonte: Traduzido e adaptado de Ball (2012).

Etapa 1 - **Concepção da pesquisa**: envolve estágio inicial da pesquisa, com a escolha da questão de pesquisa, metodologia para coleta, processamento e análise dos dados. É nessa etapa que o pesquisador também indicará as relações existentes entre os produtos de dados de pesquisa;

Etapa 2 - **Coleta de dados**: consiste na distinção das fontes de dados primárias e secundárias de coleta de dados;

Etapa 3 - **Processamento de dados**: envolve o processamento e análise de dados para produção de dados de saída, objetivando responder à questão de pesquisa. Esses dados de saída podem ser registrados em formato legível por máquina ou em formato legível por seres humanos, a exemplo de um relatório técnico;

Etapa 4 - **Arquivamento de dados**: equivale à transmissão dos dados para um arquivo que garanta a sua preservação e de seus metadados. É nesse momento que se agrega valor aos dados devido a sua preservação ao longo do tempo;

Etapa 5 - **Distribuição de dados**: consiste na distribuição dos dados aos usuários diretamente, por meio de bibliotecas ou através de arquivos de dados;

Etapa 6 - **Descoberta de dados**: envolve a divulgação dos dados por meio de publicações, *sites* ou serviços *on-line*;

Etapa 7 - **Análise de dados**: consiste na utilização dos dados por outros pesquisadores;

Etapa 8 - **Reuso de dados**: envolve o uso dos dados em estruturas de pesquisa diferentes, combinando com outros conjuntos de dados similares.

O modelo de ciclo de vida de dados desenvolvido pelo Arquivo de Dados do Reino Unido, o **UK Data Archive Data Lifecycle**, trata da importância de um armazenamento de dados organizado, bem documentado, preservado e compartilhado para a promoção da investigação científica e aumento das oportunidades de aprendizado e inovação (UK DATA ARCHIVE, 2012). O UK Data Archive Data Lifecycle consiste em seis etapas de ações, como pode ser visto na Figura 4:

Figura 4 - Modelo de ciclo de vida dos dados de pesquisa formulado pela UK Data Archive Data



Fonte: Traduzido e adaptado de UK Data Archive (2012).

Etapa 1 – **Planejamento**: consiste na concepção da gestão e armazenamento dos dados, observando seus formatos;

Etapa 2 - **Coleta de dados**: corresponde à coleta de dados com atribuição de metadados e aquisição de dados;

Etapa 3 - **Processamento e análise de dados**: corresponde à etapa em que os dados são processados e analisados através da digitalização, transcrição e tradução de dados. Também é a fase de checagem, validação, descrição, documentação, gestão, armazenamento e interpretação dos dados;

Etapa 4 - **Publicação e compartilhamento**: equivale à etapa onde são observadas questões sobre direitos autorais, criação de documentação de orientação à acesso, publicação e compartilhamento de dados;

Etapa 5 - **Preservação dos dados**: corresponde à migração de dados para um formato mais adequado, ao armazenamento e às estratégias de segurança através de *backup* dos dados, acompanhado da documentação de preservação e curadoria dos dados; e

Etapa 6 - **Reutilização de dados**: consiste na descoberta dos dados por outros pesquisadores, possibilitando que correções sejam feitas em pesquisas anteriores.

Diante dos modelos aqui expostos, pode-se observar que, embora os modelos apresentem suas particularidades, carregam similaridades entre si no entendimento e fluxo das etapas a serem seguidas na gestão dos objetos digitais.

Na seção a seguir, são abordados os temas: acesso aos dados de pesquisa, licenças de uso dos dados e privacidade no uso de dados sensíveis.

### 2.2.2 Acesso, licenciamento, embargo e privacidade de dados

O direito de acesso à informação é um direito fundamental de todos os indivíduos. Tendo por base o princípio da dignidade humana, encontra-se previsto tanto na Declaração Universal dos Direitos Humanos (DUDH) quanto na Constituição Federal Brasileira de 1988 (CF). Contudo, segundo Nascimento (2019), embora seja um direito fundamental, é preciso ter-se a noção de que, justamente por essa razão, não se trata de um direito absoluto, sendo necessário ponderar sua relação com outros direitos igualmente fundamentais.

O movimento Ciência Aberta vai ao encontro da DUDH e da CF ao preconizar a abertura da informação gerada nas atividades científicas nas diferentes áreas de domínio, envolvendo distintas instâncias de abertura da ciência, além do acesso à informação, disponibilizada pela administração pública e representada na Lei nº 12.527, de 2011, Lei de Acesso à Informação (LAI). Embora a LAI não trate especificamente de acesso à informação científica representada por dados de pesquisa, em seu *Art. nº4*, abrange-os, ao considerar informação como “dados, processados ou não, que podem ser utilizados para produção e transmissão de conhecimento, contidos em qualquer meio, suporte ou formato” (BRASIL, 2011, *online*). Nascimento (2019) também aborda essa questão ao referir que, embora nos diversos campos do conhecimento as categorias “informação” e “dado” sejam distintas e apresentem especificidades, do ponto de vista da normativa jurídica, equivalem-se, sendo sinônimas.

Embora a DUDH e a CF garantam que o acesso à informação é um direito fundamental a todos e a LAI ampare legalmente esse direito, nem todas as informações podem ser disponibilizadas publicamente, necessitando de restrições ao seu acesso, devido ao seu conteúdo sigiloso. Segundo Nascimento (2019), pode-se classificar as hipóteses de sigilo informacional em dois grandes grupos: **sigilo imposto pela LAI**, o qual envolve informações pessoais, de segurança e de apoio à tomada de decisões administrativas e informações que podem ser classificadas como reservadas, secretas e ultrassecretas; e **sigilo atribuído por outras hipóteses legais**, envolvendo informações que estão representadas em diferentes dispositivos legais e necessitam de sigilo por questões: fiscais; contábeis; contratuais; empresariais; de propriedade intelectual; de patentes; de segredo industrial ou comercial; e de segredo judicial. A aplicação do sigilo não visa proteger a informação, mas sim regula a proteção de um direito, objeto ou valor, os quais estarão em risco caso informações relacionadas a eles sejam indiscriminadamente acessadas (NASCIMENTO, 2019).

Recentemente, ao entrar para o rol de países que possuem bases normativas para a proteção de dados pessoais, o Brasil promulgou a Lei nº 13.709, de 2018, chamada Lei Geral de Proteção de Dados (LGPD), que determina quais os direitos que os indivíduos terão sobre seus dados, estabelecendo critérios e requisitos que empresas e agentes públicos deverão seguir no tratamento desses dados (DONEDA, 2019).

A utilização de dados pessoais tem crescido em diferentes áreas, fazendo-se necessário o resguardo da titularidade destes na utilização em pesquisas científicas. Dessa forma, segundo Doneda (2019), a LGPD proporciona uma maior segurança na utilização desses dados, norteadando seu uso e dirimindo dúvidas quanto às normativas a serem aplicadas. Ainda de acordo com Doneda (2019, aula 6), a LGPD “proporciona ao titular uma série de direitos para que possa conhecer e administrar livremente como seus dados são utilizados e, ao mesmo tempo, criar uma base legal para que a sua utilização se dê com segurança.”

É importante salientar que a LGPD não se aplica a dados que não identifiquem o titular, ou seja, dados anonimizados, e sim a dados pessoais, os quais são:

[...] aqueles que podem ser relacionados a uma pessoa, independentemente de sua natureza ou qualquer outro atributo seu. Assim, qualquer informação que esteja associada a uma pessoa será um dado pessoal desta, independente de eventuais qualificações posteriores (se o dado é referente à saúde, à sua qualificação pessoal, se é um dado cadastral ou outro). A noção de dado pessoal, portanto, é muito ampla. No entanto, o dado pessoal não é, por si só, sigiloso ou confidencial, porém apenas objeto de algum tipo de controle pelo titular: dados pessoais podem, inclusive, ser de acesso público em determinadas circunstâncias. (DONEDA, 2019, AULA 6)

Em seu *Art. 13*, a LGPD também estabelece regras que possibilitam que dados pessoais sejam utilizados em pesquisas em saúde pública, utilizando-se de recursos como a pseudonimização, em que os identificadores dos dados são removidos e lhes é atribuído um código chave, o que proporciona maior segurança aos indivíduos envolvidos.

Por ser um movimento emergente, a Ciência Aberta, com alguns de seus aspectos ainda em processo de adequação a normativas relacionadas a direitos e garantias individuais e coletivas, necessita que questões quanto à titularidade de direitos, licenciamentos e restrições sejam trazidas à luz das normativas legais, a fim de evitar a insegurança jurídica entre os envolvidos (DONEDA, 2019).

O recurso de utilização da Lei de Direitos Autorais (Lei 9.610/98) é uma das várias formas de proteção de direitos relativos à criatividade humana, os quais estão sob o guarda-chuva do conceito de Propriedade Intelectual. De acordo com Souza (2019), alguns desses direitos contam com legislação própria para atender suas especificidades, dentre eles:

- a) os **Direitos Conexos**, igualmente protegidos pela Lei 9.610/98, abrangem as interpretações e execuções musicais;

- b) os **Segredos de Negócios**, que abrangem informações confidenciais que envolvam negócio e têm sua violação prevista na Lei de Propriedade Industrial (Lei 9.279/96);
- c) o direito sobre **Marcas**, o qual é também protegido pela Lei de Propriedade Industrial;
- d) o direito de proteção a **Patentes**, o qual envolve a proteção a invenções e modelos de utilidade, sendo também protegido pela Lei 9.279/96; e
- e) os direitos de proteção sobre **Software**, protegidos pela Lei 9.609/98, a qual abrange questões relacionadas a computadores, aplicativos, jogos eletrônicos e bancos de dados digitais, sendo uma subsidiária à Lei de Direitos Autorais e completando a proteção em casos de omissão da própria legislação.

No contexto mundial, o surgimento do direito de autor se deu de forma lenta e progressiva. Cada país elaborou suas normativas com suas especificidades, porém mantendo a integridade intelectual dos autores (MONTEIRO, 2017). No Brasil não foi diferente: a primeira iniciativa de proteção aos direitos autorais só se deu em 1827, sendo de caráter restritivo aos professores dos cursos de Direito da época. Posteriormente, várias iniciativas foram feitas, mas somente em 1998 foi aprovada a Lei Brasileira de Direitos Autorais, Lei 9.610.

O direito autoral apresenta-se sob duas vertentes de garantias ao autor: a primeira refere-se ao **direito moral**, que garante a citação de paternidade das obras, convertendo-se em um direito inalienável e imprescritível; e a segunda é representada pelo **direito patrimonial**, o qual tem o propósito de regular as relações jurídicas envolvendo as diferentes possibilidades de exploração econômica das obras. Contudo, “[...] mesmo as obras protegidas por direitos autorais sofrem restrições obrigatórias à exclusividade e impostas em razão do interesse público”, como em casos do “direito de acesso à cultura, à educação, ao conhecimento e à liberdade de expressão” (SOUZA, 2019, AULA 7).

Quanto aos dados de pesquisa, a garantia de preservação de sua autoridade requer cuidados e práticas que vão além do direito autoral, uma vez que o seu compartilhamento e reuso exigem ações de infraestrutura, curadoria e preservação, passando por questões legais relacionadas a permissões, licenciamento, embargo e privacidade de dados, questões estas que devem estar explicitadas no plano de gestão de dados do pesquisador, bem como nas políticas dos repositórios.

A heterogeneidade dos dados cauda longa traz maior complexidade ao compartilhamento e reuso desses dados, uma vez que, para diferentes tipos de dados e coleções de dados, pode-se ter diferentes tipos de permissões de acordo com suas peculiaridades quanto a sua tipologia, proveniência e licenças de uso. De acordo com Freund, Sembay e Macedo (2019, p. 815), “[...] a proveniência de dados vem se tornando cada vez mais presente no ambiente científico, tanto para garantir a origem dos dados como para avaliar a sua acurácia.”

Deve-se considerar também a questão dos dados sensíveis, os quais envolvem dados pessoais, que permitiriam a identificação de indivíduos a partir do acesso a esses dados; dos dados confidenciais, que abarcam dados que não são de domínio público e cujo acesso pode ocasionar interferência em negócios, saúde e questões políticas; e dos dados pessoais sensíveis, que envolvem informações acerca de questões de raça, sexualidade, crença e religião, doenças físicas ou mentais, etc. (SALES; SAYÃO, 2020, *slide 92*).

Embora o movimento de abertura dos dados de pesquisa preconize o livre acesso aos dados para que eles possam ser usados, reutilizados e distribuídos sem nenhum impedimento, seja legal, tecnológico ou social, alguns dados podem requerer, ainda assim, algum tipo de restrição ao seu acesso nos repositórios. De acordo com Repositórios Científicos de Acesso Aberto de Portugal (2021), nos repositórios institucionais existem quatro tipos de direitos de acesso aos documentos: o **acesso Aberto** (*Open Access*), o **acesso Fechado** (*Closed Access*), o **acesso Restrito** (*Restricted Access*) e o **acesso Embargado** (*Embargoed Access*).

No acesso Aberto, o texto integral do documento é disponibilizado *on-line* de forma direta e imediata, sem nenhum tipo de restrições de acesso, permitindo ações de leitura, pesquisa, *download* e impressão, independentemente de sua localização geográfica. Já o acesso Fechado, ao contrário, não permite o acesso direto e imediato ao documento, somente disponibiliza acesso aos seus metadados. Nessa situação, somente o administrador do repositório institucional terá acesso ao documento. O acesso Restrito, por sua vez, permite o acesso direto e imediato somente a um grupo restrito de usuários, não sendo possível ao público em geral. Nesse caso, o acesso pode ser permitido, por exemplo, apenas para usuários com vínculo institucional por meio de identificação de usuário ou endereço de IP. Por fim, o acesso Embargado impossibilita que o texto integral do documento seja acessado por um período determinado. Esse período estará mencionado nos metadados do registro. Ainda

conforme Repositórios Científicos de Acesso Aberto de Portugal (2021), nos acessos **Embargado**, **Fechado** ou **Restrito** há a possibilidade de os usuários solicitarem acesso ao documento mediante a funcionalidade dos repositórios “Solicitar cópia ao autor”. Se permitido, recebem uma cópia do texto integral do documento via *e-mail*.

Considerando esses aspectos, as normativas de uso dos serviços dos repositórios deverão explicitar as condições e regras apropriadas para: o depósito de dados, acesso aos dados, uso, o embargo e privacidade dos dados e as recomendações e boas práticas de uso dos dados.

De acordo com Monteiro (2017), ao proporcionarem novas possibilidades por meio da sua reutilização, o avanço na ciência e o grande aumento na geração de dados científicos requerem um equilíbrio jurídico entre os autores detentores dos direitos autorais e o usuário final. Essa relação encontra amparo nas licenças abertas de uso, que, ao assegurar a publicação de dados sem ferir os direitos autorais, transferem o direito patrimonial a outrem, trazendo clareza quanto à utilização de conteúdos. Assim, oferecem segurança aos usuários na utilização dos conjuntos de dados (MONTEIRO, 2017).

As licenças abertas para dados devem garantir que os dados possam ser modificados e compartilhados sem imposição de restrições quanto ao seu acesso, propósitos ou intenções, bem como não devem discriminar ou excluir grupos ou entidades a esse contexto de reuso e partilha. A escolha da licença a ser atribuída dependerá da intenção dos criadores quanto à finalidade do compartilhamento do seu trabalho e das expectativas em relação ao seu uso por terceiros (OPEN KNOWLEDGE PORTUGAL, [20--]).

De acordo com Open Knowledge Portugal (20--), existem três abordagens possíveis para a atribuição de licenças abertas para dados: as licenças **copyleft**, que permitem atribuição e partilha dos dados nos mesmos termos; as licenças **copyfree**, as quais dispensam a premissa da partilha nos mesmos termos, estipulando apenas a obrigação de atribuição da autoria original; e as licenças **domínio público**, em que os criadores abdicam de seus direitos autorais.

Dentre essas três abordagens de licenças abertas, temos o exemplo das licenças **Creative Commons (CC)** e **Open Data Commons (ODC)**, apresentadas a seguir.

O Creative Commons (CC) é uma organização não governamental internacional sem fins lucrativos, que visa promover o compartilhamento do

conhecimento, da cultura e da criatividade por meio da atribuição de licenças públicas e ferramentas de domínio público. A organização conta com afiliados ao redor do mundo para garantir o funcionamento e a conscientização do uso de suas licenças, as quais oferecem aos criadores individuais e instituições em todo o mundo uma forma simples, padronizada e gratuita de conceder permissões em termos mais flexíveis de uso de trabalhos criativos e acadêmicos, garantindo atribuição adequada e permitindo, dessa forma, que outras pessoas copiem, distribuam e façam uso dessas obras de acordo com as licenças atribuídas pelos autores (CREATIVE COMMONS, 2019).

As licenças CC foram sendo aprimoradas desde seu surgimento em 2001 e, atualmente, encontram-se em sua versão 4.0, que foi lançada em novembro de 2013. Embora todas as versões CC anteriores sejam válidas, é recomendado que se use a versão 4.0, visto que ela incorpora as licenças mais modernas com abrangência internacional (VALENTE; HOUANG, 2020).

O projeto CC oferece um conjunto de seis tipos de diferentes licenças que permitem maior flexibilidade na utilização de obras protegidas por direitos autorais, contendo atribuições menos permissivas e outras mais permissivas. Contudo é necessária a combinação de quatro características para a montagem da licença a ser utilizada: **Atribuição (BY)**, que é comum a todas as licenças CC e exige que todos os materiais usados e compartilhados tenham sua autoria identificada; **NãoComercial (NC)**, que proíbe o uso comercial do material; **NãoDerivados (ND)**, que permite somente o uso da obra no original, sem alterações do material; e **Compartilhalgual (SA)**, que exige que as obras derivadas do material sejam lançadas sob a mesma licença da obra original.

Embora as licenças CC sejam formadas pelas recombinações dessas quatro características, existe uma única exceção à regra: as características **NãoDerivados (ND)** e **Compartilhalgual (SA)** não podem ser combinadas entre si, visto que a Compartilhalgual estabelece condições de uso da mesma licença da obra original em obras derivadas, enquanto a NãoDerivados impede a criação de obras adaptadas (VALENTE; HOUANG, 2020).

A seguir, expõem-se as características das licenças CC, iniciando com a atribuição (BY) e seguindo com as outras cinco licenças combinadas:

- a) **Creative Commons (CC-BY)**  - é a licença mais flexível de todas as licenças disponíveis, pois permite aos utilizadores a distribuição, remixagem, adaptação e criação a partir do trabalho original, mesmo que seja para fins comerciais, desde que o devido crédito pela criação original seja atribuído.
- b) **Creative Commons (CC-BY-SA)**  - permite aos reutilizadores a distribuição, remixagem, adaptação e desenvolvimento de material em qualquer meio ou formato a partir do trabalho original, desde que seja atribuído o devido crédito ao criador, e as adaptações devem ser compartilhadas nos mesmos termos. A licença também permite uso comercial.
- c) **Creative Commons (CC-BY-NC)**  - permite aos reutilizadores a distribuição, remixagem, adaptação e desenvolvimento de material em qualquer meio ou formato a partir do trabalho original apenas para fins não comerciais da obra, exigindo concessão de crédito ao criador.
- d) **Creative Commons (CC-BY-NC-SA)**  - permite aos reutilizadores, desde que concedido crédito ao criador, a distribuição, remixagem, adaptação e construção do material em qualquer meio ou formato apenas para fins não comerciais e apenas enquanto a atribuição for dada ao criador. Adaptações devem ser compartilhadas sob os mesmos termos.
- e) **Creative Commons (CC-BY-ND)**  - permite aos usuários a reutilização, cópia e distribuição de material em qualquer mídia ou formato somente enquanto a atribuição for dada ao criador, não permitindo derivados ou adaptações do trabalho. Permite uso comercial.
- f) **Creative Commons (CC-BY-NC-ND)**  - permite cópia e distribuição do material em qualquer meio ou formato, mas não permite derivações ou adaptações. É permitida apenas para fins não comerciais e apenas enquanto a atribuição for dada ao criador.

A Creative Commons também disponibiliza aos usuários uma ferramenta pública de dedicação ao **domínio público (CC0)**, ou seja, “todos os direitos concedidos”. Isso permite aos criadores abdicarem de seus direitos autorais e colocarem seus trabalhos em domínio público em todo o mundo, permitindo aos reutilizadores a distribuição, remixagem, adaptação e desenvolvimento da obra em qualquer meio ou formato, sem restrições. A organização também disponibiliza a

Marca de Domínio Público, a qual permite que usuários da internet marquem um trabalho para indicar que ele se encontra no domínio público (CREATIVE COMMONS, 2019).

As licenças CC podem ser aplicadas a qualquer tipo de obra, no entanto não se recomenda seu uso para licenciamento de *software* e *hardware* de computadores. Também não devem ser utilizadas em obras que já estejam em domínio público. Recomenda-se, nesse caso, a utilização da Marca de Domínio Público.

O projeto **Open Data Commons (ODC)** é hospedado pela Open Knowledge Foundation e gerenciado por um conselho consultivo presidido por seu idealizador, Jordan Hatcher, com a finalidade de fornecer soluções legais para dados abertos. A primeira licença a ser lançada pela ODC, em 2008, foi a **Public Domain Dedication and License (PDDL)**. A partir da PDDL, a ODC disponibiliza três licenças:

a) **Domínio Público e Licença de Dedicção Open Data Commons (PDDL):**

semelhante à ferramenta pública de dedicação ao domínio público (CC0), há renúncia de todos os direitos, não impondo restrições ao uso de dados ou banco de dados e tornando o usuário livre para criar, adaptar e compartilhar conteúdo;

b) **Licença de atribuição Open Data Commons (ODC-BY):** licença de atribuição para dados e bancos de dados, semelhante à licença de atribuição *Creative Commons* (CC-BY), permitindo cópia, distribuição, produção, uso, modificação e construção sobre os dados ou banco de dados, desde que se forneça a atribuição adequada (MONTEIRO, 2017);

c) **Licença de banco de dados aberto Open Data Commons (ODC-ODbL):**

direcionada ao compartilhamento de atribuição semelhante para dados e bancos de dados. Semelhante à licença CC BY-SA, permite cópia, distribuição, uso e produção de trabalhos a partir dos dados ou banco de dados, modificando, transformando e construindo sobre os dados ou banco de dados, desde que eles forneçam a atribuição adequada e que todos os trabalhos derivados sejam disponibilizados sob essa mesma licença (MONTEIRO, 2017).

O Open Data Commons disponibiliza para consulta quadros com a indicação de licenças que estão em conformidade com a definição de licenciamento aberto e as licenças em conformidade e recomendadas para atribuição, bem como disponibiliza um quadro indicando licenças que estão em conformidade com o licenciamento aberto, mas que não atendem aos requisitos de reutilização ou foram substituídas por

novas versões. Disponibiliza igualmente normas comunitárias de uso das ODC, por meio das “melhores práticas” para sua comunidade de usuários, estimulando o compartilhamento de trabalhos com a licença PDDL.

De acordo com o projeto ODC, o uso de uma licença de Domínio Público Licença de Dedicção Open Data Commons (PDDL) evidencia os limites do licenciamento de dados, haja vista ser um problema inclusive para advogados experientes determinarem onde os direitos começam e terminam, mesmo em licenças Creative Commons. Assim, ao utilizar a PDDL para quaisquer alterações e adições feitas, o trabalho permanecerá aberto e gratuito para todos.

As licenças projetadas para conteúdo não são as mais adequadas ao licenciamento de dados e bases de dados. Por isso, tais iniciativas devem ser vistas com cautela. Com exceção da atribuição CC0 de domínio público, as outras licenças CC provavelmente não estariam capacitadas a corrigir problemas específicos ao licenciamento de dados e bancos de dados, uma vez que os direitos em bases de dados apresentam direitos adicionais de propriedade intelectual, diferenciando-se dos direitos relativos à base de dados e ao seu conteúdo (OPEN KNOWLEDGE FOUNDATION, [20--]).

Na seção a seguir, são apresentados os diferentes tipos de repositórios de dados de pesquisa e suas características conforme seu propósito de atuação.

### 2.3 REPOSITÓRIOS DE DADOS DE PESQUISA (RDPs)

A pesquisa científica contemporânea, com o intuito de novas descobertas, ao produzir e utilizar intensamente dados de pesquisa, apresenta desafios e novas oportunidades. Os desafios consistem na inclusão da otimização no uso e reuso desses dados, e as oportunidades dizem respeito ao desenvolvimento de infraestruturas apropriadas ao seu armazenamento, gestão e preservação (SAYÃO; SALES, 2016; MONTEIRO, 2017). Para que esses dados atendam às premissas de acesso, compartilhamento e reuso, eles requerem um ambiente digital que contemple uma infraestrutura tecnológica e de gestão ao longo do seu ciclo de vida, assegurando-lhes conformidade com os princípios FAIR e permitindo, assim, que sejam localizáveis, acessíveis, interoperáveis e reusáveis.

No centro da busca por sistemas informacionais com infraestrutura tecnológica e gerencial que apoie os processos de gestão dos dados gerados nas atividades

científicas, destacam-se os RDPs. Esses repositórios são mantidos por ações que viabilizam o armazenamento de dados e incorporam uma infraestrutura necessária para a administração e preservação dos dados, assegurando, por meio de um plano de gestão e curadoria, que os dados sejam acessados e reutilizados ao longo do tempo (SAYÃO; SALES, 2016b, SANCHEZ; VIDOTTI; VECHIATO, 2017; MONTEIRO, 2017). De acordo com Monteiro (2017, p. 16), o conjunto de ações empreendidas em um repositório, ao otimizarem a recuperação de dados, amplia suas potencialidades de reuso entre os pesquisadores e, “dessa forma, agiliza os processos de investigação e, conseqüentemente, o avanço na ciência.”

Além da infraestrutura tecnológica, o desenvolvimento de repositórios deve estabelecer critérios de perenidade e confiabilidade a estes, garantindo, por longo tempo, a qualidade e a permanência dos seus depósitos. Esses repositórios são reconhecidos como confiáveis por agências certificadoras, as quais atestam o atendimento às questões técnicas de gerenciamento do repositório e dos objetos digitais, questões organizacionais de infraestrutura e de tecnologia, bem como questões de segurança. Como exemplo de agências certificadoras, tem-se a Data Seal Approval, a Core Trust Seal e Nestor e a norma ISO 16363:2012, que estabelece parâmetros a serem seguidos para garantir confiabilidade (PAVÃO *et al.*, 2018a).

Direcionados a apoiar a preservação, a integridade e a difusão dos dados científicos e tendo por princípio assegurar o acesso contínuo e aberto a esses dados, “os repositórios assumiram o centro das atividades de procura para a produção do conhecimento”, chegando a tempo de atender a demanda gerada pela grande quantidade de dados nascidos digitalmente (SILVA, 2019, p. 89). Mesmo que os pesquisadores compartilhem cada vez mais seus dados em revistas científicas, por meio de arquivos vinculados aos artigos, percebe-se uma mudança nas políticas editoriais para que os dados de pesquisa sejam depositados em repositórios. De acordo com Silva (2019, p. 89), essa mudança é um reconhecimento por parte dos editores de que os repositórios, além de disporem “[...] das características ideais para uma boa organização, preservação e difusão dos dados científicos, ainda facilitam o cumprimento dos requisitos da maioria dos organismos de financiamento.”

O armazenamento de dados de pesquisa em RDPs traz diversos benefícios aos autores, a outros pesquisadores, à pesquisa científica e à própria ciência. Esses benefícios foram elencados por Sayão e Sales (2016a), a saber:

- a) ampliam a visibilidade dos dados de pesquisa;

- b) possibilitam o compartilhamento de dados;
- c) permitem que sejam dados crédito e reconhecimento aos autores dos dados;
- d) oferecem ambiente digital adequado à preservação dos dados;
- e) contribuem para a formação da memória científica institucional referente aos dados de pesquisa e contribuem para o princípio da transparência;
- f) oferecem segurança aos dados por meio de um sistema de armazenamento seguro;
- g) permitem que os dados estejam disponíveis *on-line* para que sejam “acessados, baixados, visualizados e processados por pessoas ou sistemas”;
- h) por meio da curadoria digital, “proporcionam um ambiente apropriado para os processos de avaliação, de adição de valor, reformatação, agregação e recriação de dados”;
- i) possibilita a criação de novos “serviços de informação para pesquisadores, gestores e financiadores de pesquisa a partir da análise e integração dos dados arquivados com fontes internas e externas à instituição”;
- j) aumentam o reuso dos dados, minimizando a duplicação de esforços, e “otimizam os investimentos na coleta e geração de dados”;
- k) permitem, por meio da interoperabilidade, a formação de redes de repositórios de dados e “abre a possibilidade de inserção dos repositórios de dados às redes interoperáveis definidas pelo padrão *Linked Data*”;
- l) tornam-se um indicador de qualidade e produtividade da instituição.

Os RDPs apresentam características específicas quanto ao seu propósito de atuação. De acordo como Sayão e Sales (2016), em relação às características dos repositórios, há certo consenso entre os autores que pesquisam sobre a temática, que caracterizam-nos em: repositórios institucionais, disciplinares, multidisciplinares e repositórios orientados por projetos.

Os **repositórios institucionais de dados de pesquisa** caracterizam-se por estarem vinculados e serem gerenciados por instituições acadêmicas, como universidades ou institutos de pesquisa, e são voltados para o arquivamento de dados geralmente provenientes unicamente das atividades de pesquisa dessas instituições. Segundo Sayão e Sales (2016), algumas vezes, as instituições estendem os modelos de dados dos seus repositórios institucionais de documentos para incluir dados de pesquisa, a exemplo do **Carpe Dien** do Instituto de Engenharia Nuclear (IEN/CNEN).

Porém, ainda de acordo com Sayão e Sales (2016), na maioria dos casos, os RDPs são plataformas independentes devido à necessidade de esquemas mais ricos de metadados, mas “[...] também podem estabelecer *links* entre os seus recursos e os recursos dos repositórios de *eprints*, estabelecendo novas formulações de publicações acadêmicas” (SALES, 2014).

Ainda de acordo com Sayão e Sales (2016), por vezes, os repositórios institucionais também são considerados multidisciplinares, considerando o amplo escopo das pesquisas desenvolvidas nas universidades e institutos de pesquisa. Ademais, “na qualidade de repositórios institucionais, cumprem também o papel de registrar a memória acadêmica da instituição circunscrita pela geração de dados” (SAYÃO; SALES, 2016, p. 102). Como exemplo, tem-se o **Edinburgh Data Share** (UK), repositório institucional e multidisciplinar da Universidade de Edinburgh voltado para os dados de pesquisa gerados pelas pesquisas da Universidade.

Os **repositórios disciplinares de dados de pesquisa** são repositórios direcionados para o arquivamento de domínios específicos de pesquisa, tendo-se como exemplos: o repositório **Bio Models Database**, voltado para arquivamento, descoberta e intercâmbio de modelos computacionais na área de biologia; o **GenBank**, base de dados de sequenciamento genético; e o **PANGAEA** – Data Publisher for Earth & Environmental Science –, biblioteca digital aberta direcionada ao armazenamento, publicação e distribuição de dados georeferenciados provenientes das pesquisas no campo da Ciência do Sistema Terrestre (SAYÃO; SALES, 2016). Ainda segundo os autores, a variedade e heterogeneidade dos repositórios disciplinares evidenciam-se, visto que refletem a multiplicidade de disciplinas e a diversidade de dados gerados mundialmente no contexto da pesquisa científica.

Os **repositórios multidisciplinares de dados de pesquisa** se caracterizam por reunirem coleções de dados coletados ou gerados por atividades de pesquisa nas diferentes áreas de conhecimento. Como observado anteriormente por Sayão e Sales (2016a), alguns repositórios institucionais de universidades, devido à interdisciplinaridade das pesquisas realizadas em seu âmbito, também se enquadram como repositório multidisciplinar. Como exemplos de repositórios multidisciplinares, tem-se o **Dryad** e o **Figshare**.

Já os **repositórios de dados de pesquisa orientados por projetos** fazem o arquivamento de coleções de dados resultantes de projetos de pesquisa ou resolução de problemas específicos. Pampel *et al.* (2013) sinalizam como exemplo dessa

categoria de repositório o Scientific Drilling Database (SDDB), administrado pelo Centro Alemão de Pesquisa em Geociências da GFZ, o qual fornece dados de perfuração, abertos e reusáveis, que são criados no âmbito do Programa de Perfuração Científica Continental (ICDP).

Quanto ao seu funcionamento, os repositórios podem adotar vários modelos que levem em consideração suas relações com os produtores e os detentores da informação. Várias instituições podem utilizar os serviços de um único repositório, fazendo com que este atue como seu repositório de dados institucional. A exemplo, tem-se o Data Archiving and Network Services (DANS), que agrega várias instituições com espaço de armazenamento para seus dados de pesquisa em um serviço denominado DataverseNL, e o Texas Data Repository (TDL) nos Estados Unidos, que reúne diversas universidades (PAVÃO *et al.*, 2018a).

Diante do exposto, entende-se que os repositórios dispõem das condições ideais para armazenamento, preservação e difusão dos dados de pesquisa, contribuindo também para o cumprimento das exigências das agências de financiamento para a gestão de dados pelos pesquisadores. Cabe aos pesquisadores, com base nas características específicas relacionadas ao propósito de atuação dos repositórios, determinar qual repositório é o mais apropriado para o depósito de seus dados de pesquisa.

A seção a seguir apresenta os procedimentos metodológicos utilizados para o presente estudo.

### 3 PROCEDIMENTOS METODOLÓGICOS

O presente estudo investigou práticas de acesso aberto a dados de pesquisa, instituídas em políticas de depósito, acesso e uso de dados de pesquisa em RDPs de universidades internacionais, com ênfase nas suas permissões, restrições e licenças. Dessa forma, este capítulo apresenta as abordagens e procedimentos metodológicos adotados para o desenvolvimento desta pesquisa.

A seguir, são apresentados o delineamento da pesquisa, o procedimento de identificação dos repositórios para análise de suas políticas, descrição da etapa de coleta e análise de dados.

