

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

**ANA JULIA LOPES**

**Práticas adotadas para desenvolver Competência em Dados nas  
universidades internacionais**

Porto Alegre  
2024

ANA JULIA LOPES

**Práticas adotadas para desenvolver Competência em Dados nas  
universidades internacionais**

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

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Caterina Marta Groposo Pavão

Porto Alegre  
2024

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**

Reitor: Prof. Dr. Carlos André Bulhões Mendes

Vice-Reitora: Prof.<sup>a</sup> Dr.<sup>a</sup> Patricia Pranke

**FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO**

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

Vice-diretora: Prof.<sup>a</sup> Dr.<sup>a</sup> Vera Regina Schmitz

**DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO**

Chefia: Prof. Dr. Rene Faustino Gabriel Junior

Chefia Substituta: Prof.<sup>a</sup> Dr.<sup>a</sup> Caterina Marta Groposo Pavão

**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

Coordenador: Prof. Dr. Thiago Henrique Bragato Barros

Coordenadora Substituta: Prof.<sup>a</sup> Dr.<sup>a</sup> Jussara Borges de Lima

**CIP - Catalogação na Publicação**

Lopes, Ana Julia

Práticas adotadas para desenvolver Competência em

Dados nas universidades internacionais / Ana Julia

Lopes. -- 2024.

188 f.

Orientador: Caterina Marta Groposo Pavão.

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, 2024.

1. Competência em dados. 2. Dados de pesquisa. 3. Gestão de dados. 4. Bibliotecário de dados. I. Groposo Pavão, Caterina Marta, orient. II. Título.

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

**Universidade Federal do Rio Grande do Sul**

Faculdade de Biblioteconomia e Comunicação

Departamento de Ciências da Informação

Programa de Pós-Graduação em Ciência da Informação

Rua Ramiro Barcelos, 2705, Campus Saúde

Porto Alegre/RS – CEP 90035-007

Telefone: 51 3308 2163

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

ANA JULIA LOPES

**Práticas adotadas para desenvolver Competência em Dados nas universidades internacionais**

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

**Orientadora:** Prof<sup>a</sup>. Dr<sup>a</sup>. Caterina Marta Groposo Pavão

Aprovada em: Porto Alegre, 16 de setembro de 2024

**BANCA EXAMINADORA**

Prof<sup>a</sup>. Dr<sup>a</sup>. Caterina Marta Groposo Pavão – Orientadora  
Programa de Pós-Graduação em Ciência da Informação (PPGCIN)  
Universidade Federal do Rio Grande do Sul (UFRGS)

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Samile Andréa de Souza Vanz – Titular  
Programa de Pós-Graduação em Ciência da Informação (PPGCIN)  
Universidade Federal do Rio Grande do Sul (UFRGS)

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Flavia Maria Bastos – Titular  
Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP)

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Sônia Araújo de Assis Boeres – Titular  
Universidade de Brasília (UNB)

---

Prof. Dr. Rene Faustino Gabriel Júnior – Suplente  
Programa de Pós-Graduação em Ciência da Informação (PPGCIN)  
Universidade Federal do Rio Grande do Sul (UFRGS)

## **Agradecimentos**

À Deus toda honra e glória. Sempre. Sem Ele nada disso seria possível.

Aos meus pais, Arnei e Susana, minha gratidão por todo apoio, suporte e palavras de incentivo. Sem vocês eu não conseguiria, esse momento não é meu, é NOSSO! Só consegui, pois, vocês desde sempre me guiaram, ensinaram e batalharam para que eu tivesse essa oportunidade.

Minha gratidão aos meus familiares e amigos que me ajudaram durante essa trajetória. Minha gratidão à minha tia Beatriz por me receber em Porto Alegre, abrir as portas de sua casa e me adotar como filha. Tia, obrigada por tudo! Nice e Felipe, vocês foram incríveis. Obrigada!

Gratidão à Banca pelas contribuições, Prof<sup>a</sup> Dr<sup>a</sup> Samile, Dr<sup>a</sup> Flávia e Dr<sup>a</sup> Sônia, foi um imenso prazer conhecê-las (mesmo que à distância) e desfrutar de seus conhecimentos e experiências.

E não menos importante, meu ENORME agradecimento à minha querida Orientadora Prof<sup>a</sup> Caterina por toda paciência, conversa, incentivo e paciência (sim, paciência de novo!). Prof. você foi fundamental para essa dissertação iniciar, se desenvolver, e finalizar, você é incrível, quando “crescer” quero ser como você. Tens minha admiração. Muito obrigada, mesmo!

Novamente, obrigada Deus. Que Seu nome seja louvado e engrandecido!

*“Quanto a vocês, porém, sejam fortes e corajosos, pois seu trabalho será recompensado”*

*(2Cr 15:7)*

## RESUMO

A premissa de que os dados de pesquisa são valiosos e requerem tempo e recursos para serem criados, preservados e gerenciados, exige dos pesquisadores habilidades, contudo, nem todos reconhecem todas as etapas da gestão de dados de suas pesquisas. À vista disso, surge a competência em dados de pesquisa como uma maneira de desenvolver as habilidades, conhecimentos e rotinas relativas aos dados de pesquisa. E suas etapas. Esse estudo teve como objetivo analisar as práticas adotadas para desenvolver competências em dados de pesquisa em instituições internacionais. O *corpus* do estudo composto por documentos e informações relacionadas com o desenvolvimento de competências em dados de pesquisa encontradas nos sites das 10 primeiras universidades do *ranking Webometrics.2023*. Os documentos selecionados foram analisados com o auxílio do *software* Nvivo a partir de categorias elaboradas pela autora. Observou-se que as instituições abordaram durante suas atividades de competência em dados assuntos referentes ao manuseio dos dados de pesquisa, destacando as boas práticas da pesquisa científica voltadas a documentação, preservação e armazenamento e *backup*, e a importância de os pesquisadores conhecerem a respeito dessas ações. Trataram também aspectos sobre dados pessoais, sensíveis e anonimizados, ética na pesquisa científica, citação de dados, compartilhamento, curadoria de dados, o que constitui as boas práticas de gestão de dados, repositórios de dados e as políticas adotadas e os termos de uso. Além disso, identificou-se que nas universidades estudadas a Rede de Biblioteca e o Setor de Serviço de Dados de Pesquisa são os responsáveis por desenvolver e elaborar atividades de orientação e suporte para o avanço das práticas de competência em dados. A participação e envolvimento dos bibliotecários no oferecimento e planejamento dos serviços de competência em dados foi significativo, uma vez em todas as universidades participam ativamente das atividades de desenvolvimento das competências em dados dos pesquisadores. Conclui-se que a participação e envolvimento dos bibliotecários no oferecimento e planejamento de serviços de dados de pesquisa nas universidades internacionais selecionadas, são imprescindíveis na geração de conhecimento científico e na orientação durante todo o ciclo de vida das pesquisas. no Brasil, a realidade de alguns profissionais bibliotecários é incompatível com a realidade apresentada nas universidades estudadas, sendo necessário capacitações para que possam adquirir habilidades para cumprir seu papel como agentes da informação nesse contexto.

**Palavras-chave:** Competência em dados. Dados de pesquisa. Gestão de dados. Bibliotecário de dados.

## **ABSTRACT**

The premise that research data is valuable and requires time and resources to create, preserve and manage demands skills from researchers, but not everyone recognizes all the stages of managing research data. In view of this, research data competence emerges as a way of developing skills, knowledge and routines relating to research data and its stages. The aim of this study was to analyze the practices adopted to develop research data competencies in international institutions. The corpus of the study, made up of documents and information related to the development of competences in research data, courses, lectures, workshops, services or research centers, found on the websites of the top 10 universities in the Webometrics.2023 ranking, was analyzed with the help of the Nvivo software using categories drawn up by the author. It was observed that the institutions addressed issues relating to the handling of research data during their data competency activities, highlighting good scientific research practices relating to documentation, preservation and storage and backup, and the importance of researchers knowing about these actions. They also dealt with aspects of personal, sensitive and anonymized data, ethics in scientific research, data citation, sharing, data curation, what constitutes good data management practices, data repositories and the policies adopted and terms of use. In addition, it was identified that in universities the Library Networks and Research Data Service sectors are responsible for developing and designing guidance and support activities for the advancement of data competence practices. The participation and involvement of librarians in the provision and planning of data competency services was significant, as they provided activities on research data management, storage, preservation, among other topics. It can be concluded that the participation and involvement of librarians in the provision and planning of research data services at the selected international universities is essential for generating scientific knowledge and providing guidance throughout the research lifecycle.

Keywords: Data competence. Research data. Data management. Data librarian.

## LISTA DE ILUSTRAÇÕES

Figura 1	Modelo de ciclo de vida de curadoria da DCC	35
Figura 2	Ciclo de vida dos dados formulado pela <i>DataOne</i>	36
Figura 3	Modelo de ciclo de vida dos dados elaborado pela DDI	37
Figura 4	Modelo de ciclo de vida dos dados de pesquisa formulado pela UK Data Archive Data	38
Figura 5	Modelo de Ciclo de Vida do Dado Aberto de Pesquisa.	40
Gráfico 1	Quantidade de documentos analisados por universidade	65
Figura 6	Universidade de Harvard e os Códigos identificados nos documentos	68
Figura 7	Instituto de Massachusetts e os códigos identificados nos documentos	70
Figura 8	Universidade de Michigan e os Códigos identificados nos documentos	72
Figura 9	Universidade Cornell e os códigos identificados nos documentos	74
Figura 10	Universidade de Oxford e os códigos identificados nos documentos	76
Figura 11	Universidade Columbia - Nova York e os códigos identificados nos documentos	78
Figura 12	Universidade da Pensilvânia e os códigos identificados nos documentos	79
Figura 13	Universidade de Stanford e os códigos identificados nos documentos	81
Figura 14	Universidade da Califórnia - Berkeley e os códigos identificados nos documentos	83
Figura 15	Universidade de Washington e os códigos identificados nos documentos	84
Figura 16	Universidades onde o Código “Boas práticas de pesquisa científica” foi identifica	87
Gráfico 2	Número de Referências do Código Documentação sobre os dados, nas universidades	92
Gráfico 3	Quantidade de Referências para o Código Armazenamento e	93

	Backup, nas universidades	
Figura 17	Universidades onde o Código “Ciclo de vida dos dados” foi identificado	94
Figura 18	Universidades onde o Código “Citação de dados” foi identificado	97
Figura 19	Universidades onde o Código “Compartilhamento de dados” foi identificado	100
Gráfico 4	Quantidade de Referências para o Código Compartilhamento de dados, nas universidades	101
Figura 20	Universidades onde o Código “Curadoria de dados” foi identificado	103
Figura 21	Universidades onde o Código “Dados pessoais” foi identificado	106
Figura 22	Universidades onde o Código “Ética” foi identificado	111
Figura 23	Universidades onde o Código “Gestão de dados de pesquisa” foi identificado	113
Gráfico 5	Quantidade de Referências para o Código Gestão de Dados de Pesquisa, nas universidades	116
Figura 24	Universidades onde o Código “Metadados” foi identificado	119
Gráfico 6	Quantidade de Referências para o Código Metadados, nas universidades	123
Figura 25	Universidades onde o Código “Planos de Gestão de Dados” foi identificado	125
Gráfico 7	Quantidade de Referências para o Código Plano de Gestão de Dados, nas universidades	127
Figura 26	Universidades onde o código “Princípios FAIR” foi identificado	129
Figura 27	Universidades onde o código “Repositório de dados” foi identificado	133
Gráfico 8	Quantidade de Referências para o código Repositório de dados, nas universidades	135
Figura 28	Universidades que oferecem atividades de competência em dados por meio de oficinas	142
Figura 29	Universidades que oferecem atividades de competência em dados por meio de cursos	143
Figura 30	Universidades que oferecem atividades de competência em dados por meio de atendimento individual	145

Figura 31	Universidades que possuem setor de serviço de dados	148
Figura 32	Universidades onde a biblioteca é responsável pelos serviços de dados	151
Figura 33	Universidades que oferecem serviços de competência em dados para os pesquisadores	155

## LISTA DE QUADROS E TABELAS

Quadro 1	Objetivos específicos e métodos de pesquisa	51
Quadro 2	Universidades selecionadas para a pesquisa	53
Tabela 1	Universidades e quantidade de documentos recuperados	54
Quadro 3	Códigos e subcódigos criados no Nvivo	57
Tabela 2	Códigos e subcódigos, total de documentos codificados, e total de referências geradas	58
Tabela 3	Casos criados para a pesquisa, quantidade de documentos codificados e suas referências	60
Quadro 4	Lista de atributos associado à classificação de Caso “Instituição”	62
Quadro 5	Atributos e Valores de Atributos	63

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>15</b>
1.1	OBJETIVOS.....	20
1.2	JUSTIFICATIVA.....	20
<b>2</b>	<b>REFERENCIAL TEÓRICO.....</b>	<b>25</b>
2.1	DADOS DE PESQUISA.....	25
2.2	GERENCIAMENTO DE DADOS DE PESQUISA.....	31
2.3	CICLO DE VIDA DOS DADOS.....	33
2.4	COMPETÊNCIA EM DADOS DE PESQUISA.....	41
<b>3</b>	<b>PROCEDIMENTOS METODOLÓGICOS.....</b>	<b>50</b>
3.1	DELINEAMENTO DA PESQUISA.....	50
3.2	DESCRIÇÃO DA ETAPA DE COLETA.....	54
3.3	DESCRIÇÃO DA ETAPA DE ANÁLISE DOS DADOS.....	55
<b>4</b>	<b>ANÁLISES DOS RESULTADOS.....</b>	<b>64</b>
4.1	CONTEXTUALIZAÇÃO DAS UNIVERSIDADES E O USO DE DADOS.....	66
4.1.1	<b>Universidade de Harvard.....</b>	<b>66</b>
4.1.2	<b>Instituto de Tecnologia de Massachusetts.....</b>	<b>68</b>
4.1.3	<b>Universidade de Michigan.....</b>	<b>71</b>
4.1.4	<b>Universidade de Cornell.....</b>	<b>73</b>
4.1.5	<b>Universidade de Oxford.....</b>	<b>75</b>
4.1.6	<b>Universidade de Columbia – Nova York.....</b>	<b>77</b>
4.1.7	<b>Universidade da Pensilvânia.....</b>	<b>78</b>
4.1.8	<b>Universidade de Stanford.....</b>	<b>80</b>
4.1.9	<b>Universidade da Califórnia - Berkeley.....</b>	<b>82</b>
4.1.10	<b>Universidade de Washington.....</b>	<b>83</b>
4.2	AS UNIVERSIDADES E O DESENVOLVIMENTO DE COMPETÊNCIA EM DADOS DE PESQUISA.....	86
4.2.1	<b>Boas práticas da pesquisa científica.....</b>	<b>86</b>
4.2.1.1	<i>Boas práticas da pesquisa científica para preservação de dados....</i>	<i>89</i>
4.2.1.2	<i>Boas práticas da pesquisa científica para Documentação sobre os dados.....</i>	<i>89</i>
4.2.1.3	<i>Boas práticas de pesquisa científica para Armazenamento e</i>	

	<i>backup</i> .....	90
4.2.2	<b>Ciclo de vida dos dados</b> .....	<b>93</b>
4.2.3	<b>Citação de dados</b> .....	<b>97</b>
4.2.4	<b>Compartilhamento de dados</b> .....	<b>99</b>
4.2.5	<b>Curadoria de dados</b> .....	<b>102</b>
4.2.6	<b>Dados pessoais</b> .....	<b>105</b>
4.2.6.1	<i>Dados sensíveis</i> .....	107
4.2.6.2	<i>Anonimização</i> .....	108
4.2.7	<b>Ética</b> .....	<b>110</b>
4.2.8	<b>Gestão de Dados de Pesquisa</b> .....	<b>112</b>
4.2.9	<b>Metadados</b> .....	<b>118</b>
4.2.10	<b>Plano de Gestão de Dados</b> .....	<b>124</b>
4.2.11	<b>Princípios FAIR</b> .....	<b>129</b>
4.2.12	<b>Repositório de dados</b> .....	<b>132</b>
4.2.12.1	<i>Política de dados</i> .....	135
4.2.12.2	<i>Termo de uso de dados</i> .....	137
4.3	<b>AS UNIVERSIDADES E SUAS PRÁTICAS DE COMPETÊNCIA EM DADOS</b> .....	<b>140</b>
4.4	<b>ESTRATÉGIAS PARA O DESENVOLVIMENTO DE COMPETÊNCIA EM DADOS DE PESQUISA NAS UNIVERSIDADES BRASILEIRAS</b> .....	<b>160</b>
5	<b>CONSIDERAÇÕES FINAIS</b> .....	<b>168</b>
	<b>REFERÊNCIAS</b> .....	<b>173</b>

## 1 INTRODUÇÃO

A premissa de que os dados de pesquisa são valiosos e requerem tempo e recursos para serem criados, preservados e gerenciados com o intuito de cumprir uma das suas finalidades que é o reuso, implica que os pesquisadores devem possuir habilidades e dedicar esforços ao gerenciamento desses dados de pesquisa durante todo o seu ciclo de vida.

Com o crescimento exponencial dos dados de pesquisa surgiram diversos desafios relativos à sua gestão ao longo de todo o ciclo de vida dos dados. Um dos problemas apontados na literatura diz respeito à necessidade de um processo de gerenciamento dos dados no que tange a sua visualização e aproveitamento, considerado um pré-requisito para o seu compartilhamento eficaz entre uma comunidade científica específica e comunidades que não fazem parte do grupo alvo. Também, os cientistas de todas as áreas do conhecimento começaram a ter dificuldades para gerenciar e armazenar o enorme volume de dados gerados em suas próprias pesquisas (Monteiro; Sant'Ana, 2018; Lima; Pinto; Farias, 2020).

A respeito disso, Whitmire, Boock e Sutton (2015) afirmam que a complexidade de formatos de dados, padrões, infraestrutura, etc., representam desafios para os pesquisadores atenderem aos requisitos de gerenciamento de dados de pesquisa. A facilidade com que coletam complexos conjuntos de dados está superando seus conhecimentos e habilidades para gerenciá-los adequadamente, habilidades essas, que são cruciais para garantir a qualidade, integridade, capacidade de compartilhamento, descoberta e reutilização dos dados ao longo do tempo de forma econômica, segura e eficiente.

Assim sendo, um plano de gestão de dados (PGD) que oriente os pesquisadores durante a coleta e tratamento até a divulgação dos conjuntos de dados é imprescindível para que sejam cuidados ao longo do seu ciclo de vida. Outro aspecto importante, elencado por Veiga *et al.* (2019), é que os PGDs sejam, se possível, legíveis e acionáveis por máquina, permitindo a integração automatizada com outros sistemas de informação e facilitando o acesso aos dados.

Nessa perspectiva, o crescente anseio pelos dados gerados nas atividades científicas proporcionou, tanto por agências de fomento, instituições de pesquisa e comunidade acadêmica, a necessidade de uma gestão ativa superior ao simples

armazenamento e disponibilização na *web*, um sistema confiável para a publicação de dados baseada em uma potente infraestrutura digital viabilizando o compartilhamento de dados de forma eficaz para que esses recursos atinjam todo seu potencial de reuso (Henning *et al.*, 2019a).

Em busca de práticas para fomentar a padronização e o reuso dos dados de pesquisa é que os Princípios FAIR se apresentam como requisitos orientadores. Os dados de pesquisa devem ser localizáveis, acessíveis, interoperáveis e reutilizáveis, por homens e máquinas. A adoção dos Princípios FAIR pela comunidade de produtores e consumidores de dados de pesquisa permitirá incorporar boas práticas para a publicação, compartilhamento e gestão de dados de pesquisa (Henning *et al.*, 2019b).

À medida que as pesquisas se tornam mais abundantes em dados, pesquisadores enfrentam novos desafios no gerenciamento de informações, de tal maneira que as bibliotecas e seus profissionais começam a especializar-se a fim de promover pluralidade de serviços de suporte em dados, tais como treinamentos, instruções, apoio no gerenciamento, curadoria e visualização (Reis, 2019).

Dessa forma, surgiu, nos últimos anos, o bibliotecário de dados, àquele que está capacitado, a gerenciar e organizar dados em contextos acadêmicos e científicos auxiliando estudantes e pesquisadores que administram volumes de dados.

Os dados de pesquisa desempenham um papel crucial na sustentabilidade e na reprodutibilidade da pesquisa. Ao longo dos anos, as bibliotecas acadêmicas e de pesquisa têm desempenhado um papel fundamental como organizações de informação e um novo papel surgiu em sua evolução, definido como o papel do bibliotecário de dados. Ohaji, Chawner e Yoong (2019) afirmam que o bibliotecário de dados é um profissional dentro da biblioteconomia que realiza atividades de suporte à pesquisa e gestão de dados de pesquisa. Reis e Sena (2021) corroboram que o bibliotecário de dados, por terem uma base de conhecimento especializada no que diz respeito à descoberta, armazenamento e preservação da informação, podem reverter essa habilidade para auxiliar os pesquisadores na administração de seus dados.

A compreensão da necessidade de organização de informações baseada em padrões e as habilidades de gerenciamento de informações que os bibliotecários

possuem, como na organização de informações e recursos, podem ser estendidas à gestão de dados, como mencionado pelos autores supracitados.

Da mesma forma que, fornecer orientações e capacitações em gestão de dados de pesquisa podem ser vistas como uma extensão natural do trabalho existente em serviços de consultoria em educação informacional. Há também uma conexão potencial entre gestão de dados e a agenda de acesso aberto que as bibliotecas têm sido tão ativas na promoção (Cox; Pinfield, 2014). Além disso, as redes profissionais de bibliotecários também os tornam boa alternativa em compartilhar práticas entre instituições.

A diversidade de dados e objetos de informação é uma das principais razões para a necessidade de usar, explorar, tomar decisões, determinar tipos de dados, comunicar e desenvolver competências em dados de pesquisa, exigindo dos pesquisadores habilidades no que diz respeito ao gerenciamento dos dados.

No entanto, nem todos os pesquisadores possuem habilidades com ferramentas e *softwares*, ou reconhecem as etapas da gestão de dados de suas pesquisas, tornando-se fundamental que o pesquisador e demais profissionais que trabalham com dados tenham domínio de todas as etapas que envolvem a gestão dos dados de pesquisa, reconheça a importância da documentação, do conhecimento das mais variadas ferramentas e *software* de tratamento, análise, visualização, preservação e outros instrumentos com os quais está lidando na realização da pesquisa. Isto é, seja competente em utilizar os dados, não somente para saber coletá-los, mas gerenciá-los e disponibilizá-los para os mais diversos fins.

À vista disso, surge o termo *Data Literacy* que é um processo de construção de habilidades e competências em dados, que está relacionado à compreensão que o indivíduo tem para lidar com os dados e suas tipologias.

Steinerová e Ondrisová (2019) afirmam que *Data Literacy*<sup>1</sup> se dedica à compreensão das habilidades, conhecimentos e rotinas relativas aos dados de pesquisa e constitui-se, entre outras coisas, em atividades práticas de manipulação de dados de pesquisa. Este termo segue a tendência de ver a competência em dados como uma prática sociocultural com o uso de tecnologias digitais no ambiente da informação.

---

<sup>1</sup> Traduzido pela autora como Competência em Dados.

O termo “competência” pode ser entendido por meio de diversos conceitos. Belluzzo, Santos e Almeida Junior (2014, p. 63) consideram que competência “[...] constitui um conjunto de conhecimentos, atitudes, capacidades e aptidões que habilitam alguém para vários desempenhos da vida”. Ser competente em algo, pressupõem que a pessoa realize operações mentais, tenha capacidades para usar as habilidades e emprego de atitudes adequadas à realização de atividades e conhecimentos. Fleury e Fleury (2000) enfatizam que a competência prevê se a pessoa é qualificada para realizar algo, ou seja, se tem capacidade de mobilizar, integrar e aplicar conhecimentos em determinados contextos. Diz respeito à habilidade que o indivíduo tem de resolver determinado assunto ou realizar determinada tarefa ou função.

Nota-se que os autores apontam que habilidades estão vinculadas a competências visto que precisam ser inter-relacionadas com conhecimentos para que haja uma atuação competente. No entanto, não se pode considerar que ser competente em algo é o mesmo que ter habilidade. Muitas vezes entendidos como sinônimos, competência e habilidade devem ser diferenciados para melhor compreensão de todo o processo. Belluzzo, Santos e Almeida Júnior (2014, p. 63) dizem que as habilidades “[...] acham-se ligadas a atributos relacionados não apenas ao saber-conhecer, mas ao saber-fazer, saber-conviver e ao saber-se [...]”. Santos e Tolfo (2006) corroboram com a mesma ideia colocando que habilidade está relacionada com a forma de executar tarefas, aplicar conhecimentos, agir e pensar. Isto é, são capacidades que uma pessoa adquire para desempenhar determinado papel ou função.

No contexto dos dados de pesquisa, as competências em dados e as habilidades são imprescindíveis para o bibliotecário e o pesquisador se conscientizarem da importância de estarem sempre se atualizando e ampliando suas competências no desenvolvimento de suas atividades, possibilitando assim, um aprendizado continuado e a longo prazo para sua vida acadêmica e de pesquisa. É nisso que se constitui a competência em dados, estudo contínuo sobre os dados e seus processos.

Sendo assim, as práticas adotadas para desenvolver as competências necessárias entre os pesquisadores para trabalhar com dados de pesquisa em instituições de ensino e pesquisa se mostram importantes para que saibam efetuar a

gestão de seus respectivos dados. Trabalhar com dados envolve atividades e processos para garantir que os mesmos sejam documentados, organizados, armazenados e arquivados, de modo que estejam disponíveis para acesso, uso e reutilização sempre que necessário, durante ou após a conclusão da pesquisa, pela própria equipe ou outros pesquisadores interessados. O preenchimento de um plano de gestão de dados, no qual se descreve como os dados serão coletados, armazenados, quais mecanismos serão utilizados para compartilhamento e preservação, por exemplo, também contribui para uma boa gestão dos dados de uma pesquisa.

Ao ampliar suas competências na gestão de dados de pesquisa, o pesquisador terá autonomia e as habilidades necessárias para garantir a integridade, qualidade e usabilidade dos dados. Além disso, será capaz de adotar boas práticas internacionalmente reconhecidas para o acesso e compartilhamento de dados, como por exemplo, os Princípios FAIR, e assim seus dados estarão prontos para o reuso e suas pesquisas poderão ser replicadas em outros contextos e por outros pesquisadores.

Verifica-se hoje, no Brasil, que diversas instituições estão criando seus repositórios de dados de pesquisa e outras tantas estão estudando as possibilidades de criação. Por outro lado, muitos profissionais da Ciência da Informação buscam capacitações e aperfeiçoamento na área de Ciência Aberta, dados de pesquisa e todo o seu entorno como forma de ampliar seu conhecimento e atender às demandas institucionais e do mercado.

Porém, esses profissionais devem se preparar para ir mais além, para serem disseminadores desse conhecimento e protagonistas na qualificação dos pesquisadores e da comunidade acadêmica que produz, manuseia ou busca dados na sua instituição. Também, criando mecanismos de apoio para o planejamento e práticas de dados de pesquisa, e oferecendo serviços de acordo com as necessidades da comunidade acadêmica em diferentes fases de seus esforços de pesquisas. É necessária uma estreita colaboração entre todas as partes interessadas para iniciar, promover e fortalecer a aprendizagem/capacitação coletiva, além de superar desafios que possam surgir das atividades de pesquisas.

Tendo em mente o contexto exposto e a necessidade de apresentar alternativas que atendam a conjuntura nacional como forma de auxiliar as

instituições, bibliotecas, serviços de informação, etc. a pensar estratégias de competência em dados, formulou-se a seguinte questão de pesquisa: *quais as práticas adotadas para desenvolver competências em dados de pesquisa nas universidades internacionais?*

Para responder à questão de pesquisa proposta foram definidos os objetivos, geral e específicos que guiarão esta pesquisa.

## 1.1 OBJETIVOS

O objetivo geral consiste em *analisar as práticas adotadas para desenvolver competências em dados de pesquisa nas universidades internacionais.*

Para atingir o objetivo geral o mesmo foi desdobrado nos seguintes objetivos específicos:

- a) identificar universidades internacionais que adotam práticas para desenvolver competências em dados de pesquisa;
- b) identificar as práticas adotadas para desenvolver competências em dados de pesquisa nas universidades internacionais selecionadas;
- c) verificar, nas instituições selecionadas, quais são as estruturas de apoio às atividades de desenvolvimento de competências em dados de pesquisa;
- d) propor estratégias para desenvolver competências em dados de pesquisa em âmbito nacional.

## 1.2 JUSTIFICATIVA

Os dados sempre foram elementos importantes nas pesquisas, especialmente nas pesquisas científicas. São informações muitas vezes “brutas” que precisam ser tratadas para posterior disseminação. São um dos principais tópicos e tendências na pesquisa da Ciência da Informação. A CI está em constante evolução devido às rápidas mudanças tecnológicas e ao aumento exponencial de dados. Portanto, tópicos e tendências emergentes são frequentemente estudados para manter a relevância e a aplicação prática das pesquisas. Também, permite desenvolver teorias mais robustas e práticas inovadoras que podem ser aplicadas em bibliotecas e serviços de informação estabelecendo uma base sólida para novas pesquisas e avançar nos conhecimentos na área. O paradigma intensivo de dados da ciência e

da pesquisa tem sido explicado como uma tendência emergente que influencia o trabalho de informação.

Com o surgimento das tecnologias digitais, e das possibilidades de busca e recuperação da informação por meio de mecanismos digitais mais aprimorados surgem pesquisas que enfatizam a Ciência de Dados. Esta ciência reúne múltiplos aspectos da informação a partir de seus dados, contando com uma equipe multidisciplinar de profissionais como, por exemplo, estatísticos, analistas de dados, programadores, matemáticos e também bibliotecários (Reis, 2019). Esses dados produzem informações que são diariamente inseridas em ambientes digitais, gerando um grande volume, uma vez que a cada dia mais e mais informações são incluídos nos bancos de dados.

A intensificação de estudos que aprofundem discussões sobre dados de pesquisa é importante pois abrangem diferenças disciplinares na produção, obtenção, gerenciamento, compartilhamento e reuso de dados. Esses estudos, contribuem para os processos de tomada de decisão em instituições de pesquisa e ensino, fornecem subsídios para a criação ou aperfeiçoamento de sistemas de informação e na especialização para promover pluralidade de serviços de suporte em dados (Carvalho, 2018). O material em questão (os dados) pode ser usado e reutilizado, não mais de modo passivo, mas ativamente para novos propósitos, dando origem a novos materiais, dados e objetos de pesquisa.

Com isso, surgem desafios diante do volume de dados que emergem das pesquisas científicas, principalmente no que tange ao papel dos pesquisadores e bibliotecários ao manipular esses dados. Um desses papéis é em relação à competência em dados, na promoção de cursos sobre ferramentas tecnológicas potenciais para dados de pesquisa, uma vez que este profissional ainda se encontra necessitado de atualizações na perspectiva de dados de pesquisa (Costa; Sales; Zattar, 2020).

Uma análise preliminar da literatura científica possibilitou a identificação de estudos sobre competência em dados de pesquisa que abordam aspectos como: o desenvolvimento de competência de dados (Edwards; Cowie; Trask, 2022), conexões entre os dados de pesquisa abertos a fim descobrir as lacunas de competência em dados (Raffaghelli; Manca, 2023); competência em dados e

impacto tecnológico nas bibliotecas (Oyelude, 2022); gerenciamento de dados de pesquisa e competência em dados e serviços de dados de pesquisa (Koltay, 2022).

Entretanto, a literatura científica analisada é de âmbito internacional, não possuindo destaque brasileiro na produção de estudos e materiais teóricos sobre competências em dados de pesquisa que abordem de forma prática como desenvolver essas competências. Não encontraram-se estudos que relatem experiências nas quais profissionais da informação se envolvem ativamente condução de atividades para desenvolver habilidade e competências para que os pesquisadores, no seu dia a dia, possam fazer uma gestão mais eficiente e eficaz dos dados de suas pesquisas.

Visto isso, justifica-se a escolha do tema de pesquisa perante o déficit na literatura científica brasileira sobre a matéria, pois com o crescente interesse em dados de pesquisa e o movimento em prol da ciência e do acesso aberto, é necessário discutir em território nacional as competências que os pesquisadores precisam desenvolver ou aprimorar para trabalhar com dados de pesquisa durante todo seu ciclo de vida.

Nguyen (2020) destaca que aumentar a competência em dados de pesquisa é um esforço interdisciplinar em várias frentes. Educadores e pesquisadores desempenham um papel fundamental na identificação de problemas e na iniciação de discussões que visam soluções viáveis e valiosas: diagnosticamente, por meio da observação e descrição de problemas; prognosticamente, indicando causas e delineando consequências; e didaticamente, propondo contramedidas normativo-éticas. Tal posicionamento é fundamental para a realização de estudos que consideram as perspectivas dos indivíduos, bem como o contexto dos quais fazem parte.

Outro motivo que levou a realizar este estudo, foi o ingresso da autora no grupo de pesquisa da RDP Brasil, criado a partir de projeto da Rede Nacional de Ensino e Pesquisa (RNP), Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). O grupo de trabalho RDP Brasil foi formado em 2017, a partir de um edital de financiamento lançado pela RNP e pelo Ibict para promover pesquisas sobre compartilhamento dos dados coletados, gerados e utilizados pelos pesquisadores brasileiros.

Inicialmente o grupo foi uma parceria entre professores e alunos da Universidade Federal do Rio Grande do Sul (UFRGS) e a Universidade Federal do Rio Grande (FURG). Atualmente, o RDP Brasil é formado por pesquisadores da Faculdade de Biblioteconomia e Comunicação (FABICO) da Universidade Federal do Rio Grande do Sul (UFRGS), alunos de graduação, mestrado e doutorado. Ao longo destes anos de trabalho, o grupo realizou estudos de natureza teórica, tecnológica e empírica, com a produção de relatórios científicos e técnicos, vídeos e outros materiais instrucionais e de Pesquisa e Desenvolvimento (P&D), de forma a colaborar com a comunidade brasileira no desenvolvimento e implementação de repositórios de dados de pesquisa, em seus aspectos de governança, políticas e sustentabilidade (RDP Brasil, 2022).

Com as experiências dentro do grupo de pesquisa, considerando a crescente demanda de compartilhamento de dados de pesquisa motivada pelo novo paradigma da Ciência Aberta, as diretrizes produzidas pelo Grupo de Trabalho formado para discutir uma proposta de Política Nacional para Ciência Aberta, as recentes recomendações da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) sobre acesso a dados de pesquisas científicas com financiamento público e o trabalho realizado pelo Ibict e RNP no âmbito do Compromisso três do 4º Plano de Ação Nacional para Governo Aberto (OGP), (RDP BRASIL, 2022) surgiu o interesse de estudar sobre os dados de pesquisa, principalmente o que diz respeito às competências, pois viu-se nos estudos do grupo que os pesquisadores carecem de conhecimento sobre boas práticas aplicadas a dados de pesquisa.

Uma das comprovações disso, foi repercussão da elaboração de CARDS como a série “Aprendendo sobre Dados de Pesquisa”<sup>2</sup> e Aprende Aprendendo sobre dados de pesquisa (2023)<sup>3</sup> que visa desmistificar o uso, reuso e compartilhamento de dados. Estes materiais foram traduzidos para a língua espanhola pelo Consórcio Iberoamericano para la Educación en Ciencia y Tecnología (ISTEC)<sup>4</sup> e divulgados na Semana do Acesso Aberto para toda a América Latina.

À vista disso, com o desenvolvimento deste estudo, pretende-se oportunizar informações sobre dados de pesquisa e acompanhar o desenvolvimento e

---

<sup>2</sup> <http://hdl.handle.net/10183/252865>

<sup>3</sup> <http://hdl.handle.net/10183/256319>

<sup>4</sup> <http://sedici.unlp.edu.ar/handle/10915/147179>

entendimento sobre competências em dados de pesquisa nas comunidades acadêmicas.

## 2 REFERENCIAL TEÓRICO

Esta seção apresenta a fundamentação teórica relativa às temáticas tratadas neste estudo. Os temas abordados a seguir são: dados de pesquisa segundo sua diversidade e heterogeneidade na perspectiva da ciência aberta; ciclo de vida dos dados, boas práticas para dados de pesquisa e competências em dados de pesquisa como conceito geral que rege esta pesquisa.

### 2.1 DADOS DE PESQUISA

O acesso à informação científica tornou-se fundamental para a construção de novos conhecimentos científicos. Para acompanhar os rápidos avanços da ciência, são necessárias novas concepções de sua formalização, pois os moldes existentes já não suprem mais as necessidades dos pesquisadores perante a Sociedade da Informação. A Ciência Aberta é um movimento que incentiva a transparência da pesquisa científica desde a concepção da investigação até o uso de *softwares* abertos, além de promover esclarecimento na elaboração de metodologias e gestão de dados científicos para que possam ser distribuídos, reutilizados e estar acessíveis a todos os níveis da sociedade (Silva; Silveira, 2019).

Os dados podem ser gerados em diversos ambientes e por diversos motivos. Podem ganhar novas atribuições durante o seu ciclo de vida. Desta forma, questiona-se o que é um dado e quando ele pode ser considerado um dado de pesquisa?

O entendimento sobre o termo “dado” pode variar consideravelmente entre pesquisadores e áreas do conhecimento. Semeler e Pinto (2019) escrevem que a palavra “dado” significa uma única peça de informação. Para os autores podem ser fatos ou estatísticas reunidas para referência ou análise de algo; também podem ser uma coleção diversa incluindo imagens, fluxo de vídeos ou áudio, *software*, algoritmos, etc. Hjørland (2018), acrescenta que os dados possuem seu conceito utilizado em vários campos como banco de dados, mineração de dados, dados vinculados, arquivo de dados, metadados, e *Big Data*.

Mineração de dados, “*data mining*” em inglês, envolve o uso, principalmente, de Inteligência Artificial (IA) no processo de explorar grandes conjuntos de dados para descobrir padrões, correlações, tendências, preferências e outras informações

úteis. Dados vinculados, em inglês "*Linked Data*", é uma abordagem que facilita a interconexão de diferentes fontes de dados, permitindo conexões mais ricas e novas percepções a partir da combinação de dados de várias fontes disponíveis na *Web*. Entende-se por arquivos de dados o armazenamento de conjuntos de dados, facilitando o acesso, a reutilização e o compartilhamento dos dados entre pesquisadores, instituições e o público em geral. Já os metadados são informações contextualizadas sobre um recurso de informação e ajudam a organizar, encontrar, entender e gerenciar os dados, fornecendo detalhes sobre o conteúdo, estrutura, contexto, qualidade e outras características relevantes. O *Big Data*, produção de grande volume de dados, resultou em estudos sobre a compreensão e uso destes dados que são um dos principais tópicos e tendências na pesquisa da informação. Seu paradigma intensivo tem sido explicado como um movimento emergente que influencia o trabalho de informação.

Os dados sempre foram presentes e constantes nas pesquisas, principalmente de cunho científico. Com a expansão do uso das tecnologias e possibilidades de buscar mecanismos mais elaborados para realizar pesquisas, e a capacidade de armazenar grandes quantidades de dados em redes de infraestrutura digital, despertou nos pesquisadores das ciências naturais e sociais, das artes e humanidades o interesse nos dados de pesquisa (Koltay, 2017; Reis, 2019).

Com a consolidação da ciência aberta os dados de pesquisa começam a ser reconhecidos como elementos importantes, principalmente a partir do momento em que documentos acadêmicos se tornam objetos digitais complexos, onde novos serviços e ferramentas são desenvolvidos para produzir, publicar, disseminar e gerenciar esses conceitos de mídia acadêmica, dando a entender que os padrões e fluxos da comunicação científica estão rapidamente se ajustando às novas normas de publicação e compartilhamento de dados de pesquisa (Sayão; Sales, 2020; Silveira, 2021).

Conceituando o que são dados de pesquisa, nos apoiamos no relatório da OCDE publicado em 2007, que define os dados de pesquisa como “[...] registros factuais (pontuações numéricas, registros textuais, imagens e sons) usados como fontes primárias para pesquisa científica e que são normalmente aceitos na comunidade científica como necessários para validar os resultados da pesquisa” (Organização para a Cooperação e Desenvolvimento Econômico, 2007, p. 13,

tradução nossa). Corroborando com o ponto de vista dos autores supracitados, Sales e Sayão (2020) propõe que os dados de pesquisa são todo e qualquer tipo de registro apurado, observado, ou usado pela pesquisa científica, tratado ou aceito como necessário para validar os resultados pela comunidade científica. No entanto, de acordo com os autores, esse termo não compreende os cadernos de laboratório, análises preliminares e rascunhos de artigos científicos, planos para pesquisas futuras, revisões por pares ou comunicações pessoais com colegas ou objetos físicos.

Ainda, Rothenberg (1995) quando discutiu os desafios associados à preservação de dados digitais e a necessidade de estratégias para garantir que as informações permaneçam acessíveis e utilizáveis no futuro, coloca que alguns tipos de dados têm valores imediatos e duradouros, valor que é reconhecido tanto no momento de sua criação quanto ao longo do tempo; alguns ganham valor com o tempo, inicialmente parecer de menor importância, mas seu valor aumenta à medida que se tornam parte de um contexto mais amplo ou à medida que novas descobertas ou tecnologias aumentam sua relevância; alguns têm valor transitório como por exemplo, logs de sistema ou dados de sensores em tempo real que possuem validade por um curto período e outros ainda são mais fáceis de recriar do que de curar como as simulações computacionais que podem ser repetidas com as mesmas condições e parâmetros.

A partir disso, buscando uma categorização dos dados de pesquisa, a *National Science Board* (2005) considera que os dados encontram-se em diferentes formas, isso inclui sua natureza, reprodutibilidade e o nível de processamento a que foram submetidos, e também provêm de diferentes origens sejam eles observacionais, computacionais ou experimentais. Estas distinções são fundamentais para as escolhas feitas para arquivamento e preservação, tendo importantes implicações nas políticas adotadas.

Borgman (2012) também categoriza os dados de pesquisa em observacionais, computacionais ou experimentais, no entanto inclui os dados procedentes dos registros governamentais, que são dados de empresas públicas e privadas que fornecem dados significativos para pesquisas científicas, humanísticas e científicas sociais.

Harvey<sup>5</sup> (2010 *apud* Sales, 2014, p.52) apresenta outra classificação para os dados de pesquisa dividindo-os em 3 coleções:

[...] coleção de pesquisa - formada por dados locais gerados em um laboratório ou projeto; coleção de comunidade - formada por bases de dados internacionais geradas a partir de pesquisas integradas, como por exemplo, a MGI – Mouse Genome Informatics; e coleção de referência que reúne dados referenciais sobre determinados experimentos.

Por fim, Lyon (2007) considera que é fundamental reconhecer que os dados são heterogêneos e complexos, embora tenha várias categorias genéricas que podem ser aplicadas. A autora se respalda na classificação de duas organizações de pesquisa relacionadas à ciência, sendo elas: a *Natural Environment Research Council* (NERC), que divide os dados em dados canônicos (que tem variação mínima) e dados episódicos (que podem mudar); e o *International Union of Crystallography* (IUCR) que categoriza os dados em dados crus (imagem de arquivos), dados primários (fatores de estrutura) e dados derivados (seis dimensões do modelo estrutural).

Vista a categorização dos dados de pesquisa entende-se que o acesso a eles sustenta-se na necessidade de uso que tendem a ser mais profundas e complexas com o aumento das necessidades e demandas.

Santa Anna, Dias e Maculan (2019) declaram que o acesso aos dados das pesquisas contribui para que ações de organização sejam implementadas, para em seguida os dados serem compartilhados e novo conhecimento produzido. Verifica-se, assim, um novo fenômeno, em que os dados das pesquisas assumem importância considerável, tornando essencial seu compartilhamento de acordo com boas práticas, nacionais ou internacionais, dada a vasta quantidade de dados gerados pelas instituições de ensino e pesquisa, os quais possuem alto potencial de reuso.

Vivencia-se um cenário de mudança de paradigma no mundo científico, onde as práticas de produção, acesso e disseminação do conhecimento estão em franca revisão. Percebe-se que nunca existiram tantos pesquisadores produzindo ciência no mundo como nos últimos anos (Henning *et al.*, 2019b).

---

<sup>5</sup> HARVEY, D. R. **Digital curation**: a how-to-do-it manual. London: Facet, 2010.

As tecnologias disponíveis modificaram os meios de comunicação, como também o entendimento das boas práticas para o compartilhamento e disponibilização de recursos na *Web*, muitas vezes pautados pela abertura do acesso e pela democratização do conhecimento. Reis (2019) afirma que os pesquisadores enfrentam novos desafios no gerenciamento de informações de tal maneira que as bibliotecas e seus profissionais começam a se especializar a fim de promoverem pluralidade de serviços de suporte a dados, sendo importante que os mesmos sejam administrados com boas práticas nacionais ou internacionais.