#### 3.1 DELINEAMENTO DA PESQUISA

A presente pesquisa trata de um estudo de natureza aplicada, com abordagem qualitativa, não probabilística, de caráter descritivo, utilizando-se da pesquisa documental como procedimento metodológico. Ao buscar descrever os aspectos específicos das políticas de gestão de repositórios, o estudo está em conformidade com o pensamento de Patton (2007<sup>9</sup> *apud* Moreno 2018), que considera que uma pesquisa qualitativa produz relevantes descrições detalhadas e interpretações particularizadas de elementos que os moldam e são moldados por eles, assim como com o pensamento de Gil (2002, p. 42), para o qual um estudo descritivo “[...] tem por objetivo a descrição das características de determinada população ou fenômeno ou o estabelecimento de relações entre variáveis.”

A pesquisa considerou, para o seu *corpus*, as políticas de RDPs de universidades em âmbito internacional e, como recorte desse corpus, as 18 primeiras universidades elencadas no ranking *Webometrics.info* de 2020, buscando identificar seus repositórios de dados de pesquisa por meio de procedimento descrito a seguir.

---

<sup>9</sup> PATTON, M. Q. Qualitative Evaluation. In: LEWIS-BECK, M. S.; BRYMAN, A.; LIAO, T. F. (Eds.) **Encyclopedia of Social Science Research Methods**. Thousand Oaks, CA: SAGE Publication, 2007.

### 3.2 PROCEDIMENTO DE IDENTIFICAÇÃO DOS REPOSITÓRIOS

O presente estudo teve por *start* metodológico a pesquisa realizada por Monteiro e Sant’Ana (2018), na qual os autores buscavam investigar a presença de PGDs disponibilizados nos *websites* de RDPs de universidades para identificar possíveis benefícios gerados pelo seu uso. Na metodologia adotada, os autores, por meio do ranking *Webometrics.info* de 2016, identificaram dentre as 100 melhores universidades do mundo, 36 universidades que tinham repositórios digitais e disponibilizavam seus PGDs em seus *websites*. Esta pesquisa se diferencia do estudo de Monteiro e Sant’Ana (2018), uma vez que diferentemente dos autores, buscou-se analisar as políticas de RDPs.

A incipiência de iniciativas brasileiras na implantação de RDPs com a disponibilização de suas respectivas políticas de gestão corroborou para a decisão de seguir a metodologia adotada por Monteiro e Sant’Ana (2018), para identificação das universidades no presente estudo.

No entanto, para o presente estudo, após se considerar que a análise de aspectos muito específicos das políticas dos RDPs poderia tornar a apresentação de resultados imensamente repetitiva e cansativa ao leitor, entendeu-se ser suficiente pesquisar a metade do número de repositórios investigados por Monteiro e Sant’Ana (2018), tendo-se assim por recorte desta pesquisa, os repositórios das 18 primeiras universidades elencadas no *Webometrics.info* 2020.

A identificação dos repositórios das universidades deu-se em dois momentos: primeiro, buscou-se os *links* dos repositórios das universidades presentes no estudo de Monteiro e Sant’Ana (2018) e que permaneceram elencadas no Ranking de 2020; num segundo momento, por meio dos *weblinks* das universidades indicadas no Ranking, procurou-se identificar a localização de acesso aos repositórios das novas instituições incluídas no *Webometrics.info* de 2020 dentro dos *websites* das instituições.

Sendo assim, no Quadro 2 a seguir, estão elencadas as universidades para o presente estudo, com o nome dos respectivos repositórios e *websites*.

**Quadro 2 – Universidades selecionadas para pesquisa, nome do repositório e website**

Nº	País	Universidade	Repositório	Websites
1	EUA	Harvard University	Harvard Dataverse	<a href="https://dataverse.harvard.edu/">https://dataverse.harvard.edu/</a>
2	EUA	Stanford University	Stanford Digital Repository (SDR)	<a href="https://sdr.stanford.edu/">https://sdr.stanford.edu/</a>
3	EUA	Massachusetts Institute of Technology	DSpace@MIT	<a href="http://dspace.mit.edu/">http://dspace.mit.edu/</a>
4	EUA	University of California Berkeley	Data Repository [Berkeley]	<a href="https://researchdata.berkeley.edu/">https://researchdata.berkeley.edu/</a>
5	EUA	University of Washington	Research Works Archive	<a href="https://www.lib.washington.edu/scholpub/scholarly-publishing-services/researchworks">https://www.lib.washington.edu/scholpub/scholarly-publishing-services/researchworks</a>
6	EUA	University of Michigan	Deep Blue Data	<a href="https://deepblue.lib.umich.edu/data/?locale=en">https://deepblue.lib.umich.edu/data/?locale=en</a>
7	UK	University of Oxford	ORA-Data	<a href="http://researchdata.ox.ac.uk/">http://researchdata.ox.ac.uk/</a>
8	EUA	Johns Hopkins University	JHU Data Archive	<a href="https://archive.data.jhu.edu/">https://archive.data.jhu.edu/</a>
9	EUA	Columbia University New York	Columbia Academic Commons	<a href="https://academiccommons.columbia.edu/">https://academiccommons.columbia.edu/</a>
10	EUA	Cornell University	eCommons	<a href="http://guides.library.cornell.edu/ecommons">http://guides.library.cornell.edu/ecommons</a>
11	EUA	University of Pennsylvania	Scholarly Commons	<a href="https://repository.upenn.edu/">https://repository.upenn.edu/</a>
12	UK	University of Cambridge	Apollo	<a href="https://www.data.cam.ac.uk/repository">https://www.data.cam.ac.uk/repository</a>
13	EUA	University of California Los Angeles UCLA	Merritt	<a href="https://merritt.cdlib.org/">https://merritt.cdlib.org/</a>
14	EUA	Yale University	ISPS Data Archive	<a href="https://isps.yale.edu/research/data">https://isps.yale.edu/research/data</a>
15	EUA	University of Wisconsin Madison	MINDS@UW	<a href="http://researchdata.wisc.edu/">http://researchdata.wisc.edu/</a>
16	EUA	University of California San Diego	UC San Diego Library Digital Collections	<a href="https://library.ucsd.edu/dc">https://library.ucsd.edu/dc</a>
17	EUA	University of Minnesota System	DRUM	<a href="https://conservancy.umn.edu/handle/11299/166578">https://conservancy.umn.edu/handle/11299/166578</a>
18	UK	University College London	UCL Research Data Repository (RDR)	<a href="https://rdr.ucl.ac.uk/">https://rdr.ucl.ac.uk/</a>

Fonte: elaborado pela autora.

### 3.3 DESCRIÇÃO DA ETAPA DE COLETA E ANÁLISE DE DADOS

Para a coleta e posterior análise dos dados, considerando os objetivos propostos para o estudo, foram elaborados quadros para cada aspecto a ser identificado nas políticas dos repositórios. Cada quadro identifica os repositórios pesquisados e as informações coletadas em suas políticas referentes aos seguintes aspectos: **proveniência dos dados para depósito**, etapa em que se registrou a origem dos dados depositados nos repositórios; **tipos de acesso aos dados nos repositórios**, fase na qual se observou que alguns repositórios permitem acesso flexível aos dados; **restrições de acesso aos dados nos repositórios**, etapa em que se pôde identificar as restrições impostas ao acesso dos dados; **permissão para depósito de dados**, em que se observou as permissões explicitadas nas políticas para depósito de dados nos repositórios; **restrições para depósito de dados**, fase na qual se identificou as restrições para depósito de dados nos repositórios; **restrições aos tipos de dados para depósito**, etapa na qual observou-se se as políticas fazem alguma restrição a algum tipo específico de dados de pesquisa para depósito nos repositórios; **permissões e restrições para uso de dados baixados**, fase em que se buscou nas políticas a identificação de diretrizes para o uso dos dados baixados; **licenças de uso dos dados**, etapa se identificou dentre as políticas as orientações sobre a atribuição de licenças para o uso de dados; **privacidade no uso de dados pessoais**, fase em que se observou nas políticas analisadas como se dá a privacidade de informações sensíveis no uso de dados de pesquisa; **embargo no uso de dados**, etapa em que se buscou detectar se nas políticas consta a utilização do embargo como instrumento de restrição de acesso aos dados de pesquisa; e **responsabilidade no uso do repositório e dos dados**, fase em que se identificou nas políticas as responsabilidades atribuídas aos atores envolvidos no uso dos dados e dos serviços oferecidos pelos repositórios.

A coleta de informações nos repositórios se deu em três momentos dentro do período de 16 de maio a 23 de setembro de 2020. Na primeira etapa, buscou-se localizar todas as políticas nos *websites* dos repositórios e salvar toda documentação identificada em arquivos formato .DOCX e .PDF em pastas no Google Drive e em *pen drive*. As informações foram salvas em formato .DOCX quando dados dispersos no *site* foram compilados, e em formato .PDF os documentos já disponibilizados pelos

repositórios nesse formato. Nessa etapa, enviou-se um *e-mail* para as universidades cujas políticas disponibilizadas nos *sites* dos repositórios não puderam ser identificadas, ou tenham gerado alguma dúvida. Também se realizou um contato via *chat on-line* com uma das universidades. Na segunda etapa, iniciou-se uma compilação das informações coletadas para, então, na terceira etapa, realizar sua análise.

No decorrer da pesquisa, considerou-se necessário identificar a data de implementação dos repositórios pesquisados e, para tanto, entrou-se em contato com as universidades nas datas de 10 a 21 de janeiro de 2021 para obter essa informação, quando não foi possível localizá-la nos *websites* dos Repositórios. Os contatos realizados por *e-mail* e via *chat on-line* para obter informações sobre as políticas dos repositórios deram-se com a:

- **UC Berkeley**, contato via *chat on-line* em 19 de agosto de 2020, (<https://ucberkeley.libanswers.com/>);
- **University of Minnesota System**, contato com a Bibliotecária Jennifer Patiño, em 18/08/2020;
- **University of Cambridge**, contato com Peter Sutton-Long, Gerente de Serviços do Repositório, em 18/09/2020;
- **Johns Hopkins University**, contato com Marley Kalt, Consultor de Gerenciamento de Dados, em 23/09/2020;
- **University of California Los Angeles**, contato com Eric Lopatin, Gerente de produto de preservação digital da CDL/UC3, em 23/09/2020;
- **UC San Diego**, contato com Abigail Pennington, da Curadoria de Dados da Universidade, em 08/09/2020.

Já os contatos para obter informações sobre as datas de fundação dos repositórios foram realizados com a:

- **Columbia University New York**, Kathryn Pope, Gestora do Repositório Digital, em 12/01/2021.
- **US San Diego**, contato com Roger Smith, em 12/01/2021.
- **University of California Los Angeles**, contato com Eric Lopatin, Gerente de produto de preservação digital da CDL/UC3, em 12/01/2021.

- **UC Berkeley**, contato com Erin D. Foster, Líder do Serviço Programa de Gerenciamento de Dados de Pesquisa (RDM), UC Berkeley Libraries/ Research IT, em 12/01/2021.
- **ISPS Data Archive**, contato com Limor Peer, Diretor de Pesquisa Associado do ISPS, em 21/01/2021.

Também foi enviado um e-mail para a University of Washington, mas não houve retorno. Todas as etapas de coleta de informações foram registradas, conforme mostra o Quadro 3.

Quadro 3 – Controle da coleta de dados

DATA DA 1ª COLETA	DATA DA 2ª COLETA	DATA DE REGISTRO DAS INFORMAÇÕES COLETADAS	REPOSITÓRIOS
16/08/2020		Início em 25/08/2020	18_UCL Research Data Repository
17/08/2020	18/08/2020 e-mail do DRUM	05/09/2020	17_DRUM
17/08/2020	08/09/2020 e-mail do UCSan Diego	06/09/2020	16_UCSanDiego
17/08/2020		07/09/2020	15_MINDS@UW - Research Data Service (é o serviço de apoio)**
17/08/2020		07/09/2020	14_ISPS Data Archive
17/08/2020	12/09/2020 23/09/2020 e-mail de resposta do Merritt	09/09/2020	13_Merritt****
18/08/2020	18/09/2020 e-mail de resposta do Apollo	12/09/2020	12_Apollo_Data Repository [Cambridge] - O repositório institucional é o <b>Apollo</b> ***
18/08/2020		13/09/2020	11_Scholarly Commons
18/08/2020	13/09/2020	13/09/2020	10_eCommons
19/08/2020		13/09/2020	9_Columbia Academic Commons
19/08/2020	23/09/2020 e-mail de resposta do JHU	22/09/2020	8_JHU Data Archive****
19/08/2020		23/09/2020	7_ORA-Data - Oxford University Research Archive
16/05/2020			6_University of Washington
16/05/2020			5_Universityof Michigan
19/08/2020	19/08/2020 e-mail da UC Berkeley		4_Data Repository [Berkely] *
16/05/2020			3_Stanford University
16/05/2020			2_MIT
16/05/2020			1_Harvard University

**Notas:**

- \*E-mail recebido da UC Berkeley: “UC Berkeley uses Dryad as a data repository: <https://datadryad.org/>. The policies are linked at the bottom of the page. You can also find more information on UC Berkeley data management at <https://researchdata.berkeley.edu/>. (Em: 19/08/2020).
- \*\*E-mail recebido da Data & Digital Scholarship Librarian - University of Wisconsin-Madison: “You can find the policies for our MINDS@UW institutional repository at this [link](https://www.library.wisc.edu/digital-library-services/minds/forms-and-materials/) <https://www.library.wisc.edu/digital-library-services/minds/forms-and-materials/> (Em: 18/08/2020).
- \*\*\*Segundo e-mail, a página que responderá as perguntas sobre Política de uso de dados está em: <https://www.data.cam.ac.uk/funders> (E-mail de 18/09/2020).
- \*\*\*\*De acordo com e-mail da JHU, o *link* para as políticas é: <https://dataservices.library.jhu.edu/archiving/>
- \*\*\*\*\*De acordo com e-mail do Merritt, informações sobre termos de serviços do Dryad encontram-se em: <https://datadryad.org/stash/terms>

Fonte: elaborado pela autora.

Na etapa de análise dos dados coletados, procurou-se detectar, dentre todas as políticas dos repositórios disponibilizadas em seus *websites*, aquelas que abrangem os aspectos propostos nos objetivos específicos do presente estudo. Ou seja, buscou-se por políticas que abarcassem: permissões e restrições para depósito, acesso e uso de dados; o licenciamento no uso dos dados; a privacidade no uso de dados sensíveis; e o uso do embargo como instrumento restritivo de acesso aos dados. Também se buscou por políticas que discriminassem as responsabilidades no uso dos dados e dos serviços oferecidos pelos repositórios.

Ao se buscar identificar essas políticas, constatou-se, no entanto, que nos repositórios institucionais não exclusivos para dados de pesquisa, essas informações estavam mais dispersas, exigindo que as informações específicas para dados de pesquisa fossem filtradas. Contudo, em algumas políticas, essas informações foram apresentadas para conteúdos digitais no geral.

Para os repositórios que não possuem um serviço próprio para depósito e preservação de dados, além das políticas que constavam nos seus *sites*, analisou-se os termos de usos dos serviços contratados, como o Dryad.

Embora cada repositório apresente especificidades em relações as suas políticas, de uma forma geral, os aspectos pesquisados estavam descritos nas políticas referentes a: termos de uso; contrato de uso de dados; serviços ao depositante, termos de depósito ou política de envio de dados.

Quando necessário, devido à falta de informações referentes aos aspectos investigados nas políticas específicas para dados de pesquisa, também realizaram-se análises nas outras políticas dos repositórios.

Na seção seguinte, são apresentados os resultados, partindo de uma breve contextualização das universidades selecionadas para este estudo, assim como seus repositórios e as políticas identificadas em seus *websites*. Num segundo momento, são apresentadas as análises dos aspectos específicos das políticas dos repositórios.

## 4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Nesta seção, são apresentados os resultados e as análises dos dados encontrados sobre permissões, restrições e licenças para depósito, acesso e uso de dados de pesquisa em RDPs de universidades internacionais. A seção está organizada em três partes: a primeira traz uma breve apresentação das universidades e de seus repositórios com a indicação das políticas identificadas em seus *websites*; a segunda traz considerações sobre os repositórios; e a terceira tem como tema a análise dos aspectos específicos das políticas de RDPs indicados para este estudo e para atender os objetivos propostos.

### 4.1 APRESENTAÇÃO DAS UNIVERSIDADES E SEUS REPOSITÓRIOS

Nesta seção, faz-se uma breve apresentação e contextualização das universidades e de seus repositórios. Também são apresentadas as políticas dos RDPs identificadas em seus *websites*.

#### 4.1.1 Harvard Dataverse

O **Harvard Dataverse** é um repositório multidisciplinar de dados de pesquisa, de âmbito internacional, disponibilizado pela Harvard University, instituição privada comunitária, sem fins lucrativos, com sede em Cambridge e Boston, estado de Massachusetts. Nomeada em homenagem ao seu primeiro benfeitor, John Harvard, a universidade é a mais antiga instituição de ensino superior dos Estados Unidos, fundada em 1636. Dedicada à excelência em ensino, aprendizagem e pesquisa nas diferentes áreas do conhecimento, Harvard oferece uma educação global por meio de suas 12 unidades acadêmicas e do Radcliffe Institute for Advanced Study, instituto de estudos avançados com foco em pesquisa interdisciplinar.

O Repositório Dataverse de Harvard faz parte do Dataverse Project, aplicativo *web* de código aberto para compartilhamento, preservação, citação, exploração e análise de dados de pesquisa (HARVARD LIBRARY, 2020), desenvolvido pelo Institute for Quantitative Social Science (IQSS) de Harvard. O projeto teve início em 2006, a partir de experiência anterior com o Virtual Data Center (VDC), de 1999 a

2006, resultado de uma parceria entre o Harvard-MIT Data Center e a Biblioteca da Universidade de Harvard. Conta também com a colaboração de diferentes instituições ao redor do mundo, firmada por meio de um consórcio global, Global Dataverse Community Consortium (GDCC), entre organizações que utilizam o Dataverse e estão engajadas em prol de projetos de códigos abertos.

Contando com a colaboração da Biblioteca de Harvard, considerada a maior biblioteca acadêmica do mundo (HARVARD, 2020), e o Sistema de Tecnologia da Informação da Universidade, o IQSS disponibiliza, de forma aberta e gratuita o Harvard Dataverse para que pesquisadores de todas as disciplinas ao redor do mundo, independentemente de vínculo institucional, possam depositar seus dados de pesquisa. Contribuindo para a disponibilização e replicação de dados de pesquisa, para a visibilidade destes via *web* e para a devida atribuição de créditos aos pesquisadores, autores de dados e todos os outros atores envolvidos na publicação de pesquisas e financiamento de projetos, um repositório de dados Dataverse

consiste na instalação do *software Dataverse*, o qual pode hospedar vários arquivos virtuais por meio da criação de comunidades/coleções (Dataverses), nas quais estarão contidos os conjuntos de dados (Datasets) com seus metadados descritivos e arquivos de dados (incluindo documentação e código que acompanham os dados). Como método de organização, as coleções Dataverse também podem conter outras sub-coleções de dados (Dataverses). (HARVARD LIBRARY, 2020, *on-line*, tradução nossa)

Em constante crescimento, desde 2007, o Dataverse de Harvard vem publicando nas diferentes áreas do conhecimento, comunidades/coleções e conjuntos de dados, disponibilizando atualmente<sup>10</sup> um volume de 4.543 comunidades/coleções (*dataverses*) e 107.379 conjuntos de dados (*datasets*).

Em relação às políticas para gestão dos repositórios Dataverse, embora tenham suas individualidades e políticas específicas, por fazerem parte do Dataverse Project, todos são convidados a seguir as diretrizes descritas nas Normas Comunitárias Dataverse para *download* de conjuntos de dados pelos usuários, bem como o Guia do Usuário Dataverse.

Especificamente para o Harvard Dataverse, as políticas, termos de uso e diretrizes desenvolvidas com orientação legal da Cyberlaw Clinic no Berkman Center for Internet & Society da Harvard University se constituem em<sup>11</sup>: **Termos Gerais de**

---

<sup>10</sup> Informação coletada em 02/02/2021.

<sup>11</sup> Disponível em: <https://support.dataverse.harvard.edu/harvard-dataverse-general-terms-us>

**Uso; Política de Privacidade; Política de Preservação; Termos de uso do Application Program Interface (API); e política de Contrato de uso de dados.**

#### 4.1.2 Stanford Digital Repository (SDR)

O **Stanford Digital Repository (SDR)** é o repositório institucional da Stanford University, universidade privada situada em Palo Alto, estado da Califórnia, nos Estados Unidos. Fundada em 1885 pelo então senador Leland Stanford e sua esposa, Jane Lathrop Stanford, hoje a universidade é composta por sete escolas e 18 institutos interdisciplinares, os quais agregam estudantes de todos os 50 estados estadunidenses e de 91 outros países.

A universidade abriga, ainda, 20 bibliotecas individuais com seus respectivos acervos de acordo com área do conhecimento abrangida. Dentre os serviços oferecidos pelas Bibliotecas de Stanford, destaca-se o Serviço de gerenciamento de dados, o qual dá suporte aos pesquisadores da instituição em relação ao gerenciamento, compartilhamento e preservação de dados de pesquisa. As bibliotecas também gerenciam o Repositório digital da universidade, o SDR, por meio do seu Sistema de preservação digital, fornecendo, assim, suporte ao gerenciamento de longo prazo de informações acadêmicas em Stanford.

Em atividade desde 2005, o SDR é um repositório institucional não exclusivo para dados de pesquisa, disponível aos professores, pesquisadores, estudantes e instituições parceiras da comunidade de Stanford para o gerenciamento e preservação dos seus diferentes tipos de produção acadêmica e de pesquisa. Em constante crescimento, o SDR gerencia atualmente<sup>12</sup> mais de 500 TB de conteúdo exclusivo (1,9 milhões de objetos), sendo 11.652 registrados como *datasets*.

Quanto às políticas de gestão estabelecidas para o uso do SDR, estas apresentam-se no formato de diretrizes, a saber<sup>13</sup>: **Visão geral** do repositório; **Benefícios do SDR**; **Serviços ao depositante**; **Serviços DOI**; e, por fim, os **Termos de depósito** do SDR, o qual disponibiliza um contrato de depósito de conteúdos digitais antes do depósito destes.

---

<sup>12</sup> Informação coletada em 28/12/2020.

<sup>13</sup> Disponível em: <https://library.stanford.edu/research/stanford-digital-repository>

#### 4.1.3 DSpace@MIT

O **DSpace@MIT** é o repositório institucional do Massachusetts Institute of Technology (MIT), universidade privada de pesquisa localizada em Cambridge, Massachusetts, nos Estados Unidos. Fundada em 1861 com a missão de acelerar a revolução industrial do país, a comunidade MIT hoje é impulsionada pelo propósito de fazer um mundo melhor por meio da educação, pesquisa e inovação. Contando com 30 departamentos distribuídos em cinco escolas e uma faculdade, além de centros de pesquisas e laboratórios, o MIT é uma instituição global multidisciplinar que impulsiona o trabalho interdisciplinar através da colaboração entre seus pesquisadores e organizações locais, nacionais e internacionais.

Dando suporte aos serviços de pesquisa, o instituto conta, ainda, com a parceria das suas bibliotecas, as quais são também responsáveis pela assistência às comunidades do MIT na gestão de dados de pesquisa e pela disponibilização do DSpace@MIT. Fundado em 2002, o DSpace@MIT não é um repositório exclusivo para dados de pesquisa, mas um repositório institucional para coletar, armazenar e preservar os diferentes tipos de produção acadêmica e de pesquisa das comunidades MIT.

Disponibilizado pelos serviços das bibliotecas do MIT, o DSpace@MIT é um repositório de relevância internacional reconhecido como um dos principais repositórios acadêmicos do mundo (MIT LIBRARIES, [2021?]). Dentre sua coleção de mais de 90.000 obras e mais de 1 milhão de *downloads* mensais<sup>14</sup>, estão artigos de conferência, imagens, artigos acadêmicos revisados por pares, pré-impressões, relatórios técnicos, teses, trabalhos e conjuntos de dados de pesquisa. O DSpace@MIT divide-se em 99 comunidades DSpace em diferentes temáticas, as quais subdividem-se em coleções conforme a tipologia do material armazenado. Muitas dessas coleções são de dados relacionados a publicações, como teses e artigos científicos, e outras, referentes a conjuntos de dados, disponibilizados em comunidades específicas, como a comunidade de Conjuntos de Dados de Pesquisa Robótica (*Robotics Research Datasets*).

Embora o DSpace@MIT seja um serviço das bibliotecas do MIT, a responsabilidade pelas políticas institucionais do Repositório fica a cargo do Instituto.

---

<sup>14</sup> Informação coletada em 29/12/2020.

As políticas identificadas na pesquisa abrangem<sup>15</sup>: as **Políticas da comunidade e da coleção**; as **Políticas/Diretrizes/Licenças gerais**; e as **Diretrizes para conjuntos de dados de pesquisa**, orientando sobre os conjuntos de dados a serem enviados ao DSpace@MIT.

As comunidades DSpace@MIT são unidades administrativas do MIT que produzem pesquisa e têm por responsabilidade a definição de políticas comunitárias sobre o conteúdo dos dados, permissões e limites de acesso, retiradas de itens e coleções do repositório, bem como aprovar a inclusão ou retirada de sub-comunidades do DSpace@MIT.

Quanto às bibliotecas do MIT, cabe a estas a responsabilidade de armazenamento e preservação de todo conteúdo enviado ao DSpace@MIT, distribuindo os conteúdos de acordo com as decisões da comunidade e reportando a ela todas as alterações significativas que se fizerem necessárias referentes aos conteúdos depositados. Já ao MIT cabe a responsabilidade de definir políticas institucionais que afetem o DSpace, como regras de direitos autorais e questões de suporte às políticas instituídas.

#### 4.1.4 Data Repository [Berkeley]

O **Data Repository [Berkeley]** é o repositório de dados de pesquisa da University of California Berkeley (UC Berkeley), uma das dez universidades semiautônomas da University of California (UC). Fundada em 1869 por Frederick Low, a UC é hoje composta, além de Berkeley, por mais nove universidades, localizadas nas cidades de Davis (UC Davis), Irvine (UC Irvine), Los Angeles (UCLA), Merced (UC Merced), Riverside (UC Riverside), San Diego (UC San Diego), San Francisco (UC San Francisco), Santa Barbara (UC Santa Barbara) e Santa Cruz (UC Santa Cruz).

A UC Berkeley localizada em Berkeley, no estado da Califórnia, Estados Unidos, é uma instituição pública multidisciplinar de pesquisa, cujo cerne de suas atividades consiste na investigação científica. A preocupação com a pesquisa está presente por meio de sua estrutura organizacional, a qual conta com 125 unidades de pesquisa distribuídas pelo campus. Sendo uma instituição altamente conceituada, Berkeley consta em *rankings* universitários, assim como a Universidade dos Estados

---

<sup>15</sup> Disponível em: <https://libguides.mit.edu/c.php?g=176372&p=1158918>

Unidos, com o maior número de programas de pós-graduação bem avaliados e com uma biblioteca universitária pública de número um na América do Norte (BERKELEY UNIVERSITY OF CALIFORNIA, 2021).

O Sistema de bibliotecas em Berkeley, composto por 24 bibliotecas, atua em parceria com a Universidade na criação e implantação de serviços de pesquisa avançada, auxiliando-a a manter-se como líder na produção de conhecimento. Igualmente, por meio de parceria, o Sistema de bibliotecas, juntamente ao Serviço de Tecnologia da Universidade, oferece o Programa de Gerenciamento de Dados de Pesquisa (RDM) aos professores, funcionários e estudantes da comunidade acadêmica, bem como auxílio no gerenciamento de dados durante todo o processo de suas pesquisas.

Embora a Universidade auxilie sua comunidade de pesquisadores na gestão de seus dados de pesquisa por meio do RDM oferecido por seu Sistema de bibliotecas, o seu repositório de dados, o Data Repository [Berkeley], lançado no ano de 2009, não está localizado nas dependências físicas da UC Berkeley. A Universidade utiliza os serviços de curadoria do Repositório Digital Dryad para que seus pesquisadores depositem seus dados de pesquisa. Nesse repositório, 524<sup>16</sup> arquivos de dados de pesquisa relacionados à UC Berkeley foram recuperados mediante busca por instituição.

Quanto às políticas do Data Repository, conforme informações recebidas após contato com a Universidade, este faz uso dos Termos de Serviços do Dryad, não apresentando políticas específicas elaboradas para o Repositório.

Os Termos de Serviço são destinados aos indivíduos e organizações que utilizam ou depositam dados no Dryad, mediante aceite desses Termos. Dentre os **Termos de Serviço do Dryad**, destacam-se<sup>17</sup>: as **Políticas de Publicação**; o **Uso**; a **Privacidade**; e as **Disposições gerais**.

#### 4.1.5 Research Works Archive

O **Research Works Archive** é o repositório institucional da University of Washington (UW), instituição pública de pesquisa com campus em Seattle, Tacoma e Bothell, no estado de Washington, Estados Unidos. Fundada em 1861, a UW é uma

---

<sup>16</sup> Dados coletados em: 17/01/2021.

<sup>17</sup> Disponível em: <https://datadryad.org/stash/terms>

universidade multidisciplinar, considerada uma das mais prestigiadas instituições de ensino e pesquisa do mundo, com vasta oferta de cursos de graduação a cada trimestre, somando 1.800 cursos em suas faculdades e escolas. No campo da pesquisa, a UW conta com 284 centros de pesquisa e encontra-se entre as cinco universidades públicas dos Estados Unidos com financiamento total acima de um bilhão (UNIVERSITY OF WASHINGTON, 2019).

A Universidade conta com o suporte de suas bibliotecas na orientação ao seu corpo de professores, pesquisadores e estudantes, por meio dos Serviços de Publicação Acadêmica, onde são disponibilizadas as diretrizes sobre o acesso e uso do seu repositório digital, o Research Works Archive. O Research Works Archive não é um repositório exclusivo para dados de pesquisa, mas sim para o armazenamento e disseminação da produção acadêmica e de pesquisa da UW, permitindo o depósito de qualquer formato ou conteúdo de arquivo digital. O depósito de objetos digitais no repositório se dá por solicitação de autodepósito por parte dos depositantes ou depósito mediado pela equipe do Research Works Archive, podendo estar sujeitos à aplicação de restrições de direitos autorais. Em relação à data de criação do Repositório, não foi possível localizá-la, nem no *web site* do Research Works Archive nem no das Bibliotecas da UW.

A organização do Repositório se dá a partir de sua divisão em seis comunidades DSpace, as quais podem se dividir em subcomunidades que, por sua vez, dividem-se em coleções temáticas. Os dados e conjuntos de dados de pesquisa não estão armazenados em comunidades específicas, mas sim dispersos nas distintas coleções do Repositório.

Em relação às políticas do Repositório, estas se aplicam à gestão de todas as suas coleções digitais, sendo elas<sup>18</sup>: a **Política de Coleção Digital**; a **Política de Preservação Digital**; a **Política de Retirada**; a **Política de restrição de acesso**; e a **Política de Direitos Autorais**.

---

<sup>18</sup> Disponível em: <https://www.lib.washington.edu/scholpub/scholarly-publishing-services/researchworks>

#### 4.1.6 Deep Blue Data (DBD)

**O Deep Blue Data (DBD)** é o repositório de dados de pesquisa da University of Michigan (U-M), instituição pública de pesquisa, localizada na cidade de Ann Arbor, no estado de Michigan, nos Estados Unidos. Fundada em 1817, antes de Michigan se tornar um estado, a U-M é a universidade mais antiga do estado, sendo transferida para sua atual localização no ano de 1837, onde encontra-se o seu Campus Central.

A U-M é uma universidade multidisciplinar que abriga 19 escolas e faculdades no campus de Ann Arbor, além de contar com os campi de Dearborn e Flint. O campus de Flint é a filial mais antiga da Universidade, datando de 1956. Três anos mais tarde, em 1959, foi fundado o campus de Dearborn, por meio de doações da Ford Motor Co., ofertando, inicialmente, os cursos de pós-graduação em Engenharia Industrial, Engenharia Mecânica e Administração de Empresas. Dearborn, ao longo dos anos, foi ampliando a oferta de novos programas em outras áreas do conhecimento.

A universidade encontra-se entre as melhores universidades públicas do país no que se refere à produtividade de pesquisa, contando com mais de 263.000 produções de pesquisa e com gastos superiores a US\$ 1 bilhão. Além disso, é importante ressaltar que a pesquisa é parte central da sua missão de promoção de iniciativas de colaboração interdisciplinar envolvendo seus professores e estudantes (UNIVERSITY OF MICHIGAN, 2021a). Abrangendo pesquisas nos campos da ciência, engenharia, medicina, ciências sociais e humanidades, a U-M é a número um em volume de pesquisas entre as universidades públicas dos Estados Unidos, de acordo com a National Science Foundation (UNIVERSITY OF MICHIGAN, 2021b).

Dando suporte à disseminação e preservação das produções acadêmicas e de pesquisa da sua comunidade, a Universidade conta com os serviços oferecidos pela sua Biblioteca, a University of Michigan Library (U-M Library). Classificada como uma das principais bibliotecas de pesquisa acadêmica da América do Norte, a U-M Library (UNIVERSITY OF MICHIGAN LIBRARY, 2021) hospeda os dois repositórios da Universidade: o Deep Blue, repositório institucional da U-M, que desde 2006 preserva os diferentes tipos de documentos resultantes da produção intelectual e científica da comunidade da U-M; e o Deep Blue Data (DBD), expansão do Deep Blue e repositório de dados de pesquisa que, a partir de 2016, foi destinado ao armazenamento e preservação digital de dados de pesquisa desenvolvidos ou usados no apoio às atividades de pesquisa na U-M.

A busca no DBD permite recuperar dados por: tipo de recurso, autor, disciplina e idioma. Por tipo de recurso, tem-se a opção de recuperar por: **datasets**, com um volume de 373 itens, e por **coleção**, com um total de 15 coleções, cada uma com diferentes números de itens contemplados. As disciplinas contempladas para recuperação de dados são: Ciência; Engenharia; Humanidades; Ciências da Saúde; Ciências Sociais; Outras; Fontes de informação geral; Governo, Política e Direito; e Estudos Internacionais.

As políticas identificadas no DBD abrangem<sup>19</sup>: os **Termos de uso** do repositório; a **Política de envio e depósito** de dados; a **Política de Coleções e conteúdo** dos dados; e a **Política de preservação** de dados de pesquisa.

#### 4.1.7 ORA-Data

O **ORA-Data** é o repositório de dados de pesquisa da University of Oxford, instituição de ensino superior pública, localizada no condado de Oxfordshire, Inglaterra. Sem uma data precisa de sua fundação, contudo com evidências históricas de que no ano de 1096 já existia alguma forma de ensino em Oxford, identifica-se que o seu desenvolvimento como instituição de ensino se deu rapidamente a partir de 1167, quando Henrique II proibiu estudantes ingleses de frequentarem a Universidade de Paris (UNIVERSITY OF OXFORD, 2020b).

Oxford é uma instituição multidisciplinar de ensino e pesquisa, com uma estrutura organizacional colegiada distinta, a qual é composta por 39 faculdades e seis salas privadas de ensino, representando comunidades acadêmicas, que reúnem estudantes e pesquisadores de diferentes disciplinas, culturas e países. As faculdades são financeiramente independentes e autogovernadas, mas vinculadas à Universidade Central (UNIVERSITY OF OXFORD, 2020c).

Com mais de 250 programas de pós-graduação e um corpo discente composto por mais de 50% de estudantes vindos de diferentes partes do mundo, Oxford tem consistentemente a maior renda de pesquisa de patrocinadores externos dentre todas as universidades do Reino Unido. Também tem sido líder no Reino Unido, através da Oxford University Innovation, na criação de empresas de base tecnológica de origem acadêmica, as Spinouts (UNIVERSITY OF OXFORD, 2020a).

---

<sup>19</sup> Disponível em: <https://deepblue.lib.umich.edu/data/agreement.agreement?locale=en>

Buscando atender às necessidades de pesquisa de sua comunidade, a Universidade disponibiliza, por meio dos serviços oferecidos pelo seu Sistema de bibliotecas — as Bibliotecas Bodleian —, acesso a recursos de pesquisa impressos e *on-line* financiados ou produzidos por Oxford. Considerado o maior sistema de bibliotecas universitárias do Reino Unido, o sistema das Bibliotecas Bodleian agrega mais de 100 bibliotecas, incluindo 30 em Oxford, além de bibliotecas de pesquisa de professores e de departamentos e institutos, as quais, juntas, somam um volumoso acervo de mais de 12 milhões de itens impressos, além de revistas e coleções especiais.

Dentre os serviços oferecidos, Bodleian, por meio do Bodleian Digital Library Systems and Services (BDLSS), gerencia e disponibiliza acesso ao repositório institucional da Universidade, o Oxford University Research Archive (ORA), para os diferentes tipos de publicações, e o **ORA-Data**, repositório de dados de pesquisa de Oxford. O ORA-Data, oferece serviços de arquivamento, preservação, descoberta e compartilhamento de dados produzidos por pesquisadores de Oxford.

Tendo sua fase piloto lançada em primeiro de dezembro de 2014, o ORA-DATA continua com funcionalidades em desenvolvimento. Agregado ao ORA, possibilita que os dados possam ser facilmente vinculados e navegados junto a publicações relacionadas. Embora a Universidade disponibilize aos seus pesquisadores o ORA-Data para depósito de seus dados de pesquisa, não é uma obrigatoriedade o depósito neste, visto que cabe aos pesquisadores a decisão de depositar no ORA-Data ou em um repositório especializado de um determinado campo disciplinar que dê maior visibilidade aos seus registros de pesquisa.

As políticas do ORA-Data incluem<sup>20</sup>: as **Condições de depósito**; a **Licença de usuário do ORA**; a **Política de envio de dados**; a **Política de armazenamento e preservação de dados**; a **Política de retirada de dados**; e a **Política de contestação de dados**.

#### 4.1.8 JHU Data Archive

O **Johns Hopkins University Data Archive (JHU Data Archive)** é o repositório de dados da Johns Hopkins University (JHU), instituição privada de ensino

---

<sup>20</sup> Disponível em: [https://libguides.bodleian.ox.ac.uk/ld.php?content\\_id=26573479](https://libguides.bodleian.ox.ac.uk/ld.php?content_id=26573479)

superior, localizada em Baltimore, Maryland, nos Estados Unidos. Inaugurada em 1876, a Universidade leva o nome do filantropo de Maryland, Johns Hopkins, empresário que acreditava no investimento para melhorias em saúde e educação (JOHNS HOPKINS UNIVERSITY, 2021a).

Considerada a primeira universidade de pesquisa da América, sendo líder no país em financiamento federal de pesquisa e desenvolvimento desde 1979, a JHU é uma instituição multidisciplinar de ensino e pesquisa, presente não somente nos Estados Unidos, mas também em outros continentes. Distribuída em nove divisões acadêmicas, a JHU tem quatro campi em Baltimore; um em Washington, D.C.; um no Condado de Montgomery, Maryland; e instalações em toda a região de Baltimore-Washington, bem como na China e Itália (JOHNS HOPKINS UNIVERSITY, 2021b).

Dando suporte ao ensino e pesquisa, a Johns Hopkins conta com os serviços da sua rede de bibliotecas, a Sheridan Libraries, formada pela: Milton S. Eisenhower Library; John Work Garret Library; e George Peadody. A principal biblioteca de pesquisa da Universidade, a Milton S. Eisenhower Library, oferece ao corpo de pesquisadores, professores e estudantes da JHU os serviços de gerenciamento de dados de pesquisa, bem como serviços de arquivamento e compartilhamento de dados via JHU Data Archive.

O JHU Data Archive é um repositório de acesso aberto que aceita dados de pesquisa em qualquer formato e de qualquer disciplina, sendo, dessa forma, uma opção para áreas que não têm um repositório público específico para disciplinas. Não se localizou a data de implantação do Repositório, contudo estima-se que seja próxima ao ano de 2015, haja vista essa ser a data dos Termos de Uso do JHU Data Archive.

O Repositório faz parte do Projeto Dataverse, e, dessa forma, suas políticas estão condicionadas: às diretrizes apresentadas no **Guia do Usuário Dataverse**<sup>21</sup>; às diretrizes apresentadas nos **Serviços de Dados das Bibliotecas da Universidade**<sup>22</sup> [Johns Hopkins Libraries]; e aos **Termos de Uso**<sup>23</sup> do JHU Data Archive.

---

<sup>21</sup> Disponível em: <https://guides.dataverse.org/en/4.13/user/>

<sup>22</sup> Disponível em: <https://dataservices.library.jhu.edu/archiving/>

<sup>23</sup> Disponível em: <https://dataservices.library.jhu.edu/archiving/terms-of-use/>

#### 4.1.9 Columbia Academic Commons

O **Columbia Academic Commons** é o repositório institucional da Columbia University New York, instituição de ensino superior privada, localizada na cidade de Nova Iorque, nos Estados Unidos. Fundada em 1754, como King's College, a instituição, durante a Revolução Americana, teve suas atividades suspensas em 1776, reabrindo após oito anos, em 1784, com o nome de Columbia. Na segunda metade do século XIX, Columbia se desenvolveu como instituição de ensino e, ao adquirir o *status* de universidade, passou a se chamar Columbia University em 1896. Hoje a instituição é oficialmente conhecida como Columbia University in the City of New York (COLUMBIA UNIVERSITY, 2021a).

Considerada a instituição de ensino superior mais antiga no estado de Nova York e a quinta mais antiga dos Estados Unidos, a Universidade, além de seus campi em Nova York, tem duas instalações fora de Manhattan: o Nevis Laboratories, criado em 1947 e localizado em Irvington, sendo considerado o principal centro de estudos de partículas experimentais de alta energia e física nuclear; e o Lamont-Doherty Earth Observatory, criado em 1949 em Palisades, uma instituição de pesquisa líder focada em mudanças climáticas globais, fenômenos naturais, recursos não renováveis e riscos ambientais, examinando o planeta em todos os continentes e oceanos (COLUMBIA UNIVERSITY, 2021a).

A Universidade é uma instituição multidisciplinar, com uma longa e distinta história na descoberta do conhecimento. Considerada uma das mais importantes instituições de pesquisa do mundo, a Columbia University conta com mais de 200 centros e institutos de pesquisa e um orçamento anual de mais de 1 bilhão em pesquisa (COLUMBIA UNIVERSITY, 2021b).

Parceiras no desenvolvimento da pesquisa e do aprendizado na Instituição, destacam-se as bibliotecas da Columbia University, oferecendo diferentes serviços aos seus usuários, dentre os quais evidencia-se o Serviço de Dados de Pesquisa, que fornece apoio e consulta para as necessidades de dados de pesquisa dos membros de sua comunidade. As bibliotecas também dão acesso e gerenciam o repositório da Universidade, o Columbia Academic Commons. Fundado em 2006 com o nome de Digital Commons, o Repositório, embora tenha estado disponível desde essa data, passou por várias mudanças de plataforma e de nome.

O Repositório não é exclusivo para dados de pesquisa, fornecendo acesso aberto a diferentes tipos de produção acadêmica e científica dos pesquisadores da Columbia University e de suas afiliadas: Barnard College, Jewish Theological Seminary, Teachers College e Union Theological Seminary. Também não há uma exigência para que os dados de pesquisa sejam depositados exclusivamente no Repositório; pelo contrário, há um encorajamento para que os pesquisadores da Universidade e suas afiliadas, se desejarem, façam o depósito de seus dados de pesquisa sem custos no Dryad, especialmente se os arquivos forem de tamanho grande.

De acordo com informações publicadas na página de ‘Perguntas frequentes’ do Repositório, a equipe de bibliotecas da Universidade está trabalhando para que o Academic Commons seja conectado ao Dryad, e, assim, todos os dados de pesquisa enviados para Dryad por pesquisadores da Columbia também serão acessíveis no Academic Commons.

As políticas do Repositório identificadas na pesquisa abrangem<sup>24</sup>: **Diretrizes de participação; Termos de Uso; Copyright, e Proteção de dados.**

#### 4.1.10 eCommons

O **eCommons** é o repositório da Cornell University, universidade privada localizada em Ithaca, Nova Iorque, nos Estados Unidos. Fundada em 1865 por Ezra Cornell e Andrew Dickson White, Cornell é uma instituição federal de concessão de terras do Estado de Nova York, parceira da State University of New York. A Universidade conta com uma organização física composta por 15 faculdades e escolas, sendo localizadas não somente em Ithaca, mas também em Nova York e em Doha, no Catar (CORNELL UNIVERSITY, 2020).

Cornell é considerada uma instituição pioneira na educação dos Estados Unidos, atuando em diferentes frentes na área do ensino, seja na solução de problemas matemáticos, seja na inclusão de disciplinas nos currículos de formação universitária do país. Além de ser uma instituição multidisciplinar de ensino de graduação e pós-graduação, Cornell também fornece formação com ênfase profissionalizante por meio de parceria com as escolas de negócios credenciadas à

---

<sup>24</sup> Disponível em: <https://academiccommons.columbia.edu/policies/>

Universidade, nas áreas de: Economia aplicada à gestão, Administração de hotéis e Gestão administrativa.

Na área da pesquisa, segundo os *rankings* nacionais e estaduais da National Science Foundation de 2018, Cornell aparece como a primeira colocada entre as universidades do Estado de Nova York em gastos com pesquisa, com um orçamento anual de U\$1,071 bilhões de dólares, e na décima terceira posição em nível nacional (CORNELL UNIVERSITY, 2021).

Como parceira indispensável ao estudo, ensino e pesquisa na Universidade, a Biblioteca de Cornell, entre os serviços oferecidos a sua comunidade universitária, fornece acesso aberto a longo prazo para os conteúdos digitais relacionados à produção acadêmica e de pesquisa de Cornell, disponibilizados por meio do seu repositório institucional, o eCommons.

O e-Commons não é um repositório exclusivo para dados de pesquisa, mas sim para os diferentes tipos de conteúdo produzidos por seus professores, funcionários e alunos, bem como para os trabalhos produzidos por departamentos, programas, institutos e outras comunidades formais e informais de Cornell (CORNELL UNIVERSITY LIBRARY, 2020). Como a data de criação do Repositório não foi localizada em seu *website*, tomou-se por base o ano de 2002 como sendo o ano de sua implantação, haja vista essa ser a data que dá origem às primeiras estatísticas de *downloads* realizados no e-Commons.

A estrutura do eCommons segue a organização da plataforma DSpace, utilizada para o gerenciamento do Repositório, em que os conteúdos estão arranjados em Comunidades e coleções. O eCommons conta com 31 comunidades DSpace, subdivididas em outras tantas coleções cada uma. Dentro dessas comunidades e coleções, estão armazenados 90.986 itens. Desses itens, 202 estão indicados como 'Tipo' *datasets*, e cada registro contém seus respectivos metadados e a coleção a qual pertence.

As políticas do repositório que foram identificadas compõem-se em<sup>25</sup>: **Permissão de depósito; Política de coleta de conteúdo; Política de depósito e licença; Política de depósito de dados; Política de acesso; Política de retirada; Política de apoio à preservação; Política de privacidade; Termos de Uso; e Política de acessibilidade.**

---

<sup>25</sup> Disponível em: <https://guides.library.cornell.edu/ecommmons/policy>

#### 4.1.11 Scholarly Commons

O **Scholarly Commons** é o repositório digital da University of Pennsylvania (Penn), instituição de ensino superior privada localizada na cidade de Filadélfia, no estado da Pensilvânia, nos Estados Unidos. Fundada em 1740 como faculdade por Benjamin Franklin, transformou-se em universidade no ano de 1765.

Honrando sua tradição em transformar conhecimento em ação social, a Universidade, por meio do Penn Compact 2022, seu plano de gestão estratégica, busca tornar-se mais inclusiva, inovadora e com maior impacto para o país. Contando com quatro escolas de graduação, 12 escolas de pós-graduação e outras escolas profissionais, Penn é uma instituição multidisciplinar de ensino e pesquisa “[...] construída sobre o conceito da inovação” (UNIVERSITY OF PENNSYLVANIA, [2020? a], *on-line*).

A Universidade, que conta com o suporte de seus 189 centros de pesquisa e institutos, de caráter interdisciplinar, é considerada uma das melhores universidades de pesquisa dos Estados Unidos. Gerando e aplicando conhecimento nas áreas da medicina, tecnologia, negócios e ciência para melhorar a vida da sociedade global, a Penn conta com um orçamento anual de mais de US\$ 1,02 bilhão em pesquisa e desenvolvimento (UNIVERSITY OF PENNSYLVANIA, [2020? b], *on-line*).

Reforçando sua missão, a Universidade conta com as Bibliotecas Penn a fim de dar suporte ao ensino e à pesquisa por meio dos serviços que oferecem à comunidade. Entre esses serviços, destaca-se o acesso ao repositório digital da Universidade, o Scholarly Commons, lançado oficialmente no início de 2005. O Repositório é um serviço das Bibliotecas Penn, administrado em conjunto ao seu Centro para Dados de Pesquisa e aos administradores das várias escolas e centros da Universidade.

Inicialmente, foi desenvolvido em colaboração com a Penn's School of Engineering & Applied Science (SEAS) e, rapidamente, foi adotado por muitas outras escolas Penn. Desde o seu lançamento, a Scholarly Commons cresceu para além dos materiais de impressão tradicionais para incluir revistas de acesso aberto, conjuntos de dados e outras formas de bolsa multimodal.

O Scholarly Commons não é um repositório exclusivamente para dados de pesquisa, mas sim o repositório institucional de acesso aberto da Universidade, com o propósito de reunir, indexar, armazenar e disponibilizar amplamente a produção

acadêmica da comunidade Penn. Em pesquisa no Repositório, 2.501 *datasets* foram recuperados, vinculados aos seguintes tipos de documentos com seus respectivos quantitativos<sup>26</sup>: Series (1471); Theses/Dissertations (933); Journal (75); Book Gallery (17); Conference (3); Image Gallery (2). O Repositório disponibiliza as métricas de pesquisas nele realizadas por meio do PlumX Metrics da Elsevier.

As políticas do Scholarly Commons, identificadas na pesquisa, abrangem<sup>27</sup>: **Submissões apropriadas; Tipos de formatos suportados; Remoção de conteúdo; Considerações de direitos autorais; Acesso e Migração; e Considerações de privacidade.**

#### 4.1.12 Apollo

**Apollo** é o repositório institucional da University of Cambridge, instituição pública de ensino superior, localizada na cidade de Cambridge, na Inglaterra. Registros iniciais sobre a formação de Cambridge como instituição de ensino datam de 1209, tendo sido um refúgio de estudiosos, esta data é considerada como a de sua fundação. Contudo, foi apenas no ano de 1231 que o rei Henrique III lhe concedeu o monopólio de ensino (UNIVERSITY OF CAMBRIDGE, 2021a).

Com a missão de contribuir socialmente por meio da excelência na educação, aprendizagem e pesquisa, Cambridge, ao longo do tempo, tornou-se uma instituição global, multidisciplinar e internacional, reunindo mais de 9.000 estudantes internacionais de todas as partes do mundo. Com uma estrutura física diferenciada, organizada em confederação, Cambridge é composta por 31 faculdades autônomas com estatutos e regimentos próprios e 150 departamentos, escolas e faculdades universitárias (UNIVERSITY OF CAMBRIDGE, 2021b).

Na área da pesquisa, Cambridge é líder mundial, contribuindo socialmente por meio de inovação e parcerias estratégicas em colaborações de pesquisa com universidades de todo o mundo. Em diferentes frentes de atuação, Cambridge sustenta uma enorme gama de inovações científicas, criando prosperidade, melhoria da qualidade de vida, proteção ambiental e enriquecimento cultural (UNIVERSITY OF CAMBRIDGE, 2021c).

---

<sup>26</sup> Dados coletados em 06/01/2021.

<sup>27</sup> Disponível em: <https://guides.library.upenn.edu/scholarlycommons/policies>

Comprometida em contribuir com a missão da Universidade para a excelência em pesquisa e educação, a Biblioteca Universitária de Cambridge, por meio do seu Escritório de Comunicação Acadêmica, disponibiliza e gerencia o repositório digital de Cambridge, o Apollo.

Anteriormente denominado DSpace@Cambridge, o Apollo foi criado em 2003 como um serviço para armazenar e fornecer acesso à produção de pesquisa de membros da University of Cambridge. Não sendo um repositório exclusivo para dados de pesquisa, armazena uma gama de conteúdos e fornece diferentes níveis de acesso, mas seu foco principal é fornecer acesso aberto às publicações de pesquisa da Universidade.

O Repositório utiliza a plataforma DSpace para gerenciamento do seu conteúdo digital e, portanto, segue a estrutura de organização da Plataforma para o arranjo dos conteúdos em Comunidades e coleções. Dentro da Plataforma, a organização do Apollo se dá em oito comunidades<sup>28</sup> vinculadas às escolas e faculdades de Cambridge, que subdividem-se em distintas subcomunidades relacionadas as suas unidades administrativas. Por fim, essas subcomunidades se dividem em coleções DSpace, onde estão armazenados os conteúdos digitais.

Em pesquisa no Repositório por 'Tipo de conteúdo', recupera-se, na coleção 'dataset', 2.198 conjuntos de dados de pesquisa. Contudo, ao se realizar a pesquisa sem navegar por tipo de conteúdo, somente escrevendo o termo 'dataset' recupera-se 17.887 conjuntos de dados. A diferença entre o volume de *datasets* recuperados nas duas formas de pesquisa se deve ao fato de o primeiro volume de *datasets* recuperados estar vinculado a artigos científicos, e os outros 17.887 *datasets* recuperados não apresentarem nenhuma vinculação com outros documentos no Repositório.

Em relação às políticas referentes a dados de pesquisa, estas são disponibilizadas por meio de *links* nas páginas do *Office of Scholarly Communication* (OSC), escritório de comunicação científica da Universidade, o qual dá suporte a pesquisadores e treinamentos à equipe administrativa de sua Comunidade, em conformidade com as políticas de acesso aberto a dados de pesquisa. A página do OSC remete às páginas relacionadas<sup>29</sup>: ao **Contrato de Licença de Depósito**; à

---

<sup>28</sup> Dados de 07 jan. 2021.

<sup>29</sup> Disponível em: <https://osc.cam.ac.uk/repository>

## **Política de Preservação Digital; aos Termos de Uso do Repositório; e à Política do DOI.**

Os **Termos de Uso**<sup>30</sup> do Repositório fornecem diretrizes sobre: o tipo de produção de pesquisa que pode ser depositado; os termos de acordos entre depositantes e Repositório; limites de responsabilidade dos depositantes e do Repositório; Política de retirada de conteúdo do repositório; e outros termos de uso aplicáveis no acesso ao Repositório.

A Universidade de Cambridge também provê um Quadro de Política de Gestão de Dados de Pesquisa, com o objetivo de: fornecer orientação a sua Comunidade sobre gestão de dados de pesquisa e garantir que a Universidade empreenda medidas institucionais adequadas em prol das boas práticas de gerenciamento de dados e atenda aos requisitos políticos dos financiadores da Universidade.

### 4.1.13 Merritt

**Merritt** é o serviço de repositório de preservação digital da University of California (UC), localizada no estado da Califórnia, Estados Unidos. Fundada em 1869, a UC tornou-se um sistema de 10 universidades semiautônomas espalhadas pelo estado da Califórnia, conforme exposto anteriormente em 4.1.4.

Fazendo parte desse Sistema está a University of California, Los Angeles (UCLA), localizada na região de Westwood, na cidade de Los Angeles, Estados Unidos. Conhecida originalmente como “Ramo Sul” da UC, a UCLA deu início as suas atividades no ano de 1919. No entanto, somente no ano de 1927 é que os regentes da UC adotaram o nome University of California in Los Angeles.

Reconhecida como uma das maiores universidades de pesquisa do mundo, a UCLA é uma universidade pública, multidisciplinar e comprometida com a excelência na geração e aplicação do conhecimento em prol do aperfeiçoamento global da sociedade. Marcada por descobertas, criatividade e inovação, a Universidade conta com um portfólio de quase 3.000 invenções e mais de 140 empresas criadas com base na tecnologia desenvolvida em seu âmbito (UCLA, 2021a). Em questões de orçamento, as estatísticas fiscais do ano de 2020 revelam que a Universidade contou

---

<sup>30</sup> Disponível em: <https://www.repository.cam.ac.uk/>

com um montante de US\$1.43 bilhão em financiamento de pesquisa, US\$1.1 bilhão em despesas com pesquisa e teve 394 novas invenções divulgadas (UCLA, 2021b).

Auxiliando na efetivação da missão da Universidade, a Biblioteca da UCLA, constituindo um sistema de bibliotecas, é considerada uma das principais bibliotecas acadêmicas dos Estados Unidos. Dentre os serviços oferecidos pela Biblioteca à sua comunidade, destacam-se os Serviços de Gestão e Curadoria de Dados. O Serviço fornece orientações aos professores, funcionários e pesquisadores da UCLA sobre como fazer o gerenciamento de seus dados de pesquisa e o armazenamento e compartilhamento desses dados. Apresenta, igualmente, opções de armazenamento de dados, para que os pesquisadores possam escolher o recurso mais apropriado para depositar seus dados de pesquisa. Entre as opções de recursos indicadas pelo Serviço, estão: o Merritt, o armazenamento em nuvem e os repositórios disciplinares.

O Merritt, fundado em 2009, não é um repositório exclusivo da UCLA, mas sim um serviço de repositório disponibilizado e mantido pelo Centro de Curadoria da Universidade da Califórnia (UC3) na Biblioteca Digital da Califórnia (CDL), para que os membros da Comunidade da UC possam, se desejarem, depositar seus dados de pesquisa. Com seu nome fazendo referência ao Lago Merritt, que foi o primeiro refúgio oficial de vida selvagem nos Estados Unidos e é um Marco Histórico Nacional, o Repositório está disponível para todos os membros da comunidade da UCLA, a fim de ajudar na gestão, arquivamento e/ou compartilhamento de conteúdo digital.

Utilizando a interface do Dryad para publicação dos dados de pesquisa, o Merritt possibilita que seus pesquisadores enviem seus conjuntos de dados para o Dryad diretamente ou por meio de um periódico. Todos os dados gerenciados no Dryad são automaticamente preservados em Merritt.

Quanto às políticas do Repositório, o Merritt e o Dryad cumprem os **Termos de Uso** e a **Política de Privacidade** da CDL<sup>31</sup>. A gestão dos dados no Repositório está condicionada aos **Termos de Serviço do Dryad**<sup>32</sup> e à **Política de Depósito** do Centro de Curadoria da Universidade (UC3).

Em tempo, embora o Merritt não seja um repositório exclusivo e gerenciado pela UCLA, a Universidade está entre as 18 melhores universidades do mundo indicadas no ranking *webometrics.info* de 2020 e faz parte do sistema de universidades da UC.

---

<sup>31</sup> Disponível em: <https://cdlib.org/about/policies-and-guidelines/terms-conditions/>

<sup>32</sup> Disponível em: <https://datadryad.org/stash/terms>

#### 4.1.14 ISPS Data Archive

O **ISPS Data Archive** é o repositório da Yale University, instituição de ensino superior privada, localizada em New Haven, no estado de Connecticut, Estados Unidos. Fundada em 1701, próximo a Saybrook, Connecticut, como o Collegiate School para educar os alunos para o emprego público, teve sua sede mudada para New Haven, em 1716, e seu nome alterado para Yale College, em 1718 (YALE UNIVERSITY, 2021c).

Tendo por missão melhorar o mundo para as gerações futuras por meio da pesquisa, bolsas de estudo, educação, preservação e prática, Yale é uma instituição de alcance local e internacional. Por meio de parcerias locais, contribui para o fortalecimento da comunidade e economia local e, através de parcerias internacionais, busca promover a compreensão das diferentes culturas, melhorar as condições sociais da humanidade, inteirar-se profundamente dos segredos do universo e treinar a próxima geração de líderes mundiais (YALE UNIVERSITY, 2021a).

Yale é uma instituição multidisciplinar de pesquisa com uma grande variedade de departamentos e programas acadêmicos, escolas, centros, museus e instituições afiliadas, que recebe estudantes internacionais de diferentes países. Avançando nos campos de humanidades, artes, ciências sociais, medicina, ciência e engenharia, Yale abriga vários centros de pesquisa, laboratórios e institutos renomados (YALE UNIVERSITY, 2021b).

A Universidade, por meio do Institution for Social and Policy Studies (ISPS) disponibiliza seu repositório de dados de pesquisa, lançado em 2009<sup>33</sup>. O ISPS tem por missão fornecer aos membros da comunidade acadêmica acesso a arquivos associados a estudos acadêmicos para fins de replicação, para todos os estudos conduzidos por pesquisadores afiliados ao ISPS.

Embora a Universidade seja uma instituição multidisciplinar de pesquisa, o Repositório é disciplinar, visto que, de acordo com o ISPS (2020), seu conteúdo digital, no presente momento, em sua maioria, consiste em dados de pesquisa de experimentos em Ciências Sociais, acompanhados da documentação necessária para uso e entendimento desses dados e dos arquivos associados a eles. O acesso aos dados não tem custo e é disponibilizado somente para fins de pesquisa. Ainda

---

<sup>33</sup> Informação por e-mail em 21/01/2021.

segundo o ISPS (2020), quando possível, os dados são vinculados a projetos e publicações através do ISPS Knowledge Base, banco de dados que integra todos os produtos de pesquisa vinculados ao Instituto.

O depósito, acesso e uso de dados no ISPS Data Archive estão condicionados ao aceite dos seus **Termos de Uso**<sup>34</sup>. Outras políticas de uso do Repositório não foram localizadas, mas são oferecidas as seguintes diretrizes: de depósito; curadoria para reprodutibilidade dos dados; direitos e licenças de uso; acesso e confidencialidade dos dados; e suporte à pesquisa.

#### 4.1.15 MINDS@UW

O **MINDS@UW** é o repositório digital da University of Wisconsin-Madison (UW-Madison), localizada em Madison, no estado de Wisconsin, Estados Unidos. Fundada em 1848, a UW-Madison é uma universidade pública, multidisciplinar que segue em sua trajetória histórica a 'Wisconsin Idea'. Originada no discurso do ex-presidente da UW, Charles Van Hise, em 1905, a 'Wisconsin Idea' consiste na ampla atuação beneficente da Universidade em prol de todas as famílias do estado de Wisconsin.

Reconhecida como uma produtiva instituição de pesquisa, a UW-Madison está entre as 10 melhores universidades do país em volume de pesquisa, contando com orçamento anual de US\$ 1.2 bilhão para gastos em pesquisa. A Universidade conta, ainda, com um orçamento anual de US\$ 160 milhões destinados à sua Escola de Pós-Graduação, a qual oferece 150 cursos de mestrado e 120 cursos de doutorado (UNIVERSITY OF WISCONSIN-MADISON, 2021).

Amparando as atividades acadêmicas e de pesquisa na UW-Madison, entre os serviços oferecidos à sua comunidade, as bibliotecas da Universidade disponibilizam serviços e recursos de suporte à pesquisa, destacando-se os Serviços de Dados de Pesquisa (Research Data Services - RDS). O RDS é uma organização interdisciplinar comprometida com a prática de gerenciamento de dados de pesquisa na Universidade, buscando fornecer aos pesquisadores as ferramentas e recursos de apoio ao armazenamento, análise e compartilhamento de seus dados (UNIVERSITY OF WISCONSIN-MADISON, 2020).

---

<sup>34</sup> Disponível em: <https://isps.yale.edu/research/data/terms-of-use>

As bibliotecas da UW-Madison também oferecem os Serviços de Biblioteca Digital (Digital Library Services), por meio do qual disponibilizam acesso e armazenamento ao Multidisciplinary Institutional Network for Data and Scholarship, o MINDS@UW, repositório institucional da Universidade, sendo de responsabilidade do Centro de Coleções Digitais da Universidade de Wisconsin (UWDCC) a manutenção do *hardware*, *software* e interface da MINDS@UW.

Fundado em 2006, o MINDS@UW não é um repositório exclusivo para dados de pesquisa, mas para armazenar e preservar todo conteúdo nato digital ou digitalizado produzido pela comunidade da Universidade. Sua organização segue a estrutura da plataforma DSpace, utilizada para o gerenciamento dos conteúdos no Repositório, sendo estes arranjados em comunidades e coleções DSpace.

O MINDS@UW está organizado em 15 comunidades<sup>35</sup> subdivididas em subcomunidades e coleções. Em pesquisa no Repositório pela expressão '*dataset*', recuperou-se um total de 781 itens.

As políticas do MINDS@UW e informações sobre as melhores práticas para depósito no Repositório podem ser encontradas em<sup>36</sup>: **Políticas de depósito MINDS@UW; Guia MINDS@UW para depositar seu trabalho; MINDS@UW Melhores Práticas; Licença de Distribuição MINDS@UW; Autorização de uso de Proxy MINDS@UW; MINDS@UW de Acesso Aberto Versão flip; e Upload para MINDS@UW Vídeo.**

#### 4.1.16 UC San Diego Library Digital Collections

O **UC San Diego Library Digital Collections** é o serviço de repositório das coleções digitais da University of California San Diego (UCSD), campus de San Diego, sendo a sétima mais antiga das dez universidades semiautônomas da University of Califórnia, nos Estados Unidos. Inicialmente denominada University of California La Jolla, a UC San Diego, fundada oficialmente em 1960, tem suas origens na virada do Século XX, com a criação do Scripps Institution of Oceanography, em La Jolla, California, considerado um dos mais antigos centros de pesquisa em Oceanografia (UC SAN DIEGO, 2021a).

---

<sup>35</sup> Dados coletados em 08/01/2021.

<sup>36</sup> Disponível em: <https://www.library.wisc.edu/digital-library-services/minds/forms-and-materials/>

Contando anualmente com um financiamento em pesquisa superior a US\$ 1 bilhão, a UC San Diego é uma universidade multidisciplinar focada na inovação do país. Responsável por um impacto econômico no estado da Califórnia de US\$ 16.5 bilhões, a Universidade desempenha um papel importante na saúde fiscal da região, impulsionando empregos e o orçamento dos governos local e estadual (UC SAN DIEGO, 2021b).

Desempenhando um papel fundamental de suporte à missão da Universidade, a Biblioteca da UC San Diego (The Library UC San Diego) destaca-se como uma das 25 maiores bibliotecas acadêmicas públicas do país. O ambiente disponibiliza as Coleções Digitais da Biblioteca (*Library Digital Collections*), que dão acesso a duas grandes categorias de coleções digitais: as coleções de versões digitalizadas de documentos selecionados em distintas áreas do conhecimento e as coleções de dados de pesquisa gerados por seus pesquisadores, apoiados pelo Programa de Curadoria de Dados de Pesquisa da Instituição.

O *website* das Coleções Digitais da Biblioteca dá acesso ao Repositório, onde se encontram: as coleções de dados de pesquisa com opções de busca para recuperar os dados; informações sobre Curadoria de dados de pesquisa<sup>37</sup>, abordando, entre outras orientações, noções sobre preservação, armazenagem e preparação na Biblioteca, termo de contrato de depósito no Repositório; página para perguntas frequentes; dicas de pesquisa; e a Política de aviso e retirada da Library Digital Collections.

O UC San Diego Library Digital Collections não é um repositório exclusivo para dados de pesquisa, e, sendo assim, identificar os requisitos estabelecidos para análise das políticas não foi uma tarefa plenamente exitosa. Algumas informações, para atender aos requisitos investigativos da presente pesquisa, foram coletadas por meio de efetiva pesquisa no *site* do Repositório ou identificadas com base em informações recebidas da UC San Diego via *e-mail*. Entre essas informações, está a data de implementação do Repositório. Segundo resposta da UC San Diego, a implantação teve início em 2005, mas somente em 2010 que o sistema teve seu acesso disponibilizado ao público.