O termo “boas práticas” pode ser adotado para garantir eficiência, qualidade, segurança e consistência em várias áreas, incluindo negócios, tecnologia, educação, saúde entre outras. As instituições de ensino, institutos de pesquisas e agências financiadoras observam cada vez mais condutas e aplicação de boas práticas, e aqui cita-se como exemplos a FAPESP, a *National Science Foundation* (NSF), a *National Institutes of Health* (NIH) e a Comissão Europeia, que exigem a documentação da gestão dos dados de pesquisa, a adoção de padrões internacionais e o estabelecimento de políticas de acesso aberto (Guandalini; Furnival; Arakaki, 2019; Bonetti; Arakaki, 2022).

O compartilhamento e a gestão dos dados de pesquisa trazem inúmeros benefícios para a ciência, como a transparência, a economia de recursos, a possibilidade de reprodutibilidade, a agilização do ciclo científico, entre outros. No entanto, para alcançar o benefício máximo, não basta compartilhar os dados na *Web*, sem o preparo adequado. É preciso levar em conta boas práticas que possibilitem que os dados sejam encontráveis, acessíveis, interoperáveis e reutilizáveis, o que é proposto pelos Princípios FAIR (Bonetti; Arakaki, 2022).

Em busca de práticas para fomentar a padronização e melhorar o gerenciamento dos dados de pesquisa é que os Princípios FAIR se apresentam como um conjunto de requisitos orientadores, e são reconhecidos mundialmente como elementos-chave para boas práticas nos processos de gestão de dados.

Henning *et al.* (2019b) mencionam que esses Princípios tem o objetivo de implementar um conjunto de metadados definidos tanto para uso de mecanismos computacionais automatizados, quanto para uso por pessoas, e se forem devidamente adotados viabilizam a interoperabilidade entre diferentes ambientes de dados. Entretanto, sua implementação também carrega inúmeros desafios que

precisam ser superados, como os altos investimentos, a necessidade de capacitação, a interpretação subjetiva, a falta de guias para sua aplicação, a aceitação pela comunidade científica e assim por diante.

Outrossim, o guia *Best Practice for Researchers*, elaborado por Eynden *et al.* (2011), com boas práticas para gestão e compartilhamento de dados, aponta informações para ajudar pesquisadores e gestores de dados, em uma ampla gama de disciplinas e ambientes de pesquisa, a produzir dados de pesquisa da mais alta qualidade com o maior potencial para uso a longo prazo.

Conforme o manual supracitado, uma boa gestão de dados é fundamental para dados de pesquisa de alta qualidade e, portanto, excelência em pesquisa, é crucial para facilitar o compartilhamento de dados e garantir a sustentabilidade e acessibilidade dos dados a longo prazo. Se os dados da pesquisa forem bem organizados, documentados, preservados e acessíveis, e sua precisão e validade forem controladas durante todo seu ciclo de vida, o resultado serão dados de alta qualidade, pesquisa eficiente, descobertas baseadas em evidências sólidas e economia de tempo e recursos. Entre as práticas destacadas, está a importância de compartilhar, documentar, armazenar e gerenciar os dados de pesquisa para que os próprios pesquisadores se beneficiem.

Além disso, Guandalini, Furnival e Arakaki (2019) salientam que as boas práticas para os dados de pesquisa podem incluir, também, a elaboração de um Plano de Gestão de Dados que propõe ao pesquisador gerenciar os dados brutos de sua pesquisa, valorizando o compartilhamento em conjunto com a abertura dos dados e seu reuso em benefício aos pesquisadores e à sociedade, de modo que, o reuso dos dados de pesquisa só se torna possível se forem recuperados, acessados, legíveis e estáveis a longo prazo. Para tal, os dados de pesquisa precisam ser curados e ter um bom gerenciamento. A velocidade da produção de dados é um fator que demanda o envolvimento de bibliotecas, bibliotecários e pesquisadores na gestão desses dados para serem disponibilizados (Santanna; Dias; Maculan, 2019; Sayão; Sales, 2020).

## 2.2 GERENCIAMENTO DE DADOS DE PESQUISA

Atualmente as áreas do conhecimento têm como desafios gerenciar e armazenar o crescente volume de dados gerados nas pesquisas. As instituições acadêmicas e científicas começaram a requerer mais responsabilidades no gerenciamento dos dados de pesquisa coletados ou produzidos. A gestão apropriada dos dados é um pré-requisito para seu compartilhamento eficaz e requer planejamento e ações concretas para a coleta, armazenamento e, principalmente, para a fase de recuperação dos dados expandido sua visibilidade e possível uso. (Santanna; Monteiro, 2018).

O gerenciamento de dados de pesquisa ganhou atenção crescente nos últimos anos devido ao aumento da conscientização do valor contido nos dados de pesquisa e dos seus riscos de perda ao longo do tempo. Além da necessidade de gerenciar os dados, tem-se a necessidade de preservar e selecionar os dados com os quais se planeja trabalhar futuramente. (Sesartic; Towe, 2016). Para concretizar as ações de gerenciamento de dados de pesquisa, é imprescindível que se tenha políticas para tais atividades.

Sayão e Sales (2012) afirmam que o arquivamento eletrônico de dados começou a ser incentivado pelas agências de financiamento de pesquisa, indicando que os financiamentos e implementação das diretrizes para o setor de pesquisa começaram a traçar políticas, estratégias e prioridades que considerem os dados de pesquisa como um investimento que necessita ser preservado.

Diante disso, é fundamental que haja colaboração entre os profissionais de diversas áreas e pesquisadores mediante a abertura dos dados, sendo crucial seguir padrões e políticas que garantam o compartilhamento e reuso dos dados, do mesmo modo que seu acesso durante todo o ciclo de vida.

A complexidade dos dados e, conseqüentemente, seu crescente volume criou desafios para sua gestão ao longo de seu ciclo de vida. As principais agências de fomento do mundo, como a *National Science Foundation* e *Economic and Social Research Council*, e também do Brasil, a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), passaram a exigir que os pesquisadores submetam o Plano de Gestão de Dados (PGD) ao solicitar financiamento para pesquisa. O PGD corresponde a um documento que dispõe de informações e instruções que

aborda os critérios de organização e gerenciamento dos dados, incluindo o acesso, coleta, curadoria, disponibilização e questões relacionadas aos direitos autorais para a utilização dos dados após sua publicação. (Santanna; Monteiro, 2018; Monteiro, 2017).

Segundo Smale *et al.* (2018) os PGDs surgiram em 1966 para gerenciar a coleta e análise de dados de projetos aeronáuticos e de engenharia da NASA e para documentar as atividades de pesquisa.

[...] geralmente podem ser vistos como abordando o 'como' da implementação técnica dos processos de gerenciamento de dados de um estudo específico, geralmente nos estágios de coleta ou análise de dados. Tais publicações raramente consideram a questão do 'por que', pois há pouca consideração explícita sobre os motivos que levam os pesquisadores a usar DMPs [...] (Smale *et al.*, 2018, p. 4).

Pesquisadores e instituições de apoio à pesquisa, embora reconheçam a importância do gerenciamento e compartilhamento de dados e do PGD, infelizmente, ainda não estão dispostos a investir tempo na elaboração da documentação necessária à gestão dos dados de pesquisa e resistem à ideia de compartilhar seus dados.

Sayão e Sales (2015, p. 16), no entanto, destacam que duas das principais razões para criação de um PGD são “[...] assegurar a integridade da pesquisa e o seu potencial de replicação; assegurar que os dados e demais registros de pesquisa sejam acurados, completos, autênticos e confiáveis [...]”. Isto é, com um PGD o gerenciamento dos dados de pesquisa auxilia a organização desde a entrada no ciclo de pesquisa até a disseminação, possibilitando, também, o reuso desses dados em futuras pesquisas.

Uma vez obtidos, os dados podem ser utilizados para um fim imediato e posteriormente descartados. Para isso, torna-se fundamental conhecer as fases e fatores do processo de gestão dos dados desde o planejamento, sobre a coleta e geração dos dados, até a visualização, pois não é possível disseminar, compartilhar e reusar os dados se não estiverem adequadamente documentados ao longo de seu ciclo de vida.

## 2.3 CICLO DE VIDA DOS DADOS

Os dados podem existir como registros físicos, arquivados no computador de um pesquisador, em nuvens de armazenamento ou em outras formas de mídia digital. O gerenciamento de dados de pesquisa é sobre a organização de dados, desde sua entrada no ciclo da pesquisa até a disseminação dos resultados. Consiste em várias atividades e processos diferentes associados ao ciclo de vida da pesquisa e dos dados, envolvendo o design e criação de dados, armazenamento, segurança, preservação, recuperação, compartilhamento e reutilização, todos levando em conta as capacidades técnicas, éticas, questões legais e estruturas de governança (Cox; Pinfield, 2014).

A tarefa de estruturar esses dados buscando estudar fatores que promovam ampliação do equilíbrio entre os envolvidos no processo e a máxima otimização do uso dos dados, sugere a utilização de uma demarcação de fases envolvidas no acesso e uso dos dados como forma de destacar os diferentes momentos e fatores envolvidos no seu ciclo de vida.

De acordo com Bertin, Visoli e Drucker (2017) os dados de pesquisa têm uma vida além dos projetos que os criam, pois mesmo depois de encerrar uma pesquisa, os pesquisadores podem continuar trabalhando com os dados ou reutilizá-los para outros estudos e projetos. Isto significa que os dados de pesquisa possuem um ciclo de vida que pode ir além da pesquisa para a qual foram coletados ou gerados. Uma vez que são gerados, processados, analisados e preservados, podendo ser novamente reutilizados, os dados passam por estágios que demandam abordagens e ferramentas distintas de gerenciamento.

Aventurier (2016, *online*) destaca que o ciclo de vida dos dados de pesquisa tem diversas representações e envolve algumas etapas principais:

- a) criação dos dados - que compreende a concepção da pesquisa; o plano de gestão de dados; a localização dos dados existentes; a coleta dos dados; e a captura e criação de metadados seguindo os padrões do domínio;
- b) tratamento dos dados - que abrange a transcrição, tradução, entrada de dados, digitalização, quando for o caso, verificação, validação e limpeza

- de dados, anonimização dos dados quando necessário, descrição dos dados, e a gestão e estocagem dos dados;
- c) análise dos dados - que abarca a interpretação e derivação dos dados, produção de resultados, preparação dos dados para acompanhar as publicações científicas e preservação;
  - d) Preservação dos dados - que envolve a migração dos dados para o formato mais adequado (em geral um formato *open source* ou não proprietário), o *backup* e armazenamento de dados, a criação de metadados e documentação e o arquivamento dos dados;
  - e) Acesso aos dados - que inclui a disseminação, compartilhamento e controle de acesso aos dados, os aspectos jurídicos de reutilização, e a promoção dos dados e
  - f) Reutilização dos dados - que implica no acompanhamento, revisão ou síntese sobre a pesquisa, verificação dos resultados e o ensino e aprendizado.

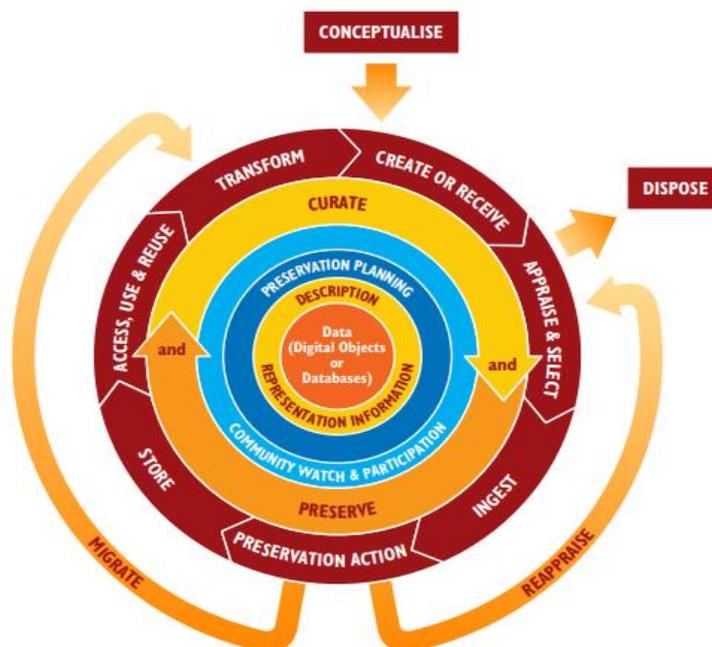
A ideia de ciclo de vida dos dados foi promovida como conceito para apoiar práticas de preservação digital e de curadoria, sendo popularizada conforme a cultura de compartilhamento de dados se torna parte da linguagem de pesquisa cotidiana. Hoje, algumas instituições adotam modelos de ciclo de vida dos dados buscando desenvolver sistemas e promover boas práticas de gerenciamento, organização e preservação de dados de pesquisa (Almeida, 2019).

Apesar de existir várias representações de ciclo de vida para dados de pesquisa, há alguns modelos que se tornaram referências para pesquisadores, bibliotecários e gestores de dados, afirma Ball (2012). De acordo com o autor esses modelos são o *DCC Lifecycle Model*, o *Dataone Data Lifecycle*, o *DDI Combined Lifecycle Model*, e o *UK Data Archive Data Lifecycle*. Cada modelo de ciclo de vida representa o curso de vida de um sistema maior por meio de uma série de etapas sequencialmente relacionadas onde a informação é produzida ou manipulada (Humphrey, 2006).

O modelo desenvolvido pelo *Digital Curation Centre* (DCC), está empenhado em promover uma abordagem de ciclo de vida para o gerenciamento de materiais digitais, com o intuito de permitir a curadoria e preservação bem sucedida dos dados, desde a fase inicial até o descarte ou seleção para reutilização e preservação

a longo prazo, declara Higgins (2008). Conforme demonstrado na Figura 1, o modelo do DCC proporciona uma visão geral de alto nível dos estágios do ciclo de vida necessários para uma curadoria bem-sucedida.

**Figura 1** - Modelo de ciclo de vida de curadoria do DCC



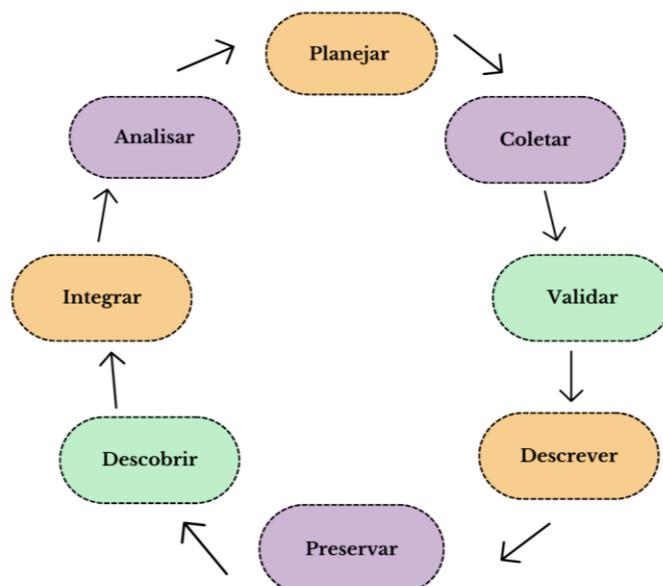
**Fonte:** Higgins (2008, p. 136).

É possível identificar, na Figura 1, três níveis de ações, sendo elas: ações de curadoria aplicáveis em todo o ciclo de vida digital como por exemplo: o planejamento de criação/coleta, armazenamento, compartilhamento e preservação dos dados; ações que necessitam ser realizadas sequencialmente para que a curadoria seja bem-sucedida, novamente verifica-se o planejamento, a coleta e criação, o processamento, o armazenamento e organização, o compartilhamento, a preservação a longo prazo e a reutilização dos dados; e ações que são realizadas ocasionalmente, conforme as circunstâncias ditarem, por exemplo, atualização de metadados, avaliação e adaptação de ferramentas para gestão de dados e revisão do PGD. Importante ressaltar que esse modelo pode ser utilizado para planejar atividades dentro de uma organização para garantir que todas as etapas necessárias sejam realizadas, cada uma na sequência correta (Higgins, 2008).

Enquanto o modelo idealizado pelo DCC dispõe de sete etapas, o modelo proposto pelo *DataOne Data Lifecycle* do *Data Observation Network for Earth*

consiste em oito etapas para disseminar as práticas de uma boa gestão e curadoria de dados, conforme mostrado na Figura 2.

**Figura 2** - Ciclo de vida dos dados formulado pela *DataOne Data Lifecycle*



**Fonte:** traduzido e adaptado de Data Observation Network for Earth (2012).

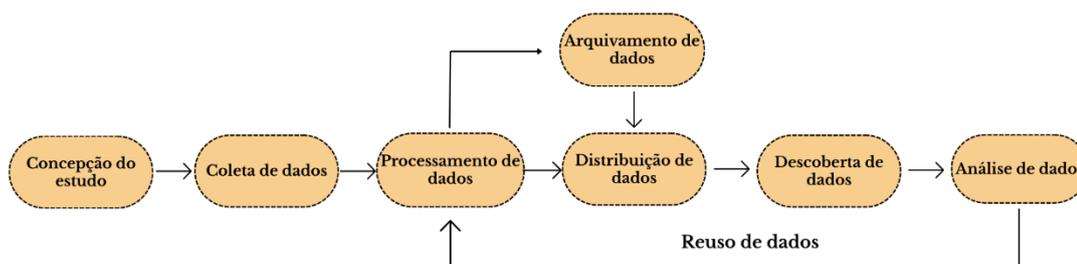
Embora a Figura 2 represente um ciclo sequencial, Bertin, Visoli e Drucker (2017) afirmam que os processos podem ocorrer simultaneamente ou se repetirem. As etapas do modelo *DataOne* abrange a estrutura que representa as operações que precisam ser realizadas sobre os registros de dados durante a sua vida, como apresentado a seguir:

- a) Planejamento: consiste na descrição de um plano de gerenciamento de dados para controlar como os dados são tratados para acesso ao longo do seu ciclo de vida;
- b) Coleta: corresponde às observações e estratégias que são feitas para a coleta de dados com o intuito de disponibilizarem os dados em formato digital legíveis e interoperáveis;
- c) Validação: refere-se à qualidade dos dados assegurados por verificações e inspeções;
- d) Descrição: corresponde à precisão de como os dados são descritos usando os padrões de metadados adequados;
- e) Preservação: consiste nos dados que são enviados para preservação apropriada de curto e longo prazo;

- f) Descoberta: refere-se aos dados potencialmente úteis que são localizados e obtidos juntamente com às informações relevantes sobre os dados (metadados);
- g) Integração: equivale aos dados de fontes diferentes que são combinados para formar um conjunto homogêneo de dados podendo ser facilmente analisado e
- h) Análise: corresponde aos dados analisados com diferentes objetivos e por diferentes pesquisadores, gerando resultados que podem ser usados para novas pesquisas.

O modelo de ciclo de vida dos dados do DDI *Combined Lifecycle Model* desenvolvido pela *Data Documentation Initiative* (DDI), também com oito etapas de ações, tem como foco os dados de pesquisa da área das ciências sociais. Em sua maior parte, o modelo é linear com caminho alternativo possibilitando chances de *feedback*. A Figura 3 apresenta o modelo descrito por Ball (2012).

**Figura 3** - Modelo de ciclo de vida dos dados elaborado pela DDI.



**Fonte:** traduzido e adaptado de Ball (2012).

As etapas do modelo DDI são:

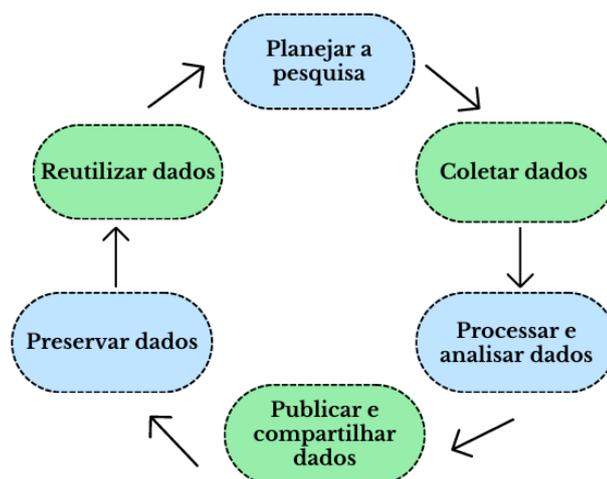
- a) Concepção do estudo: momento inicial em que o pesquisador delimita o problema de pesquisa, metodologia de coleta de dados, e também conduta de processamento e análise dos dados para responder ao problema de pesquisa;
- b) Coleta de dados: etapa que consiste na distinção das fontes de coleta dos dados, primárias e secundárias;
- c) Processamento de dados: uma vez que tenham sido reunidos, os dados passam a ser processados e analisados para produzir dados que respondam à questão de pesquisa. Esses dados podem ser registrados

em formato legível por máquina ou por seres humanos, como um relatório técnico;

- d) Arquivamento de dados: consiste na transferência dos dados para um arquivo a fim de garantir o acesso a longo prazo ao invés de serem mantidos por pesquisadores. O arquivo não apenas preserva os dados e metadados, mas também agrega valor a eles ao longo do tempo;
- e) Distribuição de dados: fase em que os dados são distribuídos aos usuários diretamente ou por meio de uma biblioteca ou arquivo de dados;
- f) Descoberta de dados: implica na divulgação dos dados por meio de livros, publicações periódicas, ou outros serviços *online*;
- g) Análise de dados: compreende a utilização dos dados por outros pesquisadores;
- h) Reuso de dados: consiste no uso dos dados em uma estrutura conceitual diferente, combinando outros conjuntos semelhantes de dados.

Outro modelo referência para pesquisadores, bibliotecários e gestores de dados citado por Ball (2012) é o *UK Data Archive Data Lifecycle* desenvolvido pelo *UK Data Archive*. Fornece um modelo de ciclo de vida de dados como uma ajuda para os pesquisadores, considerando como o gerenciamento de dados se relaciona com o ciclo de vida de um projeto de pesquisa. Esse modelo define seis etapas de ações, como ilustrado na Figura 4.

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



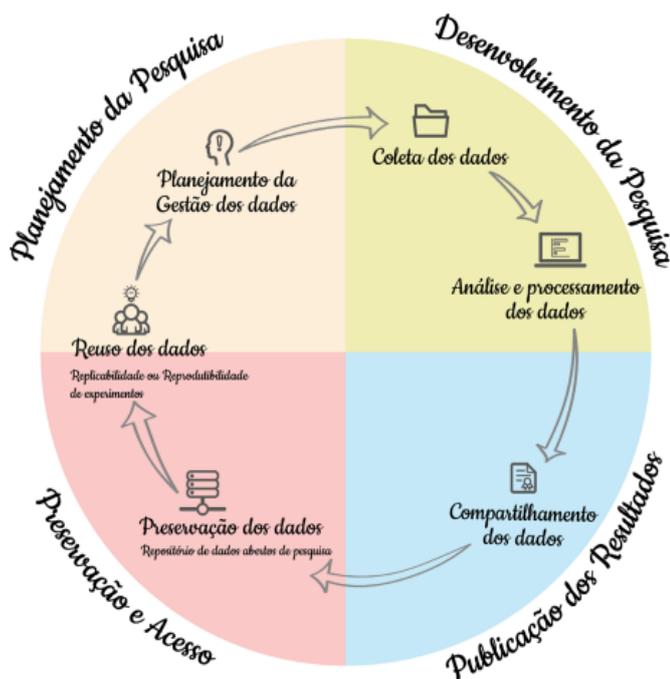
**Fonte:** traduzido e adaptado de Ball (2012).

As seis etapas do modelo são:

- a) Planejamento: compreende a criação de um planejamento de gestão e armazenamento dos dados, observando seus formatos, protocolos e modelos de processamento para explorar as fontes de dados existentes;
- b) Coleta de dados: consiste na coleta de dados com alcance de dados e atribuição de metadados;
- c) Processamento e análise de dados: equivale a fase de digitalizar, transcrever e traduzir dados, isto é, fazer a checagem dos dados;
- d) Publicação e compartilhamento: consiste em estabelecer questões de direitos autorais, criar documentação do usuário, formas de acesso apropriado aos dados para publicar e compartilhá-los;
- e) Preservação dos dados: compreende migrar os dados para o melhor formato, armazená-lo e fazer *backup*, criar documentação de preservação e curadoria;
- f) Reutilização de dados: consiste na descoberta dos dados por outros pesquisadores, possibilitando novos estudos.

Os modelos supracitados demonstram a realidade internacional e sua preocupação com a gestão dos dados de pesquisa e seu ciclo de vida. Evidenciando o contexto brasileiro no que tange ao gerenciamento de dados, o grupo de pesquisa RDP Brasil, em parceria com RNP, IBICT e CNPq, desenvolveu um modelo de ciclo de vida de dados abertos de pesquisa levando em conta a realidade dos estudos e discussões sobre acesso aos dados de pesquisa no país, relacionando as ações inerentes ao ciclo de vida do dado e às etapas de desenvolvimento da pesquisa científica. Conforme apresentado a Figura 5, o modelo trata de um processo constante dividido em quadrantes que representam etapas vitais da pesquisa científica, integrando o planejamento e desenvolvimento da pesquisa, publicação dos resultados e considerando a preservação e o acesso.

**Figura 5** - Modelo de ciclo de vida do dado aberto de pesquisa



**Fonte:** Gabriel Junior *et al.* (2019).

O modelo da Figura 5 contém quatro etapas do ciclo de vida dos dados com seis ações conforme descritas a seguir:

- Planejamento da Pesquisa: consiste no reuso de dados, considerando a reprodutibilidade dos experimentos científicos, e o planejamento da gestão dos dados, onde decide-se quais dados serão coletados, as condições para compartilhá-los ou torná-los abertos, e como esses dados serão curados e preservados;
- Desenvolvimento da pesquisa: etapa que compreende a coleta, a análise e o processamento dos dados obtidos;
- Publicação dos Resultados: equivale ao compartilhamento dos dados, se imediato ou após a aceitação da publicação associada e
- Preservação e acesso: consiste na preservação dos dados por meio dos repositórios de dados abertos de pesquisa possibilitando o reuso de dados, ação compartilhada com a etapa de planejamento da pesquisa, reiniciando-se, assim, o ciclo de vida do dado da pesquisa.

Nesses modelos observa-se que apesar de apresentarem especificidades próprias, levam semelhanças entre si no que diz respeito às etapas na gestão dos dados de pesquisa. Higgins (2008) corrobora dizendo que uma abordagem de ciclo

de vida garante que os estágios necessários para uma boa gestão sejam identificados e planejados para que as ações necessárias sejam implementadas na sequência correta, garantindo a manutenção da autenticidade, confiabilidade, integridade e usabilidade do material digital.

Os dados de pesquisa têm sido considerados um dos principais produtos das pesquisas acadêmicas nas universidades e centros de pesquisas. Nesse cenário os pesquisadores são livres para usar, reutilizar e compartilhar as informações sem restrições tecnológicas e sociais, em um ciclo de pesquisa transparente e aberto, voltado para a colaboração, onde o acesso livre é uma prática comum (Henning *et al.*, 2019b). É nesse contexto que o conhecimento visto como um bem comum auxilia e embasa as reflexões sobre as práticas da pesquisa científica, e requer iniciativas práticas, principalmente quanto ao uso desses dados para o desenvolvimento de novas pesquisas.

O planejamento e propostas de práticas para corroborar com a divulgação e desenvolvimento da ciência baseada em dados de pesquisa, direcionam para um caminho no qual novas práticas, regras e políticas estão sendo reconfiguradas. Sobre isso, Rice e Southall (2016) afirmam que os bibliotecários e pesquisadores, em seus espaços de trabalho, estão aderindo a essas perspectivas, expandindo e iniciando uma série de atividades de apoio às pesquisas, fortalecendo e conscientizando-se sobre as questões que envolvem dados dentro das instituições e os benefícios advindos de gerenciar ativamente os dados de pesquisa.

Com base nisso, desenvolver a competência em dados de pesquisa auxilia pesquisadores e bibliotecários para que acessem, interpretem, avaliem criticamente, gerenciem, manipulem e usem os dados de pesquisa de forma ética (Prado; Marzal, 2013) compreendendo as nuances da pesquisa baseada em dados.

## 2.4 COMPETÊNCIA EM DADOS DE PESQUISA

Na literatura encontram-se os termos “alfabetização em dados” e “competência em dados”, e nesta seção explicaremos o conceito de cada um, qual utilizaremos neste estudo e justificaremos a nossa escolha.

A internet, como rede de computadores, com foco na disseminação e no compartilhamento de informações, surgiu no início dos anos 1990 com a *web* 1.0, chegando aos dias de hoje com a *web* 4.0, "Web Ubíqua" ou "Web Simbiótica", que

se caracteriza pela disponibilidade a qualquer tempo e em qualquer lugar por manter usuários continuamente conectados. Considerada a *web* inteligente, será capaz de fazer deduções, usando a IA para tomar decisões, usando inferência e conteúdo anteriormente pesquisado. É também com a *web* 4.0 que surge a internet das coisas, em inglês *Internet of Things* (IoT), os dados são coletados e transmitidos para a “nuvem”, onde são processados para que rapidamente o usuário obtenha o resultado esperado, por conta própria e em seu dispositivo. Ao mesmo tempo, o *Big Data*, caracterizado por grandes quantidades de dados que não podem ser coletados, armazenados, gerenciados ou analisados por dispositivos comuns, necessita de técnicas e tecnologias avançadas para coleta, armazenamento, distribuição, gerenciamento e análise (Nazarovets; Kulyk, 2017).

Nguyen (2020) expõe algumas preocupações em relação à *Big Data* e a “onipresença” dos dados. Segundo o autor a inaptidão das pessoas que não sabem ler ou entender os dados, contribui para equívocos sobre a magnitude da pandemia, de modo que, apesar de grandes quantidades de dados, que implicam em previsibilidade, há uma noção inevitavelmente forte de incerteza sobre a crise e rumos que esta pode tomar.

Além disso, há questões técnicas: os dados precisam de avaliação em relação ao seu histórico específico, dados falhos, incompletos e tendenciosos podem levar a conclusões imprecisas, e a pouca instrução sobre o significado desses dados entre grandes parcelas do público aumenta a vulnerabilidade das pessoas.

Adquirir conhecimento prévio em dados é de suma importância na era digital uma vez que estamos rodeados por eles. Ter condições de reconhecer as más práticas intencionais, ou não, ao usar dados, e compreender as implicações da datificação por meio da mídia é crucial para navegar em um período de transformação.

Carmi, Yates e Lockley (2020) escrevem que nos últimos 50 anos houve argumento para utilizar os termos alfabetização informacional, alfabetização midiática, alfabetização digital e alfabetização de dados. Cada tipo de alfabetização surgiu em uma época em que os formuladores de políticas faziam suposições sobre a mídia e como diferentes grupos de pessoas deveriam usá-la e para quais propósitos.

Conforme os autores (2020) a “alfabetização em mídia” refletiu o crescimento da mídia de massa e audiovisual em rede, de jornais, rádio, televisão a fitas de vídeo, cabo e satélite. A “alfabetização informacional” desenvolveu-se no início da década de 1970 em torno das bibliotecas por meio do uso de computadores na escola e na educação. As definições, na literatura, em torno da “alfabetização digital” foram o foco quando a internet se tornou mais difundida no final dos anos 1990 e início dos anos 2000, quando acadêmicos e organizações tentavam entendê-la e propor/sugerir como as pessoas deveriam envolver-se com ela. A “alfabetização em dados” surgiu à medida que pesquisadores e formuladores de políticas buscavam entender as implicações de nossa sociedade dataficada.

Este último, diz respeito à maneira como os indivíduos podem envolver-se com os dados. Segundo Kjelvik e Schultheis (2019) um indivíduo alfabetizado em dados deve possuir ferramentas quantitativas e analíticas apropriadas para resolver um problema e a capacidade de aplicar essas ferramentas no contexto para analisar, interpretar e comunicar as descobertas dos dados.

González e Rodriguez (2021), no seu estudo, trazem uma proposta de alfabetização em dados, como curso zero, no qual os estudantes recebem instruções básicas ou iniciais sobre identificação, localização, visualização e recuperação de conjuntos de dados, em um contexto de dados abertos e de ciência de dados. Escrevem que a alfabetização em dados pode ser proposta em vários níveis: básica, intermediária e avançada. Em cada um dos programas de formação que contribuem para a sua aquisição, as competências básicas a desenvolver são:

[...] a) Identificar os dados que são mais apropriados em termos de propósito particular b) Interpretar vários modos de visualização de dados (gráficos, tabelas, etc.); c) Aplicar o pensamento crítico sobre a informação resultante da análise de dados; d) Conhecer e aprender a utilizar ferramentas e métodos de análise de dados; e) Reconhecer quando os dados são deturpados e/ou utilizados de uma forma enganosa ou tendenciosa; f) Comunicar dados - uma habilidade conhecida por "contar histórias de dados" - de uma forma que seja compreensível. (Gonzalez; Rodriguez, 2021, p. 321-322, tradução nossa).

Os autores ainda complementam que a alfabetização em dados pode ser entendida como a capacidade de uma pessoa para ler e compreender o significado dos dados, o que contribui para que qualquer cidadão - não apenas especialista em

*Big Data*, analista de dados, cientista de dados, etc. - possa tomar decisões que afetam seu trabalho ou sua vida cotidiana, com base em dados.

No estudo de Yang e Li (2020), os autores afirmam que a partir da literatura existente sobre alfabetização de dados, pode-se elencar três pontos sobre o assunto. Em primeiro lugar, o principal objeto de pesquisa da alfabetização em dados é focado nos alunos, mesmo do ponto de vista do professor, o objetivo final é melhorar a alfabetização em dados dos alunos. Em segundo lugar, a discussão da alfabetização de dados dos estudantes para seu sucesso se concentra principalmente no aprendizado acadêmico. E terceiro, a colaboração entre as partes interessadas contribui potencialmente para a aprendizagem dos alunos.

Conforme observado pelos autores supracitados, a alfabetização em dados é aplicada quando o indivíduo que não tem conhecimento sobre dados necessita de informações iniciais sobre o assunto, e aqui se destacam, principalmente, alunos que estão ingressando no estudo desse tema, diferente de pesquisadores e bibliotecários que já tem um conhecimento prévio, ainda que precário, e precisam de capacitação continuada para colocarem em prática o conhecimento durante seus trabalhos do dia a dia. Nesse momento é que a competência em dados de pesquisa, foco desta pesquisa, torna-se relevante no contexto desses profissionais.

Em um ambiente no qual se produz um volume de dados cada vez maior, o comportamento dos pesquisadores torna-se intensivo em dados, algumas vezes intuitivos, levando a necessidade de participar de capacitações para desenvolver ou ampliar suas competências em dados de pesquisa.

Os autores Jacobsen, Miletto e Loureiro (2022) declaram que a escolha da terminologia “competência” agrega a ideia de mobilização, compartilhamento e construção de conhecimentos, habilidades e recursos (informacionais e/ou tecnológicos) relacionados ao universo da informação com foco no “aprender a aprender” e na aprendizagem autônoma e contínua para tomada de decisões quanto ao comportamento para a identificação das necessidades informacionais.

No contexto desse estudo, a preocupação em tornar os indivíduos competentes para lidar com os dados de pesquisa é o objetivo da competência em dados, também conhecida pelos termos em inglês como *Data Literacy* ou *Data Competency*, traduzido para o português como Literacia em Dados (Silva, 2017;

Gutiérrez, 2019), Alfabetização em Dados (Reis, 2019) ou Competência em Dados (Balbinotti, *et al.*, 2022; Costal *et al.* 2020).

Em relação à tradução do termo *Data Literacy*, frequentemente encontrado na literatura, nos apoiamos no estudo de Balbinotti e colaboradores (2022) para justificar a escolha do termo competência em dados. Os Autores declaram que para o uso dessa expressão levou-se em consideração o histórico do termo Competência em Informação nas áreas da Biblioteconomia e Ciência da Informação, uma vez que diversos autores traduzem o termo *Information Literacy*, utilizado na literatura internacional, para Competência em Informação.

A respeito disso, Koltay (2015) afirma que ao falar sobre a questão de encontrar o nome próprio para o termo *Information Literacy*, é útil observar as dificuldades experimentadas nas práticas de nomear “competência informacional”.

No Brasil, o termo *Information Literacy* foi mencionado pela primeira vez pela professora Sônia Caregnato (2000), que o traduziu como Alfabetização Informacional. No ano seguinte, Dudziak (2001) defendeu a primeira dissertação de mestrado sobre o tema no país e optou por não utilizar a tradução. No entanto, na área da Biblioteconomia e Ciência da Informação, Hatschbach (2002) e Campello (2003) adotaram o termo “Competência Informacional” para se referirem ao conjunto de habilidades integradas que englobam a descoberta reflexiva de informação, a compreensão de como ela é produzida e valorizada, e o seu uso na criação de novos conhecimentos (American Library Association, 2016).

Koltay (2017a) coloca que o termo Competência Informacional (CoInfo) por si só não compreende a variedade conceitual atribuída às mudanças nos mais variados ambientes e contextos de informação, principalmente no que se refere aos dados científicos, tornando, então, a competência em dados uma dimensão ou extensão da CoInfo.

A competência em dados é uma das tentativas de trazer uma nova característica ao mundo da competência em informação no que diz respeito à contribuição para uma educação continuada que torne os indivíduos mais preparados para lidar com este tipo de informação, declara Santos (2021).

Ressalta-se ainda que a semelhança da sobrecarga de informação, trazida anteriormente pela literatura, assemelha-se à sobrecarga de dados vivenciada hoje nas ciências, e da mesma forma que foi e, ainda é, necessária a competência em

informação para compreendê-la e assimilá-la, a competência em dados torna-se importante para conscientizar os indivíduos e pesquisadores a serem autônomos nos processos e etapas que envolvem o trabalho com dados de pesquisa durante todo seu ciclo de vida (Santos, 2021; Koltay, 2017b)

Os dados de pesquisa geralmente são integrados às infraestruturas de dados e tornam-se componentes do ambiente de informação. Para o uso adequado de infraestruturas de dados e informações em disciplinas de pesquisa, é necessário o aprimoramento de conhecimento contínuo de como utilizá-los. Nesse contexto, pode-se determinar a competência em dados de pesquisa como o conhecimento de uma série de atividades de processamento de dados, desde a identificação até o planejamento, armazenamento, análises, avaliação, gerenciamento, provisão e proteção (Steinerová; Ondrisová, 2019).

Os tipos de dados de pesquisa podem incluir *Big Data*, dados abertos, dados qualitativos e quantitativos, dados experimentais, observacionais, sensoriais, simulações, referenciais, entre outros, dos mais variados tipos e formatos de acordo com as especificidades das áreas do conhecimento.

A competência em dados ajuda na resistência contra os efeitos nocivos de más práticas intencionais e não intencionais. Definir indesejabilidade é uma questão cultural ou política; contudo, reconhecer a urgência de desenvolver competência em dados e encontrar consenso sobre questões críticas sobre dados é crucial para navegar em um período de transformação (Nguyen, 2021).

As práticas em dados de pesquisa são atividades de manipulação de diferentes disciplinas. Este termo segue a tendência de ver a competência em dados como uma prática sociocultural com o uso de tecnologias digitais no ambiente da informação. No contexto deste estudo, o termo “competência” compreende, então, o sentido de “saber fazer” que requer um conjunto de “habilidades” e implica um posicionamento diante daquilo que se apresenta como necessário (Vitorino, 2008). Isto é, ser competente em dados tem a mesma relevância que ser competente em informação, pois é necessário às pessoas de todas as áreas, de todos os níveis educacionais “[...] já que tem por objetivo propiciar a capacitação contínua dos indivíduos, que, ao buscarem os conteúdos informacionais [...]” tornam-se mais autônomos (Mata, 2014, p. 16).

Dessa forma, a competência em dados faz parte do aprendizado para conduzir de forma eficiente o processo de gestão de dados de pesquisa, pois os bibliotecários e pesquisadores precisam acompanhar os estudos e efeitos dos dados na comunidade científica, levando-se em consideração a formação de novas habilidades como, por exemplo, métodos de condensar dados, combiná-los com outras fontes de informação para gerar novos conhecimento, determinar quando e por quanto tempo os dados são necessários, como manipular os dados, contextualizá-los, dentre outros processos (Costa; Sales; Zattar, 2020).

Em outros termos, tornar-se competente em dados de pesquisa é necessário para que os indivíduos que atuam diretamente com pesquisas ultrapassem os desafios de identificar os métodos e técnicas fundamentais para o gerenciamento de grandes volumes de dados. E para aqueles que utilizam dados, sejam pesquisadores ou a sociedade em geral, é essencial para identificar os bons dados, a veracidade das informações e também, saber utilizar esses recursos de forma responsável e ética.

Com essa perspectiva, surge a necessidade do profissional da informação para auxiliar no processo de desenvolvimento de competências e habilidades para que os pesquisadores adquiram maior controle sobre os dados produzidos e saibam, entre outras coisas, documentar todos os procedimentos envolvidos no ciclo de vida dos dados, possibilitando trabalhar com os dados de forma consciente em qualquer ambiente organizacional. Esse profissional é o bibliotecário de dados.

Bibliotecário de dados é um termo novo, porém, sua atuação profissional encontra-se alicerçada nos princípios das práticas tradicionais do bibliotecário aplicadas a os atuais recursos de dados. O termo emergiu conforme demanda das tecnologias digitais e a internet na década de 1990 (Reis, 2019).

No cenário internacional os bibliotecários estão cada vez mais envolvidos com uma ampla gama de iniciativas relacionadas a dados, incluindo avaliações para determinar as necessidades de instrução relacionadas a dados. Conforme Deja *et al.* (2021) eles podem usar habilidades e competências em dados para construir programas de aprendizagem para interagir com o corpo docente, alunos e pesquisadores.

Em pesquisa de Burness, Mann e Neville (2020) sobre a competência em dados por meio de uma comunidade de aprendizado de bibliotecários, os autores

destacam que o envolvimento do bibliotecário de dados com as comunidades de aprendizagem oferece oportunidades para:

- a) aumentar a compreensão das perspectivas pedagógicas do corpo docente;
- b) desenvolver relacionamentos com professores e funcionários com reuniões regulares na biblioteca e outros espaços do campus que possam levar a uma variedade de oportunidades colaborativas de ensino e aprendizagem;
- c) integrar bibliotecários como parceiros do corpo docente como desenvolvedores de currículo em áreas centrais, como alfabetização informacional e de dados; e, finalmente,
- d) influenciar o ensino e os resultados dos alunos no nível institucional e profissional.

Percebe-se que um papel importante é desempenhado pelos bibliotecários de dados que implementam princípios de competência informacional para gestão de dados no contexto acadêmico e científico. A competência em dados de pesquisa visa desenvolver e incentivar o uso de melhores práticas no que se refere a gestão dos dados de pesquisa para benefício do próprio pesquisador, do grupo de pesquisa e da instituição à qual pertence, e da comunidade científica como um todo, uma vez que com a variedade de dados e objetos de informação, é fundamental o indivíduo ter independência, saber ler, entender e interpretar os dados para reuso ou reprodutibilidade das pesquisas.

Além dos fatores já citados, a competência em dados de pesquisa visa também capacitar os profissionais de diversas áreas, e principalmente pesquisadores, a gerenciar os dados de suas pesquisas para compartilhá-los e proporcionar maior benefício para o avanço científico e tecnológico.

O intenso crescimento da quantidade de dados de pesquisa fez com que as instituições, principalmente as universidades, começassem a prestar atenção nos problemas de gerenciamento e preservação a longo prazo dos dados de pesquisa, uma vez que agora as demandas de infraestrutura para garantir acesso seguro e confiável aos dados, seu armazenamento e arquivamento requer dos pesquisadores habilidades para tratar seus próprios dados. No entanto, muitas vezes os serviços de dados das instituições são inadequados, de baixa qualidade ou simplesmente não

atendem às demandas e expectativas dos pesquisadores para organização, armazenamento e descrição de dados. Frequentemente, não recebem nenhum treinamento, o que leva à falta de confiabilidade, inacessibilidade e diminuição da vida útil de seus dados (Vilar; Zabukovec, 2019).

Diante disso, Koltay (2016) afirma que a competência em dados está intimamente relacionada aos serviços de dados de pesquisa que incluem o gerenciamento dos dados de pesquisa, ser competente em dados começa a ser amplamente aceito e esperado como uma habilidade crucial para profissionais da informação envolvidos no suporte a pesquisas intensivas em dados.

### 3 PROCEDIMENTOS METODOLÓGICOS

A metodologia, conforme Barros e Lehfeld (2007, p. 2) é “[...] o estudo da melhor maneira de abordar determinados problemas no estado atual de nossos conhecimentos [...]”, sendo aplicados processos e técnicas. Dessa forma, serão apresentados os procedimentos e métodos utilizados para a construção deste trabalho durante o desenvolvimento da pesquisa.

#### 3.1 DELINEAMENTO DA PESQUISA

A presente pesquisa trata-se de um estudo de natureza básica, com abordagem qualitativa, de caráter descritivo e exploratório, utilizando-se, principalmente, da pesquisa documental como procedimento metodológico.