---

<sup>37</sup> Disponível em: <https://library.ucsd.edu/research-and-collections/data-curation/> Acesso em: 07 set. 2020.

No *site* da Biblioteca, identificaram-se como diretrizes de uso do Repositório: o **Guia de uso das coleções digitais da Universidade**<sup>38</sup>, o **Termo de contrato de depósito no Repositório**<sup>39</sup> e a **Política de aviso e retirada do Digital Collections (DC)**<sup>40</sup>.

#### 4.1.17 Data Repository for the University of Minnesota (DRUM)

O **Data Repository for the University of Minnesota (DRUM)** é o repositório multidisciplinar de dados de pesquisa da University of Minnesota (U of M), localizada no estado de Minnesota, Estados Unidos. Com suas origens no ano de 1851, a University of Minnesota veio a se tornar uma das preeminentes universidades públicas dos Estados Unidos.

A Universidade se constitui em um sistema de cinco campi estaduais distintos, cada um com suas especificidades e pontos fortes, sendo eles: o Crookston Campus, pioneiro no país na educação *on-line* e à distância e conhecido por seu foco em aprendizagem experimental e mentoria de professores; o Duluth Campus, com uma reputação global de recursos naturais e pesquisa em água doce, além de forte ênfase no meio ambiente e sustentabilidade; o Morris Campus, com foco nas artes liberais; o Rochester Campus, com foco na área de ciências da saúde; e o Twin Cities Campus, campus principal da U of M e que se encontra entre um dos cinco campi universitários do país com escolas de engenharia, medicina, medicina veterinária, direito e agricultura em um único campus (UNIVERSITY OF MINNESOTA, 2020).

Tendo por missão o desenvolvimento do aprendizado e o compartilhamento com diversificadas comunidades do conhecimento adquirido por meio da educação, em benefício do Estado, da nação e do mundo, a Universidade tem um importante impacto econômico e social para Minnesota, contribuindo com mais de US\$ 8,6 bilhões anuais<sup>41</sup> em atividade econômica para o estado. Igualmente na casa dos bilhões, está o investimento anual de US\$ 1,01 bilhão que a Universidade faz em pesquisa.

<sup>38</sup> Disponível em: <https://ucsd.libguides.com/dc>

<sup>39</sup> Disponível em: <https://library.ucsd.edu/research-and-collections/data-curation/deposit-term-agree.html>

<sup>40</sup> Disponível em: <https://library.ucsd.edu/dc/p/takedown>

<sup>41</sup> Dados coletados em 10/01/2021 em: <https://twin-cities.umn.edu/research>

Apoiando a missão da U of M em educação, pesquisa e serviço público, estão as premiadas Bibliotecas da Universidade, consideradas um dos maiores ativos intelectuais da Universidade e do estado de Minnesota. Entre os serviços oferecidos à comunidade acadêmica da Universidade, encontra-se o acesso ao repositório de dados da Instituição, o DRUM.

O DRUM é um repositório de dados de acesso aberto com curadoria total de dados de pesquisa digital gerados por pesquisadores, estudantes e funcionários da U of M, gerenciado pelas Bibliotecas da Universidade. Lançado em 2014 pela U of M, o serviço foi desenvolvido e implementado por um grupo de bibliotecas chamado Data Management and Curation Initiative (DMCI), com patrocínio de John Butler, bibliotecário universitário para dados e tecnologia (UNIVERSITY OF MINNESOTA LIBRARIES, 2015). Todos os conjuntos de dados submetidos ao DRUM são revisados por uma equipe de curadoria de dados, a fim de garantir que eles estejam em um formato e estrutura que melhor facilitem o seu acesso, descoberta e reutilização a longo prazo.

O Repositório utiliza o DSpace para gerenciamento dos dados, seguindo, assim, a estrutura de armazenamento dos conteúdos da Plataforma, organizada em comunidades e coleções. As comunidades DSpace identificadas no DRUM estão relacionadas a cada um dos campi universitários, sendo eles: University of Minnesota Crookston; University of Minnesota Duluth; University of Minnesota Rochester; e University of Minnesota Twin Cities. Somente o campus de Morris ainda não tem uma comunidade incluída no Repositório, provavelmente devido a sua área de ensino, as artes liberais.

Em pesquisa realizada no DRUM<sup>42</sup> por *dataset*, no geral, 6.342 conjuntos de dados foram recuperados. Ademais, ao realizar-se a busca por *dataset* dentro das comunidades, os dados foram recuperados em: University of Minnesota (504); University of Minnesota Crookston (zero); University of Minnesota Duluth (243); University of Minnesota Rochester (zero); e em University of Minnesota Twin Cities (5733).

Em relação às políticas do Repositório, identificou-se que, devido ao fato de o DRUM estar inserido no Programa de Conservação Digital da Universidade, o Repositório está sujeito às políticas de conservação digital<sup>43</sup> e, também, às **Políticas**

---

<sup>42</sup> Dados coletados em 10/01/2021.

<sup>43</sup> Disponível em: <https://conservancy.umn.edu/pages/policies/>

**relacionadas ao depósito**<sup>44</sup> no DRUM, as quais envolvem: a Política de coleta de dados; a Política de acesso ao usuário final (como as pessoas podem usar os dados); a Política de preservação; a Licença de depósito; e as **Políticas relacionadas ao uso**<sup>45</sup> de materiais armazenados no DRUM, que abrangem os Termos de uso e a Política de acesso ao usuário final.

#### 4.1.18 UCL Research Data Repository (RDR)

O **UCL Research Data Repository (RDR)** é o repositório de dados de pesquisa da University College London (UCL), localizada em Londres, na Inglaterra. Fundada em 1826, a UCL é a principal universidade multidisciplinar de Londres, constituída por 11 faculdades, as quais abrangem as diferentes áreas do conhecimento. Com um histórico de excelência acadêmica e sendo consistentemente classificada entre as melhores universidades do mundo, a Universidade possui um corpo discente representativo de mais de 150 nacionalidades, distribuído entre os 440 programas de graduação e os 675 programas de pós-graduação.

Tendo o pensamento inovador como o *status quo* da Instituição desde 1826, a UCL tem sido pioneira na Inglaterra no que diz respeito aos seguintes aspectos: ao receber estudantes de qualquer religião ou origem social; ao receber mulheres na educação universitária; ao ensinar inglês, alemão, química e engenharia; ao ter uma imprensa universitária de acesso totalmente aberto; e ao ter uma abordagem corajosa e progressista até os dias atuais (UCL, 2021).

Dando suporte à aprendizagem, ao ensino e à pesquisa, a UCL conta com o apoio e serviços de suas 16 bibliotecas e espaços de aprendizagem localizados em Londres, abrangendo uma ampla gama de assuntos especializados. Entre os serviços de biblioteca oferecidos como suporte à pesquisa, são disponibilizadas à comunidade da UCL orientações sobre gestão de dados de pesquisa.

A Universidade disponibiliza aos seus funcionários e pesquisadores o seu repositório de dados de pesquisa, o RDR. Lançado em 05 de junho de 2019, o RDR utiliza o Fligshare como provedor de interface e é gerenciado pela equipe da Research Data Management (RDM), composta por profissionais especializados no suporte a dados de pesquisa.

---

<sup>44</sup> Disponível em: <https://conservancy.umn.edu/pages/drum/policies/#end-user-access-policy>

<sup>45</sup> Disponível em: <https://conservancy.umn.edu/pages/drum/policies/#end-user-access-policy>

Um diferencial do RDR está na possibilidade de filtrar a pesquisa pelo tipo de licença de uso indicada para cada conjunto de dados. Entre as licenças indicadas no filtro, estão as Creative Commons, Apache, MIT e GPL 3.0+.

A Política de Dados de Pesquisa da UCL<sup>46</sup>, com previsão de ser revista, no mínimo, a cada três anos, tem por objetivo:

[...] fornecer uma estrutura para definir as responsabilidades de todos os membros da UCL e para orientar pesquisadores e alunos sobre como gerenciar seus dados, permitindo que os dados de pesquisa sejam mantidos e preservados como um objeto de pesquisa de primeira classe em seu próprio direito e disponibilizado ao maior público possível para o maior impacto possível (AYRIS, 2018, p. 2, tradução livre da autora).

Basicamente, a Política, coordenada e implementada pelo Diretor de Serviços de Biblioteca e pelo Gerente de Registros da UCL, institui as responsabilidades em relação ao gerenciamento dos dados de pesquisa pela comunidade da UCL, envolvendo os estudantes e seus orientadores, os pesquisadores, as suas bibliotecas e os serviços de suporte e governança aos dados de pesquisa, bem como os chefes de departamentos, reitor e vice-reitor de pesquisa. Discrimina que a Política de Dados de Pesquisa não substitui nenhuma outra política da UCL e menciona que, se houver algum conflito direto com as políticas dos financiadores de pesquisa, estas terão precedência sobre a Política da UCL. Igualmente, relaciona outras três políticas da UCL: a **Política de Proteção de Dados UCL**<sup>47</sup>, a **Política de Segurança da Informação UCL**<sup>48</sup>, a **Declaração UCL sobre Integridade de Pesquisa e a Política de Gerenciamento de Registros UCL**<sup>49</sup>.

A Política não faz nenhuma menção às permissões, restrições e licenças para o acesso, compartilhamento e reuso de dados armazenados no Repositório, assim como não menciona sobre a procedência e sobre as permissões e restrições para acesso a dados, nem aborda informações sobre licenças e embargos de dados. No *site* do repositório, não há informações sobre suas políticas. Algumas dessas informações estão em *links* indicados em páginas dos serviços oferecidos pelas Bibliotecas da UCL, nos termos de uso do UCL Figshare e em 'FAQ'. Na página dos serviços de Bibliotecas da UCL, em Gestão de Dados de Pesquisa, há referência à

---

<sup>46</sup> Versão 5.7 de 12 dez. 2018, de autoria de Paul Ayris. Disponível em:

[https://www.ucl.ac.uk/isd/sites/isd/files/uclresearchdatapolicy\\_2018.pdf](https://www.ucl.ac.uk/isd/sites/isd/files/uclresearchdatapolicy_2018.pdf)

<sup>47</sup> Disponível em: [www.ucl.ac.uk/informationsecurity/policy/public-policy/DataProtectionPolicy1016.pdf](https://www.ucl.ac.uk/informationsecurity/policy/public-policy/DataProtectionPolicy1016.pdf)

<sup>48</sup> Disponível em: <https://www.ucl.ac.uk/information-security/>

<sup>49</sup> Disponível em: <https://www.ucl.ac.uk/library/about-us/records-office/university-college-london-records-management-policy>

Política de Privacidade do Serviço do Repositório de Dados<sup>50</sup>, a qual comunica sobre: a privacidade das informações coletadas ao se acessar o *site* da UCL; a privacidade de dados processados por estudantes da UCL; e a privacidade dos dados coletados dos usuários e processados por membros da equipe de funcionários da UCL.

Igualmente na página dos serviços oferecidos pelas Bibliotecas, em ‘Benefícios do repositório’, indica-se a utilização de licenças Creative Commons e sinaliza-se sobre a possibilidade de aplicação de embargos, remetendo às ‘Diretrizes sobre como aplicar um embargo ao depósito’ no Repositório<sup>51</sup>. Na mesma página, em ‘Informações importantes’, ao tratar-se sobre as responsabilidades legais sobre o depósito de dados no Repositório, remete-se às páginas onde encontram-se informações sobre os ‘Termos de uso do RDR’<sup>52</sup>, ‘Proteção de dados’<sup>53</sup> e ‘Direito autoral’<sup>54</sup>. Na mesma página, orientando sobre os dados que não devem ser depositados no RDR, faz-se *link* para informações sobre Propriedade Intelectual.<sup>55</sup>

## 4.2 CONSIDERAÇÕES SOBRE OS REPOSITÓRIOS

Nessa apresentação e contextualização das 18 universidades e seus respectivos repositórios, pôde-se observar as especificidades das distintas instituições em relação as suas estruturas organizacionais e a forma de gerenciamento das suas produções científicas e acadêmicas em seus repositórios. No tangente ao gerenciamento de dados de pesquisa produzidos nas Universidades, constatou-se que, entre as instituições analisadas, algumas implementaram repositórios exclusivos para armazenamento e preservação dos dados de pesquisa produzidos por sua comunidade científica, e outras, no entanto, passaram a fazer a gestão desses dados em seus repositórios institucionais. Há também aquelas que utilizam serviços externos de armazenamento e curadoria para o depósito de dados de pesquisa produzidos por seus pesquisadores, como a plataforma Dryad.

---

<sup>50</sup> Disponível em <https://www.ucl.ac.uk/isd/sites/isd/files/repository-privacy-notice-march2019.pdf>

<sup>51</sup> Disponível em: [https://www.ucl.ac.uk/library/sites/library/files/embargo-guidance\\_instructions\\_v2.pdf](https://www.ucl.ac.uk/library/sites/library/files/embargo-guidance_instructions_v2.pdf)

<sup>52</sup> Disponível em: <https://www.ucl.ac.uk/isd/sites/isd/files/figshare-eula-terms-conditions-ucl-26032019.pdf>

<sup>53</sup> Disponível em: <https://www.ucl.ac.uk/data-protection/data-protection-overview/understanding-data-protection-ucl>

<sup>54</sup> Disponível em: <https://www.ucl.ac.uk/library/ucl-copyright-advice/legislation-and-background-information>

<sup>55</sup> Disponível em: <https://www.ucl.ac.uk/research-services/research-studentships/intellectual-property>

A partir da análise dos 18 repositórios pesquisados apresenta-se no Quadro 4, sua caracterização, conforme propósito de atuação e classificação indicada por Sayão e Sales (2016), já exposta no referencial teórico do presente estudo. Os repositórios estão numerados de acordo com sua ordem de apresentação na seção anterior.

**Quadro 4 - Propósito de atuação dos repositórios e data de sua fundação**

PROPÓSITO DE ATUAÇÃO	REPOSITÓRIOS	UNIVERSIDADES	DATA DE FUNDAÇÃO
Institucional	2_Stanford Digital Repository	Stanford University	2005
Institucional	3_DSpace @ MIT	Massachusetts Institute of Technology	2002
Institucional	5_Research Works Archive	University of Washington	-
Institucional	9_Columbia Academic Commons	Columbia University New York	2006
Institucional	10_eCommons	Cornell University	2002?
Institucional	11_Scholarly Commons	University of Pennsylvania	2005
Institucional	12_Apollo	University of Cambridge	2003
Institucional	15_MINDS@UW - Research Data Service	University of Wisconsin Madison	2006
Institucional	16_UC San Diego Library Digital Collections	University of California San Diego	2010
Institucional de dados de pesquisa	1_Harvard Dataverse	Harvard University	2007
Institucional de dados de pesquisa	6_Deep Blue Data	University of Michigan	2006
Institucional de dados de pesquisa	7_ORA-Data	University of Oxford	2014
Institucional de dados de pesquisa	8_JHU Data Archive	Johns Hopkins University	2015
Institucional de dados de pesquisa	13_Merritt	University of California Los Angeles UCLA	2009
Institucional de dados de pesquisa	14_ISPS Data Archive	Yale University	2009
Institucional de dados de pesquisa	17_DRUM	University of Minnesota System	2014
Institucional de dados de pesquisa	18_UCL Research Data Repository	University College London	2019
Multidisciplinar de dados de pesquisa	4_Data Repository [Berkeley]	University of California Berkeley	2009

Fonte: elaborado pela autora.

Dos 18 repositórios analisados, 17 são institucionais, haja vista estarem vinculados a instituições acadêmicas e de pesquisa. Somente um entre os 18 repositórios analisados não se enquadra nessa classificação, uma vez que se trata de um serviço oferecido por um repositório multidisciplinar externo à Universidade.

Entre os 17 repositórios institucionais, nove são destinados a armazenar e preservar toda a produção acadêmica e científica da Universidade, além de dados de pesquisa, sendo eles: o **SDR**; o **DSpace@MIT**; o **Research Works Archive**; o **Columbia Academic Commons**; o **eCommons**; o **Scholarly Commons**; o **Apollo**; o **MINDS@UW**; e o **UC San Diego**. Já os outros oito repositórios analisados são exclusivamente destinados aos dados de pesquisa, a saber: o **Harvard Dataverse**; o **DBD**; o **ORA-Data**; o **JHU Data Archive**; o **Merritt**; o **ISPS Data Archive**; o **DRUM**; e o **UCL Research Data Repository**.

Somente o **Data Repository**, da University of California, Berkeley, não se enquadra como repositório institucional, visto que a Universidade faz uso dos serviços do Dryad para que seus pesquisadores depositem seus dados de pesquisa, sendo este classificado como um repositório multidisciplinar de dados de pesquisa.

Conforme pôde ser visto no Quadro 4, as datas de implantação dos repositórios foram identificadas em seus *websites* ou em contato via e-mail com as instituições. Sendo assim, relembramos que: não obteve-se a data de fundação do **Research Works Archive**, embora tenha-se enviado *e-mail* para a Universidade, solicitando informações; para o **e-Commons**, tomou-se por data inicial das suas atividades o ano de 2002, haja vista ser essa a data que dá origem às primeiras estatísticas de *downloads* realizados no Repositório; não foi possível identificar nas políticas analisadas a data atribuída ao **ISPS Data Archive** contudo, obteve-se essa informação por contato via *e-mail* com a instituição; o **UC San Diego**, de acordo com as informações recebidas da Instituição por *e-mail*, foi implementado em 2005, mas, nesse período, ainda não se encontrava disponível ao público, somente no ano de 2010.

Como pôde-se observar no Quadro 4, a maioria dos repositórios analisados iniciou suas atividades na década de 2000, como uma decorrência das estratégias implementadas pelo Movimento de Acesso Aberto à literatura científica para a criação de repositórios digitais. Posteriores desdobramentos do Movimento, como o Open Archive Initiative (OAI), a Budapest Open Access Initiative – BOAI e o OpenDOAR, contribuíram para o crescente número de repositórios institucionais e temáticos espalhados pelo mundo.

Nas políticas analisadas, observou-se que estas apresentam especificidades de acordo com o propósito de atuação do repositório. Dentre os repositórios estudados, aqueles que não são exclusivos para dados de pesquisa, disponibilizam

políticas mais generalistas para os diferentes tipos de conteúdo digital provenientes da produção gerada pela comunidade universitária a qual estão vinculados. Já os RDPs, pelo seu próprio propósito de atuação (de conservação e difusão de dados de pesquisa), dispõem políticas específicas para dados de pesquisa, com ênfase no depósito, acesso, licenças e restrições aos dados no repositório, observando embargos e privacidade dos dados; e nas diretrizes de preservação desses dados.

Apresenta-se, no Quadro 5 a seguir, a sistematização das políticas identificadas nos *websites* dos repositórios analisados.

**Quadro 5 – Políticas identificadas nos *websites* dos Repositórios analisados**

REPOSITÓRIOS	UNIVERSIDADES	POLÍTICAS
1_Harvard Dataverse	Harvard University	<b>Termos Gerais de Uso</b> ( <i>Terms of Use</i> ); <b>Política de Privacidade</b> ( <i>Privacy Policy</i> ); <b>Política de Preservação</b> ( <i>Preservation Policy</i> ); <b>Termos de uso do API</b> ( <i>API Terms of Use</i> ); <b>Política de Contrato de uso de dados</b> ( <i>Sample Data Usage Agreement</i> ).
2_Stanford Digital Repository	Stanford University	<b>Visão geral do repositório</b> ( <i>SDR overview</i> ); <b>Benefícios do SDR</b> ( <i>SDR benefits</i> ); <b>Serviços ao depositante</b> ( <i>Depositor services</i> ); <b>Serviços DOI</b> ( <i>DOI services</i> ); <b>Termos de depósito do SDR</b> ( <i>Terms of Deposit</i> ).
3_DSspace@MIT	Massachusetts Institute of Technology	<b>Políticas da comunidade e da coleção</b> ( <i>Community &amp; Collection Policies</i> ); <b>Políticas/Diretrizes/Licenças gerais</b> ( <i>General Policies/Guidelines/Licenses</i> ); <b>Diretrizes para conjuntos de dados de pesquisa</b> ( <i>Guidelines for Research Datasets</i> ).
4_Data Repository [Berkeley]	University of California Berkeley	<b>Termos de Serviço do Dryad</b> ( <i>Terms of Service</i> ), destacando-se: as Políticas de Publicação; Uso; Privacidade; Disposições gerais.
5_Research Works Archive	University of Washington	<b>Política de Coleção Digital</b> ( <i>Digital Collection Policy</i> ); <b>Política de Preservação Digital</b> ( <i>Digital Preservation Policy</i> ); <b>Política de Retirada</b> ( <i>Withdrawal Policy</i> ); <b>Política de restrição de acesso</b> ( <i>Access Restriction Policy</i> ); <b>Política de Direitos Autorais</b> ( <i>Copyright and Author Rights Policy</i> ).
6_Deep Blue Data	University of Michigan	<b>Termos de uso do repositório</b> ( <i>Terms of Use</i> ); <b>Política de envio e depósito de dados</b> ( <i>Deposit Policy</i> ); <b>Política de coleções e conteúdo dos dados</b> ( <i>Collections &amp; Content</i> ); <b>Política de preservação de dados de pesquisa</b> ( <i>Preservation Policy</i> ).
7_ORA-Data	University of Oxford	<b>Condições de depósito</b> ( <i>ORA-Data Deposit Conditions</i> ); <b>Licença de usuário do ORA</b> ( <i>ORA User License</i> ); <b>Política de envio de dados</b> ( <i>ORA-Data Submission Policy</i> ); <b>Política de armazenamento e preservação de dados</b> ( <i>ORA-Data Retention &amp; Preservation Policy</i> ); <b>Política de retirada de dados</b> ( <i>ORA-Data Withdrawal Policy</i> ); <b>Política de contestação de dados</b> ( <i>ORA-Data Complaints Policy</i> ).
8_JHU Data Archive	Johns Hopkins University	<b>Guia do Usuário Dataverse</b> ( <i>User Guide</i> ); <b>Diretrizes dos Serviços de Dados das Bibliotecas da Universidade [Johns Hopkins Libraries]</b> ( <i>Archiving</i> ); <b>Termos de Uso do JHU Data Archive</b> ( <i>Terms of Use for the JHU Data Archive</i> ).
9_Columbia Academic Commons	Columbia University New York	<b>Diretrizes de participação</b> ( <i>Participation Guidelines</i> ); <b>Termos de Uso</b> ( <i>Terms of Use</i> ); <b>Copyright e Proteção de dados</b> ( <i>Data Protection</i> ).

<b>10_eCommons</b>	Cornell University	<p><b>Permissão de depósito</b> (<i>Who may submit content to eCommons</i>);</p> <p><b>Política de coleção de conteúdo</b> (<i>Content Collection Policy</i>);</p> <p><b>Política de depósito e licença</b> (<i>Deposit and Commons license</i>);</p> <p><b>Política de depósito de dados</b> (<i>Data deposit</i>);</p> <p><b>Política de acesso e uso</b> (<i>Access and use</i>);</p> <p><b>Política de retirada</b> (<i>Alteration and withdrawal</i>);</p> <p><b>Política de apoio à preservação</b> (<i>Preservation support</i>);</p> <p><b>Política de privacidade</b> (<i>Privacy</i>);</p> <p><b>Termos de Uso</b> (<i>Terms of Use</i>);</p> <p><b>Política de acessibilidade</b> (<i>Accessibility Policy</i>).</p>
<b>11_Scholarly Commons</b>	University of Pennsylvania	<p><b>Submissões apropriadas</b> (<i>Appropriate Submissions</i>);</p> <p><b>Tipos de formatos suportados</b> (<i>Supported Format Types</i>);</p> <p><b>Remoção de conteúdo</b> (<i>Removing Content</i>);</p> <p><b>Considerações de direitos autorais</b> (<i>Copyright Considerations</i>);</p> <p><b>Acesso e Migração</b> (<i>Access and Migration</i>);</p> <p><b>Considerações de privacidade</b> (<i>Privacy Considerations</i>).</p>
<b>12_Apollo</b>	University of Cambridge	<p><b>Contrato de Licença de Depósito</b> (<i>Deposit License Agreement</i>);</p> <p><b>Política de Preservação Digital</b> (<i>Digital Preservation Policy</i>);</p> <p><b>Termos de Uso do Repositório</b> (<i>Terms of Use</i>);</p> <p><b>Política do DOI</b> (<i>DOI Policy</i>).</p>
<b>13_Merritt</b>	University of California Los Angeles UCLA	<p><b>Termos de Uso e Política de Privacidade da CDL</b> (<i>Terms of Use and CDL Privacy Policy</i>);</p> <p><b>Termos de Serviço do Dryad</b> (<i>terms of Service</i>);</p> <p><b>Política de Depósito do (UC3)</b> (<i>Deposit Policy</i>).</p>
<b>14_ISPS Data Archive</b>	Yale University	<b>Termos de Uso</b> ( <i>Terms of Use</i> ).
<b>15_MINDS@UW Research Data Service</b>	University of Wisconsin Madison	<p><b>Políticas de depósito MINDS@UW</b> (<i>MINDS@UW Deposit Policies</i>);</p> <p><b>Guia MINDS@UW para depositar seu trabalho</b> (<i>MINDS@UW Guide to Depositing Your Work</i>);</p> <p><b>Melhores Práticas MINDS@UW</b> (<i>MINDS@UW Best Practices</i>);</p> <p><b>Licença de Distribuição MINDS@UW</b> (<i>MINDS@UW Distribution License</i>);</p> <p><b>Autorização de uso de Proxy MINDS@UW</b> (<i>MINDS@UW Proxy Authorization</i>);</p> <p><b>Acesso aberto versão Flip</b> (<i>MINDS@UW Open Access Zine Flip version</i>);</p> <p><b>Depósito para MINDS@UW vídeo</b> (<i>Uploading to MINDS@UW Video</i>).</p>
<b>16_UC San Diego Library Digital Collections</b>	University of California San Diego	<p><b>Guia de uso das coleções digitais da Universidade</b> (<i>Guide to Using the UC San Diego Digital Collections</i>);</p> <p><b>Termo de contrato de depósito no Repositório</b> (<i>Deposit Terms of Agreement</i>);</p> <p><b>Política de aviso e retirada do Digital Collections (DC)</b> (<i>Take Down Policy</i>).</p>
<b>17_DRUM</b>	University of Minnesota System	<p><b>Políticas relacionadas ao depósito no DRUM:</b></p> <p>Política de submissão de dados (<i>Data Collection Policy</i>);</p> <p>Política de acesso ao usuário final (<i>End-user Access Policy</i>);</p> <p>Política de preservação (<i>Preservation Policy</i>);</p> <p>Licença de depósito (<i>Deposit License</i>).</p> <p><b>Políticas relacionadas ao uso de materiais armazenados no DRUM:</b></p> <p>Termos de uso (<i>Terms of Use</i>);</p> <p>Política de acesso ao usuário final (<i>End-user Access Policy</i>).</p>
<b>18_UCL * Research Data Repository</b>	University College London	<p><b>Política de Dados de Pesquisa da UCL</b> (<i>UCL's Research Data Policy</i>);</p> <p><b>Política de Proteção de Dados UCL</b> (<i>UCL Data Protection Policy</i>);</p> <p><b>Política de Segurança da Informação UCL</b> (<i>Information Security Policy</i>);</p> <p><b>Declaração UCL sobre Integridade de Pesquisa</b> (<i>UCL Statement on Research Integrity</i>);</p> <p><b>Política de Gerenciamento de Registros UCL</b> (<i>UCL Records Management Policy</i>).</p>

Fonte: elaborado pela autora.

\*Nota: no *site* do Repositório, não constam as políticas. Essas informações estão disponíveis nas páginas dos serviços oferecidos pelas bibliotecas da UCL.

Como pôde ser observado no Quadro 5, enquanto alguns repositórios apresentam um rol de normativas para acesso e uso dos serviços que oferecem, outros repositórios disponibilizam um sucinto conjunto de políticas. Mesmo que não estejam presentes em todos os conjuntos de diretrizes dos 18 repositórios analisados, os documentos que mais aparecem nas políticas analisadas são os '**Termos de uso**' e a '**Política de depósito ou envio de dados**'.

Na seção seguinte, são apresentados os resultados obtidos a partir das análises das políticas identificadas nos *websites* dos repositórios pesquisados quanto aos aspectos: proveniência dos dados para depósito; tipos de acesso aos dados de pesquisa; tipo de restrições de acesso aos dados; permissões para depósito de dados; restrições para depósito de dados; restrições impostas ao tipo de dados para depósito; permissões e restrições para uso de dados; licenças de uso dos dados; privacidade de dados pessoais e anonimização de dados; embargo de dados; e responsabilidades no uso de dados.

### 4.3 ANÁLISE DAS POLÍTICAS

Os resultados das análises obtidas estão em consonância com o objetivo geral da pesquisa, que busca identificar, nas políticas dos repositórios universitários elencados para o estudo, informações sobre permissões, restrições e licenciamento para o depósito, acesso e uso de dados de pesquisa. Embora todos os repositórios estudados sejam de instituições universitárias e de acesso aberto, apresentam especificidades em relação aos aspectos analisados. A seguir, dando início às análises, pode-se observar as informações coletadas sobre a proveniência de dados para depósito nos repositórios.

#### 4.3.1 Análise quanto à proveniência dos dados para depósito

Buscando caracterizar as diretrizes para depósito de dados nos repositórios, identificou-se a proveniência desses dados em cada repositório analisado, conforme pode ser visto no Quadro 6.

**Quadro 6 – Proveniência dos dados para depósito**

Repositórios	Proveniência dos dados para depósito
1_Harvard Dataverse	Dados provenientes das atividades científicas de qualquer disciplina de dentro ou de fora de Harvard.
2_Stanford Digital Repository - SDR	Dados provenientes da produção acadêmica e de pesquisas da Universidade e de instituições parceiras das Bibliotecas da Stanford.
3_DSpace @ MIT	Dados provenientes das atividades acadêmicas e científicas do corpo docente e de pesquisadores do MIT.
4_Data Repository [Berkeley]	Dados provenientes das atividades de pesquisadores afiliados à UC Berkeley.
5_Research Works Archive	Dados provenientes da produção acadêmica e de pesquisas da Universidade.
6_Deep Blue Data	Dados provenientes das pesquisas feitas na Universidade de Michigan.
7_ORA-Data	Dados provenientes das atividades de pesquisa de pesquisadores ou professores da Oxford University.
8_JHU Data Archive	Dados provenientes das atividades de pesquisa de membros da Johns Hopkins University.
9_Columbia Academic Commons	Dados provenientes de pesquisas e bolsas de estudos criadas na Columbia University e em suas instituições afiliadas.
10_eCommons	Dados provenientes da produção intelectual de professores, pesquisadores, funcionários e estudantes da Universidade de Cornell, ou de pesquisadores patrocinados por um membro atual do corpo docente ou funcionário da Cornell.
11_Scholarly Commons	Dados provenientes da produção intelectual da comunidade acadêmica da Universidade.
12_Apollo	Dados provenientes de pesquisas vinculadas à Universidade de Cambridge.
13_Merritt	Dados provenientes da Comunidade da University of California.
14_ISPS Data Archive	Dados provenientes de estudos produzidos por afiliados da Yale.
15_MINDS@UW	Dados provenientes das atividades de pesquisa de pesquisadores afiliados à University of Wisconsin Madison.
16_UCSanDiego	Dados provenientes das atividades de pesquisa de pesquisadores do campus de San Diego.
17_DRUM	Dados provenientes das atividades de pesquisas geradas por pesquisadores, estudantes e funcionários da University of Minnesota.
18_UCL Research Data Repository	Dados provenientes das atividades de pesquisa de pesquisadores da University College London e de projetos colaborativos, desde que um dos membros seja pesquisador da UCL.

Fonte: elaboração da autora.

Na análise das políticas, constatou-se que a proveniência dos dados para depósito nos repositórios, como pode ser observado no Quadro 6, em sua maioria, é exclusivamente de pesquisas oriundas da produção acadêmica e de pesquisas das universidades, contudo, alguns repositórios aceitam, igualmente, depósitos de dados de pesquisa provenientes de atividades científicas externas ao seu ambiente institucional, a saber: o **Harvard Dataverse**, da Harvard University, aceita dados de pesquisa das atividades científicas de qualquer disciplina, independentemente do pesquisador ter vínculo com a Universidade; o **Stanford Digital Repository (SDR)**, repositório da Stanford University, permite que instituições parceiras das Bibliotecas

da Stanford depositem no SDR; o **Columbia Academic Commons**, da Columbia University New York, possibilita que instituições afiliadas às Bibliotecas da Universidade também depositem dados de pesquisa; já o **eCommons**, da Cornell University, o **MINDS@UW**, da University of Wisconsin Madison, e o **UCL Research Data Repository**, da University College London, permitem o depósito de pesquisadores externos, desde que o conteúdo a ser depositado seja patrocinado ou em coautoria com um pesquisador afiliado à instituição.

Ressalta-se que, embora, os campi de Berkeley, Los Angeles e San Diego façam parte do sistema de universidades semiautônomas da University of California, cada campus faz o gerenciamento dos dados de pesquisa de seus pesquisadores. Todos os campi da University of California, por meio de contrato, têm um espaço de armazenamento destinada para sua Comunidade no Dryad. Em busca no Dryad, objetivando verificar como se dá a organização dessas Comunidades, identificou-se que, por meio de pesquisa por "Instituição", é possível recuperar dados de pesquisa provenientes dos campi University of California Berkeley, University of California Los Angeles, University of California Santa Cruz, University of California Irvine, University of California Riverside, University of California Santa Barbara, University of California San Diego, University of California San Francisco e University of California Merced.

A identificação da proveniência dos dados para depósito permite, por assim dizer, traçar o DNA desses dados, uma vez que faz um registro da sua origem, autoria e vínculo institucional, o que, para Freund, Sembay e Macedo (2019, p. 1), "[...] pode ser considerado um requisito importante para estabelecer confiabilidade e prover segurança em sistemas computacionais de informação."

Fazendo uma relação da procedência dos dados em repositórios com os critérios de seleção de uma obra, na área da Ciência da Informação, tem-se que, nos dois contextos, identificar a origem de uma obra ou de dados de pesquisa garantem confiabilidade a estes, pois considera-se a reputação de autores, pesquisadores e instituições. Referente a publicações científicas, não é diferente; pesquisadores renomados garantem credibilidade às pesquisas, bem como instituições renomadas.

O compartilhamento de dados de pesquisa ainda é uma prática em construção que requer instrumentos que a validem, garantindo-lhe confiabilidade. A identificação nas políticas dos repositórios e da proveniência dos dados aceitos para depósito pode ser um desses instrumentos, uma vez que, de acordo com Almeida (2015, p. 1), "[...] a proveniência de dados vem se tornando cada vez mais presente no ambiente

científico, tanto para garantir a origem dos dados como para avaliar a sua acurácia.” Jorge (2018) vem ao encontro do pensamento de Almeida, afirmando que “a compreensão da origem dos dados está no cerne do processo científico, pois os pesquisadores analisam cuidadosamente os dados (e qualquer documentação associada) para entender a evolução e a metodologia por trás dos dados.”

A seguir, passa-se a analisar os tipos de acesso aos dados de pesquisa identificados nas políticas dos repositórios analisados.

#### 4.3.2 Análise dos tipos de acesso aos dados nos repositórios

Embora os repositórios analisados representem práticas de acesso aberto a dados de pesquisa, procurou-se identificar nas políticas analisadas a existência de diretrizes diferentes de acesso aos dados. Para essa identificação, levou-se em consideração os quatro tipos de direito de acesso a documentos digitais sinalizados por Repositórios Científicos de Acesso Aberto de Portugal (2021), apresentados no referencial do presente estudo, ou seja: acesso aberto, acesso fechado, acesso restrito e acesso embargado.

No Quadro 7, pode-se observar o tipo de acesso aos dados de pesquisa em cada um dos repositórios pesquisados, conforme descrito em suas políticas.

**Quadro 7 – Tipos de acesso aos dados**

Repositórios	Tipos de acesso aos dados para pesquisa no repositório
1_Harvard Dataverse	Acesso aberto e gratuito na internet.
2_Stanford Digital Repository - SDR	Acesso aberto conforme as licenças estabelecidas pelo depositante.
3_DSpace @ MIT	Acesso aberto.
4_Data Repository [Berkeley]	Acesso aberto público.
5_Research Works Archive	Acesso aberto e gratuito na internet.
6_Deep Blue Data	Acesso aberto e gratuito na internet.
7_ORA-Data	Acesso aberto.
8_JHU Data Archive	Acesso aberto ao público.
9_Columbia Academic Commons	Acesso aberto e gratuito.
10_eCommons	Por padrão, todo conteúdo depositado é de acesso aberto e gratuito via <i>Web</i> .
11_Scholarly Commons	Acesso aberto e gratuito.

12_Apollo	Foco no acesso aberto, mas pode fornecer diferentes níveis de acesso.
13_Merritt	Disponível para a Comunidade da UC. Permite acesso como convidado para acessar coleções específicas.
14_ISPS Data Archive	A maioria dos arquivos são arquivos de uso público, sem restrições de acesso.
15_MINDS@UW	Acesso aberto ao público.
16_UCSanDiego	Acesso aberto para fins educacionais de pesquisa, sem necessidade de cadastro.
17_DRUM	Proporciona opções de acesso flexível aos dados: <b>acesso aberto</b> imediato aos dados para todos e <b>acesso moderado</b> aos dados mediante solicitação por <i>e-mail</i> ao autor.
18_UCL Research Data Repository	Acesso livre via internet para pesquisadores e público em geral sem necessidade de cadastro. O RDR, no momento, não tem um controle de quem acessa os dados.

Fonte: elaboração da autora.

Quanto aos tipos de acesso aos repositórios, seguindo a tipologia apresentada por Repositórios Científicos de Acesso Aberto de Portugal (2021), pode-se constatar que: dos 18 repositórios analisados, 15 são de acesso aberto ao público via *Web*, e três tem acesso diferenciado, sendo eles: o **Apollo**, da University of Cambridge, o qual, embora tenha por foco principal fornecer acesso aberto, fornece diferentes níveis de acesso aos conteúdos depositados; o **Merritt**, da University of California (UC), voltado para a Comunidade da UC, mas que permite acesso com *status* de convidado para acessar coleções específicas disponibilizadas em uma lista, podendo ser considerado de acesso restrito para algumas coleções; o **DRUM**, do University of Minnesota System, que proporciona opções de acesso flexível aos dados, como acesso aberto imediato aos dados para o público em geral e acesso moderado aos dados mediante solicitação por *e-mail* ao autor. O acesso moderado enquadra-se em acesso embargado ou restrito, levando-se em consideração as características dos tipos de acesso apresentadas por Repositórios Científicos de Acesso Aberto de Portugal (2021) no referencial teórico deste estudo.

Embora o Apollo, na 'Visão geral' do Repositório apresentada na página da comunicação acadêmica da Universidade, afirme que o Repositório fornece diferentes níveis de acesso aos conteúdos depositados, especificamente sobre o acesso aos dados de pesquisa, não se localizou, nas Políticas, nada a respeito desses diferentes níveis. Acredita-se, portanto, que essa referência esteja relacionada a outros tipos de conteúdo armazenado no Apollo, haja vista ser um repositório institucional que armazena e fornece acesso aos resultados das atividades de pesquisas da Universidade.

É importante ressaltar que o Merritt não é um repositório exclusivo da University of California Los Angeles, mas sim um serviço mantido pelo Centro de Curadoria da Universidade da Califórnia, disponível como uma opção aos seus membros para depósito de seus dados de pesquisa. Contudo, o que se observou é que, dentre os campi da University of California apresentados na presente pesquisa, o campus de Los Angeles é o único que faz menção à utilização do Merritt como opção de depósito de dados. Os campi de Berkeley e San Diego, por meio dos serviços de curadoria de dados de suas bibliotecas, remetem respectivamente ao Dryad e à Library Digital Collections.

Também é relevante observar que, no *website* do Merritt, o acesso a todos os dados é permitido somente aos membros da University of California, e acesso à convidados para dados listados no *site*. Contudo, por meio do Dryad, os usuários têm acesso à Comunidade da University of Los Angeles e a todas as outras universidades da UC.

Partindo da premissa de que pesquisas desenvolvidas em universidades públicas e financiadas com recursos públicos devam permitir livre acesso aos seus resultados, salvo impeditivos legais, entende-se que os repositórios universitários devam garantir o seu livre acesso. Os dados de pesquisa, a partir do movimento de sua abertura, entram nesse rol, devido ao entendimento de que não só as pesquisas devam estar acessíveis, mas também os dados gerados por elas. Nesse contexto, Carvalho (2018) sugere que, embora algum aspecto possa limitar sua visibilidade, em decorrência da disciplina a qual esteja vinculado, salvo questões éticas, morais e autorais, os dados de pesquisas devem ser depositados em repositórios de acesso aberto.

A universidade, como produtora de conhecimento e grande geradora de conhecimento, ao implementar repositórios de dados de pesquisa, deve assegurar que o público tenha acesso à grande variedade de documentos produzidos por sua comunidade. De acordo com Monteiro (2018, p. 162), os repositórios “[...] contribuem para assegurar que os dados sejam publicados e disponibilizados para a comunidade científica com o menor número possível de restrições.”

Dando continuidade às análises, a próxima seção trata das observações relativas ao tipo de restrições de acesso aos dados, identificadas nas políticas dos repositórios.

### 4.3.3 Análise quanto a restrições de acesso aos dados nos repositórios

Como pôde-se constatar na análise anterior, a maioria dos repositórios são apresentados em suas políticas como sendo de acesso aberto, contudo pode-se observar no Quadro 8 que, mesmo entre esses repositórios, alguns indicam a possibilidade de imposição de restrições ao acesso dos dados.

**Quadro 8 – Restrições de acesso aos dados**

Repositórios	Restrições de acesso aos dados no repositório
1_Harvard Dataverse	O depositante pode definir restrições de acesso e termos de uso específicos para uso dos dados.
2_Stanford Digital Repository - SDR	O depositante pode definir níveis apropriados de acesso aos dados.
3_DSpace @ MIT	As restrições estão relacionadas a embargos ou a algum tipo de contrato de patrocínio (com prazo estabelecido).
4_Data Repository [Berkeley]	Permite restrições de acesso a um item ou a uma coleção em diferentes níveis.
5_Research Works Archive	O depositante pode definir restrições de acesso, em diferentes níveis, aos itens individuais ou a uma coleção.
6_Deep Blue Data	Não menciona restrições.
7_ORA-Data	O depositante pode estabelecer se os dados estarão ou não acessíveis ao público.
8_JHU Data Archive	Sem restrições.
9_Columbia Academic Commons	Não menciona restrições.
10_eCommons	Permite impor restrições temporárias ao acesso aberto mundial, podendo ser restrito aos membros da comunidade de Cornell temporária ou indefinidamente.
11_Scholarly Commons	Não localizado.
12_Apollo	Os depositantes são encorajados a depositar em formato aberto, contudo o repositório aceita arquivos proprietários, o que pode ocasionar algum tipo de restrição.
13_Merritt	Restrições aos usuários externos à University of California e a conjuntos de dados do Dryad em revisão por pares (embargo de até seis meses). Metadados permanecem disponíveis.
14_ISPS Data Archive	Restrições para arquivos com informações confidenciais ou identificáveis. Restrições por parte de financiadores ou entidades. Pesquisadores podem solicitar acesso a arquivos restritos por meio de contato com o ISPS.
15_MINDS@UW	Permite restrição de acesso aos dados, mas não aos metadados, embora a orientação seja de disponibilização do conteúdo em acesso aberto.
16_UCSanDiego	Não menciona restrições.
17_DRUM	Não menciona restrições.
18_UCL Research Data Repository	As restrições se dão aos dados embargados.

Fonte: elaborado pela autora.

Em relação às restrições de acesso aos dados nos repositórios, observou-se que, entre as políticas analisadas nos 18 repositórios selecionados, 12 indicam algum tipo de restrição de acesso aos dados de pesquisa e somente seis não mencionam restrições ou não foram localizadas informações a respeito, a saber: o **Deep Blue Data**, da University of Michigan; o **JHU Data Archive**, da Johns Hopkins University; o

**Columbia Academic Commons**, da Columbia University New York; o **Scholarly Commons**, da University of Pennsylvania; o **UCSan Diego**, da University of California San Diego; e o **DRUM**, do University of Minnesota System.

Realizou-se um teste aleatório nos repositórios acima citados, buscando-se confirmar se há algum tipo de restrição de acesso aos dados. Pôde-se constatar que: no **Deep Blue Data**, dos 30 *datasets* recuperados de um total de 333 resultados, todos são de acesso aberto, permitindo *download* imediato; no **JHU Data Archive**, dos 20 *datasets* recuperados de um total de 112, constatou-se que há restrição de acesso para determinados dados, estando indicado, nos termos de acesso ao conjunto de dados, que um dos seus arquivos é restrito e deve ter o acesso solicitado (*request access*); no **Columbia Academic Commons**, num total de 20 *datasets* recuperados, todos são de acesso aberto; no **Scholarly Commons**, não se identificou restrição de acesso nos 43 resultados recuperados, estando todos disponíveis para *download* imediato; no **UC San Diego**, foram recuperados dados restritos em que constam as informações: ‘conteúdo não disponível’, ‘acesso restrito somente à Universidade’ e ‘o acesso pode ser concedido para fins de pesquisa a critério da Biblioteca da UC San Diego’; por fim, no **DRUM**, não foram identificadas restrição de acesso aos dados recuperados.

Com relação aos tipos de restrições indicadas nas políticas dos outros 12 repositórios analisados, tem-se restrições referentes: a embargos impostos por questões contratuais entre financiadores e pesquisadores; a embargos temporários de acesso aberto mundial, restringindo somente o acesso a sua comunidade; a termos de uso específico para utilização dos dados definidos pelo depositante; a níveis diferentes de acesso a itens individuais ou a uma coleção de dados; à não disponibilização ao público em geral; e a arquivos que contenham informações confidenciais ou identificáveis.

Conforme mencionado anteriormente, citando Monteiro (2018), os repositórios de dados deveriam assegurar que os dados fossem publicados e disponibilizados para a comunidade científica com o menor número possível de restrições, contudo entende-se que, por questões ético-legais, pode ser necessária a imposição de restrições de acesso a eles, embora Albagli, Clinio e Raychtock (2014, p. 435) argumentem que, segundo a definição utilizada pela Open Knowledge para Ciência aberta, “[...] o conhecimento científico deve ser livre para as pessoas usarem, reutilizarem e distribuírem sem restrições legais, tecnológicas ou sociais.” Além disso,

discute-se que, “[...] quando há limitação do acesso aos dados científicos, há de certa forma uma sabotagem no próprio processo de fazer ciência” (ALBAGLI; CLINIO; RAYCHTOCK, 2014, p. 441).

No entanto, Borgman (2010, p. 7) considera que “os períodos de embargo, quando existem, protegem o investigador, fornecendo-lhe um período de tempo para analisar os dados e publicar os resultados antes da divulgação pública de seus dados.”

Na sequência, são apresentadas as análises referentes às permissões para depósito de dados identificadas nas políticas dos repositórios analisados.

#### 4.3.4 Análise quanto as permissões para depósito de dados nos repositórios

Quanto à permissão para depósito de dados nos repositórios, embora se tenha identificado que a maioria das políticas dá permissão somente para indivíduos com vínculo institucional, observou-se também que, dentre os repositórios analisados, alguns, em suas políticas, indicam permissão para que pesquisadores externos à universidade também façam o depósito de seus dados de pesquisa. A seguir, no Quadro 9, apresentam-se os repositórios em que o depósito de dados é permitido somente para indivíduos com vínculo institucional.

**Quadro 9 – Permissão de depósito por vínculo institucional**

Repositórios	Permissão para depósito de dados
3_DSpace@MIT	Docentes e pesquisadores do MIT, mediante <i>login</i> .
4_Data Repository [Berkeley]	Professores, estudantes e funcionários da UC Berkeley, mediante <i>login</i> com ORCID no Dryad.
5_Research Works Archive	Professores, servidores e estudantes da Universidade, mediante <i>login</i> com ID institucional ou registro no DSpace com <i>e-mail</i> e senha.
7_ORA-Data	Restrito aos membros credenciados de Oxford ou a depositantes por estes designados.
8_JHU Data Archive	Alunos, professores e funcionários da Universidade, sendo necessário registro, assinatura de Contrato de Depósito e preenchimento de Formulário de Depósito. Todos os depósitos são mediados por um Consultor de Gerenciamento de Dados.
10_eCommons	Professores, pesquisadores, funcionários e estudantes de pós-graduação da Universidade de Cornell, mediante registro.
11_Scholarly Commons	Professores, funcionários e estudantes da Universidade.
12_Apollo	Restrito aos membros da Instituição, mediante cadastro e preenchimento de formulário de depósito.
13_Merritt	Disponíveis para uso bibliotecas, arquivos, museus, departamentos acadêmicos, laboratórios e outras unidades organizacionais da Universidade da Califórnia.
14_ISPSData Archive	Afiliações da Yale, mediante cadastro e <i>login</i> .

<b>16_UCSanDiego</b>	Restrito à afiliação institucional. Depósitos mediante assinatura de Termos de Contrato de Depósito. Depósito de dados sob supervisão da equipe de Curadoria.
<b>18_UCL Research Data Repository</b>	Somente para pesquisadores e doutorandos da Universidade, por meio de <i>login</i> e utilização das credenciais UCL.

Fonte: elaborado pela autora.

Dentre os repositórios cujas políticas indicam que somente indivíduos com vínculo institucional podem fazer depósito de seus dados de pesquisa, observa-se que há distinção entre esses indivíduos. Algumas políticas indicam que os repositórios estão destinados a receber depósito de dados de docentes e pesquisadores, como o **DSpace@MIT**; outros indicam depósito também de funcionários e estudantes, como o **Data Repository [Berkeley]**, o **Research Works Archive**, o **JHU Data Archive**, o **eCommons** e o **Scholarly Commons**. Outros, ainda, fazem distinção entre os estudantes, permitindo o depósito de dados somente por parte de estudantes de pós-graduação, como o **e-Commons** e o **UCL Research Data Repository**. Este último ainda faz distinção entre os estudantes de pós-graduação, permitindo o depósito somente de doutorandos. Além disso, há os que não distinguem os indivíduos dentro da instituição, estando disponível a todos com afiliação institucional, como o **ORA-Data**, o **Apollo**, o **ISPS Data Archive** e o **UC SanDiego**. Por fim, referente ao **Merritt**, suas políticas indicam permissão de depósito por parte de bibliotecas, arquivos, museus, departamentos acadêmicos, laboratórios e outras unidades organizacionais da Universidade da Califórnia.

Contudo, embora os repositórios sejam institucionais e destinados a armazenar e preservar a produção acadêmica e científica das universidades a qual pertencem, seis dos repositórios analisados, conforme Quadro 10, permitem depósito de pesquisadores externos. O depósito por parte de pesquisadores externos em alguns repositórios deve ser mediado pelo sistema de gestão de dados da instituição ou por um pesquisador participante da pesquisa que possua vínculo institucional.

**Quadro 10 – Permissão de depósito sem vínculo institucional**

Repositórios	Permissão para depósito de dados
<b>1_Harvard Dataverse</b>	Pesquisadores de qualquer disciplina, independentemente de ter vínculo com Harvard, mediante registro/criação de conta de usuário.
<b>2_Stanford Digital Repository - SDR</b>	Professores, estudantes e funcionários da Stanford. Instituições parceiras das Bibliotecas de Stanford, de pesquisa e patrimônio cultural também preservam e disponibilizam suas coleções através dos serviços do SDR.
<b>6_DeepBlue Data</b>	Equipe de pesquisa da University of Michigan. Também aceita depósito de colaborações multi-institucionais, desde que um dos membros tenha vínculo empregatício com a universidade.

<b>9_Columbia Academic Commons</b>	Aceita depósito de pesquisadores da Columbia University e de suas instituições afiliadas mediante <i>login</i> com ID da Universidade. Aqueles que não têm ID devem entrar em contato com o Academic Commons para verificar se o trabalho atende às diretrizes de participação.
<b>15_MINDS@UW</b>	Restrita a afiliados da UW por meio de <i>login</i> . Aceita depósito de colaboradores externos, desde que o conteúdo colaborativo tenha como coautor um afiliado da UW.
<b>17_DRUM</b>	Restrito a pesquisadores da University of Minnesota, mas aceita depósito de projetos de colaboração, desde que um dos pesquisadores tenha vínculo ativo com a Universidade.

Fonte: elaborado pela autora.

A partir do exposto no Quadro 10, os repositórios que permitem depósito de dados sem vínculo institucional são: o **Harvard Dataverse**, da Harvard University, o qual aceita depósito de qualquer pesquisador a partir da criação de uma conta de usuário; o **Stanford Digital Repository (SDR)**, repositório da Stanford University que permite que instituições parceiras das Bibliotecas da Universidade também depositem seus dados de pesquisa; o **Deep Blue Data**, da University of Michigan, o **MINDS@UW**, da University of Wisconsin Madison, e o **DRUM**, do University of Minnesota System, os quais aceitam que colaboradores externos incluam seus dados no Repositório, desde que tenham na equipe de pesquisa um membro com vínculo institucional com a Universidade; e o **Columbia Academic Commons**, da Columbia University New York, que permite o depósito de dados por instituições afiliadas a partir de contato com a gerência do Repositório para verificar se o conteúdo a ser depositado atende às diretrizes de participação.

Observa-se que, embora todos os repositórios identificados no Quadro 10 aceitem depósito de dados de pesquisa de membros externos à instituição a que os repositórios estão vinculados, o depósito tem uma condicionante em cinco deles. Ou seja, os membros podem não ter um vínculo direto com a universidade, mas precisam estar vinculados a instituições parceiras a ela, ou devem contar com um pesquisador vinculado a um projeto cujo um dos outros membros tenha vínculo institucional com a universidade do repositório.

O único repositório dentre os seis indicados no Quadro 10 que não condiciona o depósito de dados a vínculos e parcerias institucionais é o **Harvard Dataverse**. Tal diferencial poderia ser justificado pelo fato de o Repositório fazer parte do Dataverse Project, projeto internacional de gestão de dados de pesquisa por meio de *software* de código aberto, composto por uma comunidade de mais de 65 instituições em todo o mundo.

O Harvard Dataverse se diferencia dos outros repositórios analisados, pois, ao fazer parte de uma comunidade global de Dataverses, disponibiliza acesso a outros

*dataverses*, conjuntos de dados e arquivos de instituições e pesquisadores para além dos dados de pesquisa produzidos em Harvard.

Dando continuidade às análises, na seção a seguir, tem-se a apresentação das restrições para depósito de dados indicadas nas políticas dos repositórios analisados.

#### 4.3.5 Análise das restrições para depósito de dados nos repositórios

Embora se espere ter acesso aberto à diversidade dos dados de pesquisa sem nenhum tipo de restrição, existem condicionantes que, por diferentes razões, impedem que esses dados sejam depositados em repositórios e acessados. No Quadro 11, pode-se observar as restrições para depósito de dados identificadas nas políticas dos repositórios analisados.

**Quadro 11 – Restrições para depósito de dados**

Repositórios	Restrições para depósito de dados
<b>1_Harvard Dataverse</b>	Depósito de usuário que não esteja em conformidade com os Termos de Uso Harvard Dataverse. Na apresentação do Repositório, o tamanho dos dados é indicado.
<b>2_Stanford Digital Repository - SDR</b>	Para grandes conjuntos de dados (10GB ou mais); grandes coleções de arquivos que necessitam manipulação em massa; invenções que estejam em processo de divulgação e gerenciamento no Escritório de Licenciamento de Tecnologia; e para outros tipos de conteúdo não especificados no SDR, é necessário entrar em contato com a equipe do SDR.
<b>3_DSpace @ MIT</b>	Dados incompletos que não estejam prontos para publicação.
<b>4_Data Repository [Berkeley]</b>	Restrição a conjuntos de dados de pesquisadores relacionados ao mesmo DOI excedentes a 300GB. Restrição a conteúdos com idioma diferente do inglês.
<b>5_Research Works Archive</b>	Restrições a conteúdo não acadêmico, tamanho e formato de dados não compatíveis.
<b>6_DeepBlue Data</b>	Dados incompletos que não estejam prontos para distribuição e reutilização.
<b>7_ORA-Data</b>	Dados ainda em uso ativo (vivo) por projetos de pesquisa. Dados criados por pesquisadores exclusivamente não afiliados à Oxford University.
<b>8_JHU Data Archive</b>	Dados que não possam ser acessados publicamente. Dados de pesquisa ativa.
<b>9_Columbia Academic Commons</b>	Restrição quanto ao tamanho dos arquivos. Arquivos maiores que 100MB precisam de tutoria.
<b>10_eCommons</b>	Usuários sem afiliação à Cornell devem ser patrocinados por um membro atual do corpo docente ou funcionário da Cornell. Arquivos individuais superiores a 5GB de tamanho e arquivos vinculados a projetos de pesquisa superiores a 50GB ao ano.
<b>11_Scholarly Commons</b>	Não aceita citações bibliográficas ou resumos que não estejam acompanhados dos documentos completos. Nas FAQs, o tamanho dos dados é indicado.
<b>12_Apollo</b>	Material de pesquisa realizado na Universidade, mas que não se relaciona diretamente com pesquisa da Universidade. Nas FAQs, o tamanho dos dados é indicado.
<b>13_Merritt</b>	Não foram identificadas restrições.
<b>14_ISPS Data Archive</b>	Permite somente arquivos de replicação.
<b>15_MINDS@UW</b>	Conteúdo efêmero ou em andamento. Arquivos individuais superiores a 2GB.

<b>16_UCSanDiego</b>	Não foram identificadas restrições.
<b>17_DRUM</b>	Dados que não sejam adequados para reutilização.
<b>18_UCL Research Data Repository</b>	Alunos de graduação e mestrandos da UCL não podem fazer depósito de dados. Colaboradores externos podem ser citados nos metadados, mas não podem depositar diretamente seus dados de pesquisa.

Fonte: elaborado pela autora.

Quanto às restrições para depósito de dados encontradas nas políticas dos repositórios, identificou-se, conforme Quadro 11, que alguns repositórios, dentre eles o **Stanford Digital Repository – SDR**, o **Data Repository [Berkeley]**, o **Research Works Archive**, o **Columbia Academic Commons**, o **eCommons** e o **MINDS@UW**, impõem restrições em relação ao tamanho dos dados ou conjuntos de dados a serem depositados. Contudo, embora não esteja sinalizado nas políticas, outros repositórios indicam restrições ao tamanho dos arquivos de dados para depósito, conforme informações localizadas em seus *websites*. Passa-se a apresentar as informações identificadas nas páginas desses repositórios.

Nas políticas do **Harvard Dataverse**, não foram identificadas restrições quanto ao tamanho dos arquivos, contudo, na página de suporte do Repositório, em texto de apresentação aos pesquisadores, consta que “todos os pesquisadores de qualquer disciplina, dentro e fora da Comunidade de Harvard, podem depositar arquivos de até 2,5GB e armazenar até 1TB de dados no Repositório de Dataverse de Harvard” (HARVARD DATAVERSE SUPPORT, 2020). Sendo assim, por não constar nas políticas, não se contabilizou no quadro de restrições, mas registra-se aqui essa informação, haja vista ser uma restrição ao tamanho dos dados e ao volume de armazenagem por pesquisador.

Quanto ao **Stanford Digital Repository - SDR**, as políticas não fazem exatamente uma restrição ao tamanho dos arquivos a serem depositados, mas sim orientam que depósitos de arquivos superiores a 10GB sejam tutelados pelos serviços do SDR. Acredita-se que essa orientação se deva a questões operacionais, visto que nas orientações menciona-se que 10GB trata-se de um limite prático de tamanho para se carregar e baixar conteúdo usando um navegador da *Web*, por meio do protocolo http. E para os arquivos acima de 10GB, mediante tutoria da equipe do SDR, solicita-se um prazo mínimo de 20 dias úteis para processamento dos arquivos e publicação em uma URL persistente (STANFORD LIBRARIES, [2020?]).

O **Columbia Academic Commons**, da mesma forma que o SDR, informa a necessidade de tutoria para o depósito de arquivos de tamanhos maiores, superiores, neste caso, a 100MB. Inclusive, solicita que os pesquisadores considerem depositar arquivos grandes no Dryad, uma vez que o serviço não tem custo para os pesquisadores da Universidade e suas afiliadas. Igualmente ao SDR, o Columbia Academic Commons justifica que arquivos menores possibilitam uma melhor experiência do usuário para *download* via *Web*.

O depósito de arquivos no **eCommons** deve, primeiramente, acolher a Política de Acessibilidade *Web* do Repositório, atendendo a padrões e práticas de acessibilidade e, posteriormente, a Política de Coleta de Conteúdo, na qual está indicado que o depósito de arquivos individuais não pode exceder 5GB de tamanho e os arquivos associados a um projeto de pesquisa não podem ultrapassar o limite de 50GB ao ano.

Já o **Scholarly Commons**, em suas políticas, também não indica restrição ao tamanho do arquivo para depósito, contudo, na página de 'Perguntas frequentes (FAQs)', indica que o Repositório está bem estruturado para receber arquivos menores que 500MB, e que arquivos maiores devem ser depositados em repositórios específicos para dados de pesquisa.

Da mesma forma que os repositórios acima mencionados, o **Apollo** também não sinaliza restrições ao tamanho dos dados de pesquisa, contudo, na página de 'Perguntas frequentes (FAQs)', indica que o envio pelo *software* Symplectic Elements requer que os dados tenham um tamanho máximo de 2GB para arquivos individuais e de 20GB para todo um conjunto de dados. Indica, ainda, que se houver a necessidade de envio de arquivos maiores, há a necessidade de se entrar em contato com a equipe do Repositório. Salienta também que, para que os dados sejam reutilizados, o usuário final precisa conseguir baixá-los, o que pode ser um complicador se o arquivo for muito grande. Sendo assim, recomenda que se considere maneiras de reduzir o tamanho dos dados antes do envio.

Ainda na página de FAQs do Apollo, há a indicação de cobrança de taxa de subsídio financeiro para depósitos de conjuntos de dados com tamanho acima de 20GB. A taxa a ser subsidiada será de £4/GB, multiplicada pela quantidade total de GB de dados enviados, valor a ser faturado pelo gerente financeiro ao qual pertence o pesquisador.

O tamanho do arquivo de dados também está sinalizado nas políticas de depósito do **MINDS@UW**. O Repositório aceita o depósito de arquivos individuais de até 2GB de tamanho, sendo que, para o depósito de arquivos maiores, faz-se necessário conversar com o gerente do MINDS@UW para verificar as possibilidades.

Em relação ao **Data Repository [Berkeley]**, devido ao fato de o Repositório utilizar os serviços de curadoria do Dryad, os Termos de Serviços deste se aplicam ao Repositório de Berkeley. Sendo assim, nesses Termos, o Dryad só aceita depósito de pesquisadores afiliados a uma organização membro, de conjuntos de dados vinculados a um mesmo DOI, bem como depósitos que não excedam 300GB.

Para o **Research Works Archive**, o tamanho dos arquivos também pode ser um impeditivo para depósito de dados. Conforme indicado na Política de Coleção Digital, a submissão de arquivos com tamanho em *terabytes* é proibitiva. Da mesma forma, sinaliza que o formato do arquivo pode ser um empecilho ao seu depósito.

O **Data Repository [Berkeley]**, além da restrição quanto ao tamanho dos conjuntos de dados, impõe restrições em relação a idiomas diferentes do inglês. O **Stanford Digital Repository – SDR** também impõe restrições a dados que estejam em processo de divulgação ou cuja tipologia não esteja especificada nas políticas do Repositório. Já o **Apollo** impõe restrições para materiais de pesquisa realizados na Universidade, mas que não se relacionem diretamente com pesquisas da Universidade.

Restrições referentes a dados incompletos que não estejam prontos para publicação ou distribuição e reutilização são impostas pelo **DSpace@MIT**, **DeepBlue Data** e **DRUM**. Da mesma forma, dados que estejam em uso ativo por projetos de pesquisa não são aceitos para depósito no **ORA-Data**, no **JHU Data Archive** e no **MINDS@UW**. O **JHU Data Archive** indica, ainda, restrição a dados que não possam ser acessados publicamente, e o MINDS@UW sinaliza a necessidade de depósito intermediado pela gerência do Repositório.

O **ORA-Data** ainda impõe restrições aos dados criados por pesquisadores não afiliados à Oxford University, semelhantemente às restrições sinalizadas pelo **eCommons**, indicando que pesquisadores não afiliados necessitam ser patrocinados por pesquisadores afiliados à instituição para depósito de dados no Repositório. Em paralelo, nas políticas do **UCL Research Data Repository**, os colaboradores externos podem ser citados nos metadados, mas não podem fazer diretamente depósito de

dados no Repositório. Igualmente, o **UCL Research Data Repository** faz restrição ao depósito de dados por alunos de graduação e mestrandos da Universidade.

Já o **Harvard Dataverse**, em suas políticas, menciona como restrição o depósito de dados de usuários que não estejam em conformidade com os Termos de Uso do Repositório. O **ISPS Data Archive**, por sua vez, permite o depósito somente de arquivos de replicação. E em relação ao **Merritt** e ao **UC San Diego**, não foram localizadas menções a restrições para depósito de dados nas políticas dos Repositórios.

Em relação à limitação do tamanho dos arquivos para depósito, de acordo com Torino (2017), este é um aspecto a ser considerado por sua equipe de gestão nas políticas dos repositórios, haja vista estar relacionado à capacidade do servidor quanto ao *storage* e à memória. Considerando que os dados de pesquisa produzidos por pesquisadores de diferentes áreas do conhecimento em uma universidade não sejam semelhantes, ainda assim, todos os pesquisadores da instituição são pesquisadores em potencial para depósito de dados de pesquisa em seus repositórios. Dessa forma, impor limites para o tamanho dos dados pode ser uma estratégia importante para as universidades que contam com um volumoso corpo de pesquisadores.

O compartilhamento de dados de pesquisa ainda é uma prática em construção para os pesquisadores brasileiros, bem como as iniciativas de implementação de repositórios de dados no país. O volume anual de dados produzidos por área do conhecimento CNPq no país, de acordo com a *survey* da RDP Brasil<sup>56</sup>, aplicada a pesquisadores de todo o Brasil, não ultrapassa os 32% para o tamanho de 1 a 50GB, sendo estes dados referentes à área das Engenharias. Todas as outras áreas, de acordo com as respostas da *survey*, ficam abaixo dessa porcentagem.

Sintetizando as informações identificadas nas políticas e nas páginas dos *websites* dos repositórios sobre o tamanho dos arquivos de dados, bem como a ponderação feita por Torino (2017) e os dados apresentados na *survey* RDP Brasil, tem-se que: as políticas dos repositórios, ao delimitarem o tamanho dos arquivos de dados para depósito, devem considerar a capacidade estrutural de armazenamento e memória do servidor que atende ao repositório; os arquivos de dados devem ser passíveis de *download* via *web* pelos usuários, sendo uma questão importante a ser considerada nas políticas de depósito do repositório; a *survey* RDP Brasil, ao

---

<sup>56</sup> Disponível em:

<https://lume.ufrgs.br/bitstream/handle/10183/185195/001082283.pdf?sequence=1&isAllowed=y>

identificar que o volume de produção de dados de pesquisa anuais em GB dos pesquisadores brasileiros ainda, em sua maioria, fica dentro dos 50GB, traz subsídios aos gestores dos RDPs de universidades no que diz respeito às diretrizes de depósito a serem instituídas.

Na sequência, são apresentadas as observações referentes às restrições ao tipo de dados de pesquisa para depósito identificadas nas políticas dos repositórios analisados.