O estudo documental baseia-se em materiais que ainda não receberam um tratamento analítico, quando a fonte de coleta de dados está restrita a documentos, escritos ou não, constituindo o que se denomina de fontes primárias. A utilização desse procedimento metodológico é destacada no momento em que se pode organizar informações que se encontram dispersas, conferindo-lhe uma nova importância como fonte de consulta (Prodanov; Freitas, 2013; Gil, 2008). No que se refere a esta pesquisa, os documentos analisados foram extraídos dos *websites* de instituições internacionais, onde os dados, até então, não foram tratados e consolidados e as informações encontram-se dispersas.

As etapas realizadas durante a análise documental foram: pré-análise (baseada na definição dos objetivos da pesquisa - pré-estabelecidos); organização (seleção e classificação dos documentos conforme os códigos definidos *a priori* e *a posteriori*), e por fim, a análise e interpretação dos dados.

No quadro 1 estão representados os objetivos e como buscou-se atingi-los durante o processo desta pesquisa.

**Quadro 1 - objetivos específicos e métodos de pesquisa**

<b>Objetivo geral</b>		
Analisar as práticas adotadas para desenvolver competências em dados de pesquisa em universidades internacionais		
<b>Objetivos específicos</b>	<b>Etapas do processo de pesquisa</b>	<b>Resultados esperados e alcançados</b>
a) identificar as universidades internacionais que adotam práticas para desenvolver competências em dados de pesquisa;	Identificar as universidades classificadas no <i>Ranking Webometrics</i> 2023.	Universidades que realizam atividades para desenvolver competências em dados de pesquisa.
b) identificar as práticas adotadas para desenvolver competências em dados de pesquisa nas universidades internacionais selecionadas;	Explorar os <i>sites</i> das universidades buscando informações sobre atividades de competência em dados de pesquisa. Codificar, no NVivo, as informações encontradas.	Dados codificados que possibilitem, com o auxílio do NVivo, análise das informações encontradas nos sites das universidades.
c) verificar, nas universidades selecionadas, quais são as estruturas de apoio às atividades de desenvolvimento de competências em dados de pesquisa e	Verificar presença de setores e/ou serviços voltados ao atendimento de demandas relacionadas a dados de pesquisa por meio de exploração dos sites das universidades e bibliotecas; Descrever as atividades desempenhadas por esses setores e/ou serviços de dados.	Quem são os responsáveis pelos serviços voltados à competência em dados nas universidades estudadas, como se organizam, quais as práticas adotadas e para quem são direcionadas.
d) propor estratégias para desenvolver competências em dados de pesquisa em âmbito nacional	Sumarizar os achados de forma a relacionar aspectos que podem ser usados para criar serviços de dados de pesquisa.	Estratégias para que instituições criem setores ou serviços de dados para desenvolver competência em dados de pesquisa e atender às demandas de sua comunidade acadêmica.

**Fonte:** elaborada pela autora

Para cumprir com os objetivos deste estudo, utilizou-se o *Webometrics.info* 2023 para selecionar o *corpus* de análise por tratar-se de um *ranking* acadêmico de Instituições de Ensino Superior que oferece semestralmente um exercício científico

independente, objetivo, gratuito e aberto para fornecer informações confiáveis, multidimensionais, atualizadas e úteis sobre o desempenho de universidades de todo o mundo (*Webometrics*, 2023).

Esta investigação teve por embasamento metodológico o estudo realizado por Silveira (2021), que buscava investigar as práticas de acesso aberto a dados de pesquisa, instituídas nas políticas de depósito, acesso e uso de dados de pesquisa em repositórios de dados de pesquisa de universidades internacionais, com ênfase nas suas permissões, restrições e licenças. Na metodologia usada, a autora selecionou, por meio do *Ranking Webometrics.info* de 2020, as 18 primeiras universidades dentre as 100 melhores universidades apresentadas no *Ranking*.

Após atualizar a lista das universidades do estudo de Silveira (2021), a partir dos dados do *Webometrics Ranking of World Universities*, edição de janeiro de 2023<sup>6</sup> foi realizada a exploração dos *sites* das universidades para selecionar aquelas que possuíam as informações pertinentes ao estudo. Dessa maneira foi possível responder à questão de pesquisa: *quais as práticas adotadas para desenvolver competências em dados de pesquisa em universidades internacionais?* Assim como, atingir os objetivos propostos.

O déficit de literatura e estudos científicos brasileiros sobre competência em dados, as poucas universidades no país trabalhando de forma mais ampla no contexto dos dados de pesquisa e, conseqüentemente, com práticas contínuas para desenvolver competência em dados e verificando-se, por meio de pesquisas anteriores, a necessidade de capacitar pesquisadores para que estejam preparados para lidar com esse tipo de objeto informacional, corroborou para a decisão de seguir a metodologia adotada por Silveira (2021), no que se refere à identificação das universidades internacionais que já possuíam políticas de depósito, acesso e uso de dados de pesquisa para o presente estudo.

Inicialmente pretendia-se analisar as primeiras 20 universidades do *Ranking*, no entanto, após um estudo piloto em quatro universidades, constatou-se que as informações encontradas podem tratar de aspectos bastante semelhantes no que se refere ao tipo de estratégias adotadas para desenvolver competências em dados, tornando a apresentação dos resultados longa e repetitiva. Sendo assim, decidiu-se realizar a análise até a saturação dos dados. Considera-se saturada a coleta de

---

<sup>6</sup> <https://www.webometrics.info/en/transparent>

dados quando nenhum elemento novo é encontrado, e o acréscimo de novas informações deixa de ser necessário e relevante, uma vez que não altera a compreensão do evento estudado. Trata-se de um critério que permite estabelecer a validade de um conjunto de dados (Thirycherques, 2009).

Para atingir a saturação de dados, utilizou-se como método a amostragem teórica, que é explicada por Flick (2008) como o princípio básico de selecionar casos ou grupo de casos de acordo com critérios que digam respeito ao seu conteúdo, em vez de aplicar critérios metodológicos abstratos. Esse princípio é característico de estratégias relacionadas para a coleta de dados na pesquisa qualitativa. A estrutura e tamanho da amostra não é definida antes da coleta e análise de dados. Ela é desenvolvida, gradativamente, durante a coleta e análise de dados, sendo então complementada por novas dimensões ou limitada a dimensões e a campos determinados.

Conforme explicado por Flick (2008), a amostra (universidades internacionais) não foi pré-selecionada, mas durante a análise dos dados foi diminuída gradativamente até que se percebeu que os resultados não alteraram a compreensão do fenômeno e novas informações eram desnecessárias. Diante disso, para o presente estudo, entendeu-se ser suficiente analisar as informações encontradas nos *sites* das 10 primeiras universidades elencadas no *Webometrics.info* 2023, também considerando que elas são consideradas, pelo *Ranking*, as melhores universidades nos principais perfis do Google Scholar (Quadro 2).

**Quadro 2 - Universidades selecionadas para a pesquisa**

Nº	País	Universidades	Link
1	EUA	Universidade de Harvard	<a href="https://www.harvard.edu/">https://www.harvard.edu/</a>
2	EUA	Universidade de Stanford	<a href="https://www.stanford.edu/">https://www.stanford.edu/</a>
3	EUA	Instituto de Tecnologia de Massachusetts	<a href="https://web.mit.edu/">https://web.mit.edu/</a>
4	UK	Universidade de Oxford	<a href="https://www.ox.ac.uk/">https://www.ox.ac.uk/</a>
5	EUA	Universidade de Michigan	<a href="https://umich.edu/">https://umich.edu/</a>
6	EUA	Universidade da Califórnia - Berkeley	<a href="https://www.berkeley.edu/">https://www.berkeley.edu/</a>
7	EUA	Universidade Cornell	<a href="https://www.cornell.edu/">https://www.cornell.edu/</a>
8	EUA	Universidade de Washington	<a href="https://www.washington.edu/">https://www.washington.edu/</a>
9	EUA	Universidade Columbia - Nova York	<a href="https://www.columbia.edu/">https://www.columbia.edu/</a>
10	EUA	Universidade da Pensilvânia	<a href="https://www.upenn.edu/">https://www.upenn.edu/</a>

### 3.2 DESCRIÇÃO DA ETAPA DE COLETA

A coleta dos dados, considerando como *corpus* de análise as primeiras 10 universidades elencadas no *Ranking Webometrics 2023*, aconteceu no período de 14 de agosto a 17 de outubro de 2023. Na primeira etapa, buscou-se localizar documentos e todas as informações relacionadas com o desenvolvimento de competências em dados de pesquisa, isto é, cursos, palestras, oficinas, serviços ou centros de pesquisa oferecidos pela instituição. Todo material localizado foi salvo no *Google Drive* em pastas com o nome de cada universidade. Sempre que possível os documentos encontrados foram convertidos para o formato pdf para facilitar sua indexação e análise pelo NVivo. Cabe ressaltar que as análises ficaram restritas a documentos textuais, não foram analisados, por exemplo, materiais audiovisuais.

As informações recuperadas foram em inglês e português. No total foram recuperados 262 documentos das 10 universidades, no entanto, percebeu-se que havia documentos duplicados ou que o assunto não correspondia ao tema proposto da pesquisa - principalmente materiais utilizados em oficinas. Portanto, realizou-se uma seleção desses documentos descartando os duplicados e aqueles que não se aplicavam ao tema do estudo proposto. O total de arquivos selecionados para serem analisados foi de 146, sendo 116 descartados.

**Tabela 1** - Universidades e quantidade de documentos recuperados

Universidades	Quantidade de documentos
Universidade de Harvard	49
Instituto de Tecnologia de Massachusetts	24
Universidade de Michigan	16
Universidade Cornell	12
Universidade de Oxford	11
Universidade de Stanford	8
Universidade Columbia - Nova York	8
Universidade da Pensilvânia	8
Universidade da Califórnia - Berkeley	6
Universidade de Washington	4
<b>Total</b>	<b>146</b>

**Fonte:** dados da pesquisa

Dispondo dos documentos para iniciar a análise, percebeu-se também que para agilizar tal processo, o ideal seria traduzir os documentos que estavam em inglês para melhor compreensão e codificação. Para tal, foi utilizada a ferramenta *Online Doc Translator*<sup>7</sup>, foram traduzidos 146 documentos. Após isso, iniciou-se o processo de análise desses dados.

### 3.3 DESCRIÇÃO DA ETAPA DE ANÁLISE DOS DADOS

Se tratando desse estudo a organização dos documentos para posterior análise aconteceu em três etapas:

- a) pré-análise: que se refere a separação a partir da leitura flutuante das informações encontradas nos sites das universidades e da escolha do *corpus* documental. O universo documental foi determinado à *priori* e compõe-se de documentos sobre atividade de competência em dados promovidos pelas instituições selecionadas. A seleção do material foi realizada a partir do conteúdo que abordasse os assuntos relativo a esta pesquisa, dados de pesquisa e todo seu entorno;
- b) exploração do material: selecionou-se nos documentos temáticas pertinentes à pesquisa a fim de gerar a codificação para a análise dos dados;
- c) tratamento das informações: importação dos dados para o *software* Nvivo, para elaboração dos gráficos de análise.

Para a realização da análise, e por tratar-se de uma pesquisa qualitativa, optou-se por utilizar o *software* NVivo uma vez que ele é um dos principais *softwares* para análise de dados qualitativos, cuja finalidade é buscar, categorizar, organizar e registrar interpretações

O NVivo consiste numa ferramenta que auxilia o pesquisador nas várias etapas da pesquisa qualitativa, dentre elas a definição e organização das categorias e subcategorias analíticas; organização dos dados coletados em múltiplas fontes; análise dos dados; apresentação dos resultados; potencialização e intensificação do alcance e profundidade das pesquisas, e permissão de maior exploração da complexidade dos dados (Mozzato; Grzybovski; Teixeira, 2016).

---

<sup>7</sup> <https://www.onlinedoctranslator.com/en/>

Desenvolvido pela QSR *International*, e lançado em 2001, esse *software* gerencia quaisquer dados virtualmente, incluindo documentos do *Word*, PDFs, arquivos de áudio, tabelas de banco de dados, planilhas, vídeos, imagens e dados da Web. As três principais instâncias de gerenciamento das informações no NVivo são os arquivos, códigos e classificações.

Os arquivos<sup>8</sup> são os materiais empíricos da pesquisa, podendo ser importados arquivos de texto, imagens, sons, vídeos e conjuntos de dados, em diferentes formatos. Os códigos<sup>9</sup> são uma coleção de referências dos arquivos sobre um tema, tópico, conceito, ideia ou experiência específica. E por fim, as classificações<sup>10</sup> fornecem uma forma de registrar informações descritivas sobre os arquivos e casos do projeto. É possível usar essas informações para fazer comparações ou coletar material por atributo.

Após selecionar todos os documentos que fariam parte da pesquisa, coletados a partir da exploração dos sites das universidades, foi realizada a importação dos documentos no NVivo, criando-se pastas com o nome de cada universidade, e ali, importados todos os documentos das respectivas universidades.

Na etapa seguinte foi realizada uma leitura flutuante de cada documento e com base nos assuntos tratados elaborou-se a lista dos códigos e subcódigos que seriam usados para classificar os documentos. Os termos escolhidos foram extraídos do RDPdata: Tesouro de Dados de Pesquisa<sup>11</sup> da Universidade Federal do Rio Grande do Sul - UFRGS. Optamos por utilizar um tesouro pois este tem como função principal o controle terminológico do vocabulário utilizado em uma área específica do conhecimento (Van der Laan; Ferreira, 2000), e se tratando da abordagem do presente estudo, o Tesouro RDPdata contribuiu para a finalidade da pesquisa, uma vez que trata de termos da área de dados de pesquisa.

O Quadro 3 apresenta a codificação criada para os temas abordados nos documentos selecionados para esta pesquisa.

---

<sup>8</sup> <https://help-nv.qsrinternational.com/20/win/Content/files/files.htm>

<sup>9</sup> <https://help-nv.qsrinternational.com/20/win/Content/nodes/codes.htm>

<sup>10</sup> <https://help-nv.qsrinternational.com/20/win/Content/classifications/about-classifications.htm>

<sup>11</sup> <https://www.ufrgs.br/tesauros/index.php/thesa/terms/374>

**Quadro 3 - Códigos e subcódigos criados no Nvivo**

<b>Códigos e subcódigos</b>
• Boas práticas de pesquisa científica
o Preservação de dados
o Documentação sobre os dados
§ Armazenamento e backup
• Ciclo de vida dos dados
• Citação de dados
• Compartilhamento de dados
• Curadoria de dados
• Dados pessoais
o Dados sensíveis
o Anonimização
• Ética
• Gestão de Dados de Pesquisa
• Metadados
• Plano de Gestão de Dados
• Princípios FAIR
• Repositório de dados
o Política de dados
o Termo de uso de dados

**Fonte:** elaborado pela autora

Após concluir a etapa de criação dos códigos e subcódigos, iniciou-se o processo de codificação dos documentos, que consiste em selecionar trechos do texto e colocá-lo sob um tema (código ou subcódigo) ao qual se refere àquele fragmento textual. Todos os 146 documentos passaram por essa etapa de codificação. Salienta-se que, por tratar-se de documentos nos quais os assuntos eram sobre dados de pesquisa, a codificação dos fragmentos textuais aconteceu com base no conteúdo que tratava tanto aspectos gerais como específicos sobre o tema dados de pesquisa. Isto é, se o tema abordado no documento tratava apenas de Citação de dados e todos os trechos analisados estavam relacionados a esse aspecto, o documento inteiro foi codificado sob o código “citação de dados”. O

mesmo aconteceu com os códigos “Gestão de dados de pesquisa e “Curadoria de dados” nos quais documentos completos foram codificados por tratarem de aspectos gerais sobre as duas temáticas sem maior aprofundamento sobre as etapas, conceitos e boas práticas, por exemplo.

Apesar dos documentos selecionados tratarem de atividades desenvolvidas para ampliar a competência em dados de pesquisa, portanto, a codificação geral estar vinculada às práticas de gestão de dados de pesquisa, o intuito de utilizar códigos mais específicos e subcódigos foi pontuar os assuntos particularmente, uma vez que as universidades trabalham cada tema vinculado à gestão de dados de pesquisa de forma mais aprofundada. Esse processo de codificação permitiu agrupar os documentos de forma coerente com os temas abordados, facilitando a análise.

À medida que o conteúdo dos documentos foi sendo codificado, instantaneamente geraram-se as “Referências”, que são os trechos dos documentos marcados de acordo com um código que lhe foi atribuído durante a leitura. Quando um trecho de texto é codificado, esse conteúdo é chamado de “Referência” e está sempre vinculada a um determinado “Código”.

Esta função fornecida no NVivo permite o fácil acesso aos trechos referentes a cada um dos códigos e, a partir da leitura das referências codificadas é possível identificar os assuntos que cada universidade desenvolveu em suas atividades de competência em dados.

Na Tabela 2 estão representados, numericamente, os códigos e subcódigos, atribuídos para os documentos que passaram pelo processo de codificação e o total de referência geradas em cada um deles.

**Tabela 2** - Códigos e subcódigos, total de documentos codificados, e total de referências geradas

Códigos e subcódigos	Documentos	Referências
● Boas práticas de pesquisa científica	16	19
o Preservação de dados	13	14
o Documentação sobre os dados	44	62
§ Armazenamento e backup	32	40
● Ciclo de vida dos dados	10	12
● Citação de dados	12	14
● Compartilhamento de dados	43	63
● Curadoria de dados	8	9
● Dados pessoais	6	6
o Dados sensíveis	11	13
o Anonimização	4	4
● Ética	5	8
● Gestão de Dados de Pesquisa	31	74
● Metadados	23	29
● Plano de Gestão de Dados	37	60
● Princípios FAIR	14	14
● Repositório de dados	31	39
o Política de dados	4	7
o Termo de uso de dados	12	14

**Fonte:** dados da pesquisa

Em seguida da codificação dos materiais, foi possível, clicando sobre cada código, visualizar todas as Referências ou trechos de cada documento ao qual foi atribuído determinado Código. Ao agrupar diversos trechos de diferentes fontes que tratam de um mesmo assunto, ou seja, ao qual foi atribuído um determinado código, o *software* permite visualizar rapidamente todos esses trechos e realizar análises sobre o conteúdo.

Concluída a etapa de codificação dos documentos, foi necessário criar os “Casos”<sup>12</sup> e classificar os casos<sup>13</sup> para registrar informações descritivas sobre as entidades que seriam estudadas. Para cada classificação foi definido o valor de um “Atributo” para complementar o processo de análise dos dados.

No NVivo, os “Casos” representam as “Unidades de análise” do estudo. São pessoas, lugares, organizações, eventos ou outras entidades que representam o foco da pesquisa, são como “recipientes” que contêm todas as informações relacionadas ao estudo. Se tratando dessa investigação, os casos que foram criados referem-se a cada universidade selecionada no *Ranking webometrics.2023*, como mostra a Tabela 3.

**Tabela 3** - Casos criados para a pesquisa, quantidade de documentos codificados e suas referências

Casos	Documentos	Referências
 Universidade de Harvard	49	1.674
 Instituto de Tecnologia de Massachusetts	24	581
 Universidade de Michigan	16	127
 Universidade Cornell	12	73
 Universidade de Oxford	11	105
 Universidade de Stanford	8	35
 Universidade Columbia - Nova York	8	48
 Universidade da Pensilvânia	8	43
 Universidade da Califórnia - Berkeley	6	223
 Universidade de Washington	4	12

**Fonte:** dados da pesquisa

Os Casos correspondem às 10 universidades que compõem o *corpus* da pesquisa. Com os recursos do *software* é possível identificar em cada “Caso” as informações ali contidas por meio da codificação dos documentos e suas “Referências”. Por exemplo, acessando o Caso “Universidade de Harvard” têm-se

<sup>12</sup> <https://help-nv.qsrinternational.com/20/win/Content/cases/cases.htm>

<sup>13</sup> <https://help-nv.qsrinternational.com/20/win/Content/cases/classify-cases.htm>

acesso a 49 documentos e a 1.674 trechos codificados, as “Referências”. Isso tornou possível, em virtude da codificação dos documentos importados para o respectivo “Caso”, “Universidade de Harvard”, seguido da classificação de “Caso”, “Universidade de Harvard”.

As classificações de Casos são pertinentes para registrar informações sobre pessoas, instituições, lugares ou qualquer outro tipo de “Caso” que seja relevante para a pesquisa. Classificar um Caso, nesta investigação, envolve a descrição do tipo de instituições que desenvolvem atividades de competência em dados, e definição dos valores dos atributos, criados para qualificar os Casos.

É possível criar mais de uma classificação de Caso dependendo do tipo de pesquisa. Neste estudo criou-se apenas uma classificação de Caso, dado que se trata de um estudo descritivo e exploratório, utilizando da pesquisa documental e o *corpus* composto por 10 universidades, não envolvendo entrevistas, por exemplo, situação que demandaria a criação de mais de um caso para melhor percepção do estudo. Diante disso, formou-se a classificação de Caso “Instituições”, seguido da criação dos atributos, como está apresentado no Quadro 4.

O procedimento de criar Atributos consistiu em identificar e atribuir aspectos peculiares referente ao trabalho desenvolvido em cada instituição para tornar pesquisadores competentes em dados, indo ao encontro do objetivo b), *identificar as práticas adotadas para desenvolver competências em dados de pesquisa nas instituições internacionais selecionadas* e parte do objetivo c, que buscou *verificar, nas instituições selecionadas, estruturas de apoio às atividades de desenvolvimento de competências em dados de pesquisa*. O Quadro 4 apresenta os Atributos criados na classificação de Caso “Instituições” a partir da leitura dos 146 arquivos.

**Quadro 4** - Lista de atributos associado à classificação de Caso “Instituição”

Instituições	
	Atributos
■ ■	Oficinas
■ ■	Cursos
■ ■	Atendimento Individual
■ ■	Serviços de dados
■ ■	Biblioteca
■ ■ ■	Pesquisadores
■ ■	Alunos de Graduação
■ ■	Alunos de Pós-Graduação
■ ■	Funcionários
■ ■	Público geral

**Fonte:** dados da pesquisa

A partir desses Atributos apresentados no Quadro 4 e associando-os com os Casos, conforme a Tabela 3, foi possível responder a perguntas como: **quem** desenvolve as atividades de competência em dados? **para quem** as atividades estão disponíveis? e, **como** as universidades desenvolvem as atividades de competência em dados de pesquisa? Com base nessas indagações foram criados “Valores” para os 13 Atributos, apresentados no Quadro 4.

Quando cria-se um novo Atributo, a menos que seja criado outro valor como padrão, ele possui dois valores pré-definidos pelo sistema:

- a) Não atribuído - utilizado para indicar que o valor deste atributo ainda não foi definido.
- b) Não aplicável - utilizado para indicar que o valor não é aplicável à fonte ou Caso que se está classificando.

Com base nisso, e levando em consideração as indagações que surgiram ao associar os Atributos com os Casos, criaram-se novos Valores de Atributos pertinentes ao estudo, sendo eles: QUEM, PARA QUEM e COMO.

Quaisquer Valores de Atributo adicionados devem ser compatíveis com o tipo de dados do atributo. Se o tipo de dados for textual, como no caso desta pesquisa, é possível inserir caracteres alfanuméricos. O Quadro 5 apresenta cada Atributo associado ao seu valor.

**Quadro 5 - Atributos e Valores de Atributos**

--	Oficinas	}	Como
--	Cursos		
--	Atendimento Individual		
--	Serviços de dados	}	Quem
--	Biblioteca		
--	Pesquisadores	}	Para quem
--	Alunos de Graduação		
--	Alunos de Pós-Graduação		
--	Funcionários		
--	Público geral		

**Fonte:** dados da pesquisa

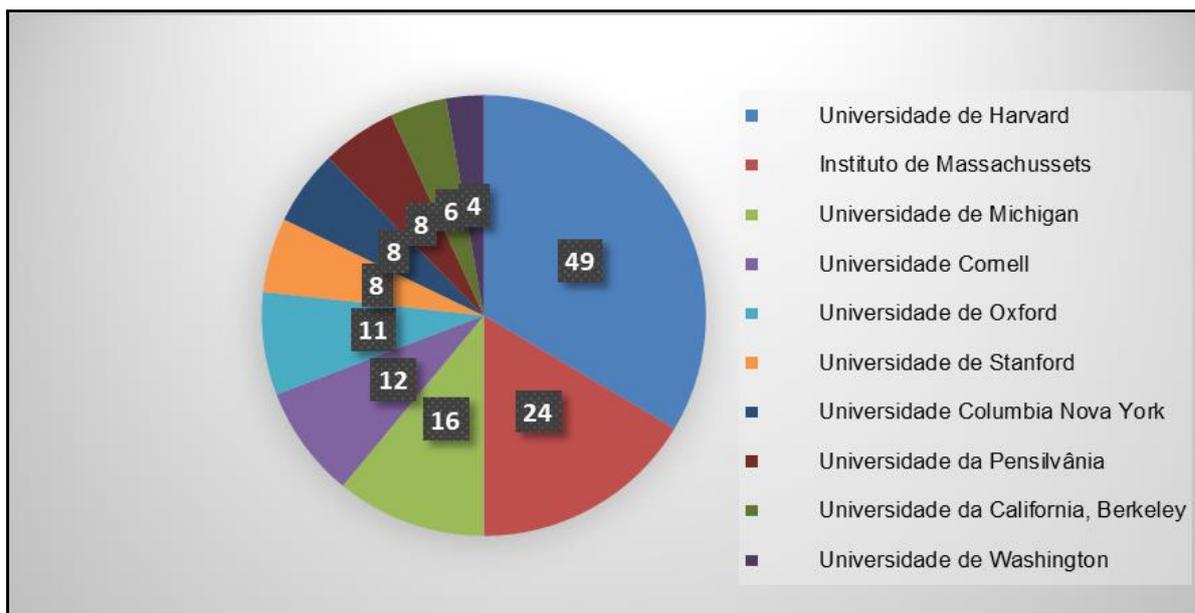
Após realizar todos os processos de codificação dos documentos recuperados, criação de Casos, classificar os Casos, identificar Atributos e Valores, foi possível visualizar por meio de gráficos as informações de forma sistematizadas e realizar as análises necessárias para atingir os objetivos a, b e c, apresentados na seção a seguir. Para, finalmente, atender ao objetivo d), propor estratégias para desenvolver competências em dados de pesquisa em âmbito nacional.

## 4 ANÁLISES DOS RESULTADOS

Nesta seção serão apresentados os resultados conforme os objetivos específicos do estudo. Os resultados das análises obtidas estão em consonância com os objetivos específicos da pesquisa: a) identificar as instituições que adotam práticas para desenvolver competências em dados de pesquisa, b) identificar as práticas adotadas para desenvolver competências em dados de pesquisa nas instituições internacionais selecionadas, c) verificar, nas instituições selecionadas, estruturas de apoio às atividades de desenvolvimento de competências em dados de pesquisa, e finalmente no objetivo específico d) serão propostas algumas estratégias para desenvolver competências em dados de pesquisa em âmbito nacional. Lembrando que o objetivo geral desta pesquisa foi analisar as práticas adotadas para desenvolver competências em dados de pesquisa em instituições internacionais.

O primeiro objetivo específico do estudo buscou *identificar as instituições que adotam práticas para desenvolver competências em dados de pesquisa*. Conforme apresentado na Tabela 1, na seção anterior, as universidades selecionadas para o estudo foram: Universidade de Harvard, Universidade de Stanford, Instituto de Tecnologia de Massachusetts, Universidade de Oxford, Universidade de Michigan, Universidade da Califórnia - Berkeley, Universidade Cornell, Universidade de Washington, Universidade Columbia - Nova York, e a Universidade da Pensilvânia.

Após acessar os *sites* de cada universidade, procurou-se localizar todas as informações relacionadas com o desenvolvimento de competências em dados de pesquisa, ou seja, cursos, palestras, oficinas, serviços oferecidos pela instituição e também, quem era o responsável por desenvolver essas atividades e serviços, como por exemplo, a biblioteca, um centro de pesquisa ou algum outro órgão ou setor da universidade. Os materiais coletados de cada instituição foram analisados e codificados de acordo com os assuntos abordados nos documentos. O total de documentos textuais analisados foi 146. O Gráfico 1 apresenta a quantidade de documentos selecionados em cada universidade:

**Gráfico 1** - Quantidade de documentos analisados por universidade

**Fonte:** dados da pesquisa

No Gráfico 1 verifica-se que a Universidade de Harvard, o Instituto de Massachussets e a Universidade de Michigan são as três instituições nas quais recuperou-se a maior quantidade de documentos para serem analisados, representando, respectivamente, 34%, 16% e 11% do total de documentos. Isso não significa que tais instituições desenvolvem mais atividades de competência em dados do que as outras, mas sim, que se destacam na disponibilização de informações sobre o assunto em seus *sites*, indicando um comprometimento com a transparência e divulgação do conhecimento na área de dados de pesquisa, permitindo àqueles que têm dúvidas quanto a essa temática tenham acesso a essas informações oportunidade de ampliar o seu conhecimento sobre dados de pesquisa.

Cada instituição que compõe esta pesquisa gera e utiliza dados de maneiras diferentes, levando em consideração sua própria missão, valores, políticas e áreas de especialização. Seu desempenho pode variar conforme os recursos disponíveis e interesses dos alunos, pesquisadores e comunidade acadêmica.

Em relação aos documentos recuperados nas instituições, conforme as observações e resultados levantados, cada uma desenvolve atividades com temas diferentes e específicos, conforme a demanda e necessidade da comunidade científica, técnica e acadêmica.

Nas próximas seções, será exposto o que foi percebido sobre as práticas de uso de dados de pesquisa de cada universidade baseado no levantamento dos

assuntos (códigos) que cada uma desenvolve em suas atividades de competência em dados a partir dos documentos analisados.

#### 4.1 CONTEXTUALIZAÇÃO DAS UNIVERSIDADES E O USO DE DADOS

Nesta seção será apresentado um pequeno histórico das universidades selecionadas para esta pesquisa, identificada em cada uma delas a relevância dos dados para o desenvolvimento das pesquisas e a sua importância institucional. Por fim, com o auxílio do NVivo, serão representados os assuntos (Códigos e Subcódigos) identificados nos documentos analisados para conhecer quais as temáticas mais recorrentes nas atividades de competência em dados realizadas por cada universidade.

##### 4.1.1 Universidade de Harvard

A Universidade de Harvard é uma das instituições mais prestigiadas do mundo em virtude da sua história e influência. Fundada em 1636 na cidade de Cambridge, Harvard é a mais antiga e conhecida universidade dos Estados Unidos. Inicialmente chamada de *New College*, a instituição passou a ter o nome atual de Harvard três anos após sua fundação em uma homenagem a John Harvard, um de seus primeiros benfeitores. Harvard conta com três *campi* principais em Cambridge, Allston e Boston, 12 unidades acadêmicas e o *Harvard Radcliffe Institute*, instituto de estudos avançados com foco em pesquisa interdisciplinar (*Harvard University*, 2024a).

Estudantes e pesquisadores de Harvard geram dados e estudos que mudam o mundo. Não há registros de quando, de fato, a instituição começou a se preocupar com o uso de dados de pesquisa, no entanto, à medida que a quantidade e a complexidade dos dados gerados pela pesquisa acadêmica aumentaram, em decorrência, principalmente, do acesso aberto, a universidade reconheceu a necessidade de desenvolver políticas, infraestrutura e recursos para garantir a integridade, acessibilidade e reprodutibilidade dos dados de pesquisa.

Muitos professores, funcionários, acadêmicos e estudantes de Harvard se envolvem em pesquisas que abrangem a coleta ou uso de dados identificáveis, confidenciais ou até privados. Com a colaboração da rede de Bibliotecas de

Harvard<sup>14</sup> a instituição também tem serviços de consultoria de dados de pesquisa, que apoiam a criação, curadoria e transformação de dados, bem como a publicação e disseminação de pesquisas. Os pesquisadores têm acesso a uma ampla gama de ofertas de serviços que auxiliam desde o planejamento de um projeto ou estudo até o compartilhamento e arquivamento de métodos ou descobertas.

Além desse setor de serviço de dados, a Universidade conta com a Política de Segurança de Dados de Pesquisa de Harvard<sup>15</sup>, que é um conjunto de diretrizes e práticas que visam proteger a integridade, confidencialidade e disponibilidade dos dados de pesquisa gerados e utilizados no âmbito da instituição. Concentram-se também no gerenciamento e administração adequados de dados de pesquisa, o que compreende pesquisas com seres humanos, dados trocados de acordo com um acordo de uso de dados e outros dados sujeitos a regulamentações estrangeiras, federais ou estaduais, requisitos de patrocinadores ou proteções à propriedade intelectual (Harvard University, 2024b).

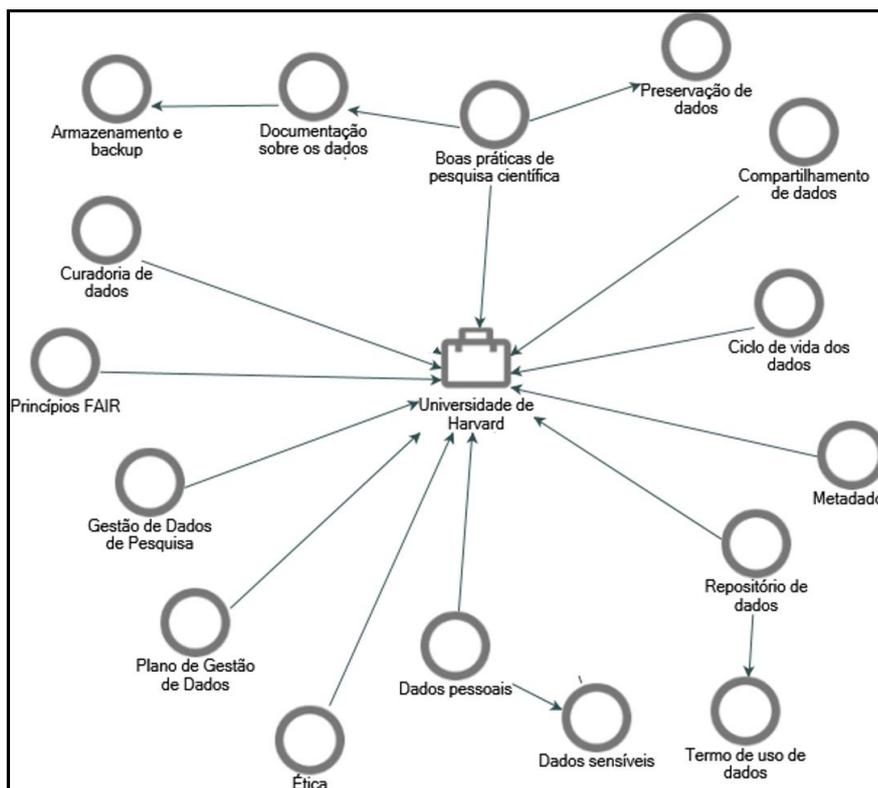
Conforme o Gráfico 1, a Universidade de Harvard teve o maior número de documentos recuperados, um total de 49. Para facilitar a análise e interpretação desses documentos, foram criados códigos e subcódigos utilizando o NVivo, como explicado na seção 3.1. Na Figura 6 apresenta-se os códigos identificados nos documentos da Universidade de Harvard, sendo eles: Boas práticas de pesquisa científica, Documentação sobre os dados, Armazenamento e *backup*, Preservação de dados, Princípios FAIR, Compartilhamento de dados, Ciclo de vida dos dados, Dados Sensíveis, Dados pessoais, Repositório de dados, Termo de uso de dados, Plano de Gestão de Dados, Gestão de Dados de Pesquisa, Curadoria de dados e Metadados.

---

<sup>14</sup> <https://www.harvard.edu/campus/libraries/>

<sup>15</sup> [https://bpb-us-e1.wpmucdn.com/websites.harvard.edu/dist/f/106/files/2023/12/HRDSP-Approved-2020.07.15\\_2023-1.pdf](https://bpb-us-e1.wpmucdn.com/websites.harvard.edu/dist/f/106/files/2023/12/HRDSP-Approved-2020.07.15_2023-1.pdf)

**Figura 6 - Universidade de Harvard e os Códigos identificados nos documentos**



**Fonte:** dados da pesquisa

Estes Códigos apresentam uma visão abrangente dos temas que são abordados nas atividades de competência em dados da comunidade de Harvard. São tratados temas que vão desde a importância de um Plano de Gestão de Dados até conteúdos que enfatizam as boas práticas da pesquisa científica que envolve, também a preservação e documentação de dados. A integração destas áreas na formação de alunos e pesquisadores é crucial para garantir que estejam bem preparados para lidar com os desafios da pesquisa moderna, especialmente em um contexto cada vez mais orientado por dados.

#### 4.1.2 Instituto de Tecnologia de Massachusetts

O Instituto de Tecnologia de Massachusetts (MIT) é uma instituição independente, mista e beneficiada por recursos privados. O Instituto admitiu seus primeiros alunos em 1865, quatro anos após a aprovação de seu estatuto de fundação. A inauguração do MIT marcou o extenso esforço de William Barton

Rogers, um ilustre cientista natural, para estabelecer um novo tipo de instituição educacional independente (MIT Facts, 2024).

O ensino e a pesquisa - com relevância para o mundo prático e para a transformação da sociedade para melhor como princípios orientadores - continuam sendo o objetivo principal do Instituto. O compromisso do MIT com o gerenciamento de dados de pesquisa é apoiado pelos serviços da MIT Library<sup>16</sup> que ajuda professores e pesquisadores a gerenciar, armazenar e compartilhar os dados que produzem. Os princípios de dados de pesquisa do MIT, aprovados em setembro de 2019, prezam pela disponibilidade, segurança e conformidade (MIT *Libraries*, 2024):

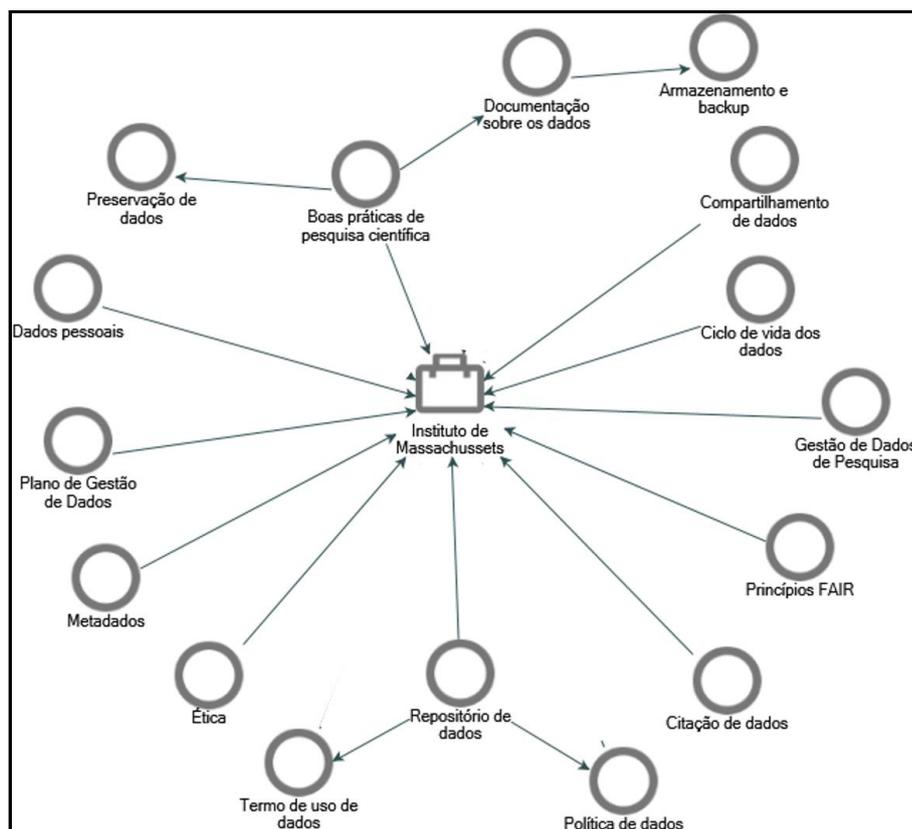
- a) disponibilidade: os pesquisadores devem disponibilizar os dados e torná-los acessíveis de forma significativa à comunidade científica e ao público na medida em que os patrocinadores, a lei e seus requisitos de pesquisa permitirem;
- b) segurança: as classes de dados que requerem atenção especial incluem dados de seres humanos, dados fornecidos por terceiros sujeitos a restrições contratuais de uso e divulgação, dados sujeitos a leis de privacidade, dados controlados por exportação, dados relativos a pesquisas não publicadas e
- c) conformidade: os pesquisadores devem gerenciar os dados de maneira consistente com as normas acadêmicas e a ética profissional, a política do MIT e as obrigações para com terceiros.

Os serviços de gerenciamento de dados estão disponíveis apenas para membros da comunidade do Instituto de Tecnologia de Massachusetts. Sendo a segunda instituição com mais documentos recuperados, 24 no total, na Figura 7 têm-se as temáticas das atividades de competência em dados encontradas nos documentos, com base nos Códigos, que foram: Boas práticas de pesquisa científica, Documentação sobre os dados, Armazenamento e backup, Preservação de dados, Ciclo de vida dos dados, Citação de dados, Compartilhamento de dados, Dados sensíveis, Ética, Gestão de Dados de Pesquisa, Metadados, Plano de Gestão de Dados, Políticas de dados, Princípios FAIR, Repositório de dados e Termo de uso de dados.

---

<sup>16</sup> <https://libraries.mit.edu/data-management/>

**Figura 7** - Instituto de Massachusetts e os códigos identificados nos documentos



**Fonte:** dados da pesquisa

Os assuntos abordados nos serviços de capacitação em dados variam de acordo com as demandas recebidas da comunidade acadêmica e até mesmo da percepção da realidade de seu público. Conforme análise nos materiais disponibilizados no *site* da biblioteca da instituição, o desenvolvimento desses assuntos apresentados na Figura 7 compreendem atividades sobre: ferramentas e técnicas que oferecem suporte ao gerenciamento prático de dados; estratégias básicas para gerenciar dados de pesquisa nos quais incluem os tópicos de práticas recomendadas para retenção e arquivamento, estruturas de diretório e convenções de nomenclatura eficazes, bons formatos de arquivo para acesso de longo prazo, opções de segurança e *backup* de dados e opções de metadados; componentes de um bom plano de gestão de dados onde apresenta-se o DMPTool, uma ferramenta *online* e personalizada pelo MIT para a elaboração de planos de gestão de dados específicos para financiadores; e o uso de metadados, nomeação de arquivo, armazenamento e compartilhamento de dados.

Alguns temas desenvolvidos nos serviços oferecidos pelo MIT evidenciam que a instituição promove reflexões e práticas no que se refere à manipulação, organização, documentação e ações construídas coletivamente que transpassam todo o ciclo de vida dos dados de pesquisa, visando capacitar sua comunidade acadêmica a usar os dados de forma correta.

### 4.1.3 Universidade de Michigan

A Universidade de Michigan (U.M) é uma instituição pública, localizada na cidade de Ann Arbor, no estado de Michigan, nos Estados Unidos. É considerada a mais antiga universidade do estado de Michigan, contando com outros dois *campi* em Flint e o Dearborn.

Como universidade pública, a U.M tem como responsabilidade expandir a acessibilidade e transparência dos seus dados de pesquisa, e o fato de agências de financiamento, editores e outros terem exigido e/ou incentivado cada vez mais a disponibilização pública de dados de pesquisa, a instituição se viu diante desafios associados a novos processos e práticas que levaram à publicação de uma série de recomendações para ajudar a comunidade da U.M a se adaptar e responder aos desafios do cenário de dados de pesquisa (*Research University Michigan*, 2024).

Como resposta a uma das demandas sobre o uso de dados de pesquisa ocorreu o lançamento de Políticas e Diretrizes que abordam aspectos do uso e gerenciamento de dados de pesquisa, dentre elas a Política de Gerenciamento de Dados de Pesquisa<sup>17</sup> e a Política de Gerenciamento de Recursos de Dados Institucionais<sup>18</sup>.