#### 4.3.6 Análise das restrições quanto aos tipos de dados para depósito

Ainda que alguns repositórios não determinem restrições aos depositantes para depósito de dados, estes podem impor restrições para tipos específicos de dados a serem depositados. No Quadro 12, pode-se observar as restrições quanto ao tipo de dados para depósito identificadas nas políticas analisadas.

**Quadro 12 – Restrições quanto ao tipo de dados para depósito**

Repositórios	Restrições quanto ao tipo de dados para depósito
1_Harvard Dataverse	Dados que infrinjam os direitos autorais, de propriedade intelectual ou quaisquer outros direitos de terceiros. Dados confidenciais e restritos.
2_Stanford Digital Repository - SDR	Dados privados e confidenciais.
3_DSpace @ MIT	Dados confidenciais e restritos.
4_Data Repository [Berkeley]	Não foram identificadas restrições.
5_Research Works Archive	Dados que infrinjam os direitos autorais.
6_Deep Blue Data	Dados que violem a ética e a conformidade com a pesquisa; dados confidenciais e dados sensíveis; dados administrativos sem o consentimento da Biblioteca da UM;
7_ORA-Data	Dados que contenham informações pessoais, confidenciais ou sensíveis (a menos que sejam suficientemente anonimizados). Dados ilegais ou que infrinjam os direitos de propriedade intelectual.
8_JHU Data Archive	Dados que contenham identificadores pessoais e de saúde.
9_Columbia Academic Commons	Não foram identificadas restrições.
10_eCommons	Dados privados, confidenciais ou que contenham informações caluniosas.
11_Scholarly Commons	Conteúdo que viole direitos autorais, de propriedade, de privacidade ou que contenham informações caluniosas (garantia dada pelo autor ao depositar no repositório).
12_Apollo	Não foram identificadas restrições.
13_Merritt	Não indica restrições, mas em Política e Procedimentos declara que conteúdos contendo informações pessoais devem ser anonimizados.

<b>14_ISPS Data Archive</b>	Não foram identificadas restrições.
<b>15_MINDS@UW</b>	Dados que contenham informações sensíveis (dados pessoais), dados restritos ou confidenciais.
<b>16_UCSanDiego</b>	Dados que infringjam direitos autorais; que contenham difamação ou matéria ilícita; que violem acordos ou leis; dados privados, confidenciais ou que contenham informações legalmente protegidas como dados pessoais identificáveis.
<b>17_DRUM</b>	Dados privados, confidenciais ou que contenham informações legalmente protegidas como dados pessoais identificáveis.
<b>18_UCL Research Data Repository</b>	Dados contendo informações pessoais ou de categoria especial; dados sobre os quais o depositante ou a UCL não detenham os direitos autorais ou de propriedade intelectual ou não tenham permissão para compartilhamento; dados que estejam em 'Acordo de compartilhamento' em vigor.

Fonte: elaborado pela autora.

Em relação às restrições impostas ao tipo de dados para depósito, observou-se que, dos 18 repositórios pesquisados, excluindo-se o **Data Repository [Berkeley]**, o **Columbia Academic Commons**, o **Apollo** e o **ISPS Data Archive**, visto que não apresentam restrições, ou não foram localizadas em suas políticas, todos os outros repositórios indicam restrições quanto ao nível de sensibilidade dos dados ou em relação às questões éticas e legais destes, como pôde ser visto no Quadro 12. Os repositórios **ORA-Data** e **Merrit** condicionam o depósito de dados sensíveis a sua anonimização.

Ao produzirem diferentes tipos de dados, as pesquisas podem, inclusive, produzir dados que contenham informações sensíveis à disponibilização e acesso dos usuários. Tanto os PGDs dos pesquisadores quanto as políticas dos repositórios devem contemplar a privacidade de dados.

Nesse sentido, a anonimização e a criptografia são estratégias a serem utilizadas na apresentação dos dados de pesquisa sem que se exponham os participantes, mantendo suas privacidades. De acordo com Monteiro (2017, p. 29), baseada em um protocolo de criptografia ou remoção de informações identificáveis a partir de conjuntos de dados, a anonimização “[...] permite a divulgação legal dos dados e dos metadados sem ferir a integridade dos participantes da pesquisa.” Já a criptografia, ainda segundo Monteiro (2017, p. 29), “[...] consiste na conversão de dados em uma forma que não pode ser facilmente compreendida por pessoas não autorizadas.”

Entre as restrições para tipos de dados, os dados que infringjam os direitos autorais passam a ser de responsabilidade dos depositantes. Nas políticas analisadas, uma das responsabilidades do autor/depositante consiste em deter direitos sobre os dados depositados, não cabendo aos repositórios essa verificação,

o qual atuará somente mediante denúncia de infração dos direitos autorais. Sendo assim, é imprescindível que, a fim de evitar futuros litígios, o pesquisador, em seu plano de gestão de dados, na seção referente às políticas de acesso, determine quem será o detentor dos direitos autorais sobre os dados da sua pesquisa (BERTIN, 2017).

Em relação a restrições aos dados que violem a ética e a conformidade com a pesquisa, pode-se salientar que a orientação à pesquisa acadêmica está presente nos serviços oferecidos pelas bibliotecas das Universidades elencadas para o estudo. Dessa forma, cabe aos autores a responsabilidade pelo descumprimento das diretrizes das Comissões de ética das Instituições.

Tratando-se das informações sigilosas, de acordo com Nascimento (2019, *online*), “o sigilo [...] não é um fim em si mesmo, a sua aplicação [...] visa à proteção de um direito, objeto ou valor que estará em risco caso a informação que lhe faça referência seja indiscriminadamente acessada.”

Dando continuidade às análises, a seguir, são apresentados os estudos sobre as permissões e restrições para uso, identificadas nas políticas dos repositórios.

#### 4.3.7 Análise das permissões e restrições para uso de dados

Para que ocorra o reuso de dados de pesquisa, estes precisam ser primeiramente compartilhados. De acordo com Borgman (2012, p. 1060), “[...] o compartilhamento de dados é a liberação destes para uso por terceiros”, podendo dar-se de diferentes formas. No caso do presente estudo, ele se dá por meio de repositórios universitários de dados de pesquisa.

Segundo Medeiros e Caregnato (2012), é a partir de repositórios que uma enorme variedade de pesquisadores, com diferentes níveis de especialização e espalhados pelo mundo, atuam no reuso de dados de pesquisa. O reuso de dados está condicionado às permissões e restrições determinadas por seus autores.

Sendo assim, a seguir, pode-se observar as permissões e restrições para uso de dados baixados identificadas nas políticas dos repositórios do presente estudo.

**Quadro 13 – Permissões e restrições para uso de dados baixados**

Repositórios	Permissão para uso de dados baixados	Restrições para uso de dados baixados
1_Harvard Dataverse	Usuários registrados no <i>site</i> ou convidados (identificação temporária) que concordem com a renúncia CC0 ou com o Contrato de Uso de Dados Personalizado de um conjunto de dados.	Restrições para <i>download</i> de conjuntos de dados ou outros materiais quando proibidos por lei.
2_Stanford Digital Repository - SDR	Alguns <i>datasets</i> têm permissão de <i>download</i> somente para usuários afiliados à Stanford; outros possuem acesso aberto mediante aceite dos Termos de Uso.	Determinadas pelas licenças de uso.
3_DSspace @ MIT	Nas Políticas, não foram localizadas informações sobre permissão para <i>download</i> , mas, ao recuperar dados de pesquisa através de busca no repositório, aparecem, em alguns registros nos Termos de Uso do dado, a Licença Creative Commons permitida.	Não foram identificadas restrições.
4_Data Repository [Berkeley]	Acesso aberto.	Não foram identificadas restrições.
5_Research Works Archive	Acesso aberto.	Não faz menção a restrições para <i>download</i> . Todo conteúdo é de acesso aberto, com exceção das restrições impostas pelo depositante.
6_Deep Blue Data	Acesso aberto.	Não menciona restrições para <i>download</i> . Todo conteúdo é de acesso aberto, respeitando os Termos de Uso e as licenças atribuídas a cada conjunto de dados.
7_ORA-Data	Acesso aberto, permitindo impressão e <i>download</i> para fins não comerciais.	Não é permitido o uso comercial dos dados e <i>download</i> que infrinja os direitos de propriedade intelectual.
8_JHU Data Archive	Acesso aberto e gratuito sem necessidade de registro ou cadastro.	Sem restrições.
9_Columbia Academic Commons	O usuário pode baixar e usar obras, com a condição de estar de acordo com os termos e condições de licença escolhidos pelo proprietário de direitos autorais.	Sem restrições.
10_eCommons	Aberto ao público em geral.	Conteúdos disponibilizados apenas para usuários da comunidade da Universidade de Cornell ou outros subgrupos de usuários especificados pelo depositante.
11_Scholarly Commons	Acesso aberto.	Não foram identificadas restrições.
12_Apollo	Não localizado.	Não foram identificadas restrições.
13_Merritt	Comunidade da UC e permite acesso público a <i>download</i> para coleções específicas para os usuários com acesso como convidado.	Não é permitido o <i>download</i> de conjuntos de dados da Dryad que estejam sendo revisados por pares (restrições de acesso por até um período de seis meses).
14_ISPS Data Archive	Permite dois níveis de acesso: público e privado.	O proprietário dos dados pode definir restrições para <i>download</i> .
15_MINDS@UW	Acesso aberto, permitindo <i>download</i> imediato.	Não foram identificadas restrições.
16_UCSanDiego	Acesso aberto sem necessidade de cadastro.	Alguns itens das coleções podem ter restrições de direitos autorais ou contratos de licenciamento anexadas ao seu uso.

<b>17_DRUM</b>	Permite que qualquer pessoa faça <i>download</i> de forma instantânea ou por solicitação via <i>e-mail</i> ao autor (para dados com acesso restringido pelo autor por um prazo máximo de dois anos).	Não foram identificadas restrições.
<b>18_UCL Research Data Repository</b>	Não menciona. Faz uma observação de que os dados devem ser usados de acordo com a licença escolhida pelo depositante.	Não foram identificadas restrições.

Fonte: elaborado pela autora.

Pode-se observar, conforme Quadro 13, que entre os 18 repositórios estudados, oito são de acesso aberto, permitindo que os usuários façam *download* para uso imediato dos dados de pesquisa e não indicando a necessidade de cadastro no repositório. São eles: o **Data Repository [Berkeley]**, o **Research Works Archive**, o **Deep Blue Data**, o **JHU Data Archive**, o **eCommons**, o **Scholarly Commons**, o **MINDS@UW** e o **UCSanDiego**. Igualmente de acesso aberto, tem-se o **Oxford University Research Archive – ORA**, contudo, sinaliza que permite que os usuários façam impressão e *download* dos dados somente para fins não comerciais. Já o **Columbia Academic Commons** sinaliza em suas políticas que os usuários podem baixar e usar obras de qualquer maneira, contanto que estejam de acordo com os termos e condições de licença escolhidos pelo proprietário de direitos autorais.

Os repositórios **Harvard Dataverse** e o **DSpace@MIT** indicam a observância das licenças Creative Commons para uso dos dados baixados. O **Harvard Dataverse** permite que os usuários façam *download* desde que sejam registrados no *site* ou tenham uma identificação temporária como convidados, além de concordarem com a renúncia CC0 ou com o Contrato de Uso de dados. Já nas políticas do **DSpace@MIT**, não foram localizadas informações referentes à permissão para *download* de dados, mas, ao realizar-se pesquisa no repositório, observou-se que aparece a indicação de uma atribuição Creative Commons em alguns registros nos Termos de uso do dado.

Comparativamente, os repositórios **Stanford Digital Repository – SDR**, **Merritt** e **ISPS Data Archive** impõem permissões diferentes para os usuários. Enquanto o **Stanford Digital Repository – SDR** e o **ISPS Data Archive** indicam níveis de acesso para *download* (público e privado), o repositório **Merritt**, destinado à Comunidade da Universidade, permite que o usuário não filiado à instituição faça *download* de dados de coleções específicas listadas no *site* como convidado. Já o **DRUM** permite que qualquer pessoa faça *download* dos dados de forma instantânea

ou por solicitação ao autor por *e-mail* para dados com restrição de acesso. Para os repositórios **Apollo** e **UCL Research Data Repository**, não foram localizadas, nas respectivas políticas, alguma menção referente à permissão para *download* de dados. No entanto, nas políticas do UCL Research Data Repository, há uma observação de que os dados devem ser usados de acordo com a licença escolhida pelo depositante.

Em relação às restrições, observou-se que, entre os 18 repositórios analisados na pesquisa, conforme Quadro 13, nove não apresentam restrições para *download* de dados, ou não foram localizadas informações referentes a restrições nas políticas observadas, sendo eles: o **DSpace @ MIT**, o **Data Repository [Berkeley]**, o **JHU Data Archive**, o **Columbia Academic Commons**, o **Scholarly Commons**, o **Apollo**, o **MINDS@UW**, o **DRUM** e o **UCL Research Data Repository**.

Já os repositórios **Research Works Archive**, **Deep Blue Data**, **Stanford Digital Repository – SDR** e **ISPS Data Archive** não mencionam restrições para *download* de dados, mas indicam que o depositante tem prerrogativa para impor tais restrições. O **Deep Blue Data** também menciona que, embora todo conteúdo depositado seja de acesso aberto, o usuário deve estar atento aos Termos de uso e às licenças atribuídas a cada *dataset*. Da mesma forma, o **Stanford Digital Repository – SDR** indica que as restrições serão determinadas pelas licenças de uso impostas pelo depositante dos dados.

Em contrapartida, os repositórios **Harvard Dataverse**, **Oxford University Research Archive – ORA**, **eCommons**, **Merritt** e **UCSanDiego** sinalizam, em suas respectivas políticas, algum tipo de restrição para *download* de dados. Enquanto para o **Harvard Dataverse**, **Oxford University Research Archive – ORA** e **UC San Diego** as restrições aos *datasets* se dão por questões legais, para o **eCommons**, as restrições estão relacionadas às imposições impostas pelo depositante referente aos grupos de usuários que podem fazer *download*. Por fim, o **Merritt**, por meio das políticas do Dryad, menciona restrições para *download* de *datasets* que estejam sendo revisados por pares.

Na análise a seguir, as permissões de uso de dados de pesquisa por meio de atribuição de licenças são apresentadas.

#### 4.3.8 Análise das permissões de uso de dados por atribuição de licenças

De acordo com Monteiro (2017), uma licença representa a cessão do direito sobre uma obra cuja exploração em outras condições seria ilícita, podendo se dar por meio de um direito exclusivo ou não exclusivo. Sendo assim, a atribuição de licenças para o uso de dados estabelece um contrato entre o depositante e o usuário/*downloader* desses dados, deixando claras as permissões para seu uso.

Segundo Gomes (2014), o uso das licenças livres foi uma forma de flexibilizar as leis de direito autoral, garantindo o acesso a uma determinada obra sem deixar de proteger os autores que publicam as suas obras em modelos abertos. O uso das licenças livres, no contexto da abertura dos dados de pesquisa, possibilita o compartilhamento e reuso de dados sem colocar em risco a propriedade desses dados, garantindo, segundo Monteiro (2017), uma segurança aos usuários na utilização de dados.

Dessa forma, buscou-se identificar, nas políticas dos repositórios analisados, como se dão as permissões para o uso de dados baixados pelos usuários dos repositórios. Tais permissões podem ser observadas no Quadro 14 a seguir.

**Quadro 14 – Atribuição de licenças de uso dos dados de pesquisa**

Repositórios	Permissões: licenças de uso
<b>1_Harvard Dataverse</b>	Estabelece contratos de: Licença de uso restrito; Licença entre depositante e <i>Downloader</i> quanto ao uso dos dados baixados; Licença CCO, domínio público (Contrato de renúncia de direitos autorais).
<b>2_Stanford Digital Repository - SDR</b>	Os depositantes podem atribuir licenças de uso aos dados. O SDR oferece licenças opcionais, Creative Commons e Open Data Commons.
<b>3_DSpace @ MIT</b>	O DSpace@MIT disponibiliza um formulário opcional para atribuição de Licenças Creative Commons.
<b>4_Data Repository [Berkeley]</b>	Ao enviar conteúdo ao Repositório, o depositante concede permissão irrevogável para tornar o Conjunto de Dados disponível ao público sob uma Renúncia CCO. Também pode ser a critério exclusivo do Dryad, licenciamento sob outros termos. Ao acessar os dados, há indicação da Licença CC permitida.
<b>5_Research Works Archive</b>	Cabe aos autores a liberação de todos os direitos necessários para uso dos dados. Nos metadados, há indicação da Licença CC permitida.
<b>6_Deep Blue Data</b>	Os depositantes atribuem licenças Creative Commons aos dados e atribuição CCO padrão para todos os metadados.
<b>7_ORA-Data</b>	Para dados que não sejam propriedade da Oxford University, o depositante concede à Universidade uma licença mundial não exclusiva. O acesso e o uso dos dados são licenciados somente para fins não comerciais.
<b>8_JHU Data Archive</b>	Por padrão, todos os novos conjuntos de dados criados recebem uma dedicação ao domínio público Creative Commons CCO. Todo o conteúdo disponibilizado é protegido por direitos autorais e/ou outros direitos de propriedade intelectual. O uso de qualquer Conteúdo disponibilizado está sujeito aos Termos Uso do JHU Data Archive.

<b>9_Columbia Academic Commons</b>	A maioria dos trabalhos no Columbia Academic Commons está sob direitos autorais com todos os direitos reservados, a menos que seja indicado de outra forma. Trabalhos podem ser licenciados com Creative Commons.
<b>10_eCommons</b>	Recomenda a aplicação de licenças. Sinaliza que, se o depositante desejar, pode aplicar licenças Creative Commons.
<b>11_Scholarly Commons</b>	Incentiva o uso de licenças Creative Commons, quando apropriado, para materiais inéditos sobre os quais os autores possuem direitos autorais.
<b>12_Apollo</b>	Salvo indicação em contrário, os conteúdos depositados são protegidos por direitos autorais com todos os direitos reservados. O uso de conteúdo está condicionado ao aceite da Licença ou dos Termos sob os quais foi liberado (por exemplo, os termos de uma Licença Creative Commons específica).
<b>13_Merritt</b>	Os dados enviados ao Dryad estão associados a licenças padrão Creative Commons CC-BY ou dedicação de domínio público CC0 que cobrem termos de acesso e uso aceitável.
<b>14_ISPS Data Archive</b>	Conteúdo digital é licenciado sob uma Creative Commons Atribuição-Nãocomercial-SemDerivações 3.0 Estados Unidos (CC BY-NC-ND 3.0 US).
<b>15_MINDS@UW</b>	Licenças Creative Commons e Open Data Commons (depositante/usuário). Licença <b>MINDS@UW</b> (depositante/Universidade).
<b>16_UCSanDiego</b>	Não localizado no <i>site</i> , mas de acordo com <i>e-mail</i> recebido da UC San Diego, as informações de Licenças Creative Commons estão anexadas aos registros dos dados.
<b>17_DRUM</b>	No momento do depósito, os autores podem optar por aplicar uma licença Creative Commons aos seus dados. No entanto, mesmo sem uma licença Creative Commons, os usuários poderão fazer <i>download</i> e usar dados.
<b>18_UCL Research Data Repository</b>	O Figshare lista as licenças Creative Commons, Open Data Commons e AusGOAL. A UCL Research Data Policy recomenda o uso da licença Creative Commons CC0 para conjuntos de dados.

Fonte: elaborado pela autora.

Nas políticas analisadas, identificou-se que todos os 18 repositórios indicam a atribuição, por parte dos depositantes, de algum tipo de licença de uso dos dados de pesquisa, sejam as licenças Creative Commons, Open Data Commons, AusGoal, ou licenças específicas do próprio repositório. Como pode ser observado no Quadro 14, cada repositório, em suas políticas, apresenta especificidades quanto à atribuição das licenças e aos contratos estabelecidos entre depositante e repositório/universidade, depositante e usuário/*downloader* para o uso dos dados.

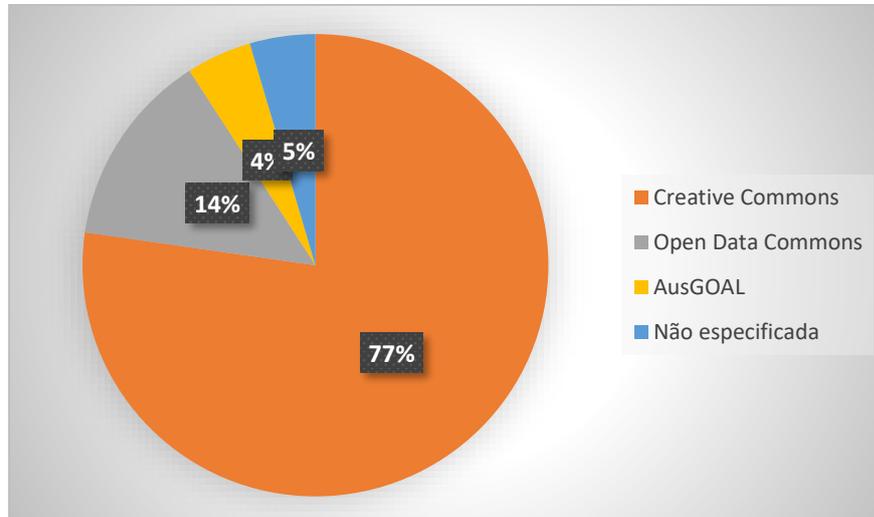
Por padrão, embora possa haver licenciamentos sobre outros termos, repositórios como o **Harvard Dataverse**, o **Data Repository [Berkeley]**, o **JHU Data Archive**, o **Merritt** e o **UCL Research Data Repository** indicam a atribuição CC0 para seus *datasets*, enquanto o **Deep Blue Data** a indica somente para seus metadados. Já o **Oxford University Research Archive – ORA**, para dados que não sejam de propriedade da Universidade, exige que o depositante conceda à Oxford University uma licença mundial não exclusiva para disponibilização dos dados no Repositório, não especificando se será uma atribuição Creative Commons ou Open Data Commons.

Nas políticas do **UC San Diego**, não foram localizadas referências quanto ao licenciamento para uso de dados, contudo, após contato com a Instituição, obteve-se a informação de que todos os registros de dados vêm acompanhados de uma atribuição Creative Commons. Já o **DRUM** sinaliza que, no momento do depósito de dados, os autores podem optar pela aplicação de uma licença Creative Commons, no entanto, mesmo que não seja atribuída uma CC, os usuários têm permissão para fazer *download* e usar os dados de pesquisa.

Os repositórios **Stanford Digital Repository – SDR**, **MINDS@UW** e **UCL Research Data Repository** indicam a possibilidade de aplicação de licenças tanto Creative Commons como Open Data Commons, sendo que o **UCL Research Data Repository** também sinaliza a utilização de licenças AusGOAL.

Como já exposto, pode-se identificar nos repositórios analisados a atribuição de licenças para uso dos dados de pesquisa, sejam elas apresentadas em suas políticas ou na recuperação dos dados na plataforma dos repositórios. No Gráfico 1, a seguir, apresenta-se a proporção entre as licenças identificadas para uso dos dados.

**Gráfico 1 – Licenças de uso indicadas nas políticas dos repositórios**



Fonte: elaborado pela autora.

Como se pode observar no Gráfico 1, a indicação de atribuição das licenças Creative Commons encontra-se presente na maioria das políticas analisadas, sendo seguida pela licença Open Data Commons, com apenas três sinalizações de uso. De acordo com Silva (2019), a CC e a Open Data Commons são as licenças mais utilizadas para dados abertos.

A indicação de atribuição das Creative Commons, em 77% dos repositórios analisados, justifica-se devido ao fato de as CC estarem há 20 anos sendo utilizadas,

desde sua fundação no ano de 2001, e também por fornecerem a possibilidade de diferentes tipos de licenças de uso. A Creative Commons, para Lemos (2005, p. 83), “[...] cria instrumentos jurídicos para que um autor, criador ou entidade diga, de modo claro e preciso, para as pessoas em geral, que uma determinada obra intelectual sua é livre para distribuição, cópia e utilização.” Sendo uma alternativa ao direito da propriedade intelectual tradicional, ainda conforme Lemos (2005, p. 83), as CC “[...] se fundamentam no exercício das prerrogativas que cada indivíduo tem, como autor, de permitir o acesso as suas obras e a seus trabalhos, autorizando que outros possam utilizá-los e criar sobre eles.”

As CC, além de permitirem uma fácil compreensão de seu conteúdo, dispõem de uma apresentação visual de fácil memorização para seu uso. De acordo com Swan (2016, p. 11), a CC se “[...] constitui na melhor solução prática [...] por ser um mecanismo de fácil compreensão, que fornece um pacote de licenças que abrange todas as necessidades dos autores, além de suas licenças poderem ser interpretadas por máquinas.”

Já a Open Data Commons foi lançada em 2008. Enquanto a CC dispõe de sete licenças distintas, a Open Data Commons disponibiliza três licenças para ajudar no gerenciamento e uso de dados abertos. Poderia justificar-se o maior número de atribuições CC nas políticas dos repositórios a partir dessa maior possibilidade de atribuições permitidas pela Creative Commons em comparativo à Open Data Commons.

Em relação à licença AusGOAL, somente o **UCL Research Data Repository** sinaliza a atribuição AusGOAL. Essa orientação se dá por meio das diretrizes do Figshare para o Repositório, que guia a escolha de licenças de uso. A Australian Governments Open Access and Licensing Framework (AusGOAL), lançada em 2011, “[...] fornece apoio e orientação ao governo e setores relacionados para facilitar o acesso aberto a informações financiadas publicamente” (WIKI AUSTRÁLIA AUSGOAL, 2015).

Dando sequência à apresentação das análises, a seguir tem-se as considerações sobre a privacidade de dados pessoais e sua anonimização, identificadas nas políticas dos repositórios.

#### 4.3.9 Análise quanto à privacidade no uso de dados pessoais

Anteriormente, na etapa de análise das restrições ao tipo de dados para depósito nos repositórios, procurou-se identificar nas políticas estudadas alguma restrição referente à tipologia dos dados de pesquisa, mostrada no Quadro 13 da referida etapa.

Nesta etapa, complementando a etapa acima citada de restrições quanto à tipologia de dados, buscou-se registrar as diretrizes adotadas nas políticas dos repositórios referentes à manutenção da privacidade de dados pessoais registrados nos dados de pesquisa. Tais diretrizes podem ser observadas pelos registros realizados por meio do Quadro 15, a ser apresentado a seguir.

**Quadro 15 – Diretrizes sobre privacidade no uso de dados pessoais**

Repositórios	Privacidade no uso de dados pessoais
1_Harvard Dataverse	É vetado o uso de informações que identifiquem seres humanos, devendo manter seu anonimato.
2_Stanford Digital Repository - SDR	O depositante garante, nos Termos de depósito, que não está infringindo a privacidade ou o direito de confidencialidade de qualquer pessoa.
3_DSpace @ MIT	Informação não localizada.
4_Data Repository [Berkeley]	O Dryad só divulgará informações pessoais individualmente identificáveis sobre seus usuários em circunstâncias restritas.
5_Research Works Archive	Não localizado.
6_Deep Blue Data	O depositante nos Termos de Uso representa e garante que seu depósito não viola a privacidade de outros.
7_ORA-Data	Todos os dados depositados devem ser anonimizados.
8_JHU Data Archive	Todos os dados depositados devem ter informações pessoais e de saúde retiradas (dados anonimizados).
9_Columbia Academic Commons	Não localizado.
10_eCommons	Os dados devem ser anonimizados.
11_Scholarly Commons	Todo conteúdo depositado não deve conter informações confidenciais do autor ou outras informações que possam comprometer a privacidade do autor.
12_Apollo	Não localizado.
13_Merritt	Na Política e Procedimentos diz que conteúdos contendo informações pessoais devem ser anonimizados.
14_ISPS Data Archive	O ISPS toma medidas para proteger informações confidenciais por meio de estratégias de orientação aos autores/colaboradores, revisão dos dados, limitação de acesso a dados e treinamentos.
15_MINDS@UW	Não localizado.
16_UCSanDiego	Não localizado.

<b>17_DRUM</b>	O usuário concorda que não fará qualquer uso de dados para identificar ou infringir os direitos de privacidade ou confidencialidade de indivíduos descobertos inadvertidamente ou intencionalmente nos dados.
<b>18_UCL Research Data Repository</b>	A Política de Privacidade do Serviço do Repositório garante: a privacidade de dados processados por estudantes da UCL, que devem registrar seus projetos no Escritório de Proteção de Dados da Universidade.

Fonte: elaboração da autora.

Quanto à privacidade de dados pessoais identificada nos dados de pesquisa, de acordo com as Políticas analisadas, em sua maioria, pôde-se constatar a presença de diretrizes que orientam os depositantes referente à necessidade de anonimização dos dados para depósito nos repositórios. Somente em seis repositórios não foram localizadas informações em suas políticas referentes à privacidade de dados pessoais, sendo eles: o **DSpace@MIT**, o **Research Works Archive**, o **Columbia Academic Commons**, o **Apollo**, o **MINDS@UW** e o **UCSan Diego**.

A seguir, no Quadro 16, faz-se um comparativo entre as restrições ao tipo de dados e as diretrizes sobre privacidade de dados pessoais, identificadas nas políticas.

**Quadro 16 – Comparativo entre restrições ao tipo de dados para depósito X diretrizes sobre privacidade de dados pessoais**

Repositórios	Restrições quanto ao tipo de dados para depósito	Privacidade de dados pessoais
<b>1_Harvard Dataverse</b>	Dados que infringam direitos autorais, direitos de propriedade intelectual ou quaisquer outros direitos. Dados confidenciais e restritos.	É vetado o uso de informações que identifiquem seres humanos, devendo manter seu anonimato.
<b>2_Stanford Digital Repository - SDR</b>	Dados privados e confidenciais.	O depositante garante nos Termos de depósito que não está infringindo a privacidade ou o direito de confidencialidade de qualquer pessoa.
<b>3_DSpace@MIT</b>	Dados confidenciais e restritos.	Não localizado.
<b>4_Data Repository [Berkeley]</b>	Não foram identificadas restrições.	As informações pessoais individualmente identificáveis sobre seus usuários só serão divulgadas em circunstâncias restritas.
<b>5_Research Works Archive</b>	Dados que infringam os direitos autorais.	Não localizado.
<b>6_Deep Blue Data</b>	Dados que violem a ética e a conformidade com a pesquisa; dados confidenciais e dados sensíveis; dados administrativos sem o consentimento da Biblioteca da University of Michigan.	O depositante, nos Termos de Uso, representa e garante que seu depósito não viola a privacidade de outros.
<b>7_ORA-Data</b>	Dados que contenham informações pessoais, confidenciais ou sensíveis (a menos que sejam suficientemente anonimizados). Dados ilegais ou que infringam os direitos de propriedade intelectual.	Todos os dados depositados devem ser anonimizados.

<b>8_JHU Data Archive</b>	Dados com identificadores pessoais e de saúde.	Todos os dados depositados devem ter informações pessoais e de saúde retiradas (dados anonimizados).
<b>9_Columbia Academic Commons</b>	Não foram identificadas restrições.	Não localizado.
<b>10_eCommons</b>	Dados privados, confidenciais ou que contenham informações caluniosas.	Os dados devem ser anonimizados.
<b>11_Scholarly Commons</b>	Conteúdo que viole direitos autorais, de propriedade, de privacidade, ou contenham informações caluniosas (garantia dada pelo autor ao depositar no repositório).	Todo conteúdo depositado não deve conter informações confidenciais do autor ou outras informações que possam comprometer sua privacidade.
<b>12_Apollo</b>	Não foram identificadas restrições.	Não localizado.
<b>13_Merritt</b>	Não indica restrições.	Na Política e Procedimentos, diz que conteúdos que contenham informações pessoais devem ser anonimizados.
<b>14_ISPS Data Archive</b>	Não foram identificadas restrições.	O ISPS toma medidas para proteger informações confidenciais por meio de estratégias de orientação aos autores/colaboradores, revisão dos dados, limitação de acesso a dados e treinamentos.
<b>15_MINDS@UW</b>	Dados que contenham informações sensíveis (dados pessoais), restritos ou confidenciais.	Não localizado.
<b>16_UCSanDiego</b>	Dados que infrinjam direitos autorais; dados que contenham difamação ou matéria ilícita; dados que violem acordos ou leis; dados privados, confidenciais ou que contenham informações legalmente protegidas como dados pessoais identificáveis.	Não localizado.
<b>17_DRUM</b>	Dados privados, confidenciais ou que contenham informações legalmente protegidas, como dados pessoais identificáveis.	O usuário concorda que não fará qualquer uso de dados para identificar ou infringir os direitos de privacidade ou confidencialidade de indivíduos descobertos inadvertidamente ou intencionalmente nos dados.
<b>18_UCL Research Data Repository</b>	Dados contendo informações pessoais ou de categoria especial; dados sobre os quais o depositante ou a UCL não detenham os direitos autorais ou de propriedade intelectual, ou não tenham permissão para compartilhamento; dados que estejam em 'Acordo de compartilhamento' em vigor.	A Política de Privacidade do Serviço do Repositório garante: a privacidade de dados processados por estudantes da UCL, que devem registrar seus projetos no Escritório de Proteção de Dados da Universidade).

Fonte: elaborado pela autora.

Observando-se o Quadro 16, percebe-se que há uma conformidade entre as restrições e as diretrizes para tratamento de dados que contenham informações sensíveis. Porém, em algumas políticas, identificou-se restrição à tipologia dos dados, mas não se localizaram diretrizes sobre o tratamento de dados com informações pessoais, sendo elas dos repositórios **MINDS@UW** e **UC San Diego**. Em outras, ao contrário, não foram localizadas restrições quanto ao tipo de dado, somente diretrizes

sobre a privacidade de dados pessoais, sendo elas dos repositórios **Data Repository [Berkeley]**, **Merritt** e **ISPS Data Archive**.

Constatou-se, nas políticas, a presença de diretrizes orientando que os conteúdos depositados não contenham informações pessoais identificáveis e que dados pessoais, quando houver, devem ser anonimizados em etapa anterior ao depósito.