Além disso, a instituição também publicou a Iniciativa de Gestão de Dados de Pesquisa (RDSI)<sup>19</sup> na qual apoia os pesquisadores da U.M na gestão e compartilhamento dos seus dados de pesquisa para melhorar a transparência e o impacto da investigação acadêmica, definindo expectativas e orientações para a gestão de dados de pesquisa. Junto a essa iniciativa, e as demais políticas, a Biblioteca de Michigan desenvolve serviços de dados para que os pesquisadores

---

<sup>17</sup> <https://spg.umich.edu/policy/303.06>

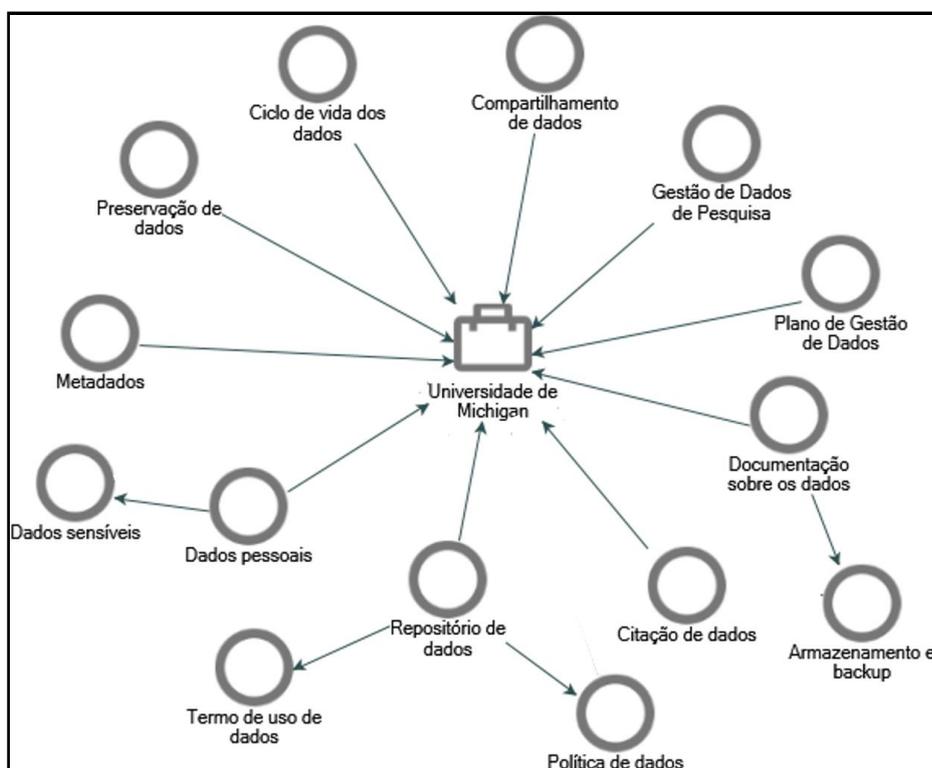
<sup>18</sup> <https://spg.umich.edu/policy/601.12>

<sup>19</sup> <https://research.umich.edu/research-data-stewardship/>

possam gerir dados de pesquisa ao longo do tempo, por meio de projetos que incluem modelos, melhores práticas e outras orientações.

Dentro desses serviços de dados que a instituição oferece, a partir dos materiais textuais recuperados, 16 no total, e dos códigos criados, os temas abordados nas atividades de competência em dados são: Documentação sobre dados, Armazenamento e Backup, Ciclo de vida dos dados, Gestão de dados de pesquisa, Plano de gestão de dados, Repositório de dados, Termo de uso de dados, Política de dados, Citação de dados, Dados sensíveis, Dados pessoais, Metadados, Compartilhamento de dados, Preservação de dados. A Figura 8 apresenta os assuntos encontrados nos documentos da Universidade de Michigan.

**Figura 8** - Universidade de Michigan e os Códigos identificados nos documentos



**Fonte:** dados da pesquisa

A iniciativa de políticas, diretrizes e serviços para apoiar a administração dos dados de pesquisa da U.M foi desenvolvida para ajudar a comunidade de pesquisa a “navegar” no cenário de dados de pesquisa em evolução. A instituição procura orientar sua comunidade a administrar de forma responsável e intencional os dados

gerados por suas pesquisas para aumentar a transparência, o rigor e o valor público da pesquisa e servir à missão mais ampla da universidade.

#### 4.1.4 Universidade de Cornell

Cornell é uma instituição privada de pesquisa e parceira da Universidade Estadual de Nova York. Desde a sua fundação em 1865 em Ithaca, NY, Cornell tem sido líder na educação fazendo contribuições em todos os campos do conhecimento. A instituição é rotineiramente classificada entre as melhores universidades do mundo (Cornell University, 2024).

A Universidade de Cornell é uma instituição com ensino voltado para alunos de graduação, pós-graduação e profissionais. As faculdades e escolas de Cornell - incluindo quatro “faculdades contratadas” (operadas pela Cornell sob contrato com o estado de Nova York) - abrangem mais de 100 campos de estudo.

Mais de 150 anos de excelência em educação, pesquisa e envolvimento público deram à Cornell uma história única, rica em inovação e tradição (Cornell University, 2024).

A universidade está comprometida com os princípios da ciência aberta e da integridade da pesquisa. As responsabilidades de Cornell com os dados de pesquisa e por seus registros científicos baseiam-se em regulamentações federais<sup>20</sup>, políticas universitárias, e outros serviços voltados ao uso e gerenciamento de dados de pesquisas.

Dentre as políticas universitária da instituição, destaca-se a Política 4.21 Retenção de Dados de Pesquisa<sup>21</sup> que define as responsabilidades compartilhadas da Universidade Cornell, e pesquisadores, na coleta, retenção, segurança, acesso, publicação e compartilhamento de dados de pesquisa. Além das políticas que deram o salto inicial na instituição a respeito do uso de dados, encontra-se também o *Cornell Data Service*<sup>22</sup>, no qual fornece serviços para todo o *campi* auxiliando na criação e implementação de gerenciamento de dados e planos de compartilhamento.

Em relação aos assuntos abordados pela instituição nas suas atividades para ampliar a competência em dados, foram identificados nos documentos os códigos:

---

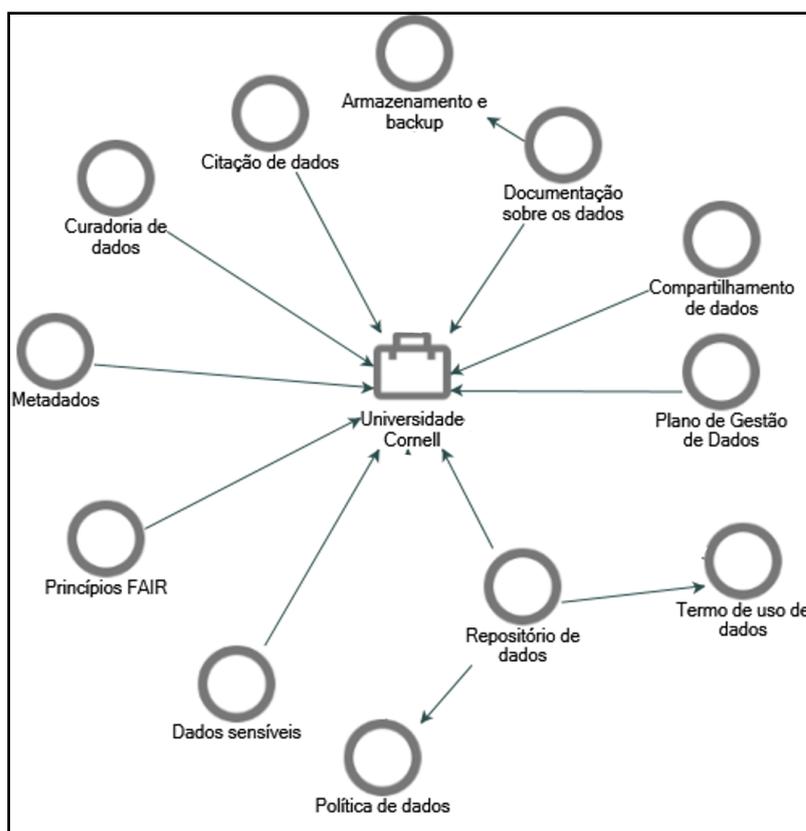
<sup>20</sup> <https://www.law.cornell.edu/cfr/text/2/200.315>

<sup>21</sup> [https://policy.cornell.edu/sites/default/files/vol4\\_21.pdf](https://policy.cornell.edu/sites/default/files/vol4_21.pdf)

<sup>22</sup> <https://data.research.cornell.edu/>

Documentação sobre dados, Armazenamento e Backup, Dados sensíveis, Compartilhamento de dados, Plano de gestão de dados, Princípios FAIR, Citação de dados, repositório de dados, Termo de uso de dados, Política de dados, Metadados, Curadoria de dados. A Figura 9 apresenta as temáticas abordadas nos documentos de Cornell.

**Figura 9** - Universidade Cornell e os Códigos identificados nos documentos



**Fonte:** dados da pesquisa

Pelas temáticas codificadas nos documentos da Universidade de Cornell, percebe-se que a instituição tem desenvolvido atividades voltadas para a conscientização e melhora da produção e organização dos dados de pesquisa, determinada a fornecer serviços e suporte à sua comunidade acadêmica desempenhando um papel fundamental no apoio às suas necessidades de dados de pesquisa.

#### 4.1.5 Universidade de Oxford

A Universidade de Oxford, a única da amostra que não está localizada nos Estados Unidos, é uma instituição de ensino superior pública situada na cidade de Oxford, Reino Unido. É a mais antiga universidade do mundo anglófono e a segunda mais antiga da Europa. Sem data certa de fundação, identifica-se que a partir de 1167 ocorreu um rápido desenvolvimento da Universidade de Oxford (*University of Oxford, 2024b*).

Oxford é uma instituição multidisciplinar de ensino e pesquisa, com uma estrutura organizacional colegiada distinta, composta por 36 faculdades que reúnem pesquisadores e alunos de diferentes disciplinas, culturas e países (*University of Oxford, 2024c*).

A instituição se destaca das demais do país quando trata-se de financiamentos e patrocínios externos de pesquisa, tendo a maior renda dentre todas as universidades do Reino Unido. Além disso, por meio da *Oxford University Innovation* (OUI), é líder no Reino Unido na geração de empresas de base tecnológica, as Spinout (*University of Oxford, 2024a*).

A Universidade de Oxford tem ciência da riqueza de dados que a sua comunidade acadêmica produz, e busca garantir o acesso mais amplo possível às suas pesquisas. Partindo disso, a instituição reconhece que as boas práticas na gestão de dados são fundamentais para um processo de investigação eficiente e eficaz, e por isso, publicou em novembro de 2023, após aprovação pelo comitê de Pesquisa e Inovação da Universidade, a Política de Gerenciamento de Dados de Pesquisa da Universidade<sup>23</sup>, substituindo as políticas de 2012 e 2018, que se aplicam a todos os membros da Universidade que estão envolvidos em pesquisas.

Além da publicação dessa política que visa promover boas práticas em relação aos dados de pesquisa, a instituição também disponibiliza, por meio dos serviços oferecidos pelo seu sistema de bibliotecas - as Bibliotecas Bodleian, apoio e orientação para quem trabalha com dados de pesquisa.

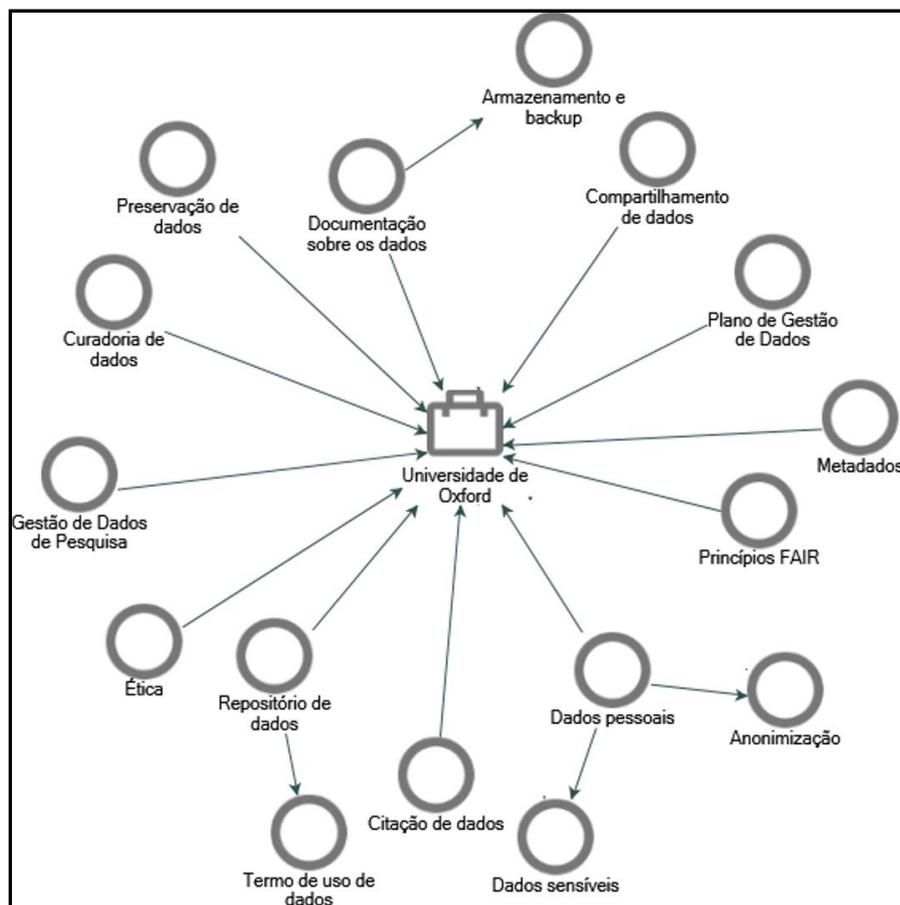
Diante da assistência em dados que Oxford fornece para sua comunidade acadêmica, nos documentos recuperados, identificou-se códigos que correspondem com as iniciativas em dados de pesquisa fornecidas pela instituição, tais como: Preservação de dados, Repositório de dados, Termo de uso de dados, Plano de

---

<sup>23</sup> <https://researchdata.ox.ac.uk/sitefiles/university-of-oxford-research-data-management-policy.pdf>

Gestão de Dados, Citação de dados, Princípios FAIR, Compartilhamento de dados, Dados Sensíveis, Dados pessoais, Anonimização, Documentação sobre os dados, Armazenamento e *backup*, Ética, Gestão de Dados de Pesquisa, Curadoria de dados, Metadados, e Preservação de dados.

**Figura 10** - Universidade de Oxford e os códigos identificados nos documentos



**Fonte:** dados da pesquisa

Destaca-se também que a Universidade de Oxford dispõe de um *site* chamado *Research Data Oxford*<sup>24</sup>, o qual é a fonte central de informações e orientações da Universidade relacionadas ao gerenciamento de dados de pesquisa, destinado a pesquisadores de Oxford de todos os níveis (incluindo estudantes que realizam projetos de pesquisa) que lidam com dados de pesquisa, e a todos aqueles envolvidos no apoio a pesquisadores.

<sup>24</sup> <https://researchdata.ox.ac.uk/>

#### 4.1.6 Universidade Columbia – Nova York

A Universidade de Columbia é uma instituição de ensino superior privada, fundada no ano de 1754, sendo a mais antiga de Nova York e a quinta mais antiga dos Estados Unidos. É uma das principais instituições de pesquisa do mundo e possui um centro médico de renome mundial. Com suas 17 escolas e quatro escolas afiliadas espalhadas por quatro *campi* localizados na cidade de Nova York e arredores, a Universidade compreende seu impacto regional e global (Columbia University, 2024).

A Universidade oferece uma gama excelente e abrangente de programas acadêmicos. Estas incluem três escolas de graduação, 15 escolas de pós-graduação e profissionais, um centro médico de renome mundial, quatro faculdades e seminários afiliados, 25 bibliotecas e mais de 100 centros e institutos de pesquisa.

As bibliotecas da Universidade se destacam por oferecer aos seus usuários Serviço de Dados de Pesquisa, que fornecem apoio às consultas para as necessidades de dados de pesquisa dos membros de sua comunidade.

Além do mais, a instituição conta com o Gabinete do Vice-Presidente Executivo de Pesquisa<sup>25</sup>, um escritório onde administram-se as políticas que regem a condução da pesquisa na Universidade e a gestão de seus programas de pesquisa. Entre as atividades desse escritório, há o programa *Research and Data Integrity* (ReaDI)<sup>26</sup>, projetado para aprimorar o gerenciamento de dados e a integridade da pesquisa na Universidade de Columbia.

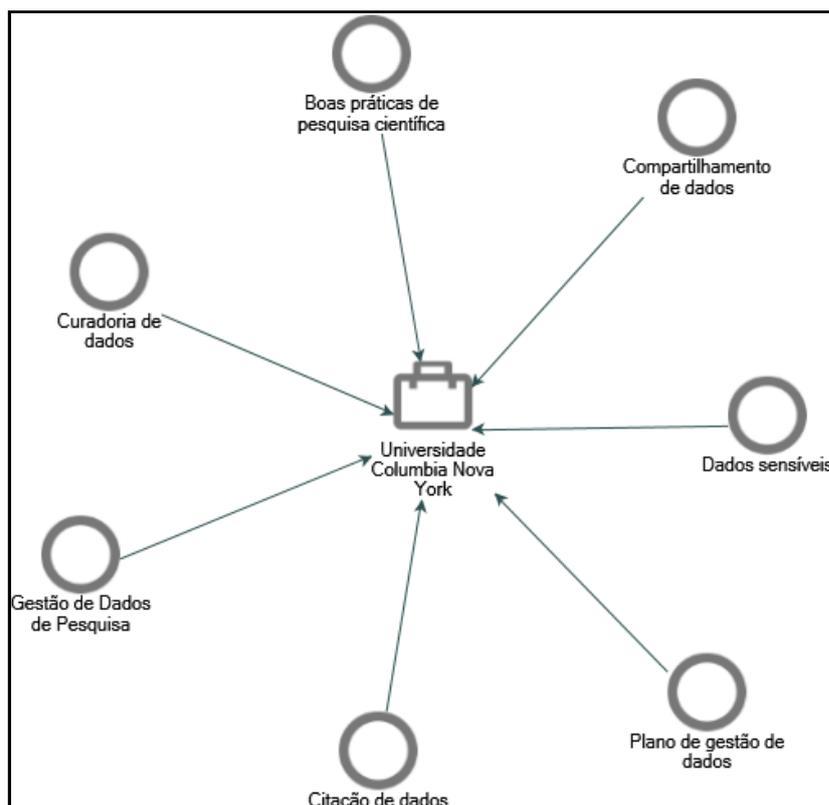
A Figura 11 mostra os assuntos que foram extraídos dos documentos recuperados da instituição. No total foram oito documentos, e após leitura e codificação obteve-se os códigos: Boas práticas da pesquisa científica, Dados sensíveis, Citação de dados, Plano de gestão de dados, Gestão de dados de pesquisa, Curadoria de dados e Compartilhamento de dados.

---

<sup>25</sup> <https://research.columbia.edu/office-executive-vice-president-research>

<sup>26</sup> [https://research.columbia.edu/ReaDI-Program#!#%2Fcu\\_accordion\\_item-12318](https://research.columbia.edu/ReaDI-Program#!#%2Fcu_accordion_item-12318)

**Figura 11** - Universidade Columbia - NY e os Códigos identificados nos documentos



**Fonte:** dados da pesquisa

Dados robustos e integridade de pesquisa são vitais para garantir que os resultados da pesquisa sejam reproduzíveis e verificáveis. As iniciativas da Universidade de Columbia - NY, visam tornar as boas práticas destacadas na Figura 11 o mais transparente e fácil de implementar possível para pesquisadores de todos os níveis na Columbia, disponibilizando os recursos mais relevantes e úteis para a comunidade de pesquisa da instituição.

#### **4.1.7 Universidade da Pensilvânia**

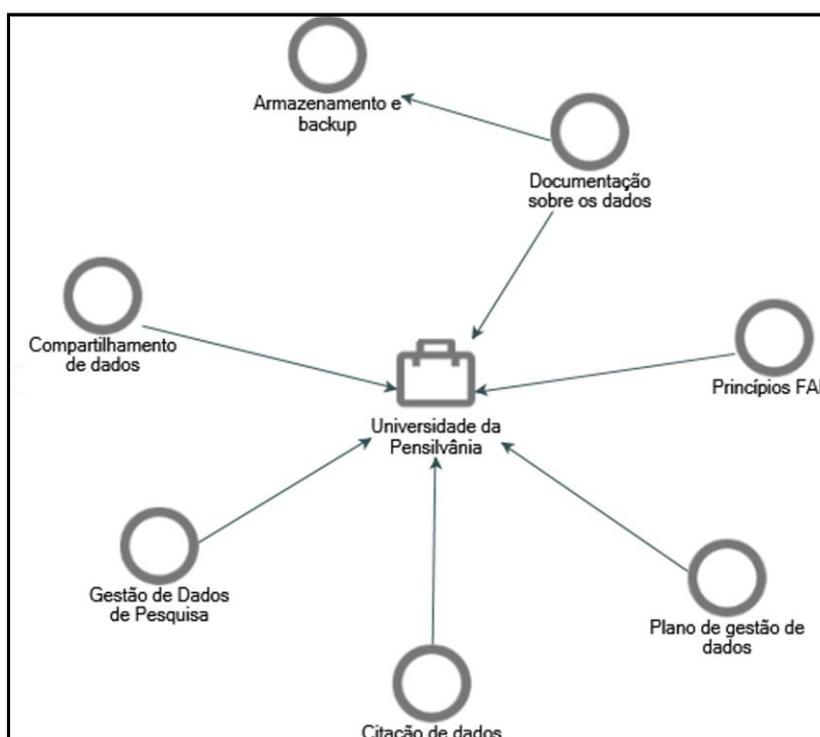
A Universidade da Pensilvânia é uma instituição privada de pesquisa, fundada em 1740, na Pensilvânia, nos Estados Unidos. É uma das nove faculdades coloniais do país e foi fundada antes da Declaração de Independência dos EUA, quando Benjamin Franklin, o fundador e primeiro presidente da Universidade, defendeu uma instituição educacional que treinasse líderes acadêmicos, de negócios e de serviço público (University of Pennsylvania, 2024).

A instituição abriga mais de 200 centros, institutos, laboratórios e iniciativas internacionais renomadas de pesquisa, além de quatro escolas de graduação e 12 escolas de pós-graduação. Também conta com um sistema de bibliotecas que incluem 19 bibliotecas físicas, e uma biblioteca digital conhecida pela inovação e riqueza de conteúdo.

As Bibliotecas Penn dão suporte ao ensino e à pesquisa por meio dos serviços que oferecem à sua comunidade acadêmica. Entre esses serviços destaca-se os da equipe da *Penn Libraries Research Data & Digital Scholarship*<sup>27</sup> que auxilia nas pesquisas e estudos baseados em dados e dá assistência para atividades de competência em dados (Penn Library, 2024).

Ao analisar os oito documentos coletados para análise, encontraram-se os assuntos abordados pela instituição: Armazenamento e *backup*, Documentação sobre os dados, Princípios FAIR, Plano de gestão de dados, Citação de dados, Gestão de dados de pesquisa e Compartilhamento de dados.

**Figura 12** - Universidade da Pensilvânia e os códigos identificados nos documentos



**Fonte:** dados da pesquisa

<sup>27</sup> <https://www.library.upenn.edu/rdds>

Os serviços disponibilizados pela Universidade da Pensilvânia buscam orientar e ajudar no processo de gestão, curadoria, publicação e compartilhamento de dados de pesquisa, seja para atender aos requisitos do financiador ou em prol da ciência aberta, fornecendo assistência para a criação e implementação de planos de gerenciamento e compartilhamento de dados, e ajudando os pesquisadores a encontrar serviços especializados relacionados a dados que eles precisam em qualquer estágio do processo de pesquisa.

#### 4.1.8 Universidade de Stanford

A Universidade de Stanford é uma instituição privada situada em Palo Alto, Califórnia, Estados Unidos, e uma das instituições mais prestigiadas do mundo. Fundada em 1885, Stanford oferece aos alunos programas acadêmicos amplos e profundos em sete escolas e diversas áreas – incluindo artes e humanidades, ciências naturais e sociais, engenharia, sustentabilidade, medicina, direito, educação e negócios (*Stanford University*, 2024).

A universidade promove práticas de pesquisa responsáveis, incluindo a proteção da privacidade dos participantes, a transparência na coleta e análise de dados, e a consideração dos impactos éticos e sociais de suas pesquisas.

Entre essas práticas, cita-se as políticas e regulamentações institucionais voltadas para o uso, retenção, acesso e compartilhamento de dados, como a Política de Retenção e Acesso a Dados de Pesquisa<sup>28</sup>, Política de Princípios Relativos à Pesquisa<sup>29</sup> e Política de Abertura na Pesquisa<sup>30</sup>.

Além disso, a instituição também conta com 20 bibliotecas com seus respectivos acervos conforme sua área de conhecimento. Entre os serviços oferecidos pelas bibliotecas, destaca-se o Serviço de Gerenciamento de Dados que dá suporte aos alunos e pesquisadores da instituição em relação ao gerenciamento, compartilhamento e preservação de dados de pesquisa (*Stanford Libraries*, 2024).

---

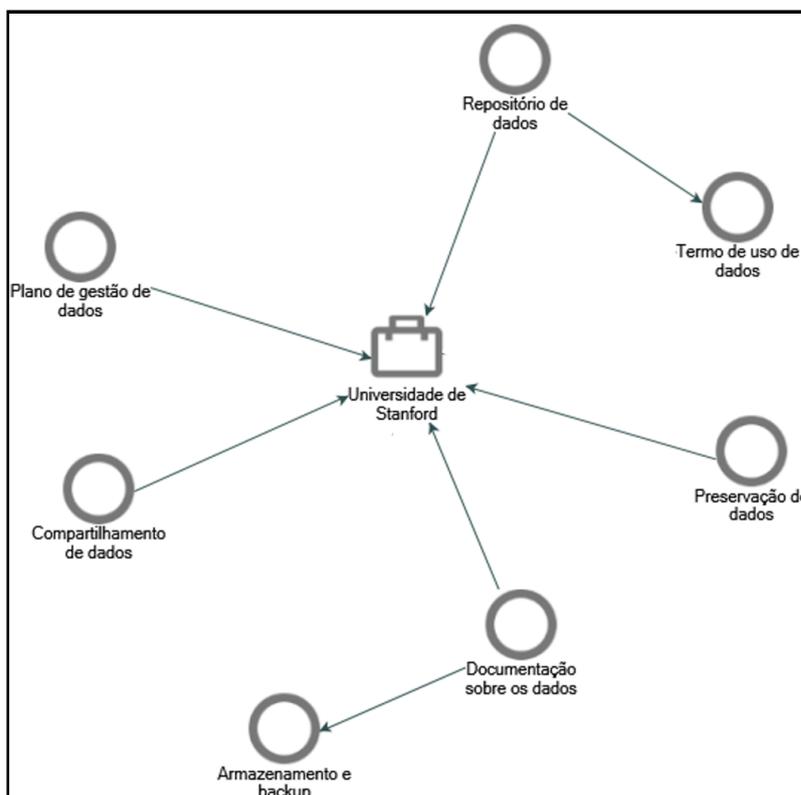
<sup>28</sup> <https://doresearch.stanford.edu/policies/research-policy-handbook/conduct-research/retention-and-access-research-data>

<sup>29</sup> <https://doresearch.stanford.edu/policies/research-policy-handbook/conduct-research/principles-concerning-research>

<sup>30</sup> <https://doresearch.stanford.edu/policies/research-policy-handbook/conduct-research/openness-research>

Nos documentos analisados da Universidade de Stanford, foi possível identificar os seguintes assuntos: Documentação sobre dados, Armazenamento e Backup, Plano de gestão de dados, Repositório de dados, Termo de uso de dados, Produção de dados, Compartilhamento de dados, e Preservação de dados, conforme a Figura 13.

**Figura 13** - Universidade de Stanford e os códigos identificados nos documentos



**Fonte:** dados da pesquisa

As políticas, regimentos e serviços de Stanford visando o uso correto dos dados por parte de sua comunidade acadêmica, tem sido um trabalho desenvolvido por profissionais que oferecem apoio técnico e recursos humanos para as orientações sobre o ato de planejar as etapas da criação dos dados, e domínio sobre as ferramentas de infraestrutura dos mesmos, bem como gerenciamento, compartilhamento e preservação de dados de pesquisa, contribuindo para o desenvolvimento de competências apoiadas em dados.

#### 4.1.9 Universidade da Califórnia - Berkeley

A Universidade da Califórnia em Berkeley (UC, Berkeley) foi fundada em 1868, nos Estados Unidos, e é uma das dez universidades semiautônomas da University of California (UC) que estão espalhadas pelo território do Estado, constituindo um dos três sistemas públicos estaduais de ensino superior da Califórnia.

A UC Berkeley é uma instituição pública multidisciplinar de pesquisa, conta com 172 programas de graduação e mais de 100 programas de pós-graduação. Sendo uma instituição altamente conceituada, em setembro de 2023 a UC Berkeley foi eleita a instituição pública número um do país, empatada com a UCLA, de acordo com a classificação das faculdades americanas de 2024 do *US News and World Report*<sup>31</sup> (Berkeley University of California, 2024).

A rede de bibliotecas em Berkeley, composta por 22 bibliotecas, oferece serviços de pesquisas avançadas para sua comunidade acadêmica. Além disso, em parcerias com outros setores da instituição, como o *Berkeley Research Data Portal*<sup>32</sup>, e o *Research Data Management Program (RDM)*,<sup>33</sup> oferece suporte, ferramentas e serviços de dados para todos os pesquisadores de Berkeley.

Dentre esses serviços e suportes oferecidos pela instituição, e com base nos materiais recuperados, os códigos identificados nos documentos foram: Documentação sobre os dados, boas práticas da pesquisa científica, Armazenamento e *backup*, Princípios FAIR, Repositório de dados, Termo de uso de dados, Dados Sensíveis, Citação de dados, Plano de gestão de dados, e Compartilhamento de dados. A Figura 14 ilustra a universidade e seus códigos.

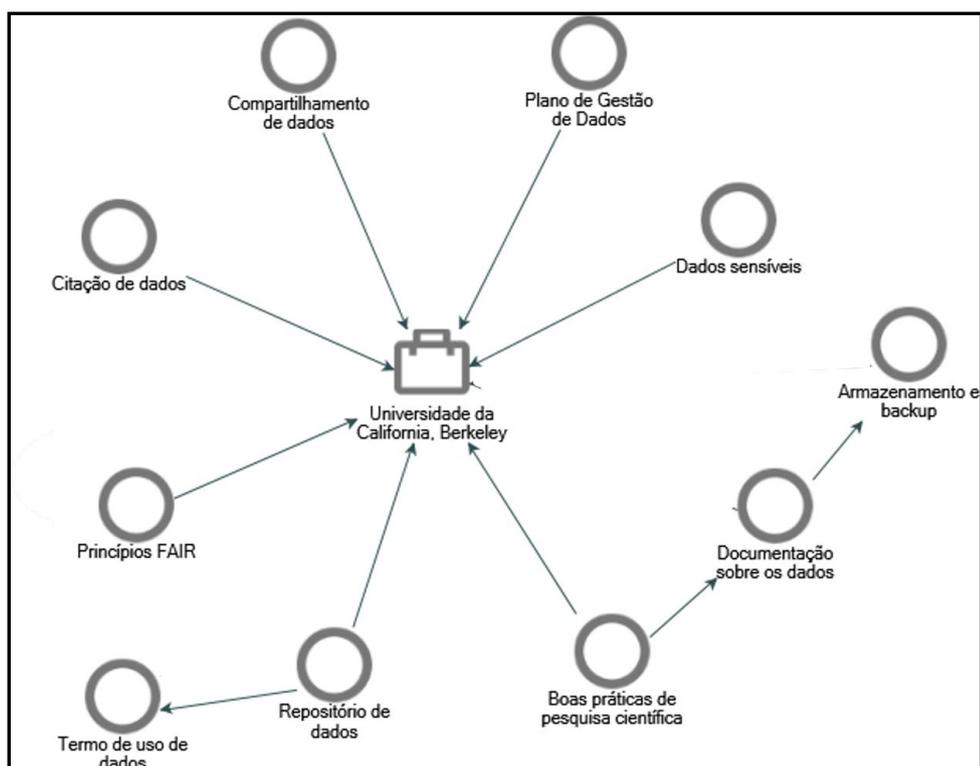
---

<sup>31</sup> <https://www.usnews.com/best-colleges>

<sup>32</sup> <https://researchdataportal.berkeley.edu/>

<sup>33</sup> <https://researchdata.berkeley.edu/home>

**Figura 14 - Universidade da Califórnia - Berkeley e os Códigos identificados nos documentos**



**Fonte:** dados da pesquisa

A instituição trabalha continuamente para identificar as necessidades não atendidas em relação ao gerenciamento de dados de pesquisa de sua comunidade acadêmica e se esforça para desenvolver soluções com parceiros do campus para atender a essas necessidades, ao mesmo tempo em que responde às mudanças nas formas como os pesquisadores trabalham com dados na UC Berkeley.

#### 4.1.10 Universidade de Washington

A Universidade de Washington, última instituição selecionada para este estudo, foi fundada em 1861 no estado de Washington, Estados Unidos, é uma instituição de pesquisa com três campus em Seattle, Bothell e Tacoma, bem como um centro médico de classe mundial que atende o estado e a região (University of Washington, 2024).

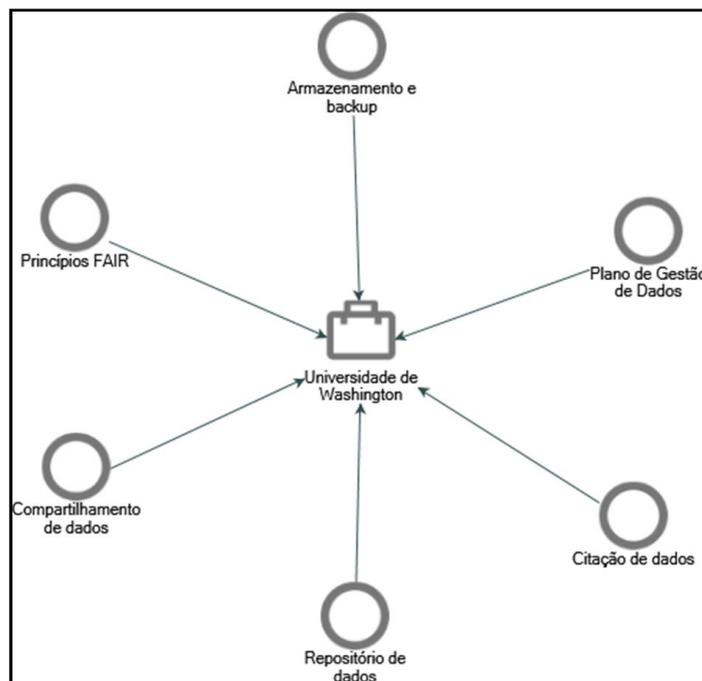
A Universidade de Washington é constantemente classificada entre as melhores universidades públicas do mundo, além de ser reconhecida por sua pesquisa estratégica e multidisciplinar. Dentro de sua extensa rede de centros e

institutos, pesquisadores, funcionários e estudantes da UW trabalham em diversas disciplinas para ampliar os limites do conhecimento

A Instituição conta com o suporte de suas bibliotecas na orientação do uso de dados de pesquisa aos alunos e pesquisadores. A Biblioteca oferece um guia de recursos relacionados aos diversos aspectos do gerenciamento de dados de pesquisa para membros da comunidade UW envolvidos em pesquisas com dados, desde introdução ao planejamento do gerenciamento de dados, requisitos para compartilhamento de dados, metadados, dicas para ajudar a mantê-los organizados, compartilhamento, arquivamento e preservação, e uma introdução às ferramentas e suporte no campus para auxiliar os pesquisadores.

Nos documentos da Universidade de Washington os temas encontrados foram: Armazenamento e backup, Compartilhamento de dados, Plano de gestão de dados, Repositório de dados, Citação de dados, Princípios FAIR e Coleta.

**Figura 15** - Universidade de Washington e os Códigos identificados nos documentos



**Fonte:** dados da pesquisa

Como pode-se observar nas análises, cada Universidade tem seus serviços, políticas ou programas que procuram ajudar alunos e pesquisadores a desenvolverem suas habilidades e competências em dados de pesquisa. Sobre isso, Goodman (*et al.* 2014) propõe algumas atitudes para tornar o gerenciamento de

dados das pesquisas mais objetivos às demandas dos pesquisadores. Os autores apresentam recomendações sobre o uso e cuidado com os dados que podem orientar os pesquisadores na tentativa de garantir que seus dados e análises continuem a ter valor, são elas:

- a) conduzir pesquisas com um determinado nível de reutilização em mente;
- b) vincular dados às próprias publicações sempre que possível;
- c) publicar uma descrição das etapas de processamento para permitir a interpretação e reutilização de dados;
- d) fomentar a utilização de repositórios de dados;
- e) compartilhar dados com um identificador permanente como, por exemplo, o *Digital Object Identifier* (DOI).

Além disso, Handerson e Corry (2020) também sugerem quatro condutas para auxiliar as práticas de competência em dados: a) criar programas nas faculdades e universidades; b) incentivar oportunidades de colaboração entre profissionais; c) incentivar o uso de dados; e d) investigar o papel da tecnologia e do *Big Data* sobre a Competência de Dados.

As práticas supracitadas elencam habilidades que o indivíduo e pesquisador devem ter ao trabalhar com conjunto de dados. Conforme os documentos analisados, cada instituição desenvolve suas atividades de competência em dados conforme o contexto que os dados são produzidos, compartilhados e reutilizados. Os códigos (temáticas/assuntos) elencados demonstram a realidade e, provavelmente, a demanda que cada universidade procura atender quando se trata de fornecer atividades e/ou serviços em dados de pesquisa.

Assuntos como *Plano de gestão de dados*, *Citação de dados*, *Gestão de dados*, *Compartilhamento de dados*, *Repositórios de dados*, *Preservação de dados*, *Metadados*, *Documentação sobre dados*, *Curadoria de dados*, por exemplo, que se destacam repetidamente nas universidades, demonstram os principais conteúdos das atividades desenvolvida pelas universidades. Isso aponta que há um esforço por parte das instituições em incluírem, nas atividades de competência em dados, noções sobre dados de maneira que o indivíduo possa pensar de forma crítica e responsável como trabalhar com dados, contribuindo para definição de uma proposta ética em uma pesquisa orientada à dados, e padrões de citação correta de

dados em determinado domínio e como compartilhá-los de maneira legal e prontos para o reuso (Costa; Salles; Zattar, 2023).

O uso do *software* Nvivo tornou possível identificar de que forma os conteúdos são abordados a partir dos documentos textuais analisados. Os códigos e subcódigos criados, viabilizou conhecer as principais temáticas abordadas nas universidades para tornar sua comunidade acadêmica competente em dados. A próxima seção apresenta essas temáticas tratadas nas atividades de competência em dados nas universidades, de acordo com os documentos recuperados, e com a descrição das atividades.

## 4.2 AS UNIVERSIDADES E O DESENVOLVIMENTO DE COMPETÊNCIA EM DADOS DE PESQUISA

As seções seguintes apresentam os conteúdos abordados e desenvolvidos nas atividades de competência em dados. Com a codificação de cada documento, mediante leitura flutuante, foi possível identificar o que foi abordado em cada universidade por meio das atividades oferecidas. Além disso, a seção também apresenta as Referências de codificação em cada universidade, ou seja, a recorrência dos Códigos em cada documento analisado, pois foi a partir desse procedimento que possibilitou-se descrever e compreender o conteúdo abordado nas atividades de competência em dados e ainda, realizar comparações entre as universidades.

Destaca-se que as universidades trabalham assuntos em comum, logo, será discorrido os assuntos de maneira geral destacando os principais pontos, e dando destaque para as diferenças encontradas.

### 4.2.1 Boas práticas da pesquisa científica

Considera-se boas práticas da pesquisa científica o conjunto de ações que envolve a ética, integridade e a gestão dos registros gerados nas pesquisas científicas, incluindo as práticas de preservação e a disponibilização dos dados, tornando-os acessíveis e facilitando a reprodutibilidade das pesquisas (Fiocruz, 2024).

A Figura 16 apresenta as instituições que nas atividades de competência em dados desenvolvem o conhecimento sobre as boas práticas da pesquisa científica.

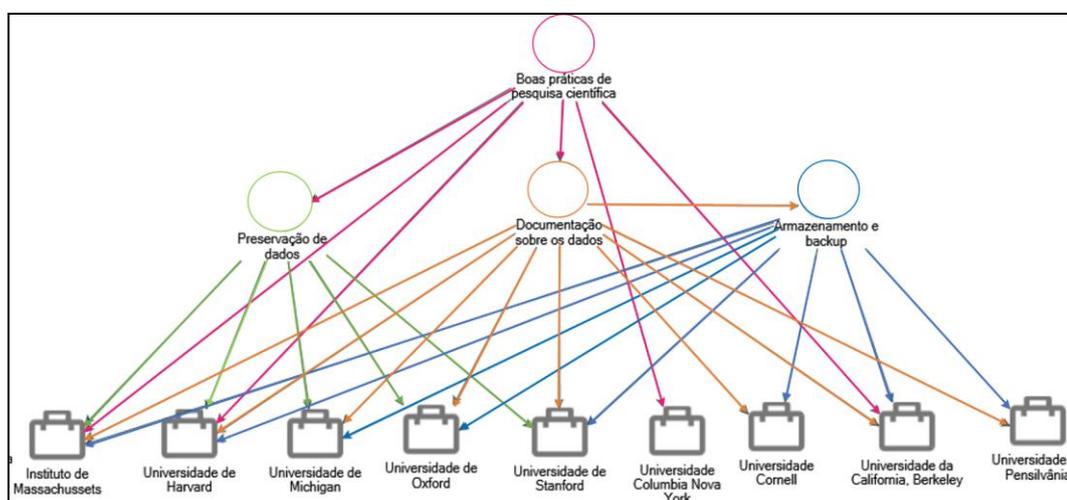
Destaca-se que durante a codificação dos documentos, percebeu-se que algumas das atividades sobre boas práticas da pesquisa científica eram conduzidas de forma mais genérica e outras de maneira específica, trazendo para a comunidade acadêmica informações mais detalhadas sobre as boas práticas da pesquisa científica para preservação de dados, para documentação de dados, e também, boas práticas de pesquisa científica para armazenamento e *backup*.

Para melhor organizar a descrição das atividades relacionadas às boas práticas de pesquisa científica, considerou-se, então, elaborar os subcódigos preservação de dados, documentação de dados e armazenamento e *backup* como forma de especificar as boas práticas de pesquisa científica tratadas nas universidades.

Dessa forma, o Código “Boas práticas de pesquisa científica” contém os Subcódigos “Preservação de dados” e “Documentação sobre os dados”, da mesma forma que o Subcódigo “Documentação sobre os dados” contém o Subcódigo “Armazenamento e *Backup*”.

Observa-se na Figura 16 que o Instituto de Massachusetts, e as Universidades de Harvard, Michigan, Oxford, Stanford, Columbia – Nova York, Cornell, Califórnia - Berkeley e Pensilvânia trazem em suas atividades de competência em dados conteúdos acerca das boas práticas de pesquisa científica que os indivíduos devem observar.

**Figura 16** - Universidades onde o Código “Boas práticas de pesquisa científica” foi identificado



Fonte: dados da pesquisa.

O Instituto de Massachusetts, a Universidade de Harvard, a Universidade de Columbia - NY, e a Universidade da Califórnia - Berkeley abordam o assunto “Boas práticas de pesquisa científica” de uma forma mais abrangente. Conforme análise dos documentos, percebeu-se que as instituições se preocupam com as boas práticas da pesquisa científica apresentando conjuntos de princípios, normas e diretrizes que orientam a condução de pesquisas baseadas em dados na comunidade acadêmica. Destaca-se as Universidade de Harvard e Colúmbia, NY, por exemplo, que orientam sobre suas políticas e programas institucionais criados para respaldar e assegurar o uso dos dados. Harvard disponibiliza a “Política de Segurança de Dados de Pesquisa de Harvard”, que trata de um conjunto de diretrizes e práticas que visam proteger a integridade, confidencialidade e disponibilidade dos dados de pesquisa gerados e utilizados no âmbito da instituição. Na Universidade de Columbia - NY, verificou-se a existência do Gabinete do Vice-Presidente Executivo de Pesquisa, um escritório onde administram-se as políticas que regem a condução da pesquisa na Universidade e a gestão de seus programas de pesquisa.

Além disso, as instituições abordaram assuntos a respeito das políticas de compartilhamento de dados das três agências de financiamento de pesquisa mais comuns dos EUA: a *National Institutes of Health* (NHI)<sup>34</sup>, *National Science Foundation* (NSF)<sup>35</sup>, e do *Office of Digital Humanities* (ODH)<sup>36</sup> do *National Endowment for the Humanities* (NEH), que exigem dos pesquisadores um plano de gestão de dados que descreva como os dados serão gerenciados e compartilhados.

Como mencionado anteriormente, essas instituições quando apresentam em suas atividades de competência em dados as boas práticas de pesquisa científica, ressaltam suas políticas institucionais que são estabelecidas para garantir à sua comunidade acadêmica que o desenvolvimento das pesquisas, bem como o uso e acesso dos dados sejam gerenciados de forma íntegra e transparente.

Outras instituições também abordaram em seus conteúdos as boas práticas de pesquisa científica, mas, apresentando aspectos mais específicos como as boas práticas para preservação de dados documentação dos dados e armazenamento e *backup*, como apresentado a seguir.

---

<sup>34</sup> <https://www.nih.gov/>

<sup>35</sup> <https://www.nsf.gov/>

<sup>36</sup> <https://www.neh.gov/divisions/odh>

#### 4.2.1.1 Boas práticas da pesquisa científica para preservação de dados

O Instituto de Massachusetts, Universidades de Harvard, Michigan, Oxford, e Stanford abordaram a temática boas práticas de pesquisa científica de maneira mais específica, voltada para a preservação de dados. A preservação dos dados envolve práticas e estratégias para garantir o armazenamento dos dados a longo prazo, de maneira segura e acessível. Com o crescente aumento das produções de dados, torna-se fundamental garantir a segurança e longevidade dos mesmos.

A respeito dessa prática as instituições apresentadas na Figura 16 mencionam em suas atividades a importância da preservação dos dados a longo prazo, principalmente visando a reutilização dos mesmos. Destacaram que para garantir que os dados sejam preservados, é preciso identificar quais conjuntos de dados produzidos são considerados de valor a longo prazo e a partir disso, delinear os planos para preparar e documentar os dados para preservação e compartilhamento, explicando o plano adotado de arquivamento e preservação, certificando, assim, o valor a longo prazo dos principais conjuntos de dados.

Além disso, as instituições abordaram os benefícios da preservação de dados, frisando que tal prática possibilita que os dados de pesquisa estejam disponíveis para futuros estudos, análises, e reutilização, conservando dados históricos, por exemplo, e culturais a longo prazo. A Universidade de Harvard apontou alguns repositórios que garantem a preservação dos dados ao longo do tempo, como o Zenodo<sup>37</sup>, OSF do *Center for Open Science* (COS)<sup>38</sup>, *Figshare*<sup>39</sup>, e seu próprio repositório institucional, o Harvard Dataverse<sup>40</sup>.

#### 4.2.1.2 Boas práticas da pesquisa científica para Documentação sobre os dados

Ainda a respeito das boas práticas de pesquisa científica, o MIT, Harvard, Michigan, Oxford, Stanford, Cornell, Califórnia e Pensilvânia, abordaram nos materiais analisados a importância da documentação de dados, destacando que para documentar corretamente os dados é essencial descrevê-los de forma clara e completa, para que outras pessoas também possam usá-los de forma consistente.

---

<sup>37</sup> <https://about.zenodo.org/>

<sup>38</sup> <https://osf.io/>

<sup>39</sup> <https://figshare.com/>

<sup>40</sup> <https://dataverse.harvard.edu/>

Além disso, orientam que métodos comuns de documentação de dados incluem dicionários de dados, cadernos de laboratório, livros de códigos qualitativos, entre outros meios e acrescentam que a documentação de dados frequentemente envolve, também, o uso de convenções padronizadas de nomenclatura e formatação, bem como padrões e ontologias de dados e metadados.

A documentação dos dados está intimamente relacionada à organização dos dados. As Universidades colocam a importância de informações a respeito da descrição dos dados e dos metadados, isto é, das informações estruturadas que descrevem o conteúdo e facilitam sua localização ou uso.

O Instituto de Massachusetts, por exemplo, destacou que uma boa documentação de dados inclui informações sobre nomes, rótulos e descrições para variáveis, registros e seus valores, explicação ou definição de códigos e esquemas de classificação usados, definições de terminologia especializada ou siglas utilizadas, entre outros fatores.

As instituições que abordaram a temática sobre as boas práticas da pesquisa científica para a documentação de dados, são as mesmas que desenvolveram tópicos sobre as boas práticas da pesquisa científica para armazenamento e *backup*. Isso pode demonstrar o comprometimento dessas instituições em apresentar à sua comunidade acadêmica um contexto completo sobre as boas práticas da pesquisa científica tratando também da documentação como uma necessidade para definir o armazenamento e *backup* dos dados.

#### 4.2.1.3 Boas práticas de pesquisa científica para Armazenamento e backup

De acordo com a Figura 16, o Instituto de Massachusetts, as Universidades de Harvard, Michigan, Oxford, Stanford, Cornell, Califórnia e Pensilvânia apresentaram uma série de informações a respeito do armazenamento e *backup* dos dados.

Sobre o armazenamento, as instituições forneceram um “macete” chamado “Regra de Três”, que consiste em TRÊS cópias, em pelo menos DOIS tipos de mídia diferentes, com UMA cópia em um local totalmente diferente:

1. “aqui” – uma cópia local no seu computador ou *desktop*, onde os arquivos foram criados ou coletados;

2. “próximo” – uma cópia externa em um local diferente, com tipo de mídia diferente do original; e
3. “distante” – uma cópia externa em um local geograficamente diferente, como um serviço de armazenamento em nuvem.

Além disso, destacaram alguns repositórios para armazenamento de dados a longo prazo, como o Harvard Dataverse, Dspace@MIT, Zenodo, entre outros.

Para o *backup*, as instituições recomendaram o uso de um serviço automatizado para criar cópias de segurança regulares. Foi destacado pelas instituições, que ao realizar os *backups*, os indivíduos considerassem criar cópias de apenas uma parte, em vez de todos os seus dados. Por exemplo, fazer o *backup* apenas do subconjunto mais valioso, importante ou vulnerável dos dados, ou apenas dos arquivos que foram alterados recentemente, pois isso pode manter os custos de armazenamento de cópias de segurança minimizados. A Universidade de Cornell, por exemplo, destacou o CrashPlan<sup>41</sup> como opção de *backup* para os departamentos e seus funcionários.

Com o uso do *software* Nvivo foi possível, ainda, identificar a quantidade de vezes que os subcódigos descritos a cima foram referenciados nos documentos analisados em cada universidade, isto é, trechos sobre o referido assunto. As Referências de codificações são essenciais para organizar e analisar as informações (dados) de maneira sistemática, podendo ser visualizadas em contextos, o que ajuda na compreensão de como e onde alguns temas aparecem. Nesse caso, possibilitou verificar o que cada instituição tratou sobre determinado assunto.

Salienta-se ainda, que esses dados não se referem à quantidade de documento, isso pode ser verificado no Gráfico 1, mas essas Referências ajudam a identificar a recorrência do tema nos documentos. Logo, nos materiais disponibilizados por essas instituições foram codificados trechos de texto a respeito dos assuntos tratados, e então, geradas as Referências de codificação.

O código “Boas práticas da pesquisa científica” e o subcódigo “Preservação de dados”, por exemplo, contém um total de 20 Referências de codificações nos 49 documentos analisados da Universidade de Harvard.

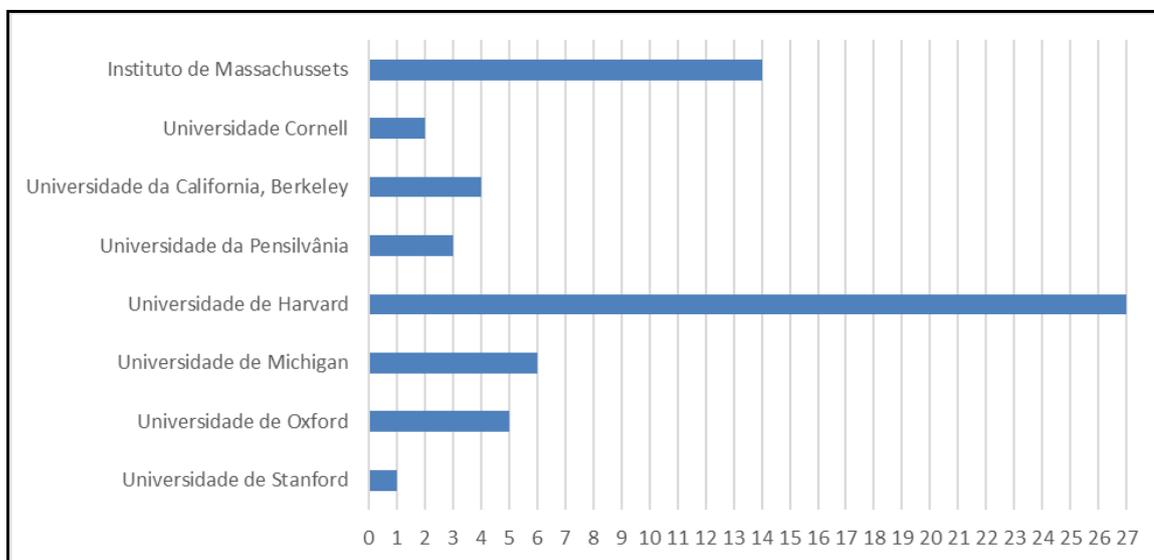
O subcódigo “Documentação de dados”, como demonstra o Gráfico 2, foi referenciado 27 vezes nos documentos da Universidade de Harvard, seguida do

---

<sup>41</sup> <https://www.crashplan.com/company/>

Instituto de Massachusetts, com 14 referências, e da Universidade de Michigan com seis.

**Gráfico 2 - Número de Referências do Código Documentação sobre os dados, nas universidades**

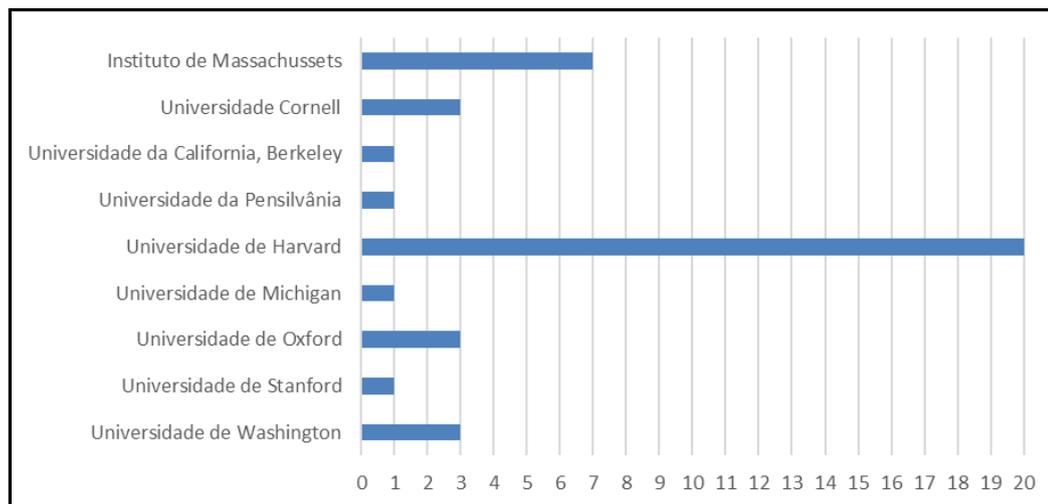


**Fonte:** dados da pesquisa.

Durante a análise dos documentos relativos às atividades oferecidas por essas instituições, observou-se que a Universidade de Harvard, por exemplo, desenvolveu mais conteúdos sobre documentação sobre os dados. Apesar das instituições terem abordado aspectos muito semelhantes sobre o assunto, Harvard se diferenciou apresentando exemplos práticos e casos de estudo, ajudando a ilustrar como a documentação de dados pode ser implementada no contexto real das pesquisas. O MIT apresentou nos materiais um compilado de orientações práticas para que o pesquisador (e demais interessados) consigam documentar corretamente os dados, e compreendam a importância dessa prática.

Igualmente, nas Referências de codificações do Subcódigo “Armazenamento e *backup*”, sobressaísse a Universidade de Harvard, com 20 Referências de codificação, seguido do MIT, com sete, como mostra o Gráfico 3.

**Gráfico 3 - Quantidade de Referências para o Código Armazenamento e Backup, nas universidades**



**Fonte:** dados da pesquisa.

Conforme a análise dos documentos, observou-se que a Universidade Harvard, diferente das demais, apresenta estratégias para garantir a longevidade e integridade dos dados ao longo do tempo. Essas estratégias incluem a implementação de políticas de *backup*, o uso de formatos de arquivo padronizados, assim como a adoção de sistemas de armazenamento. Também, salienta a importância de registrar metadados que descrevam de forma completa os conjuntos de dados, incluindo informações sobre a origem dos dados, métodos de coleta e condições de uso.

Observou-se que todas as instituições mencionadas trazem noções teóricas evidenciando primeiramente a razão dessas boas práticas serem necessárias para a pesquisa científica, para posteriormente, a partir de exemplos, levar os pesquisadores e alunos a pôr em ação o conteúdo das atividades ministradas, entendendo que se tratam de ações que visam garantir a qualidade das atividades de pesquisa por meio da aplicação das boas práticas relacionadas a dados de pesquisa.

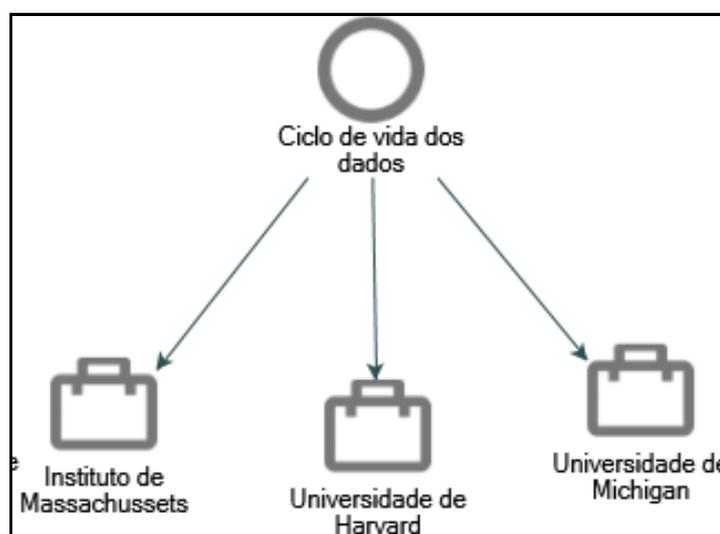
#### **4.2.2 Ciclo de vida dos dados**

O ciclo de vida de dados consiste em uma série de fases ao longo da vida útil dos dados. Cada fase é administrada por um conjunto definido de normas que visam aumentar o valor dos dados durante cada etapa do ciclo de vida, desde a fase da

criação até a preservação dos dados. Durante as análises observou-se que as instituições apresentam, em suas atividades de competência em dados, as fases e as etapas do processo de ciclo de vida dos dados.

O Instituto de Massachussets, e as Universidade de Harvard e Michigan abordaram este assunto com o intuito de situar os indivíduos quanto às etapas que envolvem o ciclo de vida dos dados de pesquisa, fomentando uma reflexão sobre os diferentes momentos e fatores envolvidos no processo da pesquisa. Destaca-se que o conteúdo dos documentos analisados possui cunhos mais conceituais e teóricos, apresentando o que constitui o ciclo de vida dos dados e suas fases (Fig. 17).

**Figura 17** - Universidades onde o Código “Ciclo de vida dos dados” foi identificado



**Fonte:** dados da pesquisa.

O Instituto de Massachusetts nas atividades ministradas sobre o ciclo de vida dos dados, trata a respeito dos requisitos para o planejamento e gerenciamento de dados, descrição de metadados, compartilhamento de dados, reutilização, descoberta de dados, análise de dados e também, sobre o ciclo de vida dos dados adaptado do modelo *Data Documentation Initiative* (DDI) versão 3.0.

A Universidade de Michigan por meio dos serviços disponibilizados pela biblioteca ajuda a identificar as fases do ciclo de vida dos dados e a considerar quais ações devem ser tomadas em cada fase para garantir que os dados sejam acessíveis, compreensíveis e adequados para reutilização.

Não encontrou-se maiores detalhes a respeito dos conteúdos abordados nas atividades de capacitação, nos materiais disponibilizados por essas duas universidades. Ambas instituições tratam de maneira genérica o assunto sobre o ciclo de vida dos dados, apenas com conceitos e exemplos práticos.

Porém, na Universidade de Harvard identificou-se, entre os conteúdos analisados, uma lista de verificação das etapas do ciclo de vida dos dados. Trata-se de um guia para que o pesquisador observe o nível em que se encontra sua pesquisa, para assim, adotar bons elementos de gerenciamento de dados de pesquisa. Essa lista de verificação também serve como documento de referência para acompanhar os elementos que compõem um bom gerenciamento de dados de pesquisa no ciclo de vida e elaboração do PGD.

Nos materiais selecionados da Universidade de Harvard verificou-se o detalhamento com que são descritas as etapas do ciclo de vida da pesquisa, abrangendo: o **planejamento** - que consiste na elaboração de um plano de gestão de dados para controlar como os dados são tratados; a **coleta** - como os dados são coletados, onde serão armazenados e por quanto tempo serão mantidos; a **análise** - etapa em que os dados passarão por uma limpeza e organização, podendo ser utilizados *software* para auxiliar nesse processo (como o *OpenRefine*) para então serem analisados; a **retenção** - analisar a necessidade de reter os dados do projeto ou pesquisa para colocá-los de acordo com as diretrizes da instituição, financiador estadual e/ou federal; o **compartilhamento** - – permite decidir a respeito da disponibilização dos dados em um repositório de dados (de acesso aberto); o **armazenamento** - delineamento de como os dados serão armazenados, pensando na sua preservação; e a **reutilização** - permitir que outros pesquisadores acessem os dados possibilitando novos estudos.

Destaca-se que nas três instituições ao tratar as questões envolvidas no ciclo de vida dos dados, as informações encontradas nos materiais analisados são em relação aos estágios iniciais do ciclo de vida dos dados de pesquisa, o que inclui, também, atividades de redação de PGD, por exemplo.

Em relação às Referências de codificação, o Código “ciclo de vida dos dados” foi referenciado uma única vez nos documentos analisados do MIT e da Universidade de Michigan, por isso a falta de informações para descrever mais

precisamente os conteúdos sobre o assunto, e, 10 vezes nos documentos da Universidade de Harvard, no qual obteve-se mais informações.

Mais uma vez a Universidade de Harvard destaca-se no número de Referências de codificação, ou seja, mais trechos codificados sobre o assunto nos documentos. Nas análises desses documentos observou-se que Harvard desenvolveu com mais detalhes aspectos relacionados ao ciclo de vida dos dados, inclusive, mais especificamente a respeito de dados biomédicos para os pesquisadores e alunos que atuam nessa área disciplinar. As demais instituições foram objetivas e generalistas a respeito do assunto.

O entendimento a respeito do processo do ciclo de vida dos dados pode contribuir para formar o olhar do pesquisador ao pensar sobre os acessos e uso dos dados a curto e a longo prazo. Os estágios do ciclo de vida dos dados variam de acordo com o tipo de pesquisa que é conduzida, os dados com os quais se está trabalhando e os objetivos e necessidades específicas para os dados. Logo, conhecer cada fase que envolve a pesquisa contribui para refletir e considerar as melhores alternativas sobre como os dados serão descritos, geridos e protegidos.

Barros *et al.* (2023) complementam que refletir sobre as etapas do ciclo de vida dos dados torna-se fundamental para garantir a qualidade, integridade e impacto dos dados de pesquisa, pois além dos indivíduos pensar sobre as etapas específicas relacionadas ao gerenciamento de dados de pesquisa, desde a coleta até a preservação a longo prazo, esse processo torna-se crucial pois envolve as práticas relacionadas aos dados coletados e gerados durante a pesquisa.

Salienta-se que, apesar de não ser observado nas demais instituições o desenvolvimento de competências sobre esse tema de forma explícita, no conteúdo dos materiais analisados neste estudo, não significa que não seja um ponto desenvolvido por elas. Representa apenas que a temática foi tratada pelas três universidades citadas de forma diferenciada, dando ênfase a algumas das etapas do ciclo de vida dos dados, como por exemplo: armazenamento, preservação, compartilhamento e reutilização dos dados.

As três instituições foram citadas intencionalmente na codificação “Ciclo de vida de dados” por mencionar o tema e explicar sobre as etapas que envolve de forma a esclarecer a comunidade acadêmica da sua importância. A disponibilização de documentação a respeito facilitou o acesso para esta pesquisa mostrando as

atividades desenvolvidas para ampliar a competência da comunidade acadêmica sobre o ciclo de vida dos dados e suas etapas.

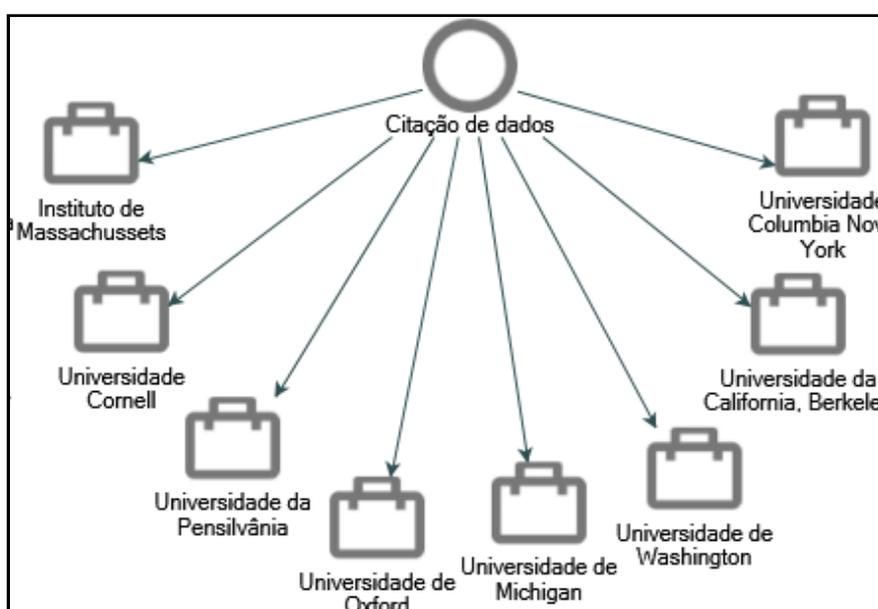
A próxima subseção apresenta o Código Citação de dados e a maneira como as instituições abordam o tema.

### 4.2.3 Citação de dados

O processo de citação de dados de pesquisa desenvolveu-se conforme os pesquisadores e demais partes interessadas perceberam que a inclusão de dados é necessária para um registo acadêmico completo entre um produto de pesquisa e as evidências em que se baseia (Oliveira *et al.*, 2022). Citar dados assegura que os pesquisadores obtenham o crédito adequado por seu trabalho.

O Instituto de Massachusetts, as Universidades de Cornell, Pensilvânia, Oxford, Michigan, Washington, Califórnia e Columbia, como mostra a Figura 18, nas suas atividades de competência em dados, apresentam informações relacionadas à citação de dados, entre elas, os benefícios para os produtores e usuários de dados, estratégias e recursos para citar dados, razões para citar dados e exemplos de como fazê-lo. O desenvolvimento dos conteúdos sobre citação de dados nas universidades é muito semelhante.

**Figura 18** - Universidades onde o Código “Citação de dados” foi identificado



**Fonte:** dados da pesquisa.

A *Penn Library* em suas atividades de capacitação elenca uma série de motivos para os indivíduos citar dados, sendo alguns deles: dar crédito aos criadores dos conjuntos de dados, permitir que aqueles que compartilham os dados meçam seu impacto, apoiar a infraestrutura de pesquisa conectando dados e pesquisas publicadas, melhorar o acesso aos dados, oferecer oportunidades para permitir a reutilização, entre outras razões. Além disso, a Biblioteca oferece um guia com os elementos centrais necessário para citar dados, com base no guia do IASSIST<sup>42</sup>, bem como orientações de COMO citar um conjunto de dados em um artigo.

As Universidades de Cornell, Columbia, Califórnia, Michigan, Washington e Oxford nos materiais analisados mencionam exemplos de alguns guias de estilos de citação de dados diante da ausência de padrões universais para a citação de conjuntos de dados, e das dificuldades que as comunidades acadêmicas têm enfrentado para desenvolver padrões de citação de dados dentro dos estilos de citação existentes, tendo a responsabilidade de criar um formato de citação de dados que esteja em conformidade com as regras de estilos existentes da melhor forma possível.

Entre os exemplos mencionados estão a norma APA (6ª edição), MLA (7ª edição), Chicago (16ª edição) (autor-data), *DataCite*, DRÍADE e *Inter-university Consortium for Political and Social Research* (ICPSR). Embora sejam recomendados estilos de citação de autores e trabalhos acadêmicos, o pesquisador é orientado que quando os formatos de citação não são especificados, é permitido seguir o estilo de citação acadêmica da sua disciplina contemplando informações de: criador ou colaborador dos dados, a data da publicação, título do conjunto de dados, editor, identificador (por exemplo, Handle, ARK, DOI) ou URL da fonte, e data de acesso, quando apropriado.

A respeito dessa dificuldade encontrada de obter padrões universais para citação de dados de pesquisa, o estudo de Oliveira *et al.* (2022) salienta que a academia ainda carece da existência de uma teoria que se proponha a explicar a citação dos dados de pesquisa que, como exposto, são cada vez mais requisitados e utilizados pela comunidade acadêmica. A citação adequada de fontes de dados traz benefícios imediatos e de longo prazo para usuários e produtores de dados.

---

<sup>42</sup> [icpsr.umich.edu/files/ICPSR/enewsletters/iassist2.html](https://icpsr.umich.edu/files/ICPSR/enewsletters/iassist2.html)

A pesar da relevância do tema, não encontraram-se muitos documentos tratando do assunto. No Instituto de Massachusetts, por exemplo, foi possível identificar cinco vezes o Código “Citação de Dados”, nas Universidades da Pensilvânia e Oxford, houve duas Referências de codificação, e em Columbia, Cornell, Califórnia, Michigan e Washington apenas uma referência nos documentos analisados.

A biblioteca do MIT (unidade responsável por fornecer capacitações) disponibiliza diversos conteúdos sobre “Citação de Dados”, é a segunda instituição com mais documentos coletados para análise, 24 no total, e desenvolveu mais conteúdos também sobre o assunto.

A MIT *Library* disponibiliza em seu *síte* um guia de estilo de citação, fornecendo instruções de como citar pesquisas bibliográficas, e como citar dados de pesquisa. Na “aba” *cite dados* são disponibilizadas diretrizes detalhadas para citar dados, e recomendações de alguns guias com orientações, como o Guia rápido para citação de dados IASSIST, e o conteúdo da DCC sobre como citar conjuntos de dados e vinculá-los às publicações.

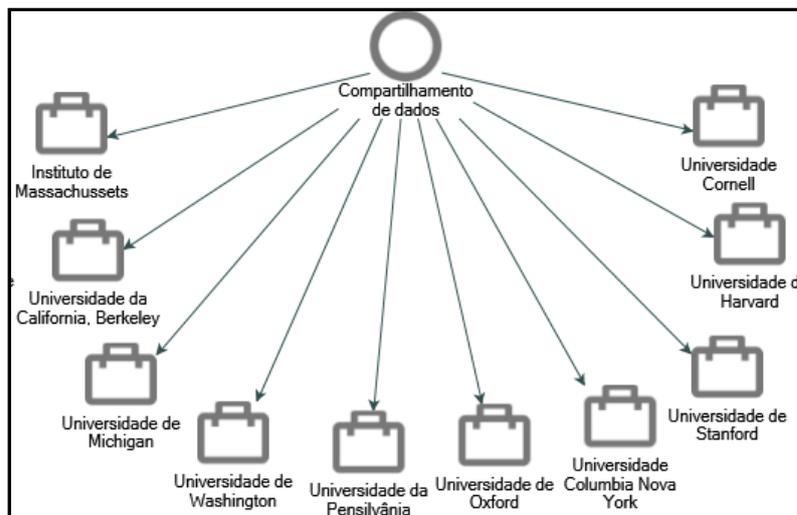
As oito universidades buscaram desenvolver na sua comunidade acadêmica noções básicas de que citar dados dá crédito apropriado ao produtor de dados, e ainda que sejam de domínio público, é uma boa prática acadêmica garantir que sejam devidamente citados.

Outra boa prática que as instituições elencam em suas atividades de competência em dados, é sobre o compartilhamento de dados e os benefícios que isso gera para a comunidade científica. A próxima subseção disserta sobre essa prática e como é desenvolvida pelos serviços de dados.

#### **4.2.4 Compartilhamento de dados**

Dados criados a partir de pesquisas são recursos valiosos que podem ser usados e reutilizados para outros fins científicos e educacionais, e o compartilhamento de dados propicia que novas pesquisas venham a ser desenvolvidas e acessadas. Conforme demonstra a Figura 19, todas as universidades selecionadas neste estudo abordam conteúdos sobre o compartilhamento de dados nas atividades que têm como objetivo tornar sua comunidade competente em dados.

**Figura 19** - Universidades onde o Código “Compartilhamento de dados” foi identificado



**Fonte:** dados da pesquisa.

Percebeu-se que as instituições ao desenvolverem seus conteúdos sobre o tópico, orientam e explicam os motivos e benefícios pelos quais o compartilhamento de dados deve ser uma prática adotada por aqueles que trabalham com pesquisas, destacando que o compartilhamento de dados proporciona:

- a) maiores chances de futuras colaborações de pesquisa;
- b) crescimento do impacto acadêmico;
- c) divulgação mais ampla dos resultados da investigação;
- d) transparência e reprodutibilidade; e
- e) cumprimento dos requisitos de editores e agências de financiamento.

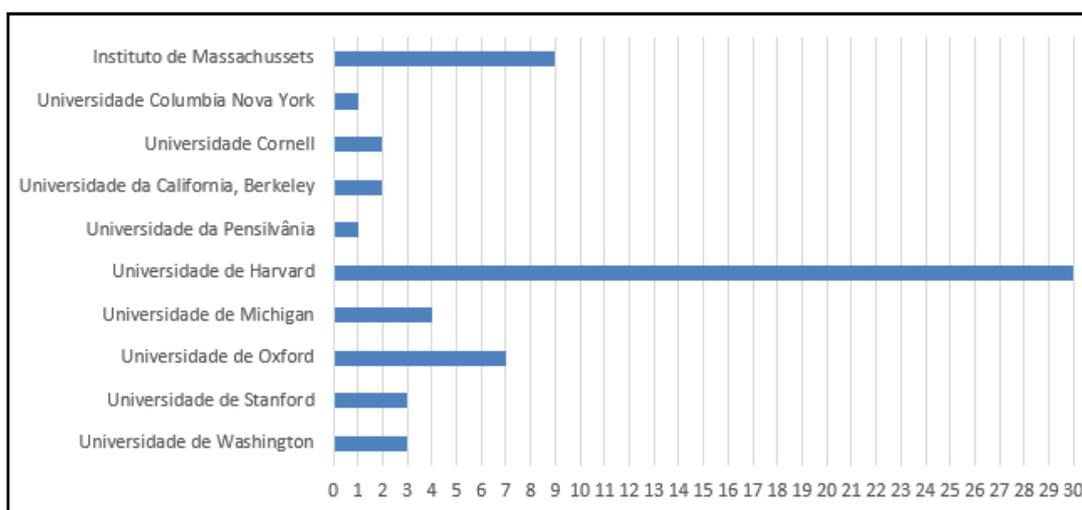
A facilidade com que os dados, graças às tecnologias disponíveis, podem ser armazenados, divulgados e facilmente acessados, propicia que muitas instituições compartilhem dados de pesquisa para aumentar o impacto e a visibilidade de suas pesquisas. Foi muito citado no conteúdo dos documentos analisados as exigências das agências de fomento à pesquisa a respeito da adoção de políticas de compartilhamento de dados como, por exemplo: a *National Institutes of Health* (NHI), a *National Science Foundation* (NSF) e o Escritório de Humanidades Digitais do *National Endowment for the Humanities* (NEH), o que, certamente, incentivam os pesquisadores a compartilhar dados e resultados. Destaca-se que as políticas de compartilhamento de dados dessas agências tendem a permitir que os pesquisadores usem e mantenham os dados de uso exclusivos para os grupos de

pesquisa por um período de tempo razoável para publicar os resultados provenientes dos dados.

Além disso, as instituições também elencam alguns repositórios recomendados para o depósito dos dados e atender a práticas de compartilhamento, como o DRYAD e o FigShare. Indicam o uso do RE3data, um diretório global de registro de repositórios de dados de pesquisa que abrange repositórios de dados de diferentes disciplinas acadêmicas, para que os pesquisadores identifiquem repositórios temáticos e se sintam encorajados a seguir a prática de compartilhamento de dados.

A Universidade de Harvard, durante a análise dos tópicos desenvolvidos sobre compartilhamento de dados, foi a que se destacou em relação ao fornecimento de informações e assessoramento sobre o tema nos materiais disponibilizados para consulta. onde O Código “Compartilhamento de dados” foi referenciado 30 vezes, como mostra o Gráfico 4.

**Gráfico 4** - Quantidade de Referências para o Código Compartilhamento de dados, nas universidades



**Fonte:** dados da pesquisa.

Diferente das demais universidades, Harvard, por meio dos Serviços de Dados da sua rede de biblioteca, foi a única instituição na qual identificou-se, entre os documentos analisados, uma descrição completa, não apenas de benefícios e motivos de compartilhar os dados de pesquisa, mas questões sobre compartilhamento de dados confidenciais e restritos e o que fazer para permitir o compartilhamento desses dados. Também, incluem informações sobre licenciamento

de dados e as possibilidades de limitar o que outras pessoas podem fazer com eles, e nesse caso esclarecem sobre a utilização das Licenças *Creative Commons*. Divulgam sua ferramenta de compartilhamento de dados, o Harvard *Dataverse*, repositório disponível para armazenamento, compartilhamento e publicação de dados de pesquisa. Finalmente, esclarecem sobre a disponibilidade de atendimento à comunidade de Harvard sobre compartilhamento e preservação de dados.

Embora o valor potencial do compartilhamento de dados seja cada vez mais reconhecido pelas instituições e pesquisadores, como percebeu-se por meio da análise dos documentos que tratam sobre o assunto, várias questões suscitam preocupações importantes e devem ser abordadas nas atividades de competência em dados para que o compartilhamento seja bem-sucedido e apropriado.

Monteiro e Sant'Ana (2016) destacam que entre essas preocupações, está a falta de entendimento dos produtores de dados em relação a sua gestão adequada que envolve uma série de atividades e práticas destinadas a garantir que os dados coletados e gerados durante a pesquisa sejam bem organizados, armazenados de forma segura e possam, então, ser compartilhados e reutilizados conforme necessário.

Araújo, Dias e Autran (2021) em pesquisa sobre o compartilhamento de dados em contexto de Ciência Aberta, observaram que a temática de compartilhamento dos dados está entrelaçada com a sua adequada gestão e publicação, visto que, na ausência de um gerenciamento efetivo e de práticas que tornem os dados públicos, a reutilização deles é afetada, logo, prejudica as novas descobertas científicas.

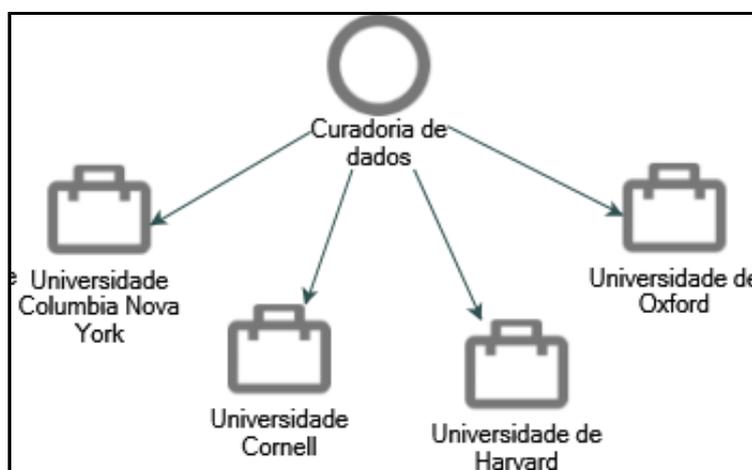
Destaca-se que esse fato pode ser superado a partir de iniciativas de boas práticas que se tornem comuns entre os pesquisadores, iniciativas essas, que já estão acontecendo nas universidades americanas e europeias por meio de políticas e atividades de capacitação para suas comunidades acadêmicas.

#### **4.2.5 Curadoria de dados**

A curadoria de dados garante que os conjuntos de dados sejam completos, bem descritos e em um formato e estrutura que melhor facilitem o acesso, a descoberta e a reutilização de longo prazo. Conforme a Figura 20, nas

Universidades de Columbia, Cornell, Harvard e Oxford identificou-se a presença de temas relacionados à curadoria dos dados de pesquisa como forma de tornar competente em dados suas comunidades acadêmicas.

**Figura 20** - Universidades onde o Código “Curadoria de dados” foi identificado



**Fonte:** dados da pesquisa.

Sobre esse tópico, as instituições abordaram questões referentes à curadoria de dados para facilitar a reutilização visando garantir que os dados sejam bem organizados, documentados e, conseqüentemente, estejam acessíveis para reuso. Uma das formas de apresentar informações pertinentes sobre os dados é a criação de um arquivo *README*, recurso valioso para quem está trabalhando com os dados, proporcionando contexto, clareza e transparência sobre a sua origem e utilização.

O arquivo *README* pode conter informações sobre: título e descrição do conjunto de dados, origem dos dados, metadados, instruções de uso, dicionário de dados com uma lista de todas as variáveis presentes no conjunto de dados, licenciamento e restrições, exemplos que ajudem a entender o conteúdo dos dados e seu contexto, referências se houver estudos ou documentos relacionados que usem esses dados, contato com informações de contato para o responsável pelo conjunto de dados, para que os usuários possam reportar problemas, fazer perguntas ou obter suporte e histórico de versões. (Digital Curation Centre, 2024)

Também, os materiais analisados sugerem estratégias às quais os pesquisadores podem aderir para realizar a curadoria de dados. As estratégias consistem em:

- a) criar um PGD incluindo de que forma os dados serão coletados, armazenados, documentados e compartilhados;
- b) coletar dados utilizando métodos que garantem a qualidade;
- c) documentar os dados e preencher os metadados (aqui indicam a criação de um arquivo *README*);
- d) determinar licença (como *Creative Commons*) e termos de acesso e reutilização; e
- e) publicar e citar.

A Universidade de Cornell, por exemplo, oferece serviços de curadoria em nível de dados para ajudar os pesquisadores a publicar seus dados e atender os Princípios FAIR e os requisitos nacionais de compartilhamento de dados<sup>43</sup>. Os curadores de dados (bibliotecários ou especialistas na área de curadoria de dados) orientam os pesquisadores a tornar seus os dados localizáveis, acessíveis, interoperáveis e reutilizáveis, alinhando-os com os Princípios FAIR, a compartilhar os dados com um identificador de objeto digital (DOI), a aumentar o alcance e impacto das pesquisa por meio da citação de dados, a refletir sobre o futuro dos trabalhos por meio do arquivamento e preservação a longo prazo e a desenvolver metadados padronizados apropriados. O processo de curadoria envolve uma revisão dos dados e da documentação de um pesquisador para garantir que os dados sejam tão completos, compreensíveis e acessíveis quanto possível. As instruções a os pesquisadores esclarecem que a extensão da revisão do conjunto de dados dependerá do tamanho do conjunto de dados depositado, da sua qualidade de documentação e da disponibilidade geral do pessoal. Essas revisões não são realizadas por pares e não julgam a relevância científica.

Nas Referências de codificação, que tornaram possível identificar e analisar cada tema desenvolvido pelas universidades, observou-se que nos documentos analisados da Universidade de Harvard e da Universidade de Cornell houve três Referências de codificação sobre “curadoria de dados” e nos documentos de Oxford e Columbia - NY, apenas uma Referência de codificação.

O tema “curadoria de dados” foi abordado, nas duas últimas universidades citadas, para explicar a respeito do processo de organizar e gerenciar um de conjuntos de dados.

---

<sup>43</sup> [eCommons Meets Federal Data Sharing Requirements – Cornell Data Services](#)

Nas Universidade de Harvard e na Universidade de Cornell suas práticas de competência em dados tratam de aspectos específicos da curadoria de dados, como por exemplo: documentação, armazenamento, preservação e compartilhamento, desenvolvendo uma série de conteúdos sobre cada uma dessas etapas, mesmo que não citem ou mencionam que estas práticas fazem parte do processo de curadoria de dados.

Silva *et al.* (2019) colocam que a curadoria de dados de pesquisa é o processo de gerenciamento de dados de pesquisa durante todo o seu ciclo de vida para disponibilidade em longo prazo e para a reutilização.

Na Universidade de Harvard o grupo de trabalho de gerenciamento de dados de pesquisa da Área Médica do campus de *Longwood* (sob responsabilidade da Biblioteca) oferece oficinas sobre curadoria de dados de pesquisa, abordando assuntos sobre gestão, documentação e metadados, armazenamento e backup, compartilhamento e publicação de dados e preservação digital, explicando para sua comunidade acadêmica sobre o que trata a curadoria de dados e todo o processo que a envolve.

O desenvolvimento desses assuntos corrobora com a visão Andrade (2023) que afirma que a curadoria de dados de pesquisa envolve atividades que inclui, também, a gestão dos conjuntos de dados, podendo assegurar que sejam pesquisados e interpretados, uma vez que o foco principal da curadoria é gerir todo o ciclo de vida dos dados para que os mesmos possam ser compartilhados, e permaneçam acessíveis e reutilizáveis, para quem deles precisarem.

As Universidades de Columbia, Harvard, Oxford e Cornell desempenham um compromisso forte com a curadoria de dados, implementando políticas, oferecendo serviços e ferramentas, e promovendo práticas que asseguram que os dados de pesquisa sejam bem gerenciados, documentados e acessíveis. Isso não só facilita a reutilização dos dados, mas também contribui para a colaboração e o avanço da ciência.

#### **4.2.6 Dados pessoais**

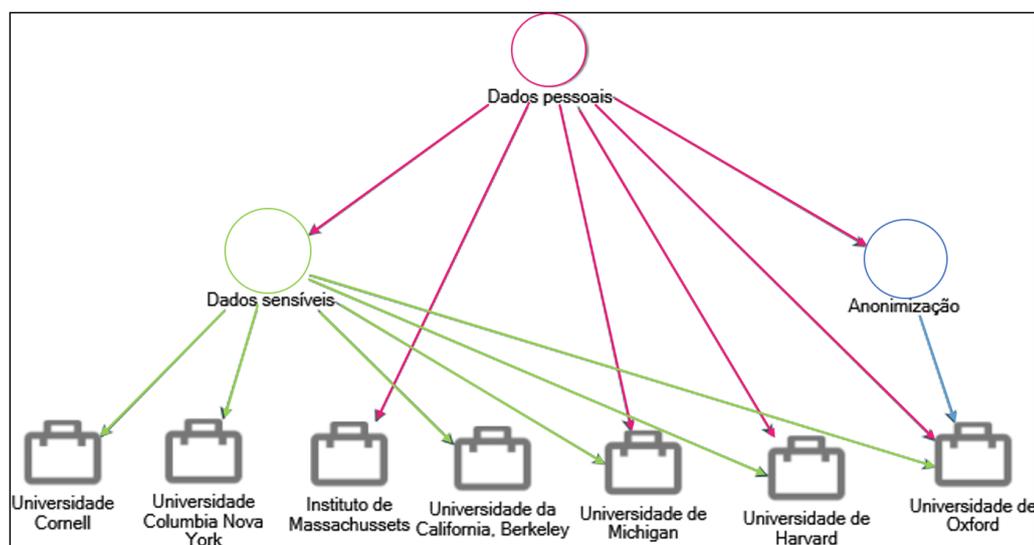
Outro assunto que as universidades desenvolvem em suas práticas de competência em dados, é a respeito dos Dados pessoais, Dados sensíveis e

Anonimização dos dados. Na Figura 21 apresentam-se as universidades nas quais identificou-se o Código “Dados pessoais” durante a análise dos documentos selecionados em cada uma das universidades da amostra deste estudo.

Devido às especificidades tratadas nos documentos analisados o Código “Dados pessoais” foi subdividido nos Subcódigos: “Dados sensíveis” e “Anonimização”. Nesse caso, ao construir as codificações entendeu-se que Dados sensíveis e Anonimização estão relacionados com o Código “Dados pessoais” ao focar no quesito de proteção da privacidade dos indivíduos e garantir a segurança das informações no que se refere às determinações de proteção de dados nas pesquisas científicas.

A Figura 21 mostra que algumas universidades abordam, em suas atividades de competência em dados, assuntos apenas sobre Dados pessoais, de forma mais genérica, outras sobre Dados pessoais e Dados sensíveis, e outras instituições desenvolvem um conteúdo mais abrangente, capacitando os pesquisadores sobre a importância e necessidade da Anonimização de dados quando trata-se de dados pessoais e dados sensíveis.

**Figura 21** - Universidades onde o Código “Dados pessoais” foi identificado



**Fonte:** dados da pesquisa.

As Universidades de Michigan, Oxford, Harvard e o Instituto de Massachussets, abordaram em suas práticas de competência em dados questões importantes sobre dados pessoais, principalmente quando envolve a coleta e o

tratamento de informações que podem ser utilizadas para identificar indivíduos, e que, portanto, devem seguir regras para a proteção desses dados.

Cada universidade fornece aos pesquisadores informações para ajudá-los a garantir que, ao lidarem com informações pessoais, elas estejam em conformidade com as orientações e políticas de sua respectiva instituição, e atendendo, sempre, aos requisitos legais sobre proteção de dados.

Nos documentos da universidade de Harvard, por exemplo, identificaram-se exemplos práticos sobre como gerenciar dados pessoais. Por serem dados que exigem um nível mais alto de proteção, levando em conta que qualquer informação da qual a perda, alteração, uso indevido ou divulgação pode afetar uma organização ou indivíduo, é necessário conhecer as regulamentações que impactam no desenvolvimento de pesquisas que utilizam dados pessoais. Por esse motivo, nas instruções para os pesquisadores os documentos citam a Lei de Portabilidade e Responsabilidade de Seguro de Saúde (*Health Insurance Portability and Accountability Act*, HIPAA), um conjunto de normas que organizações de saúde norte-americanas devem cumprir para proteger as informações dos indivíduos, e a Lei de Direitos Educacionais e Privacidade da Família (*Family Educational Rights and Privacy Act*, FERPA). Os pesquisadores de Harvard são alertados que ao receber informações ou dados de uma entidade coberta pela HIPAA ou FERPA devem estar cientes da Regra de Privacidade no que se refere à pesquisa.

#### 4.2.6.1 Dados sensíveis

Acerca de Dados sensíveis as Universidade de Cornell, Columbia e Califórnia trouxeram considerações, destacando o gerenciamento e compartilhamento desses dados, bem como diretrizes de segurança e as políticas de suas respectivas instituições. Enfatizaram que a diferença entre a gestão de dados sensíveis para os dados de pesquisa “comuns” está nas medidas de segurança e privacidade aplicadas para proteger os dados sensíveis devido ao seu potencial de causar danos se compartilhados de maneira incorreta.

Sobre isso, a Universidade de Colúmbia, NY, por exemplo, cita em seus documentos algumas políticas institucionais sobre dados sensíveis, como a Carta de

Segurança da Informação (*Information Security Charter*<sup>44</sup>), na qual especifica os deveres e responsabilidades de qualquer pessoa que use, armazene ou compartilhe dados da universidade, e também a Política de Classificação de Dados (*Data Classification Policy*<sup>45</sup>) que determina quais dados da universidade precisam ser controlados e protegidos.

As Universidade de Michigan, Harvard e Oxford são as três instituições que, nos documentos analisados, identificou-se informações sobre Dados sensíveis, e também, Dados pessoais, evidenciando o compromisso com a conscientização e qualificação de sua comunidade acadêmica com as melhores práticas no uso e tratamento desses dados.

A Universidade de Michigan, por exemplo, em uma oficina sobre “Diretrizes de segurança de dados”, destacou o trabalho do Conselho de Revisão Institucional (*Institutional Review Board - IRB*)<sup>46</sup>, um comitê independente com mandato federal responsável pela supervisão da proteção de pessoas em pesquisas conduzidas por unidades universitárias e afiliadas. Na UM qualquer pesquisador que planeje um estudo de pesquisa envolvendo seres humanos, logo, colete dados que se qualifiquem como sensíveis e pessoais, deve enviar um pedido de revisão e aprovação ao IRB antes de iniciar qualquer interação com os sujeitos ou com seus dados identificáveis. Para essa solicitação é disponibilizado um Modelo de gerenciamento e segurança de dados (*Data Management and Security Template*<sup>47</sup>), onde o pesquisador deve descrever claramente como os dados serão gerenciados, armazenados e protegidos.

#### 4.2.6.2 Anonimização

O último Subcódigo identificado nos documentos relacionado à temática Dados pessoais, sua segurança e privacidade, foi “Anonimização”. Conforme apresenta-se na Figura 21, a Universidade de Oxford foi a única que aborda esse tema específico quando trata a respeito de Dados pessoais e Dados sensíveis.

---

<sup>44</sup> <https://universitypolicies.columbia.edu/content/information-security-charter>

<sup>45</sup> <https://universitypolicies.columbia.edu/content/data-classification-policy>

<sup>46</sup> <https://orsp.umich.edu/glossary/irb-institutional-review-board>

<sup>47</sup> [https://research-compliance.umich.edu/sites/default/files/resource-download/data\\_management\\_and\\_security\\_template\\_6.3.20.docx](https://research-compliance.umich.edu/sites/default/files/resource-download/data_management_and_security_template_6.3.20.docx)

Anonimizar dados trata da ação de transformar os dados pessoais ou sensíveis, em dados que não podem ser reassociados a um indivíduo. Esse aspecto foi tratado nos documentos de Oxford, principalmente em um contexto onde o uso da anonimização dos conjuntos de dados pode ser benéfico para preservação e compartilhamento. Os instrutores explicam sobre a possibilidade de remover informações pessoais de um conjunto de dados no final da pesquisa para que uma versão anonimizada dos dados possa ser preservada. No entanto, esclarecem que a exclusão de alguns identificadores (como nomes, endereços de *e-mail*, entre outros) podem não ser suficientes para tornar totalmente anônimo um conjunto de dados: ainda pode ser possível deduzir a identidade de alguém combinando outras informações.

Discutir a respeito dos dados pessoais, sensíveis e anonimizados em ambiente acadêmico é de suma importância, pois enquanto produtor e consumidor de dados, o pesquisador deve ter consciência das leis de proteção aos dados, como a LGPD no Brasil, e até mesmo estratégias para anonimizar os dados visando a proteção dos direitos dos indivíduos, e a possibilidade de compartilhamento e reuso desses dados por outros pesquisadores, colaborando, assim, com o desenvolvimento da pesquisa e produção de novos conhecimentos.

Ao analisar os documentos de cada instituição e realizar a codificações do conteúdo verificou-se a ocorrência de seis Referências de codificação para “Dados pessoais”, em um total de 100 documentos analisados. O subcódigo “Dados sensíveis” foi referenciado treze vezes.

A respeito do assunto “Dados sensíveis” observou-se que as instituições, principalmente a Universidade de Harvard, por desenvolver serviços de dados específicos para pesquisadores da área médica, salientaram que os dados sensíveis necessitam de medidas de proteção adicionais aos dados pessoais, devido a possibilidade de causar maiores malefícios às pessoas, principalmente em caso de pesquisas médicas.

Por fim, o subcódigo “Anonimização” apresentou três referências de codificações nos 11 documentos analisados e pertencentes à Universidade de Oxford. Quando trata dos dados pessoais e sensíveis destaca que esses três elementos estão interligados possibilitando a gestão dos dados de maneira segura e ética.

As universidades que pretendem tornar sua comunidade competente em dados abordam o assunto dos dados pessoais por compreender que faz parte do desenvolvimento do indivíduo enquanto pesquisador pensar a respeito da transparência na pesquisa científica e sobre a responsabilidade de preservar e proteger essas informações usando-as apenas para os propósitos especificados e consentidos pelo titular dos dados.

Estar cientes a respeito da problemática que a privacidade de dados envolve, deve ser levando em consideração, principalmente em estudos da área da saúde. Por isso, estabelecer práticas éticas em matéria de dados e desenvolver os meios técnicos de proteção dos dados sensíveis, por exemplo, são bases fundamentais para enfrentar os desafios que envolvem a privacidade e proteção desses dados na esfera acadêmica (Conceição; Nhacuongue, 2023).

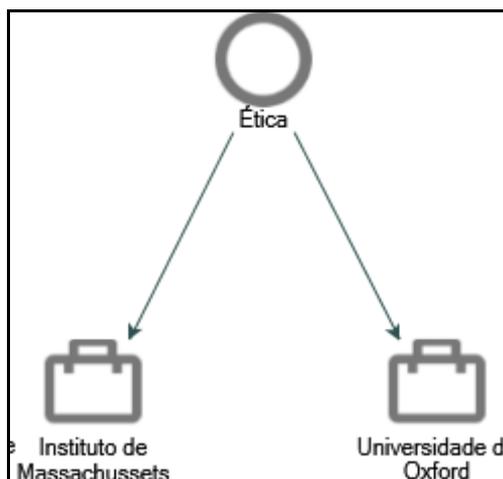
Outro aspecto importante que está alinhado com os dados pessoais diz respeito à ética na pesquisa científica, e também, sobre o uso ético dos dados, como será visto a seguir

#### **4.2.7 Ética**

Como mencionado por Conceição e Nhacuongue (2023), as práticas éticas são fundamentais no desenvolvimento de pesquisa científica, principalmente ao lidar com dados pessoais, sensíveis e anonimizados. Alguns pesquisadores precisam ponderar considerações éticas ao escrever propostas para financiamento de projetos de pesquisa, ou terão que lidar com dilemas éticos que surgem durante o desenvolvimento do estudo.

Ao analisar os documentos das universidades selecionadas e que são utilizados nas atividades de competência em dados, constatou-se que o Instituto de Massachusetts e a Universidade de Oxford foram as únicas instituições que abordaram o assunto, como ilustra a Figura 22.

**Figura 22** - Universidades onde o Código “Ética” foi identificado



**Fonte:** dados da pesquisa.

O Instituto de Massachusetts trouxe em seus documentos algumas considerações éticas para o pesquisador refletir sobre o controle de acesso aos dados sensíveis e confidenciais que podem ser preservados por meio de gerenciamento, ou restrição de acesso e utilização dos mesmos. É destacado que o controle de acesso deve ser proporcional ao tipo de dados e ao nível de confidencialidade envolvidos, e que ao planejar sobre os acessos desses dados, as seguintes observações devem ser feitas: Quem poderia acessar seus dados? O que eles podem fazer com isso? São necessárias restrições de uso específicas? e Por quanto tempo você deseja que os dados estejam disponíveis? Em alguns documentos identificou-se os deveres éticos que os pesquisadores têm em relação a suas pesquisas levando em conta coleta e manuseio desses dados, e o plano de longo prazo para lidar com eles.

A Universidade de Oxford também abordou em seus conteúdos a ética nas pesquisas científicas. Nos documentos analisados da Universidade de Oxford, observou-se que os conteúdos foram mais desenvolvidos e detalhados do que no MIT. A *Research Data Oxford*, por exemplo, apresentou a Política de Ética em Pesquisa, na qual exige que todos os estudos que envolvam dados pessoais e sensíveis sejam submetidas à revisão ética apropriada, visando respeitar a dignidade, os direitos e o bem-estar dos participantes, e minimizar os riscos não só para eles, mas também para os pesquisadores, terceiros e para a própria Universidade. Os pesquisadores são orientados a respeito da Preservação de dados

pós-projeto, principalmente nos casos onde os dados pessoais muitas vezes não são adequados para compartilhamento e para reutilização, mas é possível fornecer acesso restrito ou partilhar uma versão anonimizada ou outra forma editada dos dados. Nesses casos a Universidade indica a necessidade de pensar sobre a preservação desses dados.

Enquanto analisava-se os materiais das duas instituições percebeu-se que as questões éticas são essenciais principalmente quando se trata de instituições que trabalham com dados sensíveis e pessoais, levando em consideração a privacidade, segurança, transparência e responsabilidade do pesquisador em gerenciar e proteger esses dados. No contexto em que a coleta e produção de dados, aliada à crescente preocupação com a privacidade e a regulamentação de dados confidenciais e restritos, abordar esses assuntos com os indivíduos que estão inseridos na prática de pesquisas científicas e torna-los competentes para tomar decisões à respeito, é indispensável para o futuro das pesquisas científicas.

Observou-se sete Referências de codificação com o Código “Ética” em um total de 35 documentos do MIT e de Oxford. Por meio desse procedimento, foi possível identificar quais aspectos éticos as duas instituições têm desenvolvido em suas práticas de competência em dados.

Outrossim, o fato de não serem identificados documentos tratando das questões éticas sobre dados de pesquisa, nas demais universidades deste estudo, não significa que não seja levado em consideração o desenvolvimento ético das pesquisas e uso éticos dos dados. Pode ser que no momento desta pesquisa não tenham sido disponibilizados documentos ou atividades a respeito ou ainda que, esse tópico é abordado e desenvolvido de outra maneira, como por exemplo, em atendimentos ou consultorias individuais, onde as dúvidas do pesquisador relacionada à pesquisa que envolva dados sensíveis e pessoais, é tratado com o consultor e/ou instrutor responsável por esses serviços.

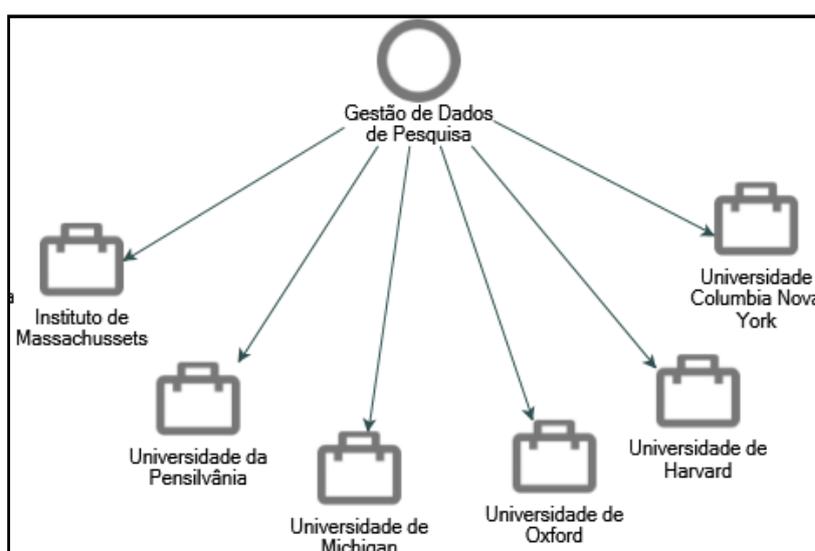
#### **4.2.8 Gestão de Dados de Pesquisa**

Seguindo a lista de codificação, a próxima figura apresenta as universidades que abordaram conteúdos sobre Gestão de Dados de Pesquisa.

A Gestão de Dados de Pesquisa diz respeito à organização dos dados desde o início do ciclo da pesquisa, passando pela divulgação e arquivamento dos

resultados. São práticas que garantem que os dados de pesquisa sejam coletados de maneira correta, organizados, tratados, armazenados, preservados e compartilhados, visando a qualidade e a integridade. A Figura 23 demonstra que o Instituto de Massachusetts, as Universidades de Pensilvânia, Michigan, Oxford, Harvard e Columbia trouxeram discussões sobre o assunto, destacando práticas para o pesquisador gerenciar os dados de suas pesquisas, bem como os benefícios para os dados.

**Figura 23** - Universidades onde o código “Gestão de dados de pesquisa” foi identificado



**Fonte:** dados da pesquisa.

Durante a análise, observou-se que o conteúdo dos documentos das seis universidades sobre o Código “Gestão de dados de pesquisa” era semelhante. Embora de maneira objetiva e simplificada, essas instituições apresentaram algumas práticas para o indivíduo começar a gestão dos seus dados de pesquisa ao iniciar um projeto e/ou pesquisa até chegar a sua conclusão.

Ao iniciar um projeto e/ou pesquisa é necessário: a) determinar as necessidades e responsabilidades da gestão de dados (definir quais serão os dados que precisarão ser gerenciados); b) armazenar e compartilhar seu trabalho durante o projeto (determinar um sistema ativo de armazenamento e compartilhamento, e configurar um sistema de *backup*); e c) documentar o projeto (determinar o que registrar sobre os dados (metadados) e como capturá-los).

Já ao encerrar o projeto, é recomendado aos pesquisadores: a) determinar o compartilhamento e as restrições de dados pós-projeto (quais dados devem ser compartilhados abertamente e como os dados confidenciais serão tratados); b) armazenar e compartilhar o trabalho a longo prazo; documentar o projeto; e c) gerenciar os registros.

Além dessas práticas mencionadas pelas instituições, elas também enfatizaram os principais benefícios de uma boa gestão de dados, uma vez que:

- a) Mantém a integridade dos dados - documentar e gerenciar adequadamente dados aumenta a reprodutibilidade do trabalho e, como resultado, aumenta a validade dos seus resultados.
- b) Melhora o impacto da pesquisa - manter dados acessíveis e confiáveis permite que o pesquisador compartilhe prontamente seus conjuntos de dados brutos e podendo melhorar o impacto da sua pesquisa, aumentando sua “relevância”.
- c) Garante a longevidade dos dados a longo prazo - a preservação adequada dos dados em um repositório de dados torna-os acessíveis e detectáveis durante anos; preservar a “contribuição de pesquisa” para a comunidade científica.
- d) Permite atender aos requisitos de financiamento das agências de fomento à pesquisa - a maioria das agências de financiamento, incluindo o NIH, exige o gerenciamento, documentação e compartilhamento adequado dos dados
- e) Permite atender aos requisitos para publicações em periódicos - muitos periódicos exigem que os artigos publicados sejam acompanhados dos dados brutos da pesquisa subjacentes.

Diante dos tópicos elencados pelas instituições, vê-se que há um reconhecimento crescente da importância de uma boa gestão de dados como um elemento essencial da pesquisa, auxiliando no desenvolvimento do estudo e influenciando em várias etapas e aspectos do processo da pesquisa, pois a medida que os pesquisadores e as instituições discutem e reconhecem essa prática, mais habilitados com a gestão dos dados a comunidade acadêmica será, melhorando a qualidade das pesquisas e a reprodutibilidade.

A Universidade de Harvard, diferente das demais, apresentou em seus conteúdos outras perspectivas sobre a gestão de dados. Além de discutir os benefícios e práticas iniciais de gestão de dados, a instituição ofertou oficinas e cursos específicos sobre esse assunto, organizados pela Biblioteca Countway<sup>48</sup> – da área médica de Harvard, abordando diversos tópicos relacionados à gestão de dados.

Na oficina “O que acontece SEM gerenciamento de dados?” colocam que não optar pela gestão de dados ao iniciar um projeto de pesquisa leva o pesquisador a:

- a) perder dados devido a falha e corrupção de arquivos por não serem armazenados, organizados, e nomeados corretamente;
- b) enfrentar inconsistência nos dados uma vez que dados coletados sem padrões ou normas podem levar a dificuldades para analisar e interpretar;
- c) enfrentar dificuldade para rastrear os dados tornando difícil localizá-los até sua origem e produzir novos estudos; e
- d) arcar com problemas de segurança e privacidade principalmente quando trata-se de dados sensíveis que envolvem regulamentações de proteção aos dados.

Identificou-se conteúdos voltados à gestão de dados clínicos, àqueles coletados durante atendimento ao paciente ou como parte de um programa de ensaio clínico. Nestes documentos os instrutores destacaram as medidas iniciais para anonimizar os dados (levando em consideração o tipo do estudo produzido), e orientando-os quanto à coleta, armazenamento, análise e compartilhamento de informações de saúde de forma segura.

Durante a análise, observou-se que a Universidade de Harvard disponibilizou mais conteúdos a respeito do assunto, desenvolvendo, conseqüentemente, mais atividades sob variadas óticas da gestão de dados.

Em relação a referência de codificação nos documentos, isto é, quantas vezes o código “Gestão de Dados de Pesquisa” foi referenciado em todos documentos recuperados, Harvard se destaca, como demonstra o Gráfico 5, em relação às demais instituições. Isso porque, além de ser a Universidade onde recuperou-se maior quantidade de documentos para a análise (como mostra no Gráfico 1), reflete, também, o comprometimento com a gestão de dados em seus

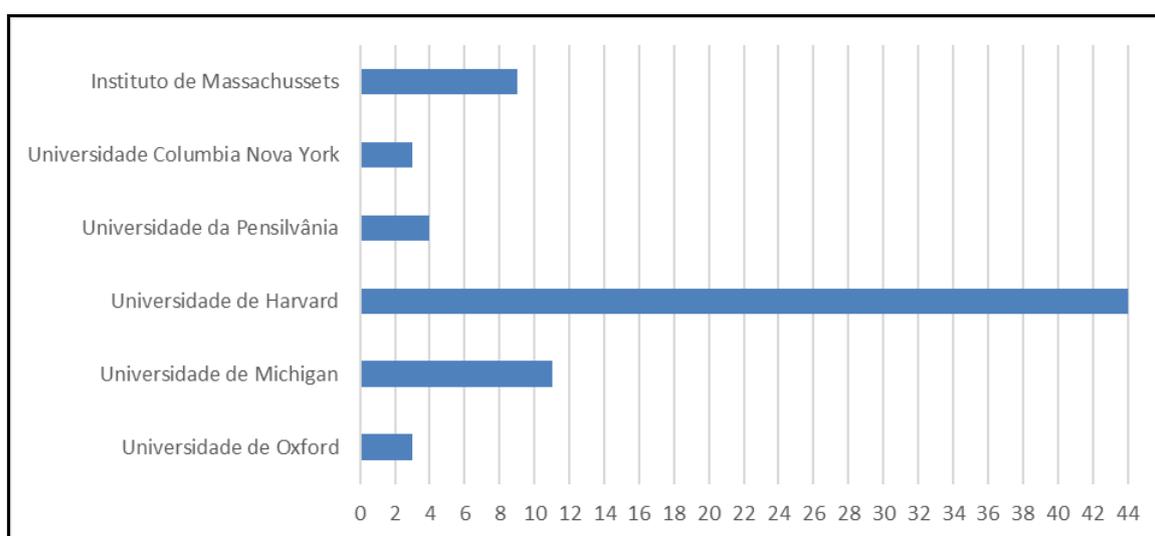
---

<sup>48</sup> <https://countway.harvard.edu/>

serviços. A Universidade de Harvard investe na capacitação de sua comunidade acadêmica, assim como no desenvolvimento de boas práticas, resultando em mais Referências de codificação.

Destaca-se que a Universidade de Michigan e o MIT, com onze e nove Referências de codificações respectivamente, também promovem, por meio de seus serviços de dados, discussões e reflexões sobre a gestão de dados de pesquisa, no entanto, de maneira genérica e objetiva, como mencionado no parágrafo anterior.

**Gráfico 5** - Quantidade de Referências para o Código Gestão de Dados de Pesquisa, nas universidades



**Fonte:** dados da pesquisa.

Cada instituição discute o tema conforme a demanda e necessidades específicas de sua comunidade interna - constituída por docentes, discentes e servidores técnicos-administrativos - muitas vezes estendendo-as também à comunidade local. As seis instituições que possuem atividades de competência em dados sobre gestão de dados de pesquisa, enfatizaram que a gestão de dados envolve uma série de atividades gerenciais práticas associadas ao ciclo de vida dos dados de pesquisa, desde a criação e/ou coleta dos dados, armazenamento, segurança, preservação, recuperação, compartilhamento e reutilização. No nível institucional, as universidades estão descobrindo como a gestão de dados de pesquisa deve ser apoiado, seja em termos de consultoria e atividades de

capacitação, ou uma infraestrutura para armazenamento, compartilhamento e curadoria (Cox; Pinfield, 2014).

Por tratar-se de um assunto que compreende a organização dos dados, desde sua entrada no ciclo da pesquisa até a disseminação e arquivamento de resultados valiosos, esperava-se que as demais universidades também estivessem presentes na Figura 23 abordando em suas práticas a importância da gestão de dados, bem como os benefícios. No entanto, observou-se que apesar das instituições não disponibilizarem documentos que tratam especificamente sobre o tema, elas abordaram questões referentes à gestão de dados de pesquisa em materiais que tratam de outros conteúdos como por exemplo: na documentação sobre compartilhamento de dados, documentação e armazenamento e preservação. As universidades apresentadas na Figura 23, trazem para sua comunidade acadêmica esclarecimento sobre o processo que envolve a gestão de dados de uma maneira mais específica, uma vez que em seus documentos tratam exclusivamente sobre a gestão de dados e todas suas etapas.

À medida que eram realizadas as análises e codificação dos materiais sobre as práticas desenvolvidas pelas instituições destacadas na Figura 23 percebeu-se que as bibliotecas novamente se destacam na organização e desenvolvimento do serviço de gestão de dados, e os bibliotecários se responsabilizaram pelas atividades. Esses profissionais têm buscado novas funções para auxiliar e colaborar com pesquisadores, e o serviço de gestão de dados é uma das atividades que podem ser oferecidas pelas bibliotecas.

Surkis e Read (2015), a respeito da atuação do profissional bibliotecário no cenário de “*Big Data*”, afirmam que cada vez mais os bibliotecários têm se motivado a oferecer serviços voltados à gestão de dados de pesquisa, que incluem a elaboração de guias e manuais sobre o tema, a realização de novas práticas de gestão de dados e a prestação de consultorias a agências de fomento à pesquisa e editores de publicações científicas. Brochu e Burns (2019) complementam ainda que os bibliotecários são os principais interessados na gestão de dados de pesquisa como apoiadores e educadores devido ao seu papel tradicional na coleta, organização, armazenamento, preservação e facilitação do livre acesso à informação.

Conforme a comunidade científica está produzindo, manipulando ou analisando dados e objetos digitais como parte de suas pesquisas o aumento na produção e utilização de dados os levam a identificar a necessidade de descrever os seus dados, a fim de torná-los localizáveis nos repositórios, compreensíveis no contexto e reutilizáveis por outros pesquisadores. Nesse sentido os metadados se tornam valiosos quando bem descritos, oferecendo sentido a um conjunto de dados que possibilitam sua recuperação. A Figura 24 apresenta as instituições que em seus documentos discutiram sobre o assunto.

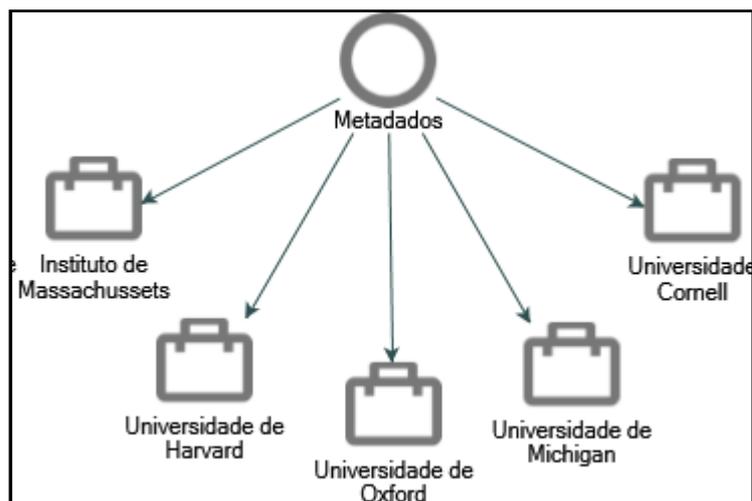
#### **4.2.9 Metadados**

Metadados são informações estruturadas que fornecem contexto para objetos de informação de todos os tipos, incluindo dados de pesquisa, e ao fazê-lo permitem o uso, preservação e reutilização desses objetos. Tenopir *et al.* (2011) em seu estudo sobre práticas de gestão de dados, concluiu que a comunidade científica, até então, possuía um déficit de consciência sobre a importância dos metadados (no que refere-se à prática), recomendando, então, que as instituições e os indivíduos que nelas atuavam com pesquisas, devem organizar atividades que auxiliam a comunidade acadêmica (principalmente pesquisadores) a preparar os metadados necessários para permitir a descoberta, preservação e reutilização dos seus dados.

A literatura científica ainda continua a destacar a importância dos metadados para a pesquisa científica abordando aspectos como: práticas e tendências atuais de metadados para dados de pesquisa (Farmel; Shiri, 2014), metadados como ferramenta para gerenciamento de dados de pesquisa (Rocha; Sales; Sayão, 2017), metadados e o impacto nas pesquisas (Conrad; Urberg, 2021); e metadados e sua importância para a pesquisa (Caliskan; Dangwal; Dandekar, 2023).

Nesse contexto de necessidades de atividades que auxiliem a comunidade a refletir a importância das práticas de descrição dos metadados, a Figura 24 exhibe as universidades que em suas atividades de competência em dados abordam e discutem sobre o papel central dos metadados nas pesquisas científicas.

**Figura 24** - Universidades onde o código “Metadados” foi identificado



**Fonte:** dados da pesquisa.

Durante a análise, observou-se que o Instituto de Massachusetts, as Universidades de Harvard, Oxford, Michigan e Cornell desenvolveram conteúdos em oficinas e cursos a respeito da importância das práticas de gestão de dados de pesquisa, incluindo a criação e utilização de metadados. Aspectos do tipo: “o que são metadados e por que são importantes?”, “o que documentar sobre seus dados”, “exemplos de padrões de metadados”, “tipos de metadados” e “metadados e descrição de dados” foram os assuntos abordados com o objetivo de tornar o seu público competente em metadados e ressaltando a importância da descrição correta dos mesmos.

O MIT trouxe em seus documentos a importância dos metadados: a definição, por que são tão importantes e como começar a usá-los. Esclarecem que os metadados permitem encontrar dados com mais facilidade e precisão, e facilitar o compartilhamento (tanto com colaboradores quanto com pesquisadores que reutilizarão os dados). O Instituto de Massachusetts também levou os indivíduos a refletir sobre a descrição de metadados por meio de perguntas e respostas com exemplos práticos de como lidar e se posicionar diante de situações reais, como: “ao fornecer dados a outra pessoa, que tipos de informações você gostaria de incluir nos dados?”, as respostas foram compiladas para auxiliar o entendimento do público sobre a atividade de descrição dos metadados. Além disso, foram elencados os três

principais tipos de metadados: descritivo, estrutural, e administrativo, destacando seus propósitos e aplicação.

A Universidade de Oxford e a Universidade de Cornell abordaram tópicos semelhantes em suas atividades de competência em dados. As instituições orientaram em seus serviços de dados que descrever e documentar dados corretamente permite que os pesquisadores, principalmente, entendam e rastreiem detalhes importantes do trabalho, e que ter metadados sobre os dados também facilita a pesquisa e a recuperação dos dados quando depositados em um repositório de dados, por exemplo. Também, frisam a importância de descrever os dados com detalhes suficientes para que outras pessoas possam avaliar se os dados são adequados para a finalidade que desejam; permitir a análise e reutilização adequada; possibilitar avaliar a qualidade dos dados; e compreender os dados, uma vez que a etapa de descrição dos metadados tem por objetivo fornecer informações detalhadas sobre o conteúdo, estrutura e contexto dos dados.

Além disso, nos documentos analisados, mostra-se que é necessário descrever algumas facetas dos dados que incluem: informações bibliográficas gerais sobre o conjunto de dados (título, autor, publicações relacionadas); tipos de arquivos usados (csv, txt, png, etc.); e informações descritivas importantes sobre o experimento (métodos de amostragem ou medição, *software* usado para análise, qualquer processamento ou transformação realizada).

Além disso, as instituições apresentaram os esquemas de metadados da *UK Data Service*<sup>49</sup>, do *Digital Curation Centre (DCC)*<sup>50</sup>, e do *FAIRsharing*<sup>51</sup> que fornecem uma ampla gama de metadados para dados de diferentes disciplinas, visando formalizar as informações necessárias para tornar um determinado tipo de dados o mais reutilizável possível.

A Universidade de Michigan oferece uma introdução básica a ferramentas, recursos, padrões e suporte para metadados e documentação de dados. Aborda o conceito teórico sobre metadados, destacando que os metadados descrevem o conteúdo, qualidade, condição e outras características dos dados, padronizando e estruturando informações que facilitam funções associadas a dados, como: organizar e gerenciar dados, preservar dados a longo prazo, garantir que os dados possam ser

---

<sup>49</sup> [Metadados — UK Data Service](#)

<sup>50</sup> [Metadados Disciplinares | DCC](#)

<sup>51</sup> [FAIRsharing | Standards](#)

indexados e descobertos em um repositório de dados, manter o contexto em torno do qual os dados foram capturados ou criados, o que é vital para facilitar a compreensão e a reutilização dos dados por outros pesquisadores.

Também foram citados pela Universidade de Michigan alguns exemplos de padrões de metadados que podem vir a auxiliar na descrição de recursos de maneira consistente, garantindo que os dados sejam compreensíveis e utilizáveis tanto no presente quanto no futuro. Os padrões listados por Michigan foram: *Dublin Core*<sup>52</sup> (um esquema amplamente utilizado e com ampla aplicabilidade); Núcleo VRA Core<sup>53</sup> (que fornece padrões que descrevem obras visuais, obras de arte e cultura); CONA - Autoridade de Nomeação de Objetos Culturais<sup>54</sup>; e Getty Thesaurus de nomes geográficos<sup>55</sup>. Além disso, apresentam algumas ontologias para ajudar a determinar um vocabulário consistente para os dados, como: o *Thesaurus* de Arte e Arquitetura<sup>56</sup>; o Serviço de Dados Vinculados da Biblioteca do Congresso<sup>57</sup>; e o Banco de dados de ontologia vegetal<sup>58</sup>.

Por fim, a Universidade de Harvard a respeito dos metadados e sua importância, assim como as demais instituições, destacaram exemplos de padrões de metadados, a importância de o indivíduo descrever corretamente os metadados, bem como os benefícios a curto e longo prazo. No entanto, por meio do Grupo de trabalho de gerenciamento de dados de pesquisa da *Longwood Medical Area*, Harvard enfatiza questões sobre a utilização dos Metadados Biomédicos uma vez que eles possuem características particulares que diferenciam de outros tipos de metadados em virtude de sua natureza e do contexto que são utilizados. Conforme as informações disponíveis nos materiais analisados as informações necessárias para capturar dados biomédicos incluem os seguintes metadados:

- a) **Metadados do reagente:** Informações sobre as amostras clínicas, reagentes biológicos (linhagens celulares, anticorpos, siRNAs), reagentes químicos (medicamentos), etc. usados para gerar os dados.
- b) **Metadados técnicos:** Informações geradas automaticamente por instrumentos de pesquisa e *softwares* associados.

---

<sup>52</sup> <https://www.dublincore.org/specifications/dublin-core/dces/>

<sup>53</sup> <https://www.loc.gov/standards/vracore/>

<sup>54</sup> <https://www.getty.edu/research/tools/vocabularies/cona/about.html>

<sup>55</sup> <https://www.getty.edu/research/tools/vocabularies/tgn/index.html>

<sup>56</sup> <https://www.getty.edu/research/tools/vocabularies/aat/about.html>

<sup>57</sup> <https://id.loc.gov/>

<sup>58</sup> <https://planteome.org/>

- c) **Metadados experimentais:** Informações sobre as condições experimentais (tipo de ensaio, pontos de tempo), o protocolo experimental e o equipamento usado para gerar os dados.
- d) **Metadados analíticos:** informações sobre métodos de análise de dados, incluindo nome e versão do software, parâmetros de controle de qualidade e detalhes do tipo de arquivo de saída.
- e) **Metadados em nível de conjunto de dados:** Informações sobre os objetivos do projeto de pesquisa, pesquisadores participantes, publicações relevantes e fontes de financiamento.

Ademais, apresentam padrões de metadados médicos que auxiliam a definir os elementos específicos de metadados a serem utilizados para descrever e organizar um conjunto de dados biomédicos. Os padrões elencados por Harvard foram: Darwin Core (dados biológicos)<sup>59</sup>; MIBBI (Informações Mínimas para Investigadores Biológicos)<sup>60</sup>; e Definições de elementos de dados do protocolo (dados de ensaios clínicos)<sup>61</sup>. Alguns exemplos de ontologias e vocabulário controlados usados em pesquisas biomédicas também foram apresentados nas atividades de competência em dados da Universidade de Harvard.

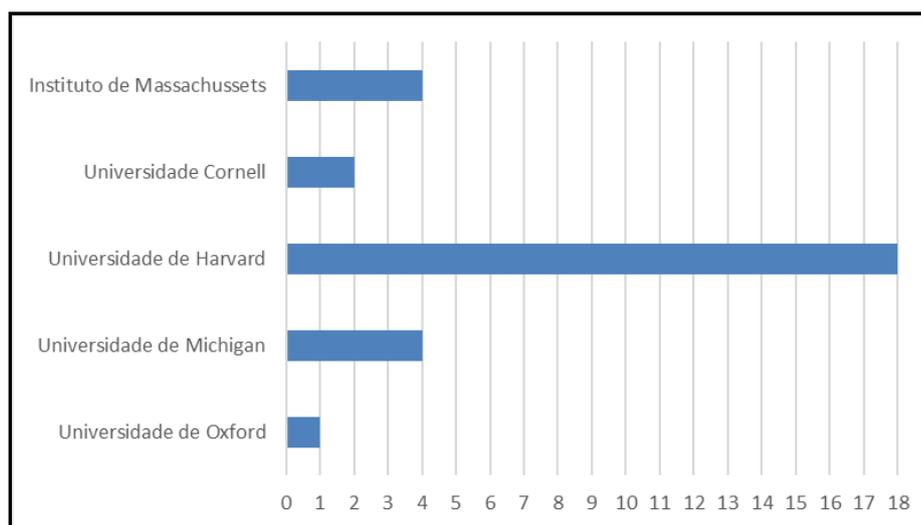
No que diz respeito às Referências de codificação nos documentos, isto é, quantas vezes o Código “Metadados” foi referenciado em todos documentos analisados, Harvard novamente se destaca, como demonstra o Gráfico 6, seguido do MIT e da Universidade de Michigan. Harvard além de abordar os conteúdos teóricos e práticos como as demais instituições, trouxe também aspectos de metadados biomédicos uma vez que sua comunidade de pesquisadores e alunos da área da saúde tem desenvolvido atividades contínuas relacionadas a gestão de dados médicos, biomédicos e clínicos visando a capacitação diante dos desafios da gestão e utilização desses dados.

---

<sup>59</sup> [Núcleo Darwin - TDWG](#)

<sup>60</sup> <https://mibbi.sourceforge.net/foundry.shtml>

<sup>61</sup> [Protocol Registration Data Element Definitions for Interventional and Observational Studies | ClinicalTrials.gov](#)

**Gráfico 6** - Quantidade de Referências para o Código Metadados, nas universidades

**Fonte:** dados da pesquisa.

Observou-se na análise dos conteúdos oriundos das atividades de competência em dados que as instituições citadas na Figura 24 percebem a necessidade da comunidade acadêmica se capacitar e considerar o contexto de como seus dados são criados, analisados e armazenados, e como serão recuperados e preservados ao longo do tempo, destacando como as práticas de descrição de metadados contribuem para esse feito. A variedade de disciplinas envolvidas e a grande quantidade de dados produzidos demanda uma abordagem específica em relação às práticas de descrição dos metadados. A identificação de elementos e práticas necessárias para descrição de metadados, conforme desenvolvidos pelas instituições desse estudo, contribui com o desenvolvimento de um conjunto de melhores práticas para implementação de capacitações a fim de preparar o caminho para o avanço de um ambiente de dados de pesquisa interoperável.

Diferentes áreas do conhecimento possuem necessidades típicas de descrição para seus dados, como observou-se nos materiais da Universidade de Harvard que abordou aspectos de metadados biomédicos. Os dados não falam por si mesmos e não têm seus conteúdos explícitos, necessitando, de ações que vão desde o planejamento no momento de sua criação, passam pela organização em coleções com referências estáveis e padronizadas e finalizam em um arquivamento de longo prazo dos dados de valor permanente. Diante disso, tem-se a necessidade de estabelecer sentido a esses dados por meio dos metadados (Rocha; Sales;

Sayão, 2014). Sem uma documentação associada aos dados, eles não passam de variáveis, cadeias numéricas, fragmentos de texto, áudio e vídeo. Isto é, os metadados permitem que os dados sejam categorizados com base nas áreas de estudo, temas e outros critérios que se julgem relevantes.

Resende (2019) coloca que os metadados são importantes para a pesquisa pois são recursos que descrevem contextos e conteúdos de arquivos de dados. Além disso, os metadados também fazem parte dos Planos de Gestão de Dados fornecendo descrição detalhada e documentação necessária para compreensão dos dados, pois os metadados tratam-se de elementos fundamentais que são detalhados dentro do PGD.

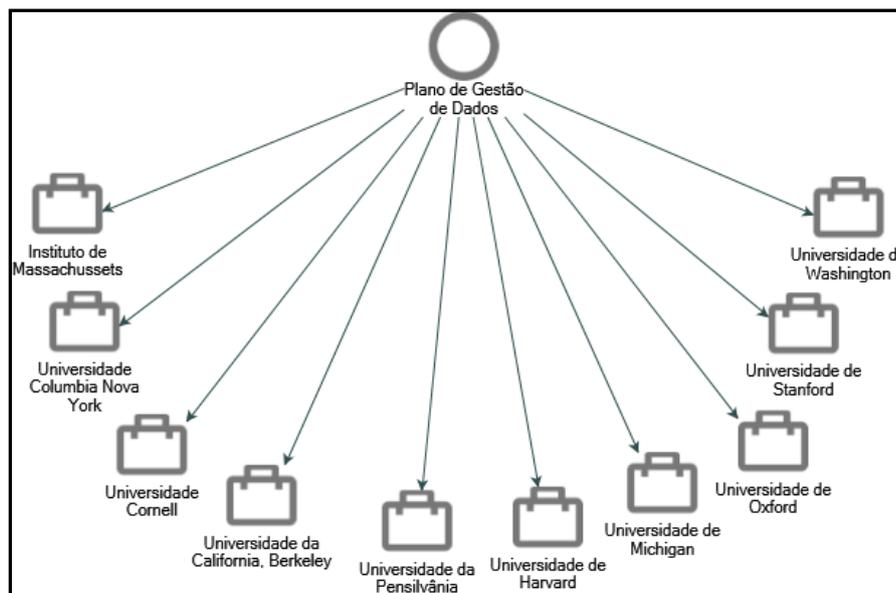
#### **4.2.10 Plano de Gestão de Dados**

Um Plano de Gestão de Dados é um documento elaborado no início do projeto de pesquisa onde são descritas as diretrizes para o ciclo de vida dos dados. Ou seja, trata-se de um documento que descreve como os dados serão tratados durante um projeto e o que acontecerá com eles após o término do projeto. (RDPdata, 2024). Na comunidade científica muitas agências de financiamento têm solicitado de seus pesquisadores um Plano de Gestão de Dados, como a NIH e a NSF no âmbito internacional, e a FAPESP no Brasil.

Neste estudo identificou-se que todas as instituições desenvolvem atividades sobre Plano de Gestão de Dados, como mostra a Figura 25. O Instituto de Massachusetts, as Universidades de Columbia - NY, Cornell, Califórnia - Berkley, Pensilvânia, Harvard, Michigan, Oxford, Stanford, e Washington abordaram em suas práticas de competência em dados aspectos do tipo: o que é um plano de gestão de dados, quais elementos se destacar na elaboração de um PGD, importância do PGD para a pesquisa, ferramentas que auxiliam na criação, e quais agências de financiamento exigem um PGD.

Conforme a análise do conteúdo disponibilizado pelas instituições, percebeu-se que todas trataram o tema com características e perspectivas muito semelhantes. A cada *slide* e informações disponíveis nos *sites* das bibliotecas e departamentos de serviços de dados, viu-se que descreviam as mesmas informações, mas de forma ou com recursos visuais diferentes

**Figura 25** - Universidades onde o código “Planos de Gestão de Dados” foi identificado



**Fonte:** dados da pesquisa.

Todos os pesquisadores das universidades, financiados ou não, são incentivados a criar um plano de gestão de dados como parte das boas práticas acadêmicas. Observou-se que as instituições apresentam conceitos teóricos sobre o PGD, destacando um plano abrange decisões iniciais de como os dados serão tratados durante a fase ativa da pesquisa e questões de preservação e compartilhamento a longo prazo.

Mesmo quando não for exigido por um órgão financiador, orientam os pesquisadores sobre a validade de criar um PGD, visto que o processo de planejamento é uma oportunidade de pensar no que é necessário para permitir que o projeto seja conduzido da maneira mais tranquila possível. Também, pode ajudar a antecipar possíveis problemas antes que estes ocorram, o que significa que as soluções podem ser encontradas em tempo útil.

Além disso, as instituições apontam os elementos que um PGD, sendo eles: qual os tipos de dados (fonte, formato, tamanho); como se dará o armazenamento, backup e a segurança; quais os métodos para divulgação; como os dados serão preservados ou descartados; qual documentação criada para os dados, e qual padrão de metadados será utilizado de acordo com a área de pesquisa; quais as proteções de privacidade e confidencialidade; como se dará o acesso e

compartilhamento dos dados; e por fim, de quem será a responsabilidade pelos dados.

Destaca-se que a Universidade de Cornell embora tenha abordado os mesmos aspectos citados referente aos elementos que contém um PGD, também, acrescenta considerações a respeito de áreas adicionais que os financiadores podem exigir que sejam explanados no PGD. Isso inclui: justificativa, monitoramento e relatórios, garantia específica de recursos para executar o Plano, fluxo de trabalho de processamento de dados (por exemplo, como planeja-se levar os dados do ponto de coleta ao ponto de acesso), garantia de qualidade de dados ou medidas de controle de qualidade, e para dados confidenciais, as medidas de segurança e quaisquer padrões formais que serão utilizados (por exemplo, permissão de agente biológico).

Outrossim, identificou-se que as Universidades e o Instituto de Massachusetts apresentaram ferramentas que auxiliam na escrita de um PDG. A ferramenta citada por todas as instituições foi a *DMPTool*<sup>62</sup>, uma ferramenta gratuita que ajuda na criação de planos de gestão de dados que atendam aos requisitos de diferentes financiadores. Notou-se que em relação a esse recurso, os instrutores mostram o passo a passo de como utilizá-la, desde o primeiro acesso com o *login*, até o preenchimento com as informações dos projetos. Os instrutores de Harvard, acrescentam que a universidade é parceira afiliada à instituição que gerencia a ferramenta, e por isso, os alunos e pesquisadores conseguem fazer *login* de usuário do *DMPTool* com o *HarvardKey*<sup>63</sup>, e recebem orientações específicas da instituição que os ajudam a concluir seu plano de gestão de dados.