Na etapa seguinte, são apresentadas as análises referentes ao embargo de dados.

#### 4.3.10 Análise quanto ao embargo de dados

O embargo refere-se à prática que impede o acesso e o uso de determinado arquivo submetido a depósito em um repositório digital, podendo este ser total ou parcial, geralmente por um tempo pré-estabelecido.

No Quadro 17, pode-se observar as informações sobre embargo de dados de pesquisa, identificadas nas políticas analisadas.

**Quadro 17 – Diretrizes sobre embargo de dados**

Repositórios	Embargo no uso dos dados
1_Harvard Dataverse	Não menciona tempo específico de embargo.
2_Stanford Digital Repository - SDR	Tempo máximo de 3 anos.
3_DSpace @ MIT	Tempo máximo de 1 ano.
4_Data Repository [Berkeley]	Por padrão, todo o conteúdo aceito é tornado público mediante aprovação do pessoal de curadoria, contudo permite embargo de um ano e no máximo 10 anos.
5_Research Works Archive	Tempo máximo de 5 anos.
6_Deep Blue Data	Não menciona.
7_ORA-Data	Permite que o depositante determine um período de embargo informado em formulário ao depositar no ORA. Permite também que um item seja mantido em acesso fechado permanentemente.
8_JHU Data Archive	Não possibilita embargo de dados.
9_Columbia Academic Commons	Não menciona.
10_eCommons	Podem ser impostos embargos temporários ao acesso aberto mundial e embargos a texto completo, por no máximo dois anos.
11_Scholarly Commons	Não menciona.
12_Apollo	Permite embargo, mas não especifica tempo.

<b>13_Merritt</b>	Embargo de até seis meses para conjuntos de dados que estejam sendo revisados por pares.
<b>14_ISPS Data Archive</b>	O depositante ao depositar dados determinará se os arquivos serão de acesso 'público' ou 'privado'.
<b>15_MINDS@UW</b>	Não menciona.
<b>16_UCSanDiego</b>	Não localizado no <i>site</i> , mas, de acordo com contato feito com a UCSD, os dados podem ser embargados. As informações de embargo ao dado são adicionadas no registro do dado. Os metadados não são embargados.
<b>17_DRUM</b>	Os autores têm a possibilidade de restringir o acesso por no máximo dois anos.
<b>18_UCL Research Data Repository</b>	O acesso aos dados pode ser embargado temporariamente antes da publicação. Permite dois tipos de embargos: <b>embargo somente de arquivo</b> , permitindo a visualização dos metadados associados ao arquivo de dados; e <b>embargo do arquivo completo</b> , aplicado aos dados e seus metadados.

Fonte: elaboração da autora.

Em relação ao embargo no uso de dados, pode-se constatar que, dos 18 repositórios que tiveram suas políticas analisadas, 13 permitem algum tipo de embargo aos dados, sendo este especificado pelo depositante. Somente um repositório não permite embargo de dados, o **JHU Data Archive**, e em outros quatro, não foram localizadas informações referentes a embargo de dados, sendo eles o **Deep Blue Data**, o **Columbia Academic Commons**, o **Scholarly Commons** e o **MINDS@UW**. Embora não se tenha localizado informações sobre embargo nas políticas do repositório **UC San Diego**, obteve-se, via contato com a instituição, a informação de que os dados podem ser embargados. Tais informações estão incluídas nos registros dos dados, mas os embargos não se aplicam aos metadados.

Alguns repositórios permitem embargo, mas não especificam o tempo. Já para os repositórios que estipulam tempo de embargo, esse varia de, no máximo seis meses a 10 anos. Embora o **ORA-Data** vá além desse tempo limite identificado nas políticas dos outros repositórios, permitindo, além de embargo temporário, que um item seja mantido embargado permanentemente. O **Data Repository [Berkeley]** é o único que indica tempo mínimo de embargo, que é de um ano.

Enquanto o **ISPS Data Archive** menciona que o embargo se dará no momento do depósito, quando o depositante especificará se o arquivo será de acesso público ou privado, o **UCL Research Data Repository** especifica que o embargo pode ser aplicado somente a um arquivo – permitindo visualização dos metadados – ou ao depósito completo, embargando dados e metadados.

Ainda que o movimento de abertura de dados de pesquisa prediga o acesso e uso de dados de pesquisa sem impeditivos de qualquer natureza, sabe-se que essa

não é a realidade vivenciada por todos os pesquisadores. O embargo, como um impeditivo ao acesso e compartilhamento de dados de pesquisa, pode dar-se por diversos motivos, “[...] entre eles por questões legais [...] e questões contratuais, como acordos firmados normalmente com editoras ou periódico científico no momento da publicação na forma de artigos” (JORGE, 2018, p. 63).

Contudo, o embargo não deve ser utilizado como um instrumento de proibição permanente ao acesso e compartilhamento de dados. De acordo com Bertin, Visoli e Drucker (2017), se existirem razões para limitar ou restringir a reutilização ou redistribuição dos dados, deve-se estimar um prazo para essa restrição. Ou seja, deve-se estimar um prazo necessário de embargo.

No Quadro 18, pode-se observar a relação entre repositórios e tempo de embargo.

**Quadro 18 – Tempo de embargo identificado nas políticas**

REPOSITÓRIOS	UNIVERSIDADES	TEMPO DE EMBARGO
13_Merritt	University of California Los Angeles UCLA	Máximo de seis meses.
3_DSpace@MIT	Massachusetts Institute of Technology	Máximo de 1 ano.
10_eCommons	Cornell University	Máximo de 2 anos.
17_DRUM	University of Minnesota System	Máximo de 2 anos.
2_Stanford Digital Repository	Stanford University	Máximo de 3 anos.
5_Research Works Archive	University of Washington	Máximo de 5 anos.
4_Data Repository [Berkeley]	University of California Berkeley	Min. 1 / Máx. 10 anos.

Fonte: elaborado pela autora.

Com base no Quadro 18, identifica-se que, entre os repositórios que permitem o embargo, constando o tempo limite de não disponibilização desses dados em suas políticas, tem-se: o **Stanford Digital Repository – SDR**, com tempo máximo de três anos para embargo de dados; **Dspace@MIT**, com tempo máximo de um ano; o **Data Repository [Berkeley]**, sinalizando tempo mínimo de um ano e máximo de 10 anos; o **Research Works Archive**, com tempo máximo de cinco anos; o **eCommons** e o **DRUM**, com tempo máximo de dois anos; e o **Merritt**, com o menor tempo de embargo, seis meses.

Buscando-se analisar tempo de embargo de dados identificados em outros estudos, recuperou-se a tese de doutorado de Costa (2017), sobre *Diretrizes para uma política de gestão de dados científicos no Brasil*. No estudo, a autora, por meio de questionário, buscou identificar a visão de servidores de agências de fomento sobre prazo para embargo de dados, considerando o processo de publicação

científica no Brasil e a nível internacional. Os respondentes tinham as seguintes opções de respostas para tempo de embargo: menos de um ano; entre um e dois anos; entre três e cinco anos; entre cinco e oito anos; entre oito e dez anos; e “não sabe”.

Segundo Costa (2017), as respostas dos servidores revelaram uma expectativa de um tempo máximo de cinco anos para embargo de dados, embora três respondentes desejassem um prazo menor, de menos de um ano de embargo. Contudo, ainda segundo a autora, um prazo menor que um ano se torna inviável, visto que, no Brasil, o Instituto Nacional da Propriedade Industrial (INPI) leva oito anos para a concessão de uma patente, com expectativa de redução desse prazo para seis anos, e a nível internacional, os Estados Unidos, por meio de seu Escritório de Patentes, leva em média dois anos para a concessão de uma patente.

Considerando o acima exposto, entende-se que, por razões políticas, comerciais, financiamento privado ou registro de patentes, alguns dados precisam que seu acesso e compartilhamento sejam restringidos por meio de embargo. Sendo assim, conforme Sayão e Sales (2015), esse período de embargo dos dados deve ser estabelecido em comum acordo entre os repositórios e seus depositantes.

Na etapa seguinte, é apresentada a análise sobre as responsabilidades no uso do Repositório.

#### 4.3.11 Análise quanto as responsabilidades no uso do Repositório

Com base nas análises realizadas, a delimitação de responsabilidades no uso dos repositórios e seus conteúdos trata-se de um acordo firmado por meio de ‘aceite’ no seu acesso e uso, geralmente sendo entre: repositório e usuário; repositório e depositante; depositante e usuário; instituição e usuário. Abaixo, no Quadro 19, apresentam-se os limites de responsabilidade identificados nas políticas analisadas.

**Quadro 19 – Responsabilidade no uso dos Repositórios**

Repositórios	Responsabilidade no uso do repositório
1_Harvard Dataverse	Discrimina responsabilidade de Usuários (depositantes) e <i>downloaders</i> , isentando o Harvard Dataverse de qualquer responsabilização e indenização.
2_Stanford Digital Repository - SDR	Os Termos de depósito indicam responsabilidade do depositante quanto à detenção dos direitos ou licença sobre o conteúdo depositado.
3_DSpace @ MIT	Discrimina responsabilidades da Comunidade DSpace@MIT, das bibliotecas do MIT e do MIT.

<b>4_ Data Repository [Berkeley]</b>	Os Termos de Serviço indicam isenção de responsabilidade do Dryad; responsabilidades dos autores, depositantes e usuários.
<b>5_ Research Works Archive</b>	Faz menção somente à responsabilidade pelo controle da qualidade dos envios.
<b>6_ Deep Blue Data</b>	Os Termos de Uso indicam responsabilidades dos depositantes e usuários ( <i>downloaders</i> ).
<b>7_ ORA-Data</b>	As Políticas do ORA-Data indicam responsabilidades da Universidade e dos pesquisadores/depositantes.
<b>8_ JHU Data Archive</b>	Indica responsabilidades dos usuários (depositante e <i>downloaders</i> ).
<b>9_ Columbia Academic Commons</b>	Indica responsabilidades dos usuários e depositantes.
<b>10_ eCommons</b>	Indica responsabilidade dos usuários (usuários externos e internos) do eCommons e da Cornell University Library.
<b>11_ Scholarly Commons</b>	Indica responsabilidade dos autores/depositantes (contribuintes do repositório).
<b>12_ Apollo</b>	Indica responsabilidade do Repositório e dos depositantes.
<b>13_ Merritt</b>	Indica responsabilidades do depositante, do usuário (quem visualiza, baixa e reutiliza conteúdo) e da CDL.
<b>14_ ISPS Data Archive</b>	Indica responsabilidade do usuário (quem usa, deposita ou baixa conteúdo), do ISPS e da Yale.
<b>15_ MINDS@UW</b>	Indica responsabilidade do depositante.
<b>16_ UC San Diego</b>	Indica responsabilidade da UCSD, dos depositantes e dos usuários.
<b>17_ DRUM</b>	Indica responsabilidade dos depositantes, dos usuários (quem acessa e baixa conteúdo) e da Universidade.
<b>18_ UCL Research Data Repository</b>	Indica responsabilidade dos usuários (depositantes) e da UCL.

Fonte: elaborado pela autora.

Observou-se que todas as políticas analisadas discriminam algum tipo de responsabilidade no uso dos repositórios, sendo geralmente indicada em seus Termos de depósito, de Uso ou de Serviços. As responsabilidades identificadas estão relacionadas aos autores, depositantes, usuários, repositórios e às próprias instituições.

Cabe salientar que o conceito de usuário, entre as políticas analisadas, apresenta uma variação conceitual. Em algumas políticas, o termo 'usuário' corresponde aos indivíduos que utilizam os serviços do repositório para depositar conteúdo, o que corresponderia ao depositante; já em outras políticas, o termo 'usuário' está relacionado aos indivíduos que acessam, baixam e reutilizam conteúdo dos repositórios; e, por fim, em outras políticas, o termo 'usuário' corresponde aos que utilizam os serviços do repositório para além de acessar e baixar, depositando conteúdo.

Para facilitar a identificação e a comparação das responsabilidades instituídas nas políticas, considerou-se o termo ‘depositante’ para o indivíduo que contribui fazendo depósito de conteúdo nos repositórios, e o termo ‘*downloader*’ para todos os indivíduos que acessam os serviços do repositório para visualizar, baixar e reutilizar seus conteúdos. Com a contribuição do Quadro 20, é possível observar as responsabilidades identificados nas políticas dos repositórios analisados.

**Quadro 20 – Comparativo das responsabilidades identificadas nas políticas**

REPOSITÓRIOS	UNIVERSIDADES	RESPONSABILIDADES			
		Depositante	<i>Downloader</i>	Repositório	Universidade e seus órgãos
1_Harvard Dataverse	Harvard University	x	x		
2_Stanford Digital Repository	Stanford University	x			
3_DSpace@MIT	Massachusetts Institute of Technology				x
4_Data Repository [Berkeley]	University of California Berkeley	x	x		
5_Research Works Archive	University of Washington	x			
6_Deep Blue Data	University of Michigan	x	x		
7_ORA-Data	University of Oxford	x			x
8_JHU Data Archive	Johns Hopkins University	x	x		
9_Columbia Academic Commons	Columbia University New York	x	x		
10_eCommons	Cornell University		x	x	x
11_Scholarly Commons	University of Pennsylvania	x			
12_Apollo	University of Cambridge	x		x	
13_Merritt	University of California Los Angeles UCLA	x	x		x
14_ISPS Data Archive	Yale University		x		x
15_MINDS@UW Research Data Service	University of Wisconsin Madison	x			
16_UC San Diego Library Digital Collections	University of California San Diego	x	x		x
17_DRUM	University of Minnesota System	x	x		x
18_UCL Research Data Repository	University College London	x	x		x

Fonte: elaborado pela autora.

Pode-se observar, com base no Quadro 21, que a indicação de responsabilidade por parte dos depositantes está presente nas políticas: do **Harvard Dataverse**; do **Stanford Digital Repository-SDR**; do **Data Repository [Berkeley]**; do **Research Works Archive**; do **Deep Blue Data**; do **ORA-Data**; do **Columbia Academic Commons**; do **Scholarly Commons**; do **Apollo**; do **Merritt**; do **MINDS@UW**; do **UC San Diego**; do **DRUM**; e do **UCL Research Data Repository**.

Aos depositantes, recai a responsabilidade de garantir que os conteúdos depositados estejam em conformidade com: a legislação sobre direitos de propriedade; os termos de depósito no repositório; as diretrizes de anonimização de dados, quando necessário; as regulamentações federais, estaduais, institucionais e de agentes financiadores; as diretrizes institucionais de ética na pesquisa; e as diretrizes de qualidade dos arquivos.

A responsabilidade envolvendo *downloaders* encontra-se presente nas políticas dos seguintes repositórios: **Harvard Dataverse**; **Data Repository [Berkeley]**; **Deep Blue Data**; **JHU Data Archive**; **Columbia Academic Commons**; **eCommons**; **Merritt**; **ISPS Data Archive**; **UC San Diego**; **DRUM**; e **UCL Research Data Repository**. Aos *downloaders*, recai a responsabilidade: pelo uso do repositório e possíveis danos causados; pela observância das licenças de uso dos conteúdos baixados; e pelo uso indevido dos conteúdos baixados.

A responsabilidade dos repositórios foi identificada somente nas políticas do **eCommons** e **Apollo**. O que se pode observar nas análises é que em todas as políticas há isenção de responsabilidade por parte dos repositórios em relação a conteúdos depositados, direitos de propriedade, divulgação de dados privados, uso de conteúdos baixados, contratos estabelecidos entre depositante e *downloader* por meio de aceite nos Termos de uso, ou seja, todas as responsabilidades recaem sobre depositantes e *downloaders*. As responsabilidades dos repositórios restringem-se à curadoria, preservação dos conteúdos e garantia de manutenção dos serviços.

As responsabilidades das Universidades ou de seus órgãos foram detectadas nas políticas do: **DSpace@MIT**; **ORA-Data**; **eCommons**; **Merritt**; **ISPS Data Archive**; **UC San Diego**; **DRUM**; e **UCL Research Data Repository**. Como ocorre com os repositórios, as universidades e seus órgãos são isentos de responsabilidade sobre danos causados por acesso e uso do conteúdo armazenado nos seus repositórios, contudo têm responsabilidade corporativa no cumprimento da lei de proteção de dados, como no caso da University College London ou da University of Minnesota,

que indica na sua política a responsabilidade de fornecer acesso perpétuo ao conteúdo armazenado em seu repositório, o DRUM. Ou ainda, como o caso da Biblioteca Digital da University of California, que ao fazer a gestão do conteúdo, tem por responsabilidade não explorar conteúdo gerenciado para obter lucro sem permissão expressa de seus proprietários legais.

De acordo com Soares (2013?), juridicamente, uma relação contratual é o resultado de um “[...] encontro de vontades, aceito pelos manifestantes dessas vontades”, no qual, sem a manifesta aceitação destes, a relação jurídica contratual não se formará. Esse contrato entre as partes, no caso dos repositórios, dá-se por meio de aceite nos ‘Termos de uso’, nos ‘Termos de depósito’ e em outros termos definidos em suas políticas por parte dos depositantes e usuários/*downloaders*. Como as relações se dão de forma virtual, a conformidade com o contrato se dá por meio do aceite desses Termos. Tais contratos sinalizam responsabilidades ou isenções para todos os envolvidos.

De acordo com Michaelis (2021, *online*), entende-se por ‘responsabilidade’ a “obrigatoriedade de responder pelos próprios atos ou por aqueles praticados por algum subordinado.” E, na responsabilidade contratual, como sendo a responsabilidade “[...] que deriva do descumprimento de uma obrigação contratual” (MICHAELIS, 2021, *online*). Ainda de acordo com Michaelis (2021), por ‘limite’, compreende-se uma linha real ou não, cuja ultrapassagem não pode ou não deve ser realizada. Já por ‘isenção’, entende-se juridicamente como a “ação ou efeito de isentar algo ou alguém de cumprimento de obrigação, encargo ou responsabilidade” (MICHAELIS, 2012, *online*).

Parafraseando Sayão e Sales (2015) e trazendo para o contexto da limitação de responsabilidade no uso dos repositórios e seus conteúdos, argumenta-se que, embora o pesquisador tenha um papel crucial na pesquisa, muitas outras pessoas estão envolvidas no processo de acesso, compartilhamento e reuso de dados de pesquisa. Por essa razão, tornam-se imprescindíveis a identificação de papéis dentro desse processo e suas devidas responsabilidades.

A seguir, tem-se a apresentação das diretrizes para a elaboração de políticas de depósito, acesso e uso de dados de pesquisa para repositórios universitários brasileiros e, por fim as considerações finais.

#### 4.4 DIRETRIZES PARA A ELABORAÇÃO DE POLÍTICAS DE DEPÓSITO, ACESSO E USO DE DADOS DE PESQUISAS EM RDPS

Com base no referencial teórico estudado e nas análises das políticas dos 18 repositórios pesquisados, seguem abaixo considerações para uma proposta de política de depósito, acesso e uso de dados de pesquisa para RDPS de universidades brasileiras.

Uma política consiste em um “[...] conjunto de regras e de procedimentos para serem cumpridos [...]” (GOMES; MORGADO, 2012). Nos repositórios, as políticas são documentos que estabelecem diretrizes a serem observadas e atendidas por quem acessa e faz uso dos serviços disponibilizados. Também orienta-se sobre as práticas adotadas pelo repositório para o armazenamento e preservação dos seus conteúdos e sobre as ações de gestão que consideram questões técnicas, legais e de infraestrutura.

As políticas devem refletir e estar em consonância com os princípios e diretrizes organizacionais da instituição à qual o repositório está vinculado, uma vez que ele é uma extensão da organização. Com base na pesquisa realizada, observou-se que geralmente a gestão dos repositórios fica a cargo das bibliotecas, por meio de seus sistemas de preservação digital, em conjunto com os setores de TI das Universidades. Em alguns casos como do Harvard Dataverse, as políticas foram desenvolvidas sob orientação legal de um dos centros de pesquisa da Universidade, o Cyberlaw Clinic no Berkman Center for Internet & Society da University of Harvard.

Aqui cabe uma observação em relação a políticas e termos de uso. Os termos de uso estariam relacionados ao estabelecimento de regras de utilização de um serviço, contudo, na maioria dos repositórios analisados, estão incluídos nas políticas sem distinção. Sendo assim, para esta proposta, os termos de uso são considerados uma de suas políticas.

Na análise das políticas, observou-se que, em sua maioria, não consta a data de sua criação ou atualização. Uma exceção é a Política de Dados de Pesquisa da University College London, que indica que a política será revista, no mínimo, a cada três anos. Considera-se importante constar nas políticas uma previsão para futuras atualizações, visto que isso servirá para estabelecer parâmetros sobre as mudanças implementadas pela gestão do repositório.

Outro elemento a ser considerado nas políticas é a proveniência dos dados para depósito no repositório. É importante especificar: se os dados depositados são exclusivos das atividades científicas de pesquisadores com vínculo institucional ou se o repositório também aceita depósito de pesquisadores externos ou afiliados; se há restrição de depósito para determinados indivíduos dentro da instituição ou se abrange toda a sua comunidade; se os dados abrangem todas as disciplinas ou se o repositório receberá somente o depósito de determinadas áreas. Essas informações podem estar incorporadas na ‘Visão geral do repositório’, que trará uma breve apresentação do repositório, contextualizando-o dentro da instituição e informando seu propósito de atuação e a quem se destina; e, por fim, a qual órgão da universidade cabe sua gerência e a quais diretrizes institucionais o repositório está vinculado.

Observou-se também, durante as análises das políticas, que se faz necessária uma prévia conceituação de alguns termos empregados, que pode materializar-se por meio de um ‘glossário de termos’. Justifica-se a inclusão de um glossário pelo fato de diferentes repositórios poderem adotar o mesmo termo para estabelecer significados diferentes. Por exemplo, alguns repositórios empregam o termo ‘usuário’ para descrever todos os indivíduos que utilizam os serviços do repositório, enquanto outros aplicam o termo ‘usuário’ para designar somente o indivíduo que acessa os serviços, mas não deposita dados, sendo utilizada, nesse caso, a designação de ‘depositante’. O Harvard Dataverse e Deep Blue Data são exemplos de repositórios que apresentam um glossário de definição de termos empregados.

Outros dois componentes a serem incorporados são: a apresentação dos ‘benefícios do repositório’, distinguindo os serviços oferecidos pelo repositório em relação a outras opções de armazenamento e preservação de dados; e as ‘melhores práticas’, adotadas para o compartilhamento de dados, abordando recomendações para: o preparo de arquivos para depósito no repositório; a escolha dos tipos e formatos de arquivos aceitáveis para depósito; a forma adequada de nomear arquivos para evitar que sejam corrompidos ao serem baixados pelos usuários; a inclusão de metadados e preparo da documentação para envio de arquivos de dados ao repositório; e a observância das questões de propriedade intelectual.

Com base nas análises feitas, a seguir, destacam-se quatro políticas essenciais para a gestão dos repositórios. Com exceção da política de restrições, todos os outros documentos estão presentes na maioria dos 18 repositórios analisados. As políticas consistem em documentos separados, não extensos, com informações pontuais.

- **Termos de uso do repositório:** apresenta os termos e condições que controlam o uso do repositório e os serviços disponibilizados e esclarece que o uso dos serviços está condicionado ao aceite desses termos e condições pelo usuário, firmando, assim, um contrato entre as partes. Os termos de uso podem especificar condições em separado para depositantes e usuários, indicando quais as atribuições de cada um e quais suas responsabilidades no uso do repositório e dos serviços oferecidos. Nos repositórios analisados, os termos de uso são disponibilizados em formato de documento para leitura, não sendo necessária nenhuma ação anterior para o seu acesso.
- **Política de acesso e uso dos dados:** expondo informações sobre:
  - Padrão de acesso aos dados no repositório: indica o padrão estabelecido de acesso aos dados, se há outros tipos de acesso (fechado, restrito ou embargado) e quando serão aplicados. O acesso embargado nas políticas analisadas está relacionado a questões contratuais de financiamento de pesquisa, sendo necessário o estabelecimento de prazo para a liberação de acesso ao dado. Assinala, ainda, se o repositório dispõe do serviço de acesso moderado por meio do preenchimento de formulário, que consiste em 'Solicitar ao autor' acesso aos dados que não estão abertos;
  - Licenças de uso: expõe se o repositório atribui uma licença padrão ao uso dos dados ou metadados (ex: dedicação domínio público ou outra), informando também a quem cabe a atribuição de licenças para os conteúdos depositados e quais licenças são utilizadas (Creative Commons, Open Data Commons, outras). Indica, ainda, se mesmo sem uma licença atribuída, é permitido baixar dados;
  - Contrato de uso dos dados: estabelece um contrato entre o *downloader* e o repositório, no qual é explicitado que o *downloader* é conhecedor das políticas de uso dos dados e de suas responsabilidades no uso dos dados baixados, comprometendo-se no adequado uso dos serviços oferecidos pelo repositório. Um exemplo de contrato de uso dos dados é o fornecido pelo Harvard Dataverse, que consiste em um documento para leitura sem exigência de nenhuma ação extra por parte do usuário para o seu acesso.

- Proteção e privacidade dos dados: descreve as diretrizes adotadas pelo repositório para garantir proteção e confidencialidade aos dados que contenham informações sensíveis. Informa quais estratégias de orientação aos autores o repositório utiliza e se recomenda a anonimização de dados.
- **Política de depósito**: esta política abrangerá orientações sobre:
  - Quem pode depositar dados: embora na visão geral do repositório já esteja indicada a proveniência dos dados, é importante explicitar na política de depósito quem são os indivíduos ou instituições com permissão para depositar dados, bem como se é necessário algum tipo de registro ou cadastro. Indica também se será por autodepósito ou depósito mediado pela equipe de gestão do repositório.
  - Coleções e conteúdo: cabem aqui uma definição de dados de pesquisa de acordo com os propósitos do repositório, orientações sobre os tipos e formato dos dados aceito para depósito e o tamanho dos dados e conjuntos de dados (especificar limite de tamanho, se houver);
  - Termos de depósito: corresponde à formalização da responsabilidade do depositante pelo conteúdo depositado no repositório, indicando que este detém os direitos autorais ou tem permissão para depositá-los. Este documento também visa garantir que os conteúdos não contenham informações privadas, restritas e sigilosas ou apresentem qualquer outro impedimento que proíba sua divulgação e compartilhamento. O depositante também se compromete em enviar arquivos que estejam livres de vírus. Em sua maioria, nas políticas analisadas, os termos de depósito são disponibilizados em documento para leitura nos *sites* dos repositórios, contudo o Stanford Digital Repository mostra-se uma exceção, visto que disponibiliza um termo em formato PDF com espaço para assinatura do depositante.
- **Política de restrições para**:
  - Acesso aos dados: especificar quais restrições de acesso podem ser aplicadas; a que se relacionam (ex.: exigência de financiadores); a quem se aplicam (ex.: usuários sem vínculo institucional); quem as impõe (normalmente são impostas pelo depositante); se existem níveis de

restrições (ao item individual ou ao conjunto de dados); e limite de tempo da restrição (se temporárias ou por prazo indeterminado);

- Depósito de dados: indicar se o repositório faz algum tipo de restrição para depósito dos dados, como de pesquisadores sem vínculo institucional, dados incompletos para publicação, dados com tamanho superior ao permitido, dados inadequados para reutilização e dados de disciplinas não incorporadas ao repositório.
- Tipos de dados: especificar as restrições impostas aos tipos de dados, como a dados privados, confidenciais ou restritos; a dados que violem a ética e a conformidade com a pesquisa; e a dados que violem os direitos autorais.
- Download de dados: indicar o tipo de restrição caso haja algum impedimento para uso imediato dos dados e estimar período máximo caso a restrição seja por embargo. Por exemplo: pode determinar que somente usuários com vínculo institucional podem baixar dados ou especificar restrição para determinados conteúdos.

Procurou-se apresentar uma proposta de política considerando os aspectos abordados no estudo e as análises realizadas. A partir da experiência vivenciada na coleta das políticas nas páginas dos repositórios, considera-se importante que elas sejam pensadas com base na experiência do usuário, evitando que sejam necessários vários *clicks* para que se tenha acesso às informações desejadas. As políticas devem estar visivelmente acessíveis aos usuários, de forma clara e pontual.

## 5 CONSIDERAÇÕES FINAIS

Tendo em vista os aspectos apresentados neste estudo, considera-se que os objetivos específicos foram atingidos. Com isso, viabilizou-se o alcance do objetivo geral que visou identificar práticas de acesso aberto a dados de pesquisa instituídas em políticas de gestão de RDPs de universidades internacionais, referente às permissões, restrições e licenciamento para depósito, acesso e uso de dados de pesquisa. O objetivo específico, relacionado à apresentação de diretrizes para a elaboração de políticas de depósito, acesso e uso de dados de pesquisa para repositórios universitários brasileiros, contemplando os aspectos referentes a permissões, restrições e licenças, também foi atendido e apresentado na seção 4.4.

Os resultados das análises demonstram que os aspectos, propostos nos objetivos para identificação nas políticas dos repositórios, encontram-se presentes, em sua maioria, nas políticas analisadas. Contudo, percebeu-se que o propósito de atuação dos repositórios influencia suas políticas, uma vez que, para os repositórios não exclusivos para dados de pesquisa, os aspectos relacionados a permissões, restrições e licenciamento abrangem, em alguns casos, todos os conteúdos digitais armazenados. No entanto, observou-se também que, em outros casos, esses repositórios apresentam políticas específicas para dados de pesquisa, o que pode ser um reflexo de demanda institucional e de agências de fomento, por gestão desses dados.

Embora a pesquisa tenha se proposto a identificar práticas de acesso aberto a dados de pesquisa em repositórios, considerou-se importante verificar se haveria nas políticas alguma diretriz diferente ao acesso aberto e gratuito via *internet* entre os repositórios elencados para a pesquisa. Também considerou-se importante verificar se, nesses repositórios, constaria algum tipo de restrição de acesso aos dados de pesquisa.

Ao procurar-se identificar a proveniência dos dados para depósito nos repositórios, buscou-se traçar um perfil dos depositantes e seus vínculos institucionais. Quanto ao aspecto 'restrições', os elementos embargo e privacidade de dados encontram-se presentes nas políticas analisadas, como imposições ao depósito, acesso e uso dos dados, de forma que identificá-los permitiu-nos, primeiramente, fazer um comparativo entre o tempo estabelecido pelos repositórios,

com o tempo de embargo identificado na expectativa de servidores de agências de fomento para a publicação de dados de pesquisa, apresentado na pesquisa de Costa (2017). Em segundo lugar, identificar as diretrizes expressas nas políticas para o tratamento de dados que contenham informações sensíveis, sendo recomendado, na maioria das políticas, a anonimização de dados. Já quanto ao aspecto ‘responsabilidade no uso do repositório e seus conteúdos’, fazer essa identificação possibilitou verificar que os repositórios, em sua maioria, atribuem as responsabilidades de uso dos serviços e do conteúdo armazenado aos seus usuários, sejam eles depositantes ou *downloaders*, resguardando para si o direito de isenção dessas responsabilidades.

Em relação aos aspectos permissões, restrições e licenças para depósito, acesso e uso de dados de pesquisa, contemplados nos objetivos específicos, sua identificação permitiu identificar as peculiaridades de cada repositório, expostas por meio de suas políticas. Referente às permissões para depósito de dados, pôde-se detectar, que alguns dos repositórios, além do depósito de seus pesquisadores, aceita submissão de dados por pesquisadores externos, pesquisadores afiliados ou mesmo por instituições afiliadas, enquanto outros são exclusivos para sua comunidade.

Já em relação às permissões para uso de dados, constatou-se que algumas políticas estabelecem permissões distintas para usuários distintos, enquanto outras permitem acesso imediato aos dados por meio de *downloads*. Referente às restrições, as análises permitiram identificar que a maioria das políticas faz restrições quanto à tipologia dos dados para depósito e algumas impõem restrições para uso de dados.

Em relação à atribuição de licenças para uso de dados, as análises possibilitaram constatar que as políticas sinalizam o uso de licenças e, em sua maioria, indicam o uso das Creative Commons e das Open Data Commons, sendo que as CC estão contempladas em 77% das políticas.

As análises ainda permitiram fazer um comparativo entre as restrições à tipologia de dados para depósito e a privacidade de dados pessoais, constatando-se, assim, que há uma conformidade, na maioria das políticas, entre as restrições e as diretrizes para tratamento de dados que contenham informações sensíveis.

Uma curiosidade entre as políticas analisadas está na informação vinculada pelo UCL Research Data Repository (RDR), da University College London, de que a Política de Dados de Pesquisa não substitui nenhuma outra política da Universidade,

e se houver algum conflito direto com as políticas dos financiadores de pesquisa, estas terão precedência sobre a Política da UCL.

Por limitações da pesquisa, teve-se a dificuldade em localizar as políticas de alguns repositórios ou informações específicas das políticas em seus *websites*, fazendo-se necessário o contato com as instituições para a obtenção dessas informações. Outro aspecto limitante – provavelmente por desconhecimento das estruturas organizacionais das universidades estrangeiras – foi a identificação de informações para contextualização e apresentação das universidades, bem como a apresentação das informações referentes às políticas nos *websites* dos repositórios.

Faz-se necessário relatar também que algumas informações relacionadas aos aspectos indicados para análises nas políticas encontravam-se dispersas em outras páginas dos repositórios que não a das políticas, como em páginas de Perguntas frequentes (FAQs), dificultando muito a identificação dessas informações para análise. Outro aspecto, exterior às políticas, mas igualmente limitante, foi a escassa literatura sobre ‘responsabilidade no uso de repositórios e de conteúdos depositados’.

Em relação às políticas analisadas, faz-se, ainda, necessário apresentar algumas outras considerações. Nas políticas, não identificou-se alguma referência sobre a gestão diferenciada de dados que considerasse as especificidades das diferentes disciplinas, embora os repositórios sejam multidisciplinares. Isso poderia ser justificado pela prerrogativa de que as permissões, restrições e licenças para depósito, acesso e uso de dados de pesquisa são responsabilidades do depositante, e ele seria a pessoa indicada para levar em consideração as especificidades de sua área de atuação no momento do depósito dos dados no repositório.