Outras instituições como, a Universidade de Michigan, listam organizações que orientam a escrita do PGD. Foi destacada por Michigan as Diretrizes para planos eficazes de gestão de dados da *Inter-university Consortium for Political and Social Research (ICPSR)*<sup>64</sup> que fornecem orientações sobre a criação de tais planos, e a *Checklist for a Data Management Plan* do DCC<sup>65</sup> que sintetiza os requisitos para os planos de gestão de dados e as melhores práticas para a comunidade em geral.

---

<sup>62</sup> <https://dmptool.org/>

<sup>63</sup> sistema de autenticação de Harvard que permite aos estudantes, professores, funcionários, ex-alunos ..., acessar serviços e recursos online com um único login

<sup>64</sup> [Impacto da pesquisa de excelência em dados do ICPSR \(umich.edu\)](https://www.umich.edu/impacto-da-pesquisa-de-excelencia-em-dados-do-icpsr)

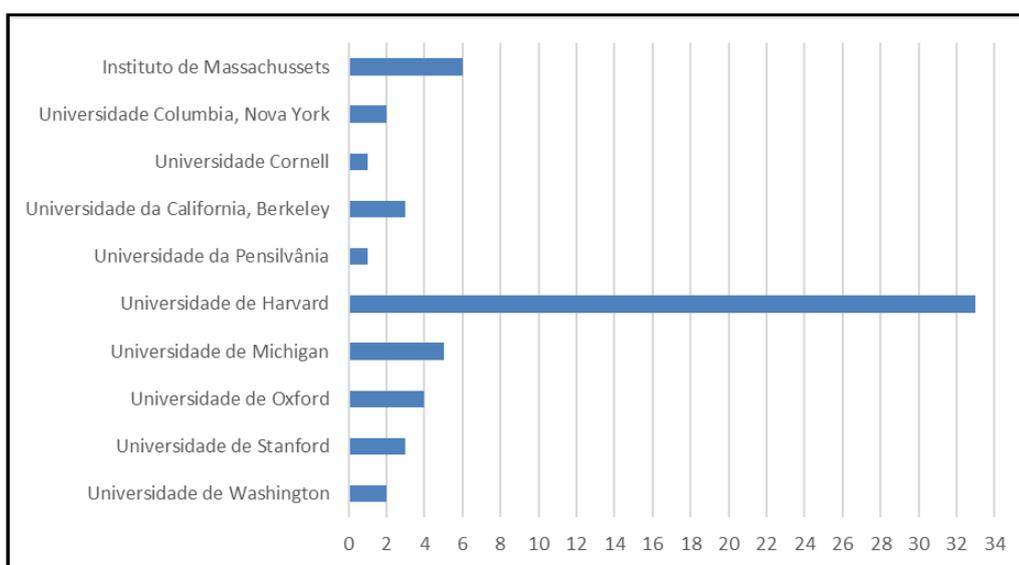
<sup>65</sup> [DMP\\_Checklist\\_FINAL \(dcc.ac.uk\)](https://www.dcc.ac.uk/dmp-checklist-final)

Salienta-se que também é comum que as agências de financiamento exijam a inclusão de um PGD como parte de qualquer pedido de financiamento, e a maioria delas tem o seu próprio conjunto de políticas, embora muitas vezes semelhantes. Nos documentos das instituições foi apresentada a nova política do NIH para gerenciamento e compartilhamento de dados, que exige a inclusão de um Plano de Gerenciamento e Compartilhamento de Dados (DMSP) em todas as inscrições novas e concorrentes desde janeiro de 2023. Essa política substituiu a Política de Compartilhamento de Dados do NIH de 2003. Além da política do NIH, foram citados os Requisitos da *National Science Foundation (NSF)*<sup>66</sup> que desde 2011 exige um PGD.

O Gráfico 7 a seguir, demonstra as Referências de codificação do Código “Plano de Gestão de Dados” nos documentos das universidades deste estudo.

É possível perceber que, novamente, a Universidade de Harvard se destaca com as codificações em seus documentos sobre o tema “Plano de Gestão de Dados”, seguido do MIT, Universidade de Michigan, Stanford e Oxford.

**Gráfico 7** - Quantidade de Referências para o Código Plano de Gestão de Dados, nas universidades



**Fonte:** dados da pesquisa.

Até o momento, verificou-se que as universidades têm implementado práticas e estratégias de competência em dados para sua comunidade acadêmica por meio de atividades voltadas à necessidade de aprimoramento da competência dos indivíduos

<sup>66</sup> <https://new.nsf.gov/funding/data-management-plan>

em relação às práticas que envolvem o uso dos dados de pesquisas, onde cada atendimento individual, oficinas ou cursos abordaram temas com o objetivo de capacitá-los a lidar com as responsabilidades que o aumento de dados provenientes das pesquisas acarreta. E as comunidades acadêmicas normalmente desenvolvem seus próprios métodos e abordagens para gerenciar e disseminar dados, como foi possível ver no que se refere ao desenvolvimento de planos de gestão de dados.

Considerando o contexto apresentado nessa seção, compreende-se que os PGDs têm uma função importante na comunicação científica e influenciam o relacionamento entre autores, instituições, e demais entidades de pesquisa. Os desafios encontrados pelas instituições no que diz respeito à gestão de dados de pesquisa, decorrente da quantidade de dados produzidos, provocou a necessidade da elaboração de políticas para, justamente, gerenciar esses dados nas instituições. Monteiro (2017) afirma que um elemento relevante na gestão de dados científicos é certificar que eles sejam compreendidos e interpretados por outras pessoas ao longo do tempo, e para isso, torna-se fundamental uma descrição nítida e detalhada dos dados, com anotações que os contextualizem. Essas ações se efetivam por meio dos planos de gestão de dados.

O PGD é uma atividade que auxilia as boas práticas científicas, uma vez que propõe que os dados sejam utilizados e disponibilizados de forma coerente com as novas circunstâncias, advindas do meio digital e de produção científica que se vivencia. No contexto da pesquisa científica, um plano de gestão de dados se faz necessário por ser considerado um recurso que possibilita que os dados sejam acessados, preservados e reusados em novos estudos (Guandalini; Furnival; Arakaki, 2019).

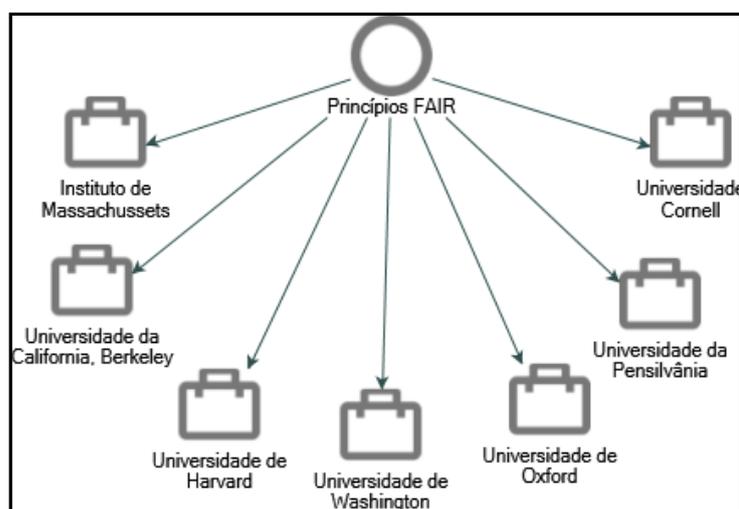
Há também a preocupação de possibilitar que os dados sejam mais encontráveis, acessíveis, interoperáveis e reusáveis. Com isso, surgiram em 2014 os princípios orientadores FAIR, considerados elementos norteadores das boas práticas de todo o processo de gestão de dados de pesquisa. Permitindo que os dados possam ser entendidos por outros pesquisadores e sistemas de computador. A próxima seção descreve as instituições que discutiram aspectos relacionados a esse assunto.

#### 4.2.11 Princípios FAIR

Publicar e compartilhar dados de pesquisa é uma ótima maneira de incorporar o acesso aberto, a Ciência Aberta e os Princípios FAIR nas jornadas de pesquisa. O crescente anseio pelos dados gerados pelas atividades científicas proporcionou, tanto por agências de fomento, instituições de pesquisa e comunidade acadêmica, a necessidade de uma gestão ativa superior ao simples armazenamento e disponibilização na *web*. Um sistema confiável para a publicação de dados baseada em uma potente infraestrutura digital, viabiliza o compartilhamento de dados de forma eficaz para que esses recursos atinjam todo seu potencial de reuso. (Henning *et al.* 2019).

Nas instituições elencadas para essa pesquisa, observou-se que sete desenvolvem atividades de competência em dados destacando informações sobre os Princípios FAIR, como mostra a Figura 26. O MIT, e as Universidades da Califórnia - Berkeley, Harvard, Washington, Oxford, Pensilvânia e Cornell apresentaram uma série de atividades visando aumentar a conscientização sobre as práticas necessárias para tornar os dados FAIR, incluindo introdução geral aos princípios FAIR e o conceitos dos acrônimos. Aponta-se que no que diz respeito a essas discussões, novamente, as universidades trouxeram conteúdos muito semelhantes, pois o intuito foi abordar a importância dos Princípios FAIR nas práticas científicas

**Figura 26** - Universidades onde o Código “Princípios FAIR” foi identificado



**Fonte:** dados da pesquisa.

À medida que ocorreram as análises dos documentos, observou-se que as instituições explicaram o significado da sigla FAIR, “*Findable, Accessible, Interoperable, Reusable*”, e também os motivos pelos quais deve-se adotar esses Princípios. Conforme destacam, existem vários benefícios para o pesquisador. Primeiro: é mais provável que se obtenha o crédito apropriado pelos dados, metadados e outros produtos criados a partir dos dados. Como os Princípios FAIR promovem o uso de identificadores persistentes, aqueles que utilizam dados de outra pessoa saberão como citar o trabalho com precisão. Em segundo lugar: os Princípios FAIR promovem a elaboração de documentação acessível e interoperável, o que significa que os dados serão mais reutilizáveis no futuro. E terceiro: os financiadores de pesquisa solicitam que os projetos de pesquisa e os dados sejam consistentes com os Princípios FAIR.

Salienta-se que as Universidade de Cornell, Pensilvânia e Harvard, além de abordarem os benefícios dos Princípios e o significado dos acrônimos, listam maneiras práticas de tornar os dados FAIR, ou seja e orientações para tornar os dados encontráveis, acessíveis, interoperáveis e reusáveis.

Essas instituições forneceram dicas para analisar se os arquivos de dados e a sua documentação (ou seja, metadados) suportam os Princípios de Dados FAIR, seguida de instruções adicionais sobre como preparar os dados adequadamente. Por meio de uma série de perguntas incentivam o pesquisador a refletir sobre importantes questões relativas aos dados da sua pesquisa:

- a) O conjunto de dados está em um repositório aberto e confiável (se disponível)?
- b) O conjunto de dados possui DOI registrado?
- c) Os arquivos de dados estão em formatos abertos padrão e/ou comumente disponíveis (tanto quanto possível)?
- d) Os dados e/ou metadados podem ser recuperados por meio de uma API e/ou descobertos por meio de um protocolo de pesquisa aberto (por exemplo, por meio do Google)?
- e) Os metadados incluem padrões legíveis por máquina quando disponíveis (por exemplo, ORCID's para autores e/ou contribuidores de dados)?
- f) Os artigos relacionados são referenciados e vinculados nos metadados?
- g) É fornecido um formato de citação para o conjunto de dados?

h) Os termos de licença, atribuição ou termos de uso estão claramente indicados?

As orientações de preparação dos dados, visam o compartilhamento dos dados de acordo com os Princípios FAIR, frisando que isso facilita a descoberta e a reutilização das pesquisas científicas.

Em relação às Referências de codificação do Código “Princípios FAIR” nos documentos das universidades, a Universidade de Harvard obteve oito referências entre seus 49 documentos, enquanto que nas demais instituições houve apenas uma referência de codificação nos respectivos materiais analisados. Harvard, como demonstrado no Gráfico 1, foi a instituição na qual foi possível identificar a maior quantidade de documentos para análise, é por esse motivo que frequentemente a Universidade aparece como àquela que dispõe de mais Referências de codificação sobre alguns assuntos.

Cada universidade abordou os aspectos dos Princípios FAIR, sua importância e benefício para o pesquisador e para os dados da pesquisa científica. A aplicação dos Princípios de Dados FAIR se dá com a utilização das boas práticas de gestão e documentação dos dados usados durante a pesquisa, com foi muito bem apontado nos conteúdos analisados. Também, identificou-se a ênfase à capacidade de ação das máquinas no atendimento dos Princípios FAIR. O crescimento em volume e complexidade significa que o suporte computacional é cada vez mais necessário na localização e no tratamento de dados. Além disso, os Princípios promovem a utilização de metadados ricos, identificadores persistentes, licenças de dados e vocabulários compartilhados e padrões comunitários.

Wilkinson *et al.* (2016) afirmam que a verdadeira intenção do Princípios FAIR, é que eles possam funcionar como uma diretriz para aqueles que desejam melhorar a reutilização de seus acervos de dados. Ao contrário das iniciativas de pares que se concentram no “fazer” humano, os Princípios FAIR enfatizam, especificamente, o aprimoramento da capacidade das máquinas de encontrar e usar automaticamente os dados, além de apoiar sua reutilização por indivíduos. Nesse contexto de possibilidade dos mecanismos automatizados reconhecerem os dados para que possam ser combinados automaticamente com outros, a interoperabilidade dos dados pode ser vista como um problema de longo prazo, que exige mais esforços na criação dos dados FAIR.

Nas análises dos documentos não foram encontradas referências a respeito da interoperabilidade dos dados, ou maneiras práticas de torná-los interoperáveis. A interoperabilidade, quando se trata dos elementos e das orientações para tornar dados FAIR deve envolver padrões, formatos e colaboração entre sistemas o que pode dificultar o atendimento a esse Princípio.

Outrossim, a maioria das atividades, identificadas nos documentos analisados, sobre os princípios FAIR foram introdutórias, mostrando para a comunidade acadêmica os benefícios que a adoção dessa prática traz, principalmente para o compartilhamento de dados.

A próxima seção apresenta as universidades que realizaram práticas de competência em dados sobre o tema “Repositório de dados”.

#### **4.2.12 Repositório de dados**

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 Repositórios de dados. Para que os dados atendam às premissas de acesso, compartilhamento e reuso, eles necessitam de 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, por exemplo, e permitindo, assim, que sejam localizáveis, acessíveis, interoperáveis e reusáveis (Silveira, 2020).

Algumas agências de financiamento ou editoras de periódicos exigem o uso de um repositório específico para o depósito dos dados de pesquisa. Alguns desses repositórios são integrados aos fluxos de trabalho dos editores, são apoiados pelas próprias agências financiadoras ou são simplesmente o repositório preferido para pesquisas em um determinado domínio. Outras sugerem o depósito dos dados e indicam diversos repositórios generalistas para tal fim.

O Código “Repositório de dados” trata-se do último tema deste estudo identificado nas atividades de competência em dados das instituições e documentos analisados. A Figura 27 apresenta o Código “Repositório de dados” seguido de seus Subcódigos “Política de dados” e “Termo de uso de dados”.

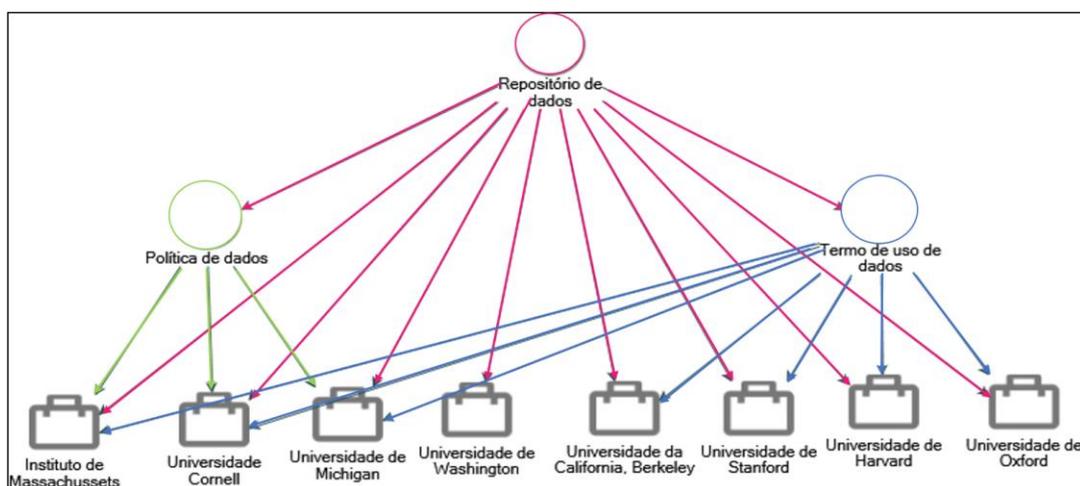
Ao elaborar as codificações observou-se que os instrutores além de desenvolverem conteúdos que discorriam sobre repositórios de dados enquanto

sistema informatizado para armazenamento e preservação dos dados, também trouxeram aspectos relacionados às políticas de dados e sobre termos de uso dos dados.

Os dois subcódigos foram criados ao analisar o conteúdo dos documentos para que fosse possível identificar mais precisamente o que cada instituição considerou destacar nas atividades de competência em dados, uma vez que entender sobre o funcionamento dos repositórios de dados de pesquisa (RDP) é de suma importância para o pesquisador que pretende depositar seus dados em uma ferramenta dessa natureza, seja ela institucional ou não.

A Figura 27 apresenta as universidades que oferecem atividades para o entendimento dos conceitos e objetivos de um repositório de dados, as formas de publicação, a política de dados e os termos de uso dos dados. Cada instituição desenvolveu as atividades de uma forma distinta, algumas dando destaque para os benefícios que o RDP oferece para os pesquisadores e comunidade acadêmica, outras elencando os tipos e apresentando seus próprios repositórios de dados, e outras apontando as características de um bom RDP.

**Figura 27** - Universidades onde o Código “Repositório de dados” foi identificado



**Fonte:** dados da pesquisa.

O MIT e a Universidade de Oxford, entre seus conteúdos, destacaram os benefícios que um RDP oferece para aqueles que publicam seus dados. Mencionam que os RDP's auxiliam no gerenciamento dos dados, na preservação ao longo prazo, maximizam a visibilidade e o impacto acadêmico, e viabilizam o cumprimento dos requisitos das agências de financiamento.

Já as Universidades de Harvard, Michigan e Cornell apontam os três tipos de repositórios como opções viáveis para que a comunidade possa escolher onde depositar seus dados. São mencionados os seguintes tipos repositórios:

- a) Repositórios Institucionais - disponíveis para pesquisadores afiliados às instituições que os administram;
- b) Repositórios Disciplinares - configurados para acomodar as necessidades de dados de uma determinada área do conhecimento ou de pesquisa e
- c) Repositórios Generalistas - aceitam dados independentemente do tipo, formato, conteúdo ou foco disciplinar.

Os RDPs têm como objetivo comum armazenar, organizar, preservar e fornecer acesso a dados de diferentes tipos e origens.

A Universidade de Cornell aproveita as atividades desenvolvidas para apresentar o seu repositório institucional, o *eCommons*, que fornece acesso de longo prazo a uma ampla gama de conteúdo digital de valor duradouro relacionado a Cornell, assim como Harvard apresenta o Harvard Dataverse.

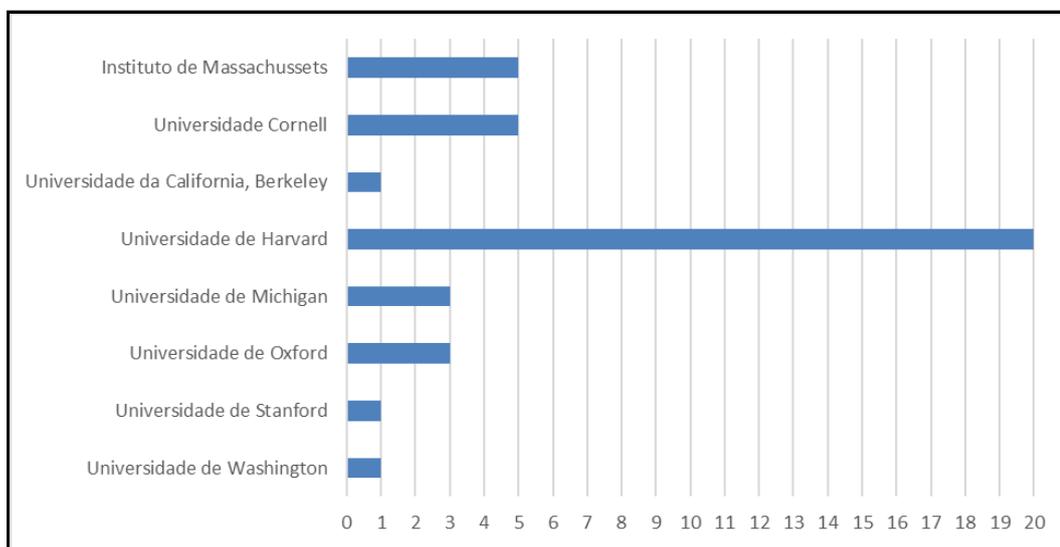
A Universidade de Washington destacou o uso do R3data para que seus pesquisadores e alunos tenham conhecimento da ampla gama de repositórios de dados disponíveis, uma vez que se trata de um diretório de registro global de repositórios de dados de pesquisa e que abrange diferentes disciplinas acadêmicas e repositórios generalistas como o Figshare, Zenodo e Dryad, mencionados também nos documentos analisados da Universidade de Harvard.

Além disso, o MIT, a Universidade de Harvard, Stanford e Califórnia - Berkeley, listaram para os pesquisadores e alunos características de um bom repositório de dados para subsidiar a análise de escolha de um RDP antes de publicar seus dados. As características citadas foram: acessibilidade, segurança e privacidade, orientações de uso claro (em relação às licenças CC), e preservação a longo prazo.

De modo geral, as instituições ao desenvolverem as atividades de competência em dados para sua comunidade acadêmica, apresentam informações sobre os repositórios de dados de pesquisa, conforme sua realidade institucional, isto é, de acordo com as demandas que surgem na instituição, na biblioteca e nos serviços de apoio a dados.

A respeito das Referências de codificação, a Universidade de Harvard foi a instituição em que o Código “Repositório de dados” foi referenciado 20 vezes, como mostra o Gráfico 8, seguido do MIT e Cornell, com cinco referências, e Michigan e Oxford com três.

**Gráfico 8** - Quantidade de Referências para o código Repositório de dados, nas universidades



**Fonte:** dados da pesquisa.

A Universidade de Harvard novamente destacou-se em relação às Referências de codificação nos documentos. O fato pode não significar que as demais universidades desenvolverem menos atividades de competência em dados sobre o tema, mas sim, que o conteúdo das atividades, oficinas e cursos, por exemplo, estão disponíveis para consultas.

#### 4.2.12.1 Política de dados

Conforme apresentado na Figura 27 o Instituto de Massachusetts, as Universidades de Cornell e Michigan abordaram em suas atividades de competência em dados informações relacionadas às políticas de dados que estão presentes nas diretrizes dos Repositórios de dados. O Tesouro de Dados de Pesquisa (2024) da Universidade Federal do Rio Grande do Sul (UFRGS) define que as políticas de dados são diretrizes de um conjunto de princípios que estabelecem orientação para o gerenciamento de dados, podendo ser usada para tratar de questões relacionadas ao acesso de dados, aquisição de dados, depósito, entre outros aspectos.

À vista disso, observou-se que nos documentos analisados, quando são discutidas as políticas de dados dos repositórios, são apresentadas informações a respeito dos formatos de arquivos de dados recomendados pelos repositórios. Destacam que os formatos de arquivo “funcionais” (aqueles usados ao coletar e trabalhar com dados do projeto) nem sempre são ideais para reutilização ou preservação a longo prazo, podendo não cumprir os requisitos dos arquivos dos RDP’s, ou não cumprir com os requisitos dos financiadores da pesquisa. A Universidade de Cornell, citou o exemplo de seu repositório institucional que além das políticas de dados oferece aos pesquisadores diretrizes gerais para a seleção de formatos de arquivo para preservação e reutilização.

Nos documentos que tratam sobre políticas de dados, percebeu-se também que as instituições abordaram aspectos sobre a segurança e privacidade dos dados, isto é, medidas que os repositórios de dados assumem para garantir a segurança de dados sensíveis, por exemplo. O MIT apresentou estratégias de privacidade para que os pesquisadores reflitam sobre a sensibilidade dos seus dados antes de depositar no repositório. Disponibilizam para a comunidade recursos de privacidade, que consiste em *sites* com informações sobre a proteção de informações relevantes para o MIT e outras organizações ou indivíduos que lidam com informações confidenciais.

Esses dois elementos que foram destacados pelo MIT, Michigan e Cornell a respeito das políticas de dados dos repositórios, demonstram que as instituições consideram essenciais que os pesquisadores tenham conhecimento dos aspectos que norteiam uma política de dados. Mesmo que o aprofundamento do conteúdo desenvolvido não tenha sido detalhado nos documentos disponíveis, reflete a dedicação das instituições em estabelecer uma base sólida de compreensão sobre as políticas de dados para garantir práticas de gestão de dados responsáveis.

Em relação às Referências de codificação, nos documentos das três universidades, 52 no total, o Subcódigo “Política de dados” foi referenciado ao todo, sete vezes, sendo cinco vezes nos 12 documentos de Cornell, uma única vez nos 24 documentos do MIT, e, também, uma vez nos 16 documentos de Michigan. Esses dados demonstram que as instituições, ao abordarem aspectos relacionados às políticas de dados dos repositórios, foram mais objetivas e não desenvolveram

muitos tópicos sobre o tema. Embora tenha sido possível identificar os elementos levantados pelas instituições, a profundidade das discussões foi escassa.

Destaca-se que o MIT, Cornell e Michigan foram as únicas universidades que apresentaram em seus documentos conteúdos que abordam aspectos relacionados aos repositórios de dados e suas políticas. As demais instituições abordaram os repositórios de dados e os termos de uso.

#### 4.2.12.2 *Termo de uso de dados*

O último Subcódigo atribuído nos documentos analisados foi “Termo de uso de dados”. Um termo de uso de dados trata das diretrizes que expressam as condições para que um conjunto de dados seja utilizado de maneira responsável. Pode ser um termo formal, como em uma licença ou contrato, ou uma informal expressando as preferências de uso do proprietário dos dados.

O termo de uso de dados aplicado aos repositórios é imprescindível para definir condições em que os dados serão acessados, compartilhados e usados, mediante a responsabilidade assumida pelo pesquisador que vierem a acessar e reutilizar os dados disponíveis no repositório.

A Figura 27 mostra que o MIT, as Universidades de Cornell, Califórnia-Berkeley, Harvard, Michigan, Oxford e Stanford desenvolveram conteúdos a respeito dos termos de uso de dados em RDP's, dando ênfase às questões de propriedade intelectual relacionadas a dados de pesquisa, principalmente quando a propriedade pode ser atribuída ao pesquisador, à instituição ou ao financiador, dependendo das condições do contrato de concessão e da existência de implicações de patente.

A partir disso, verifica-se que a compreensão das nuances a respeito da privacidade dos dados é uma preocupação de sete das 10 instituições estudadas. Abordaram esse assunto em suas atividades de competência e dados, destacando, principalmente, o uso das licenças que possibilitam ao detentor dos dados decidir e ter maior controle sobre as condições de utilização dos dados.

Observou-se que as instituições utilizam as mesmas licenças nos seus repositórios, a *Creative Commons*, a qual possui tipos de licenças padronizadas e se aplicam a dados de pesquisas e bancos de dados. As instituições destacam os seis principais tipos apresentando suas restrições e permissões.

A Universidade de Stanford, por exemplo, salientou nas suas atividades as duas licenças CC de maior relevância para a gestão de dados do seu repositório, são elas:

- a) CC0<sup>67</sup>: utilizada quando o proprietário deseja renunciar aos seus direitos autorais e/ou direitos de banco de dados, isto é, coloca o banco de dados e os dados em domínio público, e
- b) Marca de Domínio Público (PDM)<sup>68</sup>: usada para identificar obras que estão em domínio público, sem Direito de Autor nem Direitos Conexos, e para as quais não há restrições conhecidas de direitos autorais ou de banco de dados.

A Universidade de Stanford também apresentou as licenças criadas pelo grupo *Open Data Commons* (ODC)<sup>69</sup> que tem desenvolvido conjunto de ferramentas legais e licenças para auxiliar a publicar, fornecer e usar dados abertos. As licenças da ODC destacadas pela instituição foram:

- a) Dedicção e Licença de Domínio Público (PDDL): dedica o banco de dados e seu conteúdo ao domínio público, de forma gratuita para que todos possam usar da maneira mais adequada;
- b) Licença de Atribuição (ODC-By): os usuários são livres para usar o banco de dados e seu conteúdo de maneiras novas e diferentes, desde que forneçam atribuição à fonte dos dados e/ou ao banco de dados; e
- c) Licença de Banco de Dados Aberto (ODC-ODbL): indica que qualquer uso subsequente do banco de dados deve fornecer atribuição, uma versão irrestrita do novo produto deve estar sempre acessível e quaisquer novos produtos feitos com material ODbL devem ser distribuídos usando os mesmos termos. É a mais restritiva de todas as licenças ODC.

Ressalta-se ainda, que nas atividades desenvolvidas pelo Instituto de Massachusetts, identificaram-se não somente as licenças atribuídas para dados e textos, mas, também, explicações sobre as licenças de código aberto. Nesse caso, especificamente, observou-se que o conteúdo era focado para pesquisadores que já tinham experiência com práticas de dados. Destaca-se que esse foi o único documento com conteúdo mais técnicos que os demais documentos analisados.

---

<sup>67</sup> <https://creativecommons.org/publicdomain/zero/1.0/deed.pt-br>

<sup>68</sup> <https://creativecommons.org/publicdomain/mark/1.0/deed.pt-br>

<sup>69</sup> [Home — Open Data Commons: ferramentas legais para dados abertos](#)

A respeito das Referências de codificação, o Subcódigo “Termo de uso de dados” foi referenciado quatro vezes nos 24 documentos do Instituto de Massachusetts, três vezes nos 11 documentos de Oxford, e quatro vezes em 61 documentos das Universidades de Harvard e Cornell.

Notou-se que o tema “Termos de uso de dados”, embora mencionado nas atividades de competência em dados de sete instituições, com exceção da Universidade de Washington, não é explorado mais detalhadamente. São enfatizados os benefícios e a importância dos repositórios de dados, deixando de lado os aspectos técnicos e aprofundada as questões sobre as políticas de dados e os termos de uso que, sem dúvida, fazem parte do bom funcionamento de um RDP.

Os termos de uso dos dados de um repositório representam uma forma de instruir aqueles que vão acessar e utilizar os conjuntos de dados, da mesma maneira que as políticas de dados, orientam os indivíduos que vão depositar os dados, por exemplo. No caso das instituições estudadas observou-se que destacaram, em suas atividades de competência em dados, as licenças de uso e compartilhamento de dados.

Com a codificação dos documentos de acordo com os temas desenvolvidos, foi possível conhecer o que as instituições priorizam nas suas atividades de competência em dados. Observou-se que o conteúdo ministrado contempla aspectos conceituais e teóricos. Além disso, o principal objetivo é fornecer informações sobre as práticas que envolvem todo o processo de utilização dos dados de pesquisa, tanto pelos produtores como pelos consumidores dos dados, exigindo o conhecimento das ações destacadas nas atividades identificadas.

Ter conhecimento das melhores práticas de dados influencia o comportamento do pesquisador diante das suas práticas científicas. Teal *et al.* (2015) afirma que muitos pesquisadores infelizmente não estão familiarizados com as melhores práticas e ferramentas do ciclo de vida dos dados: a maior parte ou tudo o que sabem sobre gerenciamento, análise e compartilhamento de dados foi aprendido aos poucos ou nem foi aprendido. O autor ainda complementa dizendo que isto é algo preocupante pois limita a capacidade dos pesquisadores de progredir em questões importantes, ou resulta em análises imprecisas ou incompletas que podem levar a conclusões erradas. Também, leva à geração de dados que, em

última análise, não podem ser utilizados para responder as questões da investigação, desperdiçando ou subutilizando os recursos disponíveis.

No contexto das universidades acadêmicas, oferecer recursos e serviços para tornar a comunidade competente em dados e, entre outras habilidades, saber manusear os dados que produzem, torna-se essencial para o futuro da pesquisa científica. Desenvolver essas habilidades e qualificar os indivíduos que estão inseridos no processo de elaboração de uma pesquisa, com o intuito de capacitá-los para compreenderem os processos e ações que envolvem a coleta, armazenamento, preservação, compartilhamento e reuso dos dados é fundamental em um contexto onde os dados de pesquisa tornaram-se “matéria-prima” das pesquisas e da geração de novos conhecimentos.

Diante disso, a próxima seção apresenta, a partir da análise das universidades selecionadas para este estudo, os setores e/ou serviços que desenvolvem práticas de competência em dados para a comunidade acadêmica, destacando algumas características desses serviços em cada uma das universidades onde foram identificados.

#### 4.3 AS UNIVERSIDADES E SUAS PRÁTICAS DE COMPETÊNCIA EM DADOS

Para uma proposta de ação de competência em dados, Risdale *et al* (2015) afirmam que é necessário o desenvolvimento de cursos, disciplinas, e um processo intencional de aprendizagem ao longo da vida do indivíduo, isto é, além dos espaços tradicionais de aprendizagem. No entanto, para isso acontecer é importante o comprometimento entre estudantes, pesquisadores, instituições e organizações, consumidores da informação em dados, dentre outros agentes da sociedade, fomentando práticas formativas e inovadoras de competência em dados nas universidades.

O segundo objetivo da pesquisa buscou identificar as práticas adotadas para desenvolver competências em dados de pesquisa nas universidades internacionais selecionadas. Para tal, os *sites* de cada instituição foram explorados a fim de identificar informações sobre as atividades voltadas à capacitação em dados de pesquisa.

Ainda utilizando o Nvivo, os valores de atributos COMO, QUEM e PARA QUEM, apresentados no Quadro 5, foram pertinentes para a elaboração dos gráficos

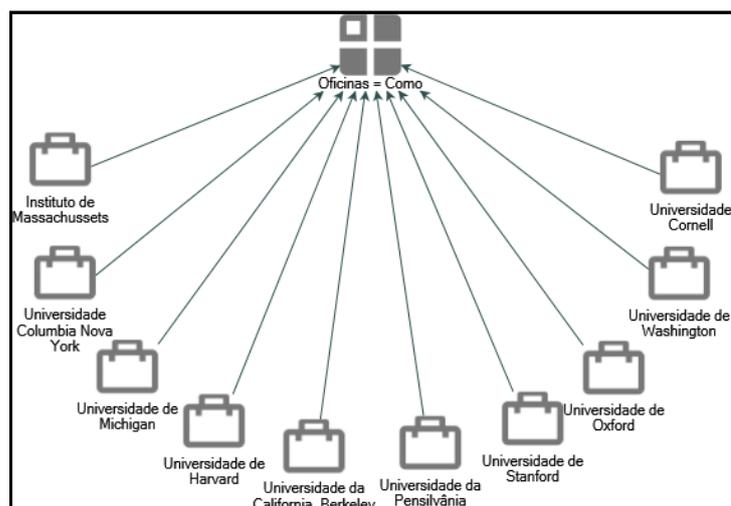
e posteriores análises, pois levando em consideração as indagações que surgiram ao associar os Atributos com os Casos, conforme a Tabela 3 e Quadro 4, da seção 3.3, eles contribuíram para melhor compreensão dos resultados.

Se tratando do objetivo b) *identificar as práticas adotadas para desenvolver competências em dados de pesquisa nas universidades internacionais selecionadas*, o valor de atributo que melhor compreende a elaboração da análise é o COMO. Assim, buscou-se identificar COMO as universidades buscam desenvolver ou aprimorar as competências em dados de pesquisa entre os membros da sua comunidade.

A aquisição de competências em dados é uma questão importante para os indivíduos que desejam ou necessitam tornar-se indivíduos e pesquisadores com habilidades na área de dados de pesquisa. Handel, Vasilevsky e Wirz (2012) afirmam que criar uma cultura de pessoas com habilidade em dados, requer um acompanhamento na sua formação científica com educação em competência em dados. Assim, o papel das universidades torna-se fundamental pois são espaços de desenvolvimento acadêmico, científico e formativo de pessoas, onde o conhecimento prático baseado em dados e as competências que se devem desenvolver, exigem um conjunto de habilidades.

Nas Universidades elencadas neste estudo, observou-se, a partir do material analisado, que o desenvolvimento das atividades com dados de pesquisa para promover o uso correto e ético dos mesmos, foram realizadas por meio de cursos, oficinas e atendimento individual. As Figuras a seguir apresentam as instituições que utilizam essas estratégias para desenvolver as competências em dados.

**Figura 28** - Universidades que oferecem atividades de competência em dados por meio de oficinas



**Fonte:** dados da pesquisa

As instituições que mais se destacam nas atividades ministradas por meio de oficinas são o Instituto de Massachusetts, a Universidade de Cornell e a Universidade da Pensilvânia, como será relatado a seguir.

No Instituto de Massachusetts as oficinas são oferecidas para ajudar a comunidade acadêmica a adquirir novas habilidades em gerenciamento de dados de pesquisa. As oficinas são ministradas de forma presencial, e desde a pandemia do COVID-19, passaram, também, para a forma *online* por meio da plataforma ZOOM, com os vídeos ficando disponíveis para posterior consulta na plataforma do *Youtube*. Os bibliotecários são os responsáveis por organizar e ministrar as oficinas.

A Universidade de Cornell também atua com oficinas para seus pesquisadores, alunos de graduação e pós-graduação durante o ano letivo. Por meio de treinamentos personalizados para turmas, equipes de projeto ou departamentos, a instituição fornece instruções sobre ferramentas e *software* de dados, além de tópicos que abordam a elaboração de plano de gerenciamento de dados conforme requisitos das agências financiadoras, e como compartilhar e publicar dados. A Universidade também possibilita ao indivíduo solicitar oficinas conforme necessidade ou demanda. Arelado a essas atividades, foi criada uma série semanal de discussões informais para percorrer o ciclo de vida dos dados, desde as etapas de planejamento de um projeto de pesquisa, até o encerramento e depósito dos dados, discutindo as melhores práticas e apontando recursos disponíveis no campus e fora dele. Alguns vídeos ficam disponíveis no *Youtube* para consulta posterior, assim

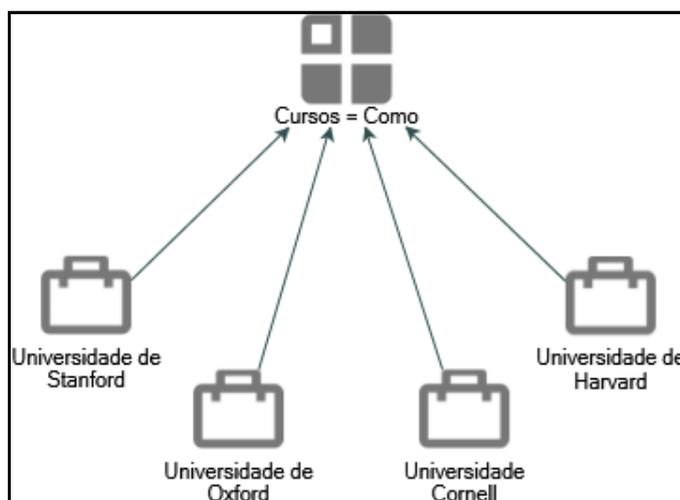
como os materiais textuais que são disponibilizados na página da biblioteca da Universidade de Cornell.

Na Universidade da Pensilvânia as oficinas são ministradas de forma presencial e *on-line*. Conforme observação direta no *site*, a instituição elabora oficinas para instigar seu público acadêmico a refletir sobre aspectos básicos necessários para a adoção de práticas de gerenciamento de dados.

As Universidades de Michigan, Washington, Harvard, Califórnia-Berkeley, Stanford, e Columbia - Nova York, também utilizam o método de oficinas como meio de capacitar e tornar seu público e principalmente os pesquisadores competentes em dados, porém, não foram identificadas características de destaque nessas práticas.

Além das oficinas, conforme a Figura 29, foi identificado que as instituições também disponibilizam cursos acerca dos dados de pesquisa. A maioria dos cursos tem um foco mais teórico do que prático, com um conteúdo mais detalhado, tendo geralmente como objetivo a formação mais completa e aprofundada do que as oficinas.

**Figura 29** - Universidades que oferecem atividades de competência em dados por meio de cursos



**Fonte:** dados da pesquisa

Na Figura de 29 percebe-se que quatro universidades oferecem cursos sobre dados de pesquisa, sendo elas: Universidade de Stanford, Universidade de Oxford, Universidade de Cornell e Universidade de Harvard.

A Universidade de Cornell e a Universidade de Stanford, além das oficinas, oferecem cursos com o intuito de melhorar as práticas de gerenciamento de dados de seus pesquisadores. Ao contrário das oficinas em que os materiais textuais, e até mesmo audiovisuais ficam disponíveis para posterior consulta no *site* da instituição, e são de fácil acesso, o mesmo não acontece com os cursos e não foi possível localizar e acessar esse tipo de recurso na sua totalidade, sendo necessário vínculo institucional para acessar as informações completas. Portanto, foram obtidas apenas informações parciais sobre datas, locais e súmula.

Na Universidade de Harvard os cursos são voltados para membros da comunidade acadêmica que têm interesses e atuam com dados de pesquisa. Os cursos abordam práticas recomendadas para facilitar a descoberta, o acesso, a integridade e o valor de reutilização dos dados de pesquisa, contendo de cinco a dez módulos, e variando entre presenciais e *on-line*. Ressalta-se também que na instituição os cursos são organizados por profissionais da *Harvard Medical School*, área médica da instituição, sendo voltados para pesquisadores que trabalham com dados médicos, de forma que, os cursos são ministrados por especialistas da área.

Muitas oportunidades de capacitação estão disponíveis na Universidade de Oxford para adquirir competências em dados de pesquisa e desenvolver uma melhor compreensão de como encontrar e trabalhar com dados. Além das oficinas, disponibiliza cursos presenciais e *on-line* mediante inscrição ou agendamento. Os cursos encontrados no *site* da instituição são organizados pela Biblioteca, e os títulos disponíveis são: “Fontes de dados para pesquisa - descoberta, acesso e uso”; “Gerenciando dados de pesquisa e planejamento de gerenciamento de dados (DMPs)”; e “Trabalhando com dados de pesquisa pessoais”. Além disso, é disponibilizado e divulgado pela instituição, o MANTRA<sup>70</sup>, um curso *on-line* gratuito hospedado pela Universidade de Edimburgo, que consiste em sete módulos interativos, além de um conjunto de tutoriais de manipulação de dados.

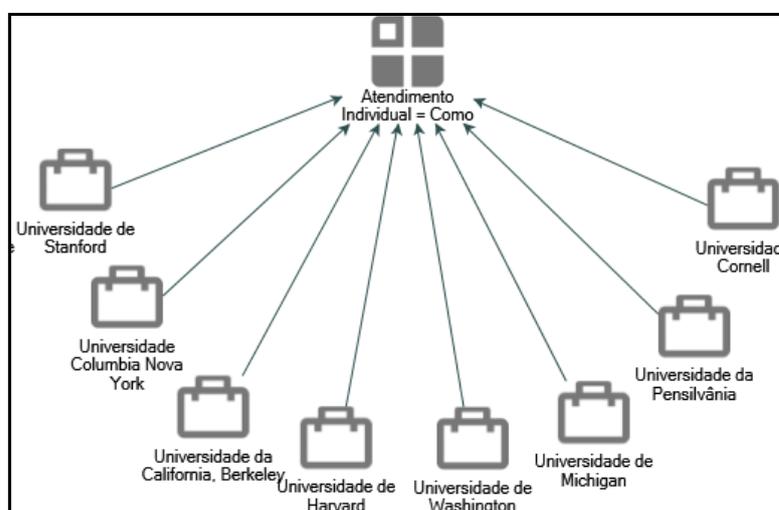
Outra forma que as instituições utilizam para preparar a comunidade acadêmica para trabalhar e desenvolver suas competências em dados, é por meio de atendimentos individuais, como mostra a Figura 30. Esses atendimentos são específicos em cada uma das instituições estudadas uma vez que permitem que os pesquisadores recebam orientações individuais e detalhadas conforme suas

---

<sup>70</sup> <https://researchdata.ox.ac.uk/mantra>

necessidades durante a pesquisa. As Universidades de Stanford; Columbia - Nova York, Califórnia-Berkeley, Harvard, Washington, Michigan, Pensilvânia e Cornell oferecem esse tipo de serviço para suas comunidades.

**Figura 30** - Universidades que oferecem atividades de competência em dados por meio de atendimento individual



**Fonte:** dados da pesquisa

Os atendimentos individuais nessas instituições podem ser agendados para uma única pessoa que necessita de orientação mais pormenorizada a respeito do seu projeto de pesquisa, ou para um grupo de alunos, por exemplo, que está em busca de orientação menos detalhadas e mais introdutórias. Esses atendimentos individuais podem ser adaptados para atender as diversas necessidades da comunidade acadêmica.

Na Universidade da Pensilvânia encontrou-se informações detalhadas a respeito do processo de solicitação e agendamento de atendimento individual. O agendamento é efetuado por ordem de chegada, dependendo da disponibilidade do instrutor responsável. Os pedidos devem ser realizados, de preferência, com um mês de antecedência em relação às datas pretendidas. Para aqueles que solicitam cursos ou oficinas para um grupo específico, orienta-se que o melhor momento é quando o plano de estudo ou projeto está finalizando, pois isso permite que o profissional responsável, na maioria das vezes um bibliotecário, prepare uma experiência de aprendizagem de alta qualidade para os participantes, um processo que geralmente envolve uma conversa prévia sobre os objetivos da sessão.

As demais instituições ao disponibilizarem os serviços de atendimento individual, na página inicial das bibliotecas, por exemplo, disponibilizam a opção “enviar *e-mail*” ou “agendar atendimento”, e assim os indivíduos podem entrar em contato com os responsáveis pelo serviço.

A partir das análises percebeu-se que o atendimento individual é uma modalidade disponível em todas as instituições, principalmente, para pesquisadores que necessitam de informações mais específicas acerca de diversos aspectos da gestão de seus dados. Os atendimentos individuais permitem por parte do mentor uma abordagem personalizada direcionada às necessidades específicas do pesquisador, oferecendo orientações detalhadas e suporte direcionado. Além disso, esses atendimentos são uma oportunidade para os pesquisadores discutirem suas dúvidas e desafios, recebendo *feedback* e soluções práticas que podem ser aplicadas diretamente em suas pesquisas, aprimorando a qualidade e também capacitando-os em aspectos essenciais para desenvolverem trabalhos baseados em dados.

As iniciativas das instituições demonstram o apoio e preocupação no que se refere ao conhecimento de dados de seus pesquisadores a partir do desenvolvimento de atividades de gestão de dados, criação de ferramentas e orientações práticas por meio de guias de pesquisa, tutorias *on-line*, cursos, oficinas e atendimento individual, como citado. As oficinas e os cursos propostos pelas instituições são os meios pelos quais buscam inserir seu público no ambiente da gestão dos dados de pesquisa.

Os dados de pesquisa são a base do registo acadêmico e são importantes para o avanço do conhecimento. Na perspectiva da pesquisa científica a competência em dados é um tipo de educação que direciona os pesquisadores para construir diálogos a partir de treinamentos e reflexões sobre o uso, descoberta e toda a estrutura orientada aos dados de pesquisa.

Whitmire (2015) coloca que à medida que a pesquisa orientada por dados se torna regular, o conhecimento prático em dados e todo seu ciclo torna-se fundamental para os pesquisadores. Todo conhecimento produzido no espaço acadêmico passa por um dos processos mais importantes da universidade, que é a pesquisa.

Quanto às competências dos pesquisadores, pode haver necessidade de aprimorar as habilidades tecnológicas que vão desde a concepção de bases e gestão de dados até à extração de dados e visualização. As universidades se alinham a essas demandas, definem e assumem a responsabilidades pela boa gestão dos seus dados de pesquisa, ampliando suas perspectivas de atuação no que diz respeito às práticas de competência em dados, sejam oficinas, cursos, atendimento individual, ou outra abordagem, que por meio dessas estratégias capacitam sua comunidade acadêmica.

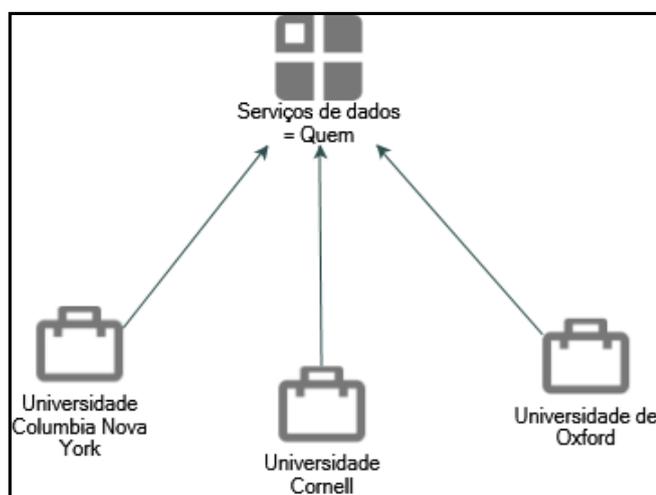
À medida que essas atividades de competência em dados se tornam frequentes, compreende-se que as instituições desempenham um papel crucial no suporte e no avanço dessas práticas. Atrelado a isso, está a criação de ambientes propícios que facilitam o desenvolvimento, a aplicação e a integração das competências em dados. Esses ambientes são caracterizados por vários elementos-chave, incluindo infraestrutura adequada, redes de colaboração e suporte, setores ou departamentos específicos voltados para a análise e gestão de dados que garantem o desenvolvimento dessas competências.

Diante disso, buscou-se atender o objetivo c), *identificar nas universidades estruturas de apoio às atividades de desenvolvimento de competências em dados de pesquisa*, isto é, a presença de setor, centro ou departamento responsável por desenvolver atividades para ampliar a competência em dados (QUEM), e ao mesmo tempo identificar o público para o qual estão voltadas essas atividades (PARA QUEM).

Para isso utilizou-se, novamente, os valores de atributo QUEM e PARA QUEM, apresentados no Quadro 5 da seção 3.3. À medida que os *sites das universidades* foram analisados, observou-se que estas possuíam serviços e desenvolviam atividades que eram oferecidas pela Rede de Biblioteca ou por um setor de serviço de dados de pesquisa, onde são planejados e elaborados projetos, programas e consultorias voltadas às necessidades e demandas dos indivíduos da instituição. Essas duas unidades, tanto as Bibliotecas como os Serviço de Dados de Pesquisa, têm sido responsáveis por desempenhar as atividades de orientação e suporte para o avanço das práticas de competência em dados nas 10 universidades analisadas.

As Universidades de Columbia, Cornell e Oxford, como mostra a Figura 31, contam com um Serviço de Dados com foco nas necessidades dos pesquisadores, desenvolvendo funções que incluem educação e capacitação, suporte técnico especializado, gestão de dados e segurança, colaboração em projetos de pesquisa e criação de programas e políticas institucionais visando apoiar a gestão de dados de pesquisa promovendo uma cultura de boas práticas em toda a instituição. Alguns desses Serviços de Dados trabalham em colaboração com outros departamentos da instituição, visando agregar conhecimentos multidisciplinares na formação de competência em dados de da comunidade acadêmica.

**Figura 31** - Universidades que possuem Serviço de Dados



**Fonte:** dados da pesquisa

A Universidade de Columbia conta com o Departamento de Pesquisa que fornece vários serviços, muitas vezes em conjunto com a Rede de Bibliotecas, para auxiliar a comunidade acadêmica a desenvolver competências em relação aos requisitos de gerenciamento de dados de pesquisa. Nesse departamento criaram o Programa *Research and Data Integrity* (ReaDI) projetado para aprimorar o gerenciamento de dados e a integridade da pesquisa na Universidade de Columbia. O Programa oferece recursos, divulgação e atende consultas de pesquisadores de todos os níveis, auxiliando no entendimento do ciclo de vida dos dados de pesquisa, e criando diversos tutoriais e *templates* para auxiliar no gerenciamento dos dados durante a fase de coleta da pesquisa e preparação para publicação.

Os materiais desenvolvidos não se destinam a ditar práticas específicas, em vez disso, fornecem sugestões e recomendações sobre formas de estruturar e gerir

aspectos da pesquisa, para que os pesquisadores e demais interessados identifiquem as abordagens que melhor se adaptam a sua realidade.

Esse é um setor específico da Universidade de Columbia que fornece informações, tutoriais e práticas para àqueles que têm dúvidas e necessidades sobre gerenciamento de dados. A biblioteca, como centro de informação e documentação atua em parceria com o Departamento de Pesquisa no oferecimento de serviços de dados de pesquisa e com atividades voltadas a esse fim.

Na Universidade de Cornell quem desenvolve as práticas de competência em dados é a *Cornell Data Services*<sup>71</sup>, uma organização colaborativa que abrange todo o *campus* e que auxilia na criação e implementação de gerenciamento de dados e planos de compartilhamento, aplicando melhores práticas para gerenciamento de dados e proporcionando serviços relacionados a dados em qualquer estágio do processo de pesquisa.

A *Cornell Data Services* oferece consultas, oficinas e atendimento durante todo o ano. A diversidade de especialistas na área de dados fornece assistência com uma variedade de tópicos de dados, ferramentas e *software* para atender às necessidades de pesquisa dos indivíduos. Assim como a Universidade de Columbia, identificou-se a presença de colaboração de outros setores, como por exemplo da biblioteca. Entre os profissionais especialistas que oferecem atendimento sobre dados estão presentes os bibliotecários que colaboram e contribuem com os conteúdos e atendimentos.

Na Universidade de Oxford o *Research Data Management Delivery Group* é responsável pelas atividades disponíveis no *site Research Data Oxford*, no qual encontram-se informações e orientações para toda a Universidade de Oxford relacionadas ao gerenciamento de dados de pesquisa. O *site* é destinado a pesquisadores de Oxford de todos os níveis, incluindo estudantes que realizam projetos de pesquisa, que lidam com dados de pesquisa, e a todos aqueles envolvidos no apoio a pesquisadores.

O serviço *Research Data Oxford* proporciona um ponto de contato único para orientação e apoio na gestão de dados de pesquisa. A equipe inclui pessoas de vários departamentos e grupos envolvidos no suporte de gerenciamento de dados

---

<sup>71</sup> <https://data.research.cornell.edu/>

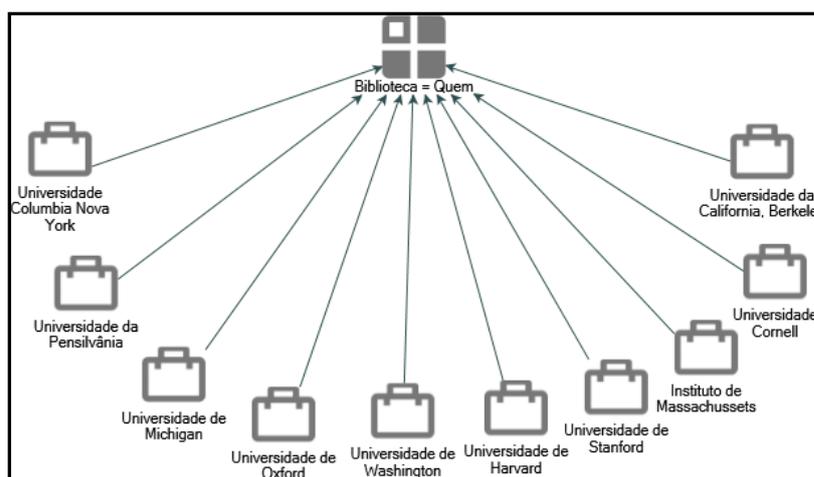
de pesquisa, incluindo Serviços de TI, as Bibliotecas Bodleian e Serviços de Pesquisa.

Além das oficinas, cursos e atendimento individual para aqueles que têm interesses relacionados a vários aspectos dos dados de pesquisa, os serviços da *Research Data Oxford* disponibilizam apresentações de *slides* projetados para cobrir o essencial do gerenciamento de dados de pesquisa, bem como vídeos no *Youtube.*, também, desenvolveram um glossário dos termos usados para descrever o armazenamento e o gerenciamento de dados. Muitas informações estão disponíveis para consultas iniciais, mas também com opções de aprofundamento no assunto. Assim como as instituições anteriores, a *Research Data Oxford* também conta com o apoio da Biblioteca e seus profissionais para desenvolver os serviços sobre dados de pesquisa.

A presença desses setores e departamentos que oferecem serviços de dados de pesquisa evidencia a importância crescente da gestão eficiente e segura dos dados nas instituições acadêmicas, o que demonstra o compromisso com a excelência na pesquisa e a valorização dos dados como ativos críticos para o avanço do conhecimento científico, principalmente, levando em consideração que se trata de um ambiente em que há uma intensa geração e consumo de dados.

Oferecer serviços de dados se torna cada vez mais uma atividade fundamental que deve ser implementada e intermediada, principalmente, pelas unidades de informação das organizações de ensino e pesquisa, isto é, as bibliotecas, que podem ampliar o escopo dos serviços já prestados no suporte à pesquisa, desempenhando atividades de apoio, instrução, acompanhamento e, conseqüentemente, ampliação de competências em dados de pesquisa. Durante a análise das competências desenvolvidas, identificou-se a biblioteca como responsável no planejamento e na prestação de serviços de competência em dados nas 10 universidades analisadas, como mostra a Figura 32.

**Figura 32** - Universidades onde a biblioteca é responsável pelos serviços de dados



**Fonte:** dados da pesquisa

Cada Universidade estudada disponibiliza serviços de dados de pesquisa para sua comunidade acadêmica, adaptando as ofertas às necessidades específicas dos pesquisadores. Destacam-se, as Universidades de Stanford, Michigan, Califórnia-Berkeley, Harvard e Oxford, pois foram as instituições em que se encontraram mais informações a respeito dos serviços oferecidos pela biblioteca, como será relatado a seguir.

A Rede de Bibliotecas da Universidade de Stanford é a responsável pelos serviços de dados de pesquisa. Elas oferecem diferentes serviços para apoiar as necessidades de dados dos pesquisadores, fornecendo orientação durante todo o ciclo de vida da pesquisa. Entre esses serviços estão oficinas e consultas de desenvolvimento de habilidades, apoio as plataformas emergentes baseadas na *web* para exploração e computação, acordos de uso de dados, descrição de conjunto de dados (para descoberta e reutilização) e suporte prático para projetos com uso intensivo de dados. As bibliotecas contam com especialistas em suporte estatístico, mineração de texto, aprendizado de máquina, ciência de dados espaciais (GIS), curadoria de dados, gerenciamento de dados e humanidades digitais, dispondo de bibliotecários no oferecimento desses serviços.

Na Universidade de Michigan a biblioteca disponibiliza um guia com informações sobre todos os aspectos do gerenciamento e administração de dados, incluindo localização, planejamento, organização, documentação, compartilhamento e preservação de dados de pesquisa. O guia é organizado pelo *Research Data Services*, uma rede de serviços oferecidos pela biblioteca para auxiliar o pesquisador

durante todas as fases do ciclo de vida dos dados de pesquisa. Os serviços de dados de pesquisa da biblioteca ajudam os pesquisadores a planejar o gerenciamento, o compartilhamento e a curadoria de dados, a desenvolver Planos de Gerenciamento de Dados (PGDs), na descoberta, acesso e aquisição de diferentes tipos de materiais de pesquisa, incluindo dados, auxilia os pesquisadores a compreender, desenvolver e aplicar estratégias para organizar e gerenciar seus dados. Os recursos de pesquisa e guias de ajuda são criados por bibliotecários especialistas da Biblioteca da Universidade de Michigan, de forma que para sanar dúvidas específicas sobre dados de pesquisa ou para agendar um atendimento individualizado, deve-se entrar em contato com o bibliotecário da área por meio da “Lista de funcionários da biblioteca”<sup>72</sup>

Na Universidade da Califórnia - Berkeley, a biblioteca também fornece serviços de dados de pesquisa, orientando os pesquisadores a descobrir, acessar, compartilhar e preservar dados por meio de aquisição e descoberta de conjuntos de dados e oferece instruções e consultas conduzidas por bibliotecários. Seus serviços incluem oficinas e atendimentos individuais que apresentam informações sobre dados espaciais, ferramentas e métodos em todos os níveis de uso de Sistemas de Informação Geográfica (GIS), além de esclarecer dúvidas sobre métodos computacionais e digitais.

Aponta-se que a biblioteca da UC Berkeley atua em conjunto com a equipe de consultoria da *Research IT*<sup>73</sup> liderando o Programa de Gerenciamento de Dados de Pesquisa (*Research Data Management Program*) que auxilia professores, funcionários e alunos a gerenciar dados de pesquisa durante todo o processo de pesquisa. Com parceiros em todo o *campus* da instituição, o Programa oferece um serviço independente de disciplinas que apoia os pesquisadores no trabalho com seus dados. O Programa RDM também aborda questões atuais e emergentes de gestão de dados, conformidade com requisitos impostos pelos financiadores e pela Universidade, e redução do risco associado aos desafios da gestão de dados.

As bibliotecas da Universidade de Harvard possuem uma rede na qual fornecem serviços de dados especializados às suas comunidades locais de pesquisa. O *Harvard Library Research Data Services* conecta membros da

---

<sup>72</sup> <https://www.lib.umich.edu/about-us/staff-directory>

<sup>73</sup> <https://research-it.berkeley.edu/>

comunidade de Harvard oferecendo serviços e recursos que abrangem o ciclo de vida dos dados, para garantir que os dados de pesquisa multidisciplinar de Harvard sejam localizáveis, acessíveis, interoperáveis e reutilizáveis (FAIR).

O Serviço de Dados de Pesquisa da Biblioteca de Harvard tem parcerias com bibliotecários de pesquisa, escritórios de computação de pesquisa e outros provedores de serviços que auxiliam, também, a revisar PGDs criados usando o DMPTool.

Na Rede de Bibliotecas de Harvard, chama-se a atenção para a *Countway Library's Publishing & Data Services*<sup>74</sup> que tem desenvolvido programas e grupos de trabalho para oferecer auxílio e suporte de gerenciamento de dados, acesso, disseminação e reprodutibilidade da pesquisa biomédica e em ciências da saúde. Nesse setor, foi formado o *Longwood Medical Area Research Data Management Working Group (LMA RDMWG)*<sup>75</sup> em resposta às necessidades não atendidas identificadas no gerenciamento de dados de pesquisa biomédica e biológica. Dentro desse GT, são mantidos vários subgrupos destinados a trabalhar em projetos e programas com base nas necessidades de gerenciamento de dados de pesquisa dos pesquisadores da área médica de Harvard. O Serviço de Dados da *Countway Library's* é mantido por três bibliotecárias.

Por fim, em Oxford, as bibliotecas são as responsáveis pelo *Bodleian Data Service*<sup>76</sup> que fornece uma gama de serviços para pesquisadores e estudantes da Universidade que precisam fazer uso secundário de estatísticas e dados. Existem várias fontes de apoio e orientação nas bibliotecas Bodleianas para os indivíduos que trabalham com dados de pesquisa, sendo útil quando este se encontra nas fases iniciais da pesquisa, durante a fase de coleta e análise de dados, ou quando a pesquisa está quase concluída. As bibliotecas Bodleian possuem uma rede de bibliotecários especializados que auxiliam por meio de consultas individuais o desenvolvimento de ideias de pesquisa e na sugestão de bancos de dados úteis ou recursos relacionados. A biblioteca também conta com programas de treinamento e ferramentas de autoestudo, e aqui cita-se a série de oficinas oferecidas pela *Bodleian iSkills*<sup>77</sup>, com foco no desenvolvimento de habilidades em descoberta de informações e comunicação científica, abrangendo recursos gerais que podem ser

---

<sup>74</sup> <https://countway.harvard.edu/services/publishing-data-services>

<sup>75</sup> <https://datamanagement.hms.harvard.edu/about-rdmwg>

<sup>76</sup> <https://www.bodleian.ox.ac.uk/collections-and-resources/data>

<sup>77</sup> <https://www.bodleian.ox.ac.uk/ask/workshops#/>

usados em uma ampla gama de disciplinas. Os serviços de dados oferecidos pela Rede de Bibliotecas atuam em parceria com o setor *Research Data Oxford*<sup>78</sup>, site dedicado ao gerenciamento de dados de pesquisa.

A atuação das bibliotecas com serviços de gestão de dados de pesquisa é um esforço internacional que tem tido êxito, conforme identificou-se nas instituições estudadas nesta pesquisa. Tradicionalmente, as bibliotecas têm desempenhado um papel contínuo no fornecimento de serviços para apoiar atividades de pesquisa, bem como na educação para a competência informacional e na consulta de estratégias de pesquisa. À medida que as tecnologias mudaram a forma como os pesquisadores interagem com o acervo físico e virtual da biblioteca, o papel tradicional se ampliou (Bresnahan; Johnson, 2013).

O fato das bibliotecas desempenharem funções que auxiliam na gestão, curadoria, acesso, coleta e compartilhamento de dados, demonstra que elas têm cumprido seu papel de maneira relevante, se adaptando às demandas atuais, e oferecendo serviços de gerenciamento de dados que ajudam pesquisadores e instituições a organizar e tratar grandes volumes de dado de maneira eficiente. Além disso, ao fornecer à sua comunidade o suporte necessário à pesquisa baseada em dados por meio dos serviços de dados disponibilizados em bibliotecas ou setores e departamentos específicos, contribuem também com o trabalho de mediação do bibliotecário e das formas de comunicação que estabelecem com os diferentes públicos que alcançam com esses serviços para atingir o objetivo de ampliar a competência em dados de pesquisa.

A respeito disso, Choudhury *et al.* (2018) declara que as instituições de pesquisa são capazes de promover uma variedade de serviços de dados que se diferenciam em relação a outros tipos de serviços (oficinas, cursos, treinamentos, palestras), em relação à complexidade (básico, intermediário ou avançado), e em relação ao público alvo desses serviços (pesquisadores, alunos de graduação, alunos de pós-graduação, etc.). Com base nisso, conforme identificava-se às unidades responsáveis pelas atividades de competência em dados nas instituições, observou-se também PARA QUEM esses serviços eram oferecidos e direcionados.

Compreende-se que no sistema do ensino superior gerar e consumir dados torna-se inevitável diante do processo de desenvolvimento de pesquisas e busca por

---

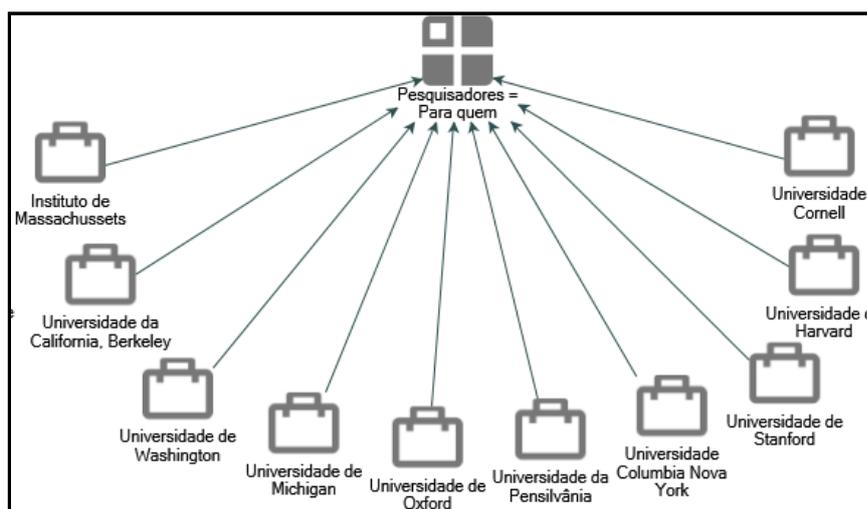
<sup>78</sup> <https://researchdata.ox.ac.uk/>

novos conhecimentos. Pesquisadores, alunos de pós-graduação e graduação, principais agentes produtores e consumidores de dados, têm enfrentado alguns desafios quanto à gestão e partilha de dados de pesquisa, como viu-se ao longo deste estudo.

Conforme informações encontradas nas universidades, as práticas de competência em dados das 10 instituições selecionadas neste estudo tinham como alvos em comum pesquisadores, alunos da pós-graduação e da graduação.

Neste estudo entende-se que o pesquisador e o aluno de pós-graduação desempenham funções diferentes quanto a responsabilidades e atuação na carreira acadêmica que se encontram. Isto é, o pesquisador é aquele que já está estabelecido em sua carreira acadêmica e contribui dentro de sua instituição com projetos, pesquisas, avançando em sua área de estudo já estabelecida. Diferente disso, o aluno de pós-graduação é aquele que está no processo de formação de uma carreira acadêmica ou de pesquisa, na busca da obtenção de seu título de mestrado e doutorado. Por isso, a separação desses dois papéis.

**Figura 33** - Universidades que oferecem serviços de competência em dados para os pesquisadores



**Fonte:** dados da pesquisa.

A medida que os *sites* das bibliotecas, setores e departamentos de serviços de dados eram explorados, identificava-se que os pesquisadores eram citados frequentemente como alvo desses serviços, tendo como principal objetivo dessas unidades qualificá-los para gerenciar, armazenar e compartilhar os dados que produziram, levando em consideração os papéis que desempenham gerando e

consumindo dados, além do fato de necessitar conhecimentos sobre o processo de elaboração de planos de gestão de dados diante das exigências das agências de fomento à pesquisa.

Em relação às abordagens das atividades de capacitação em dados, majoritariamente as atividades voltadas a esse público fornecem informações sobre as exigências das agências de fomento relativas à pesquisa e a melhor forma de respondê-las, levando em consideração o conhecimento prévio dos pesquisadores sobre o assunto. Para os indivíduos que tinham noções básicas sobre gestão de dados de pesquisa e o seu entorno, era possível fornecer informações mais introdutórias, que conceituam os termos. Além das oficinas e dos cursos, por meio dos atendimentos individuais cada pesquisador podia levar para o instrutor (bibliotecários em sua maioria) dúvidas e questões pontuais de acordo com sua área de pesquisa, e assim, o profissional especialista na área, fornecia o atendimento mais qualificado.

A compreensão dos pesquisadores acerca dos dados provenientes de suas pesquisas mudou ao ponto de perceberem que, se corretamente tratados, preservados e gerenciados, os dados não se perdem e podem ser reutilizados por eles mesmos e também por outros pesquisadores, evitando gastos desnecessários de tempo e recursos refazendo e coletando os mesmos dados.

Não só os pesquisadores contribuem com pesquisas, mas também, os alunos de pós-graduação, nas 10 universidades, foram citados como alvo das atividades de competência em dados.

Os alunos de pós-graduação desempenham um papel importante nos ambientes de pesquisa. Eles desenvolvem diversas tarefas de pesquisa, desde a coleta até a análise de dados, de forma que devem estar cientes e capacitados para lidar com os desafios complexos relativos à gestão dos dados de pesquisa. Geralmente espera-se que os pós-graduandos realizem a maioria ou todas as tarefas de gerenciamento de dados para suas próprias pesquisas e frequentemente participem de atividades para apoiar projetos de laboratório e as equipes.

No entanto, de acordo com a pesquisa de Carlson *et al.* (2013) sobre educação em gestão de dados, foi observado que os alunos de pós-graduação compõem uma comunidade em constante circulação e rotatividade de alunos que, infelizmente, chegam com habilidades desiguais nos processos de gestão de dados. A partir das

abordagens adotadas pelas instituições que desenvolvem atividades práticas em dados de pesquisa voltadas a esse público, percebe-se que elas estão alinhadas às necessidades disciplinares de seus alunos de pós-graduação, e já desenvolvem métodos para capacitá-los no que tange aos dados de pesquisa.

Igualmente, os alunos da graduação também têm acesso às ofertas de capacitações em dados por meio dos serviços disponibilizados pelas instituições pesquisadas, com exceção da Universidade de Washington na qual não se identificou atividades de competência em dados voltada aos alunos de graduação.

Destaca-se que no caso dos alunos de graduação que estão iniciando sua jornada científica e acadêmica, as atividades de competência em dados têm outro viés a respeito de como devem desenvolver essas competências. Nas universidades analisadas, observou-se que as atividades de competência em dados voltadas para esse público eram de cunho mais teórico e inicial, com conteúdo que abordam o que são dados de pesquisa, por exemplo, como citar de dados, locais para armazenamento e compartilhamento, documentação dos dados e nomenclaturas, entre outros aspectos.

No estudo de Maybee e Zilinski (2015) é demonstrado que as atividades de competência em dados, no caso de alunos que estão iniciando sua jornada acadêmica, devem ser orientados de acordo com os contextos disciplinares em que estão inseridos em vez de instruídos como em uma habilidade geral, isto é, se o aluno é da área biomédica, ciências humanas ou sociais, é importante que tenha habilidade em gerenciar dados dessas áreas nas quais acontecerão as suas pesquisas. Desta forma, a competência em dados pode não só ajudá-los a melhorar o desempenho acadêmico que está se iniciando, mas, também, fornecer uma base sólida para o seu desenvolvimento profissional, para chegar nos níveis de “pesquisadores” e “pós-graduando” sabendo fazer a curadoria de seus dados.

Destaca-se que a Universidade de Harvard oferece atividades de competência em dados para sua comunidade acadêmica voltada a gestão de dados clínicos para indivíduos da área biomédica da instituição, evidenciando que Harvard já orienta sua comunidade acadêmica a gerenciar seus dados de acordo com as disciplinas, como sugerem Maybee e Zilinski (2015).

Ao mesmo tempo que os serviços, projetos e programas de dados são oferecidos pelas instituições para capacitar esses grupos mencionados, algumas

universidades também disponibilizam seus serviços de dados para funcionários, como é o caso das Universidades de Washington e Cornell, e também para público em geral, no caso das Universidades de Columbia, Harvard, Michigan e Washington.

As áreas de envolvimento das bibliotecas acadêmicas e de pesquisa no fornecimento de serviços de dados de pesquisa abrangem todo o ciclo de vida dos dados, incluindo planejamento, curadoria e criação e conversão de metadados. À medida que a pesquisa orientada por dados se torna mais frequente, o conhecimento prático na gestão de dados é fundamental para os pesquisadores e requer das instituições o desenvolvimento de serviços voltados para sanar as necessidades desses indivíduos.

Conforme visto nas análises das atividades de competência em dados, as bibliotecas são majoritariamente as responsáveis por oferecer serviços de dados de pesquisa, muitas vezes apoiando e auxiliando no desenvolvimento de programas, oficinas, cursos, atendimentos e consultorias para suas comunidades. Observou-se também que o foco de todas as atividades de competência em dados, em nenhum dos casos estudados, contemplou os Bibliotecários, logo, entende-se que eles são os agentes responsáveis por organizar, gerenciar e manter esses serviços. Portanto, a capacitação desses profissionais deve ser realizada em momento anterior, para que quando eles assumem setores e serviços de dados, dentro ou fora da biblioteca, tenham toda a capacidade técnica necessária para orientar e desenvolver a competência em dados dos pesquisadores e da comunidade acadêmica.

As bibliotecas acadêmicas, juntamente com setores e serviços de dados e departamentos de apoio à pesquisa, têm se destacado no fornecimento de apoio à informação e formação de indivíduos em aspectos do ciclo de vida da pesquisa. A missão, o papel e a experiência das bibliotecas universitárias como organizações de informação ao longo dos anos as colocam como unidades primordiais na criação, expansão e avanço dos serviços bibliotecários relacionados com dados.

No estudo de Ashiq e Warraich (2022) os autores colocam que as bibliotecas acadêmicas são consideradas cruciais ao envolver a comunidade de pesquisa por meio dos seus serviços. Considera-se que as possibilidades de oferecer serviços de apoio à pesquisa em bibliotecas universitárias são variadas e esse suporte é fundamental para fomentar a produção de novos conhecimentos.

Verificou-se que os bibliotecários acadêmicos das universidades estudadas estão desempenhando um papel de liderança no desenvolvimento e aprimoramento de competências e apoio na gestão de dados para pesquisadores, estudantes de graduação e pós-graduação. De acordo com Koltay (2012) a necessidade de um melhor desenvolvimento de gestão de dados de pesquisa também deu origem a mais um papel para os bibliotecários, fornecer orientação especializada sobre gestão e preservação de dados.

Observou-se também, que as bibliotecas e os diversos departamentos acadêmicos têm compartilhado e feito parcerias em trabalhos de coleta de requisitos para gestão de dados. As unidades de apoio à pesquisa das instituições têm sido parceiras na avaliação do impacto dos dados de pesquisa e no planejamento e gestão de dados por meio de cooperação com centros de formação e outras unidades com funções semelhantes aos serviços de dados das universidades estudadas.

As universidades têm se empenhado em disponibilizar atividades de capacitação em dados por meio de três principais formas: oficinas, cursos e atendimento individual, auxiliando seus pesquisadores, alunos de graduação e pós-graduação, funcionários e demais interessados a ampliar ou adquirir novas competências em dados de pesquisa. Cada instituição tem desenvolvido práticas que correspondem às demandas e realidade do público, de forma que os conteúdos e temáticas abordadas podem variar de uma instituição para outra.

A criação e implementação de programas, cursos e demais atividades que se dedicam à gestão de dados contribuem para o desenvolvimento de competências apoiadas em dados. Os serviços de dados de pesquisa desenvolvidos pelas universidades analisadas, sendo coordenados, em sua maioria, pelas bibliotecas, demonstra o ponta pé inicial que essas instituições deram para que essas práticas fossem, de fato, aplicadas e cultivadas como essenciais no dia a dia do pesquisador.

Com base nas atividades de competência em dados de pesquisa oferecidas pelas universidades, identificando como, quem e para quem são desenvolvidas foi possível atingir o objetivo específico d) *propor estratégias para desenvolver competências em dados de pesquisa em âmbito nacional*. Buscou-se, a partir da análise dos *sites* das universidades e das bibliotecas encontrar subsídios que pudessem ser úteis para que as universidades brasileiras criarem, nas suas

bibliotecas, um serviço de dados para tornar alunos, pesquisadores e sua comunidade competente em dados.

#### 4.4 ESTRATÉGIAS PARA O DESENVOLVIMENTO DE COMPETÊNCIA EM DADOS DE PESQUISA NAS UNIVERSIDADES BRASILEIRAS

Com base na revisão bibliográfica e nas análises das atividades de competência em dados desenvolvidas pelas 10 universidades pesquisadas, seguem abaixo as considerações sobre as estratégias que as universidades brasileiras podem adotar para desenvolver competências em dados de pesquisa na comunidade acadêmica.

A formulação de uma estratégia tem como objetivo implementar metas, planos, ou ações para alcançar determinado propósito. Observou-se que nas universidades estudadas algumas estratégias foram adotadas para desenvolver as competências em dados de toda a comunidade acadêmica, como por exemplo, a criação de serviços de dados nas bibliotecas para oferecer cursos, oficinas e atendimentos individuais tanto para pesquisadores, alunos de graduação e pós-graduação, como para os funcionários e o público em geral.

Outros setores das instituições, como o Departamento de Pesquisa, da Universidade de Colúmbia - NY, com o Programa *Research and Data Integrity* (ReaDI), a *Cornell Data Services*, e o *Research Data Management Delivery Group* da Universidade de Oxford, também se destacaram no desenvolvimento de serviços de dados com o intuito de preparar a comunidade acadêmica a lidar com os desafios que as pesquisas intensivas em dados trazem.

Verificou-se que as práticas de competência em dados abordaram temas relacionados aos dados de pesquisa, tais como boas práticas da pesquisa científica com orientações para documentar, armazenar e preservar os dados, destacando que essas são práticas que ajudam na reprodutibilidade dos estudos e integridade à pesquisa. Também trouxeram informações a respeito do ciclo de vida dos dados de pesquisa e suas etapas, enfatizando que os dados têm uma vida útil que ultrapassa os projetos de pesquisa, e por isso cada etapa é importante e necessita de ferramentas e habilidades específicas para trabalhar com os dados.

O compartilhamento de dados foi outro tópico muito mencionado nas atividades de capacitação em dados, principalmente em contextos das exigências

das agências de financiamento à pesquisa (NHI e NSF) que solicitam um plano de compartilhamento e gestão de dados. Os Princípios FAIR, repositório de dados, manipulação de dados pessoais, sensíveis e anônimos, descrição de metadados, e o desenvolvimento de PGD's foram temáticas também apresentadas nas práticas de dados de pesquisa nas instituições.

Todos os temas abordados em cada instituição demonstram o comprometimento com a implantação de um conjunto de serviços para desenvolver a competência em dados de suas respectivas comunidades acadêmicas, principalmente no que diz respeito às boas práticas de gestão desses dados. Durante a análise percebeu-se que cada oficina e curso ministrado teve objetivos e estratégias específicas para tornar as comunidades competentes no uso dos dados. Nas universidades internacionais essas atividades de competência em dados já são uma realidade de longa data, e até mesmo uma prática comum entre os profissionais que atuam nessa frente, como por exemplo, os bibliotecários.

No início do estudo, acreditou-se que as atividades de capacitações em dados seriam direcionadas a pesquisadores e bibliotecários. Essa hipótese foi pensada, especialmente, pelo contexto nacional no qual presencia-se que a capacitação, ainda bastante insipiente, de bibliotecários e pesquisadores acontecem simultaneamente nas instituições, e pelo fato de universidades e conseqüentemente as bibliotecas acadêmicas, não disponibilizarem serviços de dados para sua comunidade científica e acadêmica, possivelmente pela falta de conhecimento ou demanda a respeito do assunto.

No entanto, conforme as análises ocorriam, verificou-se que os responsáveis por ministrar, organizar e apoiar os serviços e atividades de capacitação em dados para ampliar as competências da comunidade, eram, em sua maioria, bibliotecários que tinham expertise em gestão de dados. Os cursos, oficinas, e os atendimentos individuais, por exemplo, eram acompanhados e desenvolvidos por bibliotecário de dados, sobretudo, quando se tratava do desenvolvimento e preenchimento de PGD's ou outro assunto relacionado às práticas em dados de pesquisa.

Observou-se também que as bibliotecas universitárias são as responsáveis por desenvolver a maioria das atividades de competência em dados ou estão envolvidas efetivamente no apoio de projetos e programas relacionados à competência em dados. Sua colaboração com outros departamentos e setores das

instituições, são essenciais para o sucesso dessas iniciativas, participando ativamente da organização de serviços de dados oferecidos por outros setores, apoiando e colaborando com o conhecimento que envolve coleta, organização, arquivamento, preservação, acesso e compartilhamento de dados, e dando suportes às atividades que envolvem a pesquisa científica.

A exemplo disso, cita-se a Universidade de Harvard, com o programa de gerenciamento de dados de pesquisa da área médica, desenvolvido pela *Library Publishing & Data Services*, com suas atividades voltadas para práticas que envolvem dados biológicos, biomédicos e clínicos. Os bibliotecários ministravam as atividades sobre gestão de dados de pesquisa, armazenamento, preservação, principalmente assuntos relacionados à ética nas pesquisas quando envolvem dados pessoais e sensíveis, bem como ações de compartilhamento desses dados. Contudo, sobre temas como implementações de modelos para visualização de dados, desenvolvimento de algoritmos de *machine learning*, programação e conhecimento estatístico, entre outros aspectos mais técnicos, contam com profissionais de outros departamentos, especialistas de diversas áreas, para desenvolver as atividades.

O bibliotecário enquanto profissional da informação está apto para atuar e desenvolver suporte e assistência a pesquisadores e outros interessados no que concerne a gestão e utilização de dados por meio de orientações, capacitações e consultorias. A biblioteca como unidade de informação, e como àquela que sempre facilitou o acesso à informação e aos documentos, hoje, como notou-se no decorrer do estudo, têm facilitado o acesso aos dados, sua gestão e todos os processos envolvidos mediante os serviços oferecidos pelos bibliotecários. Segundo Tenopir *et al.* (2011) as bibliotecas universitárias tradicionalmente têm um papel no fornecimento de acesso ao registro da produção acadêmica em muitas formas, e não é surpreendente que a gestão de dados de pesquisa seja uma questão para as bibliotecas acadêmicas se envolverem também.

No contexto brasileiro, as bibliotecas universitárias disponibilizam diversos serviços para estudantes, pesquisadores e docentes. Entre esses serviços destacam-se os empréstimos de livros, acesso a base de dados, serviços de referências e de pesquisa, consultas à catálogo *on-line*, catalogação na publicação, entre outros. Entretanto, em relação ao oferecimento de serviços de dados de

pesquisa, infelizmente, essa não é uma prática comum. De acordo com o estudo de Bonetti e Moreno (2021) intitulado “*Gestão de dados de pesquisa em bibliotecas universitárias brasileiras*”, cujo objetivo foi analisar o cenário das bibliotecas universitárias brasileiras quanto à oferta de serviços de gestão de dados de pesquisa, os serviços de gestão de dados ofertados pelas bibliotecas ainda estão em estágio inicial no país.

Diante disso, ao elaborar estratégias para a competência em dados nas universidades brasileiras, levando em consideração todos os fatores supracitados, será possível que as bibliotecas das universidades brasileiras possam criar serviços de dados que venham a auxiliar pesquisadores, alunos, bibliotecários e a comunidade como um todo a desenvolver competências que permitam trabalhar com a diversidade de processos e etapas que envolvem os dados de pesquisa nas mais variadas áreas do conhecimento.

Para melhor desenvolvimento dessas estratégias, os valores de Atributos apresentados no Quadro 5 da seção 3.3, e os objetivos específicos a) e b) foram essenciais para sua elaboração, iniciando com QUEM poderá desenvolver as atividades de competência em dados, COMO poderão ser desenvolvidas e acontecer de forma a atender às necessidades da comunidade e PARA QUEM as atividades de ampliação de competência em dados podem ser disponibilizadas.

Primeiramente, recomenda-se que na biblioteca universitária seja criado um Núcleo de Serviços de Dados, no qual bibliotecários e pesquisadores de diversas áreas, com habilidades em dados de pesquisa, trabalhem em conjunto para planejar e desenvolver atividades voltadas para a competência em dados da comunidade acadêmica. O principal objetivo desse Núcleo deverá ser capacitar a comunidade científica a lidar com os aspectos que envolvem o planejamento, criação, organização, preservação e compartilhamento dos dados provenientes das pesquisas científicas, fornecendo assistência para atender às expectativas das agências de financiamento à pesquisa (FAPs e CNPq) e às boas práticas de pesquisa.

O Serviço de Dados deve auxiliar a gerenciar as pesquisas durante todo o ciclo de vida do projeto, desde a localização e verificação de fontes de dados até a organização dos dados e o compartilhamento dos resultados das pesquisas. Gerenciamento de dados de pesquisa é um termo amplo que abrange como

organizar, estruturar, armazenar e cuidar das informações digitais geradas ou usadas durante um projeto de pesquisa e de acordo com [Research Data Oxford](#) inclui:

- a) planejar com antecedência como seus dados serão gerenciados;
- b) documentar as práticas de trabalho;
- c) considerar como as informações serão tratadas no dia a dia;
- d) tomar decisões sobre o que acontece com os dados no longo prazo – após a conclusão do projeto;
- e) preparar os dados para serem preservados e
- f) considerar se os dados podem ser reprodutíveis.

O Núcleo de Serviços de Dados poderá oferecer oficinas, cursos, *sites* e portais informativos, bem como consultorias em grupo e/ou individual sobre as diversas etapas que envolvem a gestão de dados de pesquisa e demais assuntos que vimos durante este estudo, principalmente na seção 4.2. Além disso, sugere-se a criação de Grupos de Trabalho no intuito de organizar e distribuir o planejamento das atividades, como por exemplo:

- a) **Grupo de Trabalho 1:** Responsável pelo desenvolvimento de atividades voltadas ao desenvolvimento de planos de gestão de dados, bem como, consultorias de elaboração e preenchimento de PGD's
- b) **Grupo de Trabalho 2:** Responsável por desenvolver atividades de descrição de metadados, nomenclatura de arquivos e estratégias para organização de dados.
- c) **Grupo de Trabalho 3:** Responsável por desenvolver atividades de preservação de dados, com métodos e estratégias visando o armazenamento dos dados e manutenção a longo prazo.
- d) **Grupo de Trabalho 4:** Responsável por desenvolver atividades sobre acesso e compartilhamento de dados de pesquisa, orientado sobre as boas práticas de compartilhamento, Princípios FAIR e a respeito dos repositórios de dados.

O desenvolvimento das atividades de um Núcleo de Serviços de Dados necessitará de colaboração de outros profissionais da instituição, sendo cruciais para o êxito dessas atividades a participação de profissionais da Tecnologia da

Informação para auxiliar na criação dos *sites*, portais e repositórios, na utilização de ferramenta de análise e visualização de grandes volumes de dados.

As Pró-Reitorias de Pesquisa e de Pós-graduação poderão auxiliar na criação de políticas institucionais de acesso aberto, ciência aberta, acesso aos dados e compartilhamento, entre outras. É fundamental a colaboração interdepartamental da biblioteca e Núcleo de Serviço de Dados com outros setores, departamentos e unidades acadêmicas.

Também, necessita-se delimitar o público alvo das atividades. Por se tratar de um assunto com muitas especificidades e que exige um preparo especializado por parte dos profissionais dedicados à tarefa de tornar outros profissionais competentes em dados de pesquisa, sugere-se que tais práticas sejam voltadas, preferencialmente, para pesquisadores, corpo docente da instituição e acadêmicos de graduação e pós-graduação. Diferente das Universidades de Cornell, Washington, Michigan, Harvard e Columbia - NY, que disponibilizam suas atividades de competência em dados para funcionários e público em geral. Assim, as atividades devem ter como objetivo atender às necessidades das pessoas que trabalham diretamente com os dados, pelo teor científico da abordagem dos conteúdos que deverá ser direcionado para o uso, coleta, documentação, tratamento, preservação e reuso dos dados em contextos das pesquisas científicas.

Após identificar QUEM realizará o planejamento e colocará em prática as atividades de competência em dados, COMO essas atividades serão concretizadas