Quanto à identificação das políticas nos *websites* dos repositórios e a estrutura de apresentação das informações nelas contidas, pode-se elencar o **Harvard Dataverse**, o **Research Works Archive**, o **Deep Blue Data**, o **Columbia Academic Commons**, o **eCommons**, o **MINDS@UW** e o **DRUM** como os melhores repositórios para análise na presente pesquisa.

Como sugestões para futuros estudos, após a realização deste trabalho, percebeu-se a necessidade de se explorar mais questões referentes às restrições de acesso e uso de dados de pesquisa, tendo-se como instrumentos de restrição o embargo e o seu tempo de vigência, haja vista não se ter recuperado expressiva literatura sobre essa temática.

## REFERÊNCIAS

ALBAGLI, Sarita. Ciência aberta como instrumento de democratização do saber. **Trabalho, Educação e Saúde**, Rio de Janeiro, v. 15, n. 3, p. 659-664, set./dez. 2017. Disponível em: <http://dx.doi.org/10.1590/1981-7746-sol00093>. Acesso em: 22 jan. 2020.

ALBAGLI, Sarita; CLINIO, Anne; RAYCHTOCK, Sabryna. Ciência Aberta: correntes interpretativas e tipos de ação. **Liinc em Revista**, Rio de Janeiro, v.10, n.2, p. 434-450, nov. 2014. Disponível em: <http://revista.ibict.br/liinc/article/view/3593/3072> Acesso em: 20 jan. 2021.

ALBAGLI, Sarita (Org.); MACIEL, Maria Lucia; ABDO, Alexandre Hannud. **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. Disponível em: [https://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta\\_questoes%20abertas\\_PORTUGUES\\_DIGITAL%20\(5\).pdf](https://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20(5).pdf) Acesso em: 20 jan. 2020.

ALMEIDA, Rodrigo Pinheiro de. **Gerenciamento de Dados de Proveniência de Workflow de Bioinformática com banco de dados baseados em grafo**. 2015. Dissertação (Mestrado em Informática) – Departamento de Ciência da Computação, Instituto de Ciências Exatas, Universidade de Brasília, Brasília, 2015. Disponível em: [https://repositorio.unb.br/bitstream/10482/22029/1/2015\\_RodrigoPinheirodeAlmeida.pdf](https://repositorio.unb.br/bitstream/10482/22029/1/2015_RodrigoPinheirodeAlmeida.pdf) Acesso em: 19 jan. 2021.

AVENTURIER, Pascal. Dados de pesquisa: descrição do ciclo de vida de gestão de dados. **A Publicação científica** [Blog]. 2016. Disponível em: <https://publicient.hypotheses.org/1359>. Acesso em: 24 maio 2020.

BALL, Alex. **Review of data management lifecycle models**. Bath: University of Bath, 2012. Disponível em: <https://purehost.bath.ac.uk/ws/portalfiles/portal/206543/redm1rep120110ab10.pdf> Acesso em: 10 jul. 2020.

BERKELEY UNIVERISTY OF CALIFORNIA. **Research overview**. 2021. Disponível em: <https://www.berkeley.edu/research> Acesso em: 02 fev. 2021.

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**, Salvador, v. 11, n. 2, p. 34-48, ago. 2017. Disponível em: <https://portalseer.ufba.br/index.php/revistaici/article/view/21449/15200> . Acesso em: 21 jan. 2020.

BORGMAN, Christine L. **Research Data**: who will share what, with whom, when an why. (RatSWDWorking Paper n.. 161. Oct. 2010). Disponível em: [https://papers.ssrn.com/sol3/Delivery.cfm/SSRN\\_ID1714427\\_code1215910.pdf?abstractid=1714427&mirid=1](https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID1714427_code1215910.pdf?abstractid=1714427&mirid=1). Acesso em: 19 maio 2013.

BORGMAN, Christine L. The Conundrum of Sharing Research Data. **Journal of the American Society for Information Science and Technology**, v. 63, n. 6, p. 1059-1078, June 2012. Disponível em: <http://onlinelibrary.wiley.com.ez45.periodicos.capes.gov.br/doi/10.1002/asi.22634/pdf> Acesso em: 20 jan. 2021.

BRASIL. Congresso Nacional. **Lei nº 12.527**, de 18 nov. 2011, regula o acesso a informação [...]. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/12527.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/12527.htm). Acesso em: 14 jun. 2020.

CARVALHO, Catarina Prestes de; CARVALHO, Rodrigo Aquino de. Construção de políticas para repositórios institucionais: análise da ferramenta do OpenDOAR. **Biblos: Revista do Instituto de Ciências Humanas e da Informação**, v. 26, n. 2, p.105-138, jul./dez. 2012. Disponível em: <https://periodicos.furg.br/biblos/article/download/3450/2211/10380> Acesso em: 28 jan. 2021.

CARVALHO, Érika Rayanne Silva de. **Diferenças na produção, compartilhamento e (re)uso de dados**: a percepção de pesquisadores da química, antropologia e educação. 2018. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2018. Disponível em: <https://repositorio.unb.br/handle/10482/32559>. Acesso em: 31 out. 2019.

COLUMBIA UNIVERSITY. **History**. 2021a. Disponível em: <https://www.columbia.edu/content/history> Acesso em: 05 jan. 2021.

COLUMBIA UNIVERSITY. **Research**. 2021b. Disponível em: <https://www.columbia.edu/content/research> Acesso em: 05 jan.2021.

CORNELL UNIVERSITY. **Cornell research**. 2021. Disponível em: <https://research.cornell.edu/content/2019-research-stats-faculty-distinctions> Acesso em: 06 jan. 2021.

CORNELL UNIVERSITY. **University facts**: Cornell by the numbers. [2020]. Disponível em: <https://www.cornell.edu/about/facts.cfm> Acesso em: 06 jan. 2021.

CORNELL UNIVERSITY LIBRARY. **eCommons**: Cornell's Digital Repository. 2020. Disponível em: <https://guides.library.cornell.edu/ecommmons> Acesso em: 06 jan. 2021.

COSTA, Maíra Murrieta. **Diretrizes para uma política de gestão de dados científicos no Brasil**. 2017. Tese (Doutorado em Ciência da Informação) - Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2017. Disponível em: <https://repositorio.unb.br/handle/10482/24895>. Acesso em: 27 out. 2019.

COX, A. M.; PINFIELD, S. Research data management and libraries: current activities and future priorities. **Journal of Librarianship and Information Science**, London, v. 46, n. 4, p. 299-316, 2014. Disponível em:

<https://journals.sagepub.com/doi/pdf/10.1177/0961000613492542>. Acesso em: 23 maio 2020.

CREATIVE COMMONS. **About licenses CC**. 2019. Disponível em: <https://creativecommons.org/about/ccllicenses/>. Acesso em: 18 jun. 2020.

DDI-Lifecycle. 2020. Disponível em: <https://ddi-lifecycle-documentation.readthedocs.io/en/latest/User%20Guide/Introduction.html>. Acesso em: 25 maio 2020.

DIGITAL CURATION CENTER (DCC). **O que é curadoria digital?** 2020. Disponível em: <http://www.dcc.ac.uk/digital-curation/what-digital-curation>. Acesso em: 09 jan. 2020.

DONEDA, Danilo. **Proteção de dados pessoais e a pesquisa científica**. (Formação modular sobre Direito de Acesso a Informação). Fiocruz, 2019.

FECHER, Benedikt; FRIESIKE, Sascha. **Open science: one term, five schools of thought**. May 30, 2013. RatSWD\_WP\_218. Disponível em: [http://ssrn.com/abstract=2272036srn.com/sol3/papers.cfm?abstract\\_id=2272036](http://ssrn.com/abstract=2272036srn.com/sol3/papers.cfm?abstract_id=2272036). Acesso em: 20 jan. 2020.

FREUD, Gislaine Parra; SEMBAY, Márcio José; MACEDO, Douglas Dyllon Jeronimo de. Proveniência de dados e segurança da informação: relações interdisciplinares no domínio da Ciência da Informação. **RICI – Revista Ibero-Americana de Ciência da Informação**, Brasília, v. 12, n. 3, p. 807-825, set./dez. 2019. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/21203/23548> Acesso em: 18 jan. 2021.

GIL, Antonio Carlos Gil. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002. Disponível em: <http://www.urca.br/itec/images/pdfs/modulo%20v%20-%20como%20elaborar%20projeto%20de%20pesquisa%20-%20antonio%20carlos%20gil.pdf>. Acesso em: 10 dez. 2020.

GOMES, Eugênio Maria; MORGADO, Almir. **Compêndio de administração**: administração pública e privada de A a Z. Rio de Janeiro: Elsevier, 2012. 427 p.

GOMES, Sandra Lúcia Rebel. O Acesso Aberto ao conhecimento científico: o papel da universidade brasileira. **RECIIS – Rev. Eletron. de Comun. Inf. Inov. Saúde**, v. 8, n. 2, p. 93-106, jun. 2014. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/618/1584> Acesso em: 21 jan. 2021.

HARVARD DATAVERSE. **Join a growing community of Harvard and worldwide researchers who share data in Harvard Dataverse**. 2020. Disponível em: <https://support.dataverse.harvard.edu/>. Acesso em: 01 jun. 2020.

HARVARD DATAVERSE SUPPORT. **For Researchers**. 2020. Disponível em: <https://support.dataverse.harvard.edu/researchers> Acesso em: 17 jan. 2021.

HARVARD LIBRARY. Harvard Dataverse. Disponível em: <https://library.harvard.edu/services-tools/harvard-dataverse> Acesso em: 26 dez. 2020.

HARVARD UNIVERSITY. **About Harvard**. Disponível em: <https://www.harvard.edu/about-harvard/harvard-glance> Acesso em: 25 dez. 2020.

HARVEY, Douglas Ross. **Digital curation: a how-to-do-it manual**. London: Facet, 2010.

HEY, Tony (Edt.); TANSLEY, Stewart; TOLLE, Kristin. **The fourth paradigm: data-intensive scientific discovery**. Redmond, Washington: Microsoft, 2009. Disponível em: [https://www.microsoft.com/en-us/research/wp-content/uploads/2009/10/Fourth\\_Paradigm.pdf](https://www.microsoft.com/en-us/research/wp-content/uploads/2009/10/Fourth_Paradigm.pdf) . Acesso em: 22 jun. 2020.

HIGGINS, Sarah. The DCC Curation Lifecycle Model. **The International Journal of Digital Curation**, v. 3, n. 1, p. 134-140, 2008. Disponível em: <http://www.ijdc.net/article/view/69/48> Acesso em: 24 maio de 2020.

JOHNS HOPKINS UNIVERSITY. 2021a. **History & mission**. Disponível em: <https://www.jhu.edu/about/history/> Acesso em: 04 jan. 2021.

JOHNS HOPKINS UNIVERSITY. 2021b. **Schools & divisions**. Disponível em: <https://www.jhu.edu/schools/> Acesso em: 04 jan. 2021.

JORGE, Vanessa de Arruda. **Abertura e compartilhamento de dados para pesquisa nas situações de emergência em saúde pública: o caso do vírus Zika**. 2018. Tese (Doutorado em Ciência da Informação) – Instituto Brasileiro de Informação em Ciência e Tecnologia/Universidade Federal do Rio de Janeiro/Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2018. Disponível em: <http://ridi.ibict.br/handle/123456789/993>. Acesso em: 31 out. 2019.

KURAMOTO, Hélio. Réplica – Acesso livre: caminho para maximizar a visibilidade da pesquisa. **RAC**, Curitiba, v. 12, n. 3, p. 861-872, jul./set. 2008. Disponível em: [https://www.researchgate.net/publication/250993051\\_Acesso\\_livre\\_caminho\\_para\\_maximizar\\_a\\_visibilidade\\_da\\_pesquisa/fulltext/0373da030cf227a117e2267d/Acesso-livre-caminho-para-maximizar-a-visibilidade-da-pesquisa.pdf](https://www.researchgate.net/publication/250993051_Acesso_livre_caminho_para_maximizar_a_visibilidade_da_pesquisa/fulltext/0373da030cf227a117e2267d/Acesso-livre-caminho-para-maximizar-a-visibilidade-da-pesquisa.pdf). Acesso em: 21 jan. 2020.

LEMOS, Ronaldo. **Direito, tecnologia e cultura**. [Rio de Janeiro: FGV, 2005]. Disponível em: <http://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/2190/Ronaldo%20Lemos-Direito-Tecnologia-e-Cultura.pdf?sequence=1> Acesso em: 21 jan. 2021.

LYON, Liz. **Dealing with data: role, rights, responsibilities and relationships consultancy report**. p. 1-65, jun. 2007. Disponível em: [http://opus.bath.ac.uk/412/1/dealing\\_with\\_data\\_report-final.pdf](http://opus.bath.ac.uk/412/1/dealing_with_data_report-final.pdf). Acesso em: 17 maio 2020.

MEDEIROS, Jackson da Silva; CAREGNATO, Sônia Elisa. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **Liinc em Revista**, Rio de Janeiro, v. 8, n. 2, p. 311-322, set. 2012. Disponível em: <http://revista.ibict.br/liinc/article/view/3368/2969>. Acesso em: 06 mar. 2020.

MELIS, Maria Fernanda Mascarenhas dos Santos. **Acesso aberto aos dados de pesquisa nas universidades brasileiras e os indicadores de CT&I**. 2018. Dissertação (Mestrado Profissional em Propriedade Intelectual e Transferência de Tecnologia para a Inovação) – Programa de Pós-Graduação em Propriedade Intelectual Transparência de Tecnologia para a Inovação, Universidade de Brasília, Brasília, 2018. Disponível em: <http://repositorio.unb.br/handle/10482/34537>. Acesso em: 31 out. 2019.

MICHAELLIS. **Responsabilidade** [verbetes]. 2021. Disponível em: <https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/responsabilidade/> Acesso em: 23 jan. 2021.

MICHENER, William K. Ten Simple Rules for Creating a Good Data Management Plan. **PLOS Computational Biology**, v. 11, n. 10, p. 1-9, oct. 2015. Disponível em: <https://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1004525&ty=printable>. Acesso em: 23 maio 2020.

MIT LIBRARIES. **About DSpace@MIT**. [2021?]. Disponível em: <https://libguides.mit.edu/c.php?g=176372&p=1158829>. Acesso em: 07 abr. 2021. MONTEIRO, Elizabete Cristina de Souza de Aguiar. **Direitos autorais nos repositórios de dados científicos: análise sobre os Planos de Gerenciamento dos Dados**. 2017. Dissertação (Mestrado em Informação e Tecnologia) – Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2017. Disponível em: <https://repositorio.unesp.br/handle/11449/149748>. Acesso em: 23 maio 2020.

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: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 23, n. 53, p. 160-173, set./dez. 2018. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p160/37296>. Acesso em: 01 fev. 2020.

MONTEIRO, Elizabete Cristina de Souza de Aguiar *et al.* 2017. A privacidade e os planos de gerenciamento de dados de repositórios de dados científicos. **Informação&Tecnologia (ITEC)**, Marília, v. 4, n. 1, p. 35-53, jan./jun. 2017. Disponível em: <https://periodicos.ufpb.br/index.php/itec/article/view/37586/21006>. Acesso em: 23 maio 2020.

MONTEIRO, Gabriela; LUCAS, Elaine Rosangela de Oliveira. Dados científicos abertos: identificando o papel das políticas de gestão e das agências de fomento. **AtoZ: novas práticas em informação e conhecimento**, Curitiba, v. 8, n. 1, p. 13-20, jan./jun. 2019. Disponível em: <https://revistas.ufpr.br/atoz/article/view/67253/38725>. Acesso em: 14 mar. 2020.

MORENO, Fernanda Passini. Repositórios de dados de pesquisa na Espanha: breve análise. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 23, n. 53, p. 52-63, set./dez. 2018. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/download/1518-2924.2018v23n53p52/37376>. Acesso em: 05 fev. 2020.

NASCIMENTO, Francisco José Tavares do. **Direito de acesso à informação e proteção de dados pessoais**. (Formação modular sobre Direito de Acesso a Informação). Fiocruz, 2019.

National Science Foundation. **Long-Lived Digital Data Collections**: enabling research and education in the 21st century. (2005). Disponível em: <http://www.nsf.gov/pubs/2005/nsb0540/> Acesso em: 17 maio 2020.

OCDE. **OECD Principles and Guidelines for Access to Research Data from Public Funding**. OECD, 2007. Disponível em: <http://www.oecd.org/sti/sci-tech/38500813.pdf> Acesso em: 20 jan. 2020.

OLIVEIRA, Adriana Carla Silva de; SILVA, Edilene Maria da. Ciência aberta: dimensões para um novo fazer científicos. **Informação & Informação**, Londrina, v. 21, n. 2, p.5-39, maio/ago. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27666/20113>. Acesso em: 16 jul. 2019.

OLIVEIRA, Larissa Melo Bezerra de. **Elementos estruturantes para a elaboração de política de direito autoral para repositórios institucionais de acesso aberto**. 2015. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação da Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, DF, 2015. Disponível em: [https://repositorio.unb.br/bitstream/10482/19208/1/2015\\_LarissaMeloBezerradeOliveira.pdf](https://repositorio.unb.br/bitstream/10482/19208/1/2015_LarissaMeloBezerradeOliveira.pdf). Acesso em: 25 maio 2020.

OPEN KNOWLEDGE FOUNDATION. **Licenses FAQ**. [20--]. Disponível em: <https://opendatacommons.org/faq/licenses/index.html>. Acesso em: 26 jun. 2020.

OPEN KNOWLEDGE PORTUGAL. **Dados abertos**: [licenças]. [20--]. Disponível em: <http://dadosabertos.pt/licencas.php>. Acesso em: 19 jun. 2020.

PAMPEL, Heinz *et al.* **Making research data repositories visible**: the re3data.org Registry. PLoS One, San Francisco, v. 8, n.11, 2013. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0078080>. Acesso em: 15 maio 2016.

PAVÃO, Caterina Groppo *et al.* 2018a. Acesso aberto a dados de pesquisa no Brasil: práticas e soluções tecnológicas. *In*: SIMPÓSIO INTERNACIONAL NETWORK SCIENCE, 2., 2018, Rio de Janeiro. **Anais eletrônicos...** Rio de Janeiro, MPRJ, 2018, *on-line*. Disponível em: [http://networkscience.com.br/wp-content/uploads/2018/11/IISINS\\_EScience\\_DadosAbertos\\_Artigo\\_AcessoAberto.pdf](http://networkscience.com.br/wp-content/uploads/2018/11/IISINS_EScience_DadosAbertos_Artigo_AcessoAberto.pdf). Acesso em: 26 jun. 2020.

PAVÃO, Caterina Groposo *et al.* **Acesso aberto a dados de pesquisa no Brasil: repositórios brasileiros de dados de pesquisa: relatório 2018.** 2018b. Disponível em: <http://hdl.handle.net/10183/185138>. Acesso em: 04 mar. 2020.

PATTON, M. Q. Qualitative Evaluation. In: LEWIS-BECK, M. S.; BRYMAN, A.; LIAO, T. F. (Eds.) **Encyclopedia of Social Science Research Methods**. Thousand Oaks, CA: SAGE Publication, 2007.

PRÍNCIPE, Pedro. **Sabe que a partilha de dados torna a ciência mais aberta?** Formação modular sobre Ciência Aberta, 2019.

REPOSITÓRIOS CIENTÍFICOS DE ACESSO ABERTO DE PORTUGAL. **Acesso Aberto, Embargado, Restrito e Fechado: relembrar conceitos...** 2021. Disponível em: <http://projeto.rcaap.pt/index.php/lang-pt/noticias/206-acesso-aberto-embargado-restrito-e-fechado-relembrar-conceitos> Acesso em: 14 fev. 2021.

ROCHA, Rafael Port da *et al.* **Acesso aberto a dados de pesquisa no Brasil: soluções tecnológicas: relatório 2018.** Porto Alegre, RS: UFRGS, 2018. Disponível em: <http://hdl.handle.net/10183/185126>. Acesso em: 21 maio 2020.

RODRIGUES, Eloy *et al.* **Dados de pesquisa aberto.** (Formação modular sobre Ciência Aberta). Fiocruz, 2019. Disponível em: <https://mooc.campusvirtual.fiocruz.br/rea/ciencia-aberta/serie1/curso1/aula3.html>. Acesso em: abr. 2019.

SALES, Luana Farias; SAYÃO, Luis Fernando. **Dados de pesquisa: gestão e curadoria.** [Curso]. 2020. Disponível em: <https://pt.slideshare.net/luanasales/dados-de-pesquisa-gesto-e-curadoria>. Acesso em: 28 maio 2020.

SALES, Luana Farias. **Integração semântica de publicações científicas e dados de pesquisa: proposta de modelo de publicação ampliada para a área de Ciências Nucleares.** 2014. Tese (Doutorado em Ciência da Informação) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014. Disponível em: <http://repositorio.ibict.br/bitstream/123456789/874/1/LUANA%20SALES%20D.pdf>. Acesso em: 19 dez. 2019.

SALES, Luana Farias; SAYÃO, Luis Fernando. Uma proposta de taxonomia para dados de pesquisa. **Conhecimento em Ação**, Rio de Janeiro, v. 4, n. 1, p. 31-48, jan./jun. 2019. Disponível em: <https://revistas.ufrj.br/index.php/rca/article/download/26337/14573>. Acesso em 05 mar. 2020.

SANCHEZ, Fernanda Alves; VIDOTTI, Silvana Aparecida Borsetti Gregório; VECHIATO, Fernando Luiz. A contribuição da curadoria digital em repositórios digitais. **Revista Informação na Sociedade Contemporânea**, Natal, número especial, p. 11-17, 2017. Disponível em: <https://periodicos.ufrn.br/informacao/article/download/12280/8508> Acesso em: 26 jun. 2020.

SANTA ANNA, Jorge; DIAS, Célia da Consolação; MACULAN, Benildes Coura. A gestão dos dados de pesquisa nas universidades e o papel dos serviços informacionais oferecidos nas bibliotecas: uma revisão narrativa. **Múltiplos olhares em Ciência da Informação**, Belo Horizonte, v. 9, n. 2, p. 1-16, 2019. Disponível em: <https://periodicos.ufmg.br/index.php/moci/article/view/19169>. Acesso em: 21 maio 2020.

SANTOS, Jean Carlos Ferreira dos. **Estudo sobre o movimento Open Access e de suas implicações para a comunicação científica**. 2014. Dissertação (Mestrado em Política Científica e Tecnológica) – Instituto de Geociências, Universidade Estadual de Campinas, Campinas, 2014. Disponível em: [http://repositorio.unicamp.br/jspui/bitstream/REPOSIP/286633/1/Santos\\_JeanCarlos\\_Ferreirados\\_M.pdf](http://repositorio.unicamp.br/jspui/bitstream/REPOSIP/286633/1/Santos_JeanCarlos_Ferreirados_M.pdf) Acesso em 19 maio 2020.

SANTOS, Paula Xavier dos (coord.). **Livro Verde – ciência aberta e dados abertos**: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional. Rio de Janeiro: Fiocruz, 2017. Disponível em: <https://www.arca.fiocruz.br/handle/icict/24117>. Acesso em: 14 out. 2019.

SAYÃO, Luís Fernando; SALES, Luana Farias. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 1, n. 21, p. 90-115, maio/ago., 2016a. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/download/27939/20122> Acesso em: 09 nov. 2019.

SAYÃO, Luís Fernando; SALES, Luana Farias. Curadoria digital de dados de pesquisa: desafios informacionais, gerenciais, tecnológicos e políticos. In: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO, 6. Londrina, 2016b. **Slides...** Disponível em: <http://carpedien.ien.gov.br:8080/handle/ien/1803>. Acesso em: 16 jan. 2020.

SAYÃO, Luís 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**, Belo Horizonte, v. 6, n. 1, 2013. Disponível em: <http://www.brapci.inf.br/index.php/article/download/48943>. Acesso em: 21 jan. 2020.

SAYÃO, Luís Fernando; SALES, Luana Farias. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. Disponível em: [http://www.cnen.gov.br/images/CIN/PDFs/GUIA\\_DE\\_DADOS\\_DE\\_PESQUISA.pdf](http://www.cnen.gov.br/images/CIN/PDFs/GUIA_DE_DADOS_DE_PESQUISA.pdf). Acesso em: 09 nov. 2019.

SILVA, Fabiano Couto Corrêa da. **Gestão de dados científicos**. Rio de Janeiro: Interciência, 2019. 128 p.

SILVA, Faysa de Maria Oliveira; SIEBRA, Sandra de Albuquerque. Análise de modelos de ciclos de vida para curadoria de objetos digitais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18. Marília, SO, 2017.

**Anais eletrônicos.** Disponível em: <https://www.brapci.inf.br/index.php/res/v/105514>  
Acesso em: 24 maio 2020.

SOARES, Luís Alfredo Macedo. **Aspectos gerais da revisão judicial dos contratos.** 2013? Disponível em:  
[https://www.academia.edu/download/45383111/Aspectos\\_Gerais\\_da\\_Revisao\\_Judicial\\_dos\\_Contratos.pdf](https://www.academia.edu/download/45383111/Aspectos_Gerais_da_Revisao_Judicial_dos_Contratos.pdf) Acesso em: 23 jan. 2021.

SOUZA, Rosali Fernandez. Universo de ciência e tecnologia: organização e representação em classificações do conhecimento. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13. Rio de Janeiro, RJ, 2012. [Trabalho apresentado]. Disponível em:  
<http://www.eventosecongressos.com.br/metodo/enancib2012/arearestrita/pdfs/19371.pdf> Acesso em: 07 abr. 2021.

SOUZA, Allan Rocha de. **Propriedade industrial e Direitos autorais.** (Formação modular sobre Ciência Aberta). Fiocruz, 2019.

STANFORD LIBRARY. Depositor services. [2020?]. Disponível em:  
<https://library.stanford.edu/research/stanford-digital-repository/depositor-services>  
Acesso em: 17 jan. 2021.

STANFORD UNIVERSITY. **SDR overview.** 2020. Disponível em:  
<https://library.stanford.edu/research/stanford-digital-repository/sdr-overview>. Acesso em: 09 jun. 2020.

SWAN, Alma. **Diretrizes para as Políticas de desenvolvimento e promoção do Acesso Aberto.** Brasília: Unesco, 2016. Disponível em:  
[https://unesdoc.unesco.org/ark:/48223/pf0000246018\\_por](https://unesdoc.unesco.org/ark:/48223/pf0000246018_por) Acesso em: 21 jan. 2021.

TORINO, Emanuelle. Políticas em repositórios digitais: das diretrizes à implementação. In: VECHIATO, Fernando *et al.* **Repositórios digitais: teoria e prática.** 2017. Disponível em: <https://portolivre.fiocruz.br/redirecionamento?nid=1694>  
Acesso em: 17 jan. 2021.

UCL. **About UCL.** 2021. Disponível em: <https://www.ucl.ac.uk/about/what/key-statistics> Acesso em: 10 jan. 2021.

UCLA. **Impact & accomplishments.** 2021a. Disponível em:  
<https://www.ucla.edu/about/impact-and-accomplishments> Acesso em: 07 jan. 2021.

UCLA. **Research & creative activities.** 2021b. Disponível em:  
<https://www3.research.ucla.edu/> Acesso em: 07 jan. 2021.

UC SAN DIEGO. **Campus timeline.** 2021a. Disponível em:  
[https://ucsd.edu/timeline/index.html?\\_ga=2.73362581.702871740.1610137023-1697103937.1607000877](https://ucsd.edu/timeline/index.html?_ga=2.73362581.702871740.1610137023-1697103937.1607000877) Acesso em: 08 jan. 2021.

UC SAN DIEGO. **Economic impact report.** 2021b. Disponível em:  
<https://ucsd.edu/about/economic-impact-report.html> Acesso em: 08 jan. 2021

UFRGS. **Lume mantém liderança no Brasil e na América Latina**. 2020. Disponível em: <http://www.ufrgs.br/ufrgs/noticias/lume-mantem-lideranca-no-brasil-e-na-america-latina>. Acesso em: 06 mar. 2020.

UNIVERSITY OF CAMBRIDGE. **About the University**. 2021a. Disponível em: <https://www.cam.ac.uk/about-the-university/history/early-records> Acesso em: 07 jan. 2021.

UNIVERSITY OF CAMBRIDGE. **Global Cambridge**. 2021b. Disponível em: <https://www.cam.ac.uk/a-global-university> Acesso em: 07 jan. 2021.

UNIVERSITY OF CAMBRIDGE. **Research impact**. 2021c. Disponível em: <https://www.cam.ac.uk/research/impact> Acesso em: 07 jan. 2021.

UNIVERSITY OF MICHIGAN. **Research**. 2021a. Disponível em: <https://umich.edu/research/> Acesso em: 01 jan. 2021.

UNIVERSITY OF MICHIGAN. **Research at U-M**. 2021b. Disponível em: <https://research.umich.edu/research-u-m> Acesso em: 01 jan. 2021.

UNIVERSITY OF MICHIGAN LIBRARY. **Our organization**. 2021. Disponível em: <https://www.lib.umich.edu/about-us/about-library/our-organization> . Acesso em: 07 abr. 2021.

UNIVERSITY OF MINNESOTA. **University of Minnesota System**. 2020. Disponível em: <https://system.umn.edu/campuses> Acesso em: 10 jan. 2021.

UNIVERSITY OF MINNESOTA LIBRARIES. **Business and Service Model for the University of Minnesota Libraries' Data Repository & Curation Service**: a report from the University of Minnesota's Data Management and Curation Initiative. 2015. Disponível em: <https://conservancy.umn.edu/bitstream/handle/11299/171761/DMCIServicesBusinessModelforSharing3.pdf?sequence=6&isAllowed=y> Acesso em: 10 jan. 2021.

UNIVERSITY OF OXFORD. **Facts and figures** – full version. 2020a. Disponível em: <https://www.ox.ac.uk/about/facts-and-figures/full-version-facts-and-figures> Acesso em: 02 jan. 2021.

UNIVERSITY OF OXFORD. **Introduction and history**. 2020b. Disponível em: <https://www.ox.ac.uk/about/organisation/history> Acesso em: 02 jan. 2021.

UNIVERSITY OF OXFORD. **Organisation**. 2020c. Disponível em: <https://www.ox.ac.uk/about/organisation> Acesso em: 02 jan. 2021.

UNIVERSITY OF PENNSYLVANIA. **A tradition of excellence**. [2020?a]. Disponível em: <https://www.upenn.edu/about> Acesso em: 06 jan. 2021.

UNIVERSITY OF PENNSYLVANIA. **Research**. [2020?b]. Disponível em: <https://www.upenn.edu/researchdir> Acesso em: 06 jan. 2021.

UNIVERSITY OF WASHINGTON. **Annual report FY 2019**. 2019. Disponível em: [https://www.washington.edu/research/wp-content/uploads/2020/03/ARFY2019\\_v8.pdf](https://www.washington.edu/research/wp-content/uploads/2020/03/ARFY2019_v8.pdf) Acesso em: 30 dez. 2020.

UNIVERSITY OF WASHINGTON LIBRARIES. **Research Works access restriction policy**. 2020. Disponível em: <https://www.lib.washington.edu/scholpub/scholarly-publishing-services/researchworks/researchworks-access-restriction-policy>. Acesso em: 07 jun. 2020.

UNIVERSITY OF WISCONSIN-MADISON. **About**. 2021. Disponível em: <https://research.wisc.edu/about/> Acesso em: 08 jan. 2021.

UNIVERSITY OF WISCONSIN-MADISON. Research Data Services. **About RDS**. 2020. Disponível em: <https://researchdata.wisc.edu/?s=> Acesso em: 08 jan. 2021.

UK DATA ARCHIVE. **Research data lifecycle**. 2012. Disponível em: <https://www.ukdataservice.ac.uk/manage-data/lifecycle.aspx>. Acesso em 25 maio 2020.

VALENTE, Mariana G.; HOUANG, André. **Creative Commons br**: o que você precisa saber sobre licenças CC. 2020. Disponível em: <https://br.creativecommons.net/wp-content/uploads/sites/30/2021/02/CartilhaCCBrasil.pdf> Acesso em: 11 fev. 2021.

VANZ, Samile Andrea de Souza *et al.* **Acesso aberto a dados de pesquisa no Brasil**: prática e percepções dos pesquisadores: relatório 2018. Porto Alegre, RS: UFRGS, 2018. Disponível em: <http://hdl.handle.net/10183/185195>. Acesso em: 05 mar. 2020.

VICENT-LAMARRE, Philippe; BOIVIN, Jade; GARGOURI, Yassine; LARIVIÈRE, Vicent; HARNAD, Stevan. Estimating Open Access Mandate Effectiveness: the MELIBEA Score. **Journal of the Association for Information Science and Technology**, v. 67, n. 11, p. 2815–2828, 2016. Disponível em: <http://hdl.handle.net/10.1002/asi.23601>. Acesso em 21 jan. 2020.

WIKI AUSTRÁLIA AUSGOAL. **Austrália**: AUSGoal - Governos Australianos abrem o quadro de acesso e licenciamento. 2015. Disponível em: [https://wiki.creativecommons.org/wiki/Australia:\\_AUSGoal\\_-\\_Australian\\_Governments\\_Open\\_Access\\_and\\_Licensing\\_Framework](https://wiki.creativecommons.org/wiki/Australia:_AUSGoal_-_Australian_Governments_Open_Access_and_Licensing_Framework) Acesso em: 21 jan. 2021.

YALE UNIVERSITY. **About Yale**. 2021a. Disponível em: <https://www.yale.edu/about-yale> Acesso em: 07 jan. 2021.

YALE UNIVERSITY. **Centers & institutes**. 2021b. Disponível em: <https://www.yale.edu/research-collections/centers-institutes> Acesso em: 07 jan. 2021.

YALE UNIVERSITY. **Traditions & history**. 2021b. Disponível em: <https://www.yale.edu/about-yale/traditions-history> Acesso em: 07 jan. 2021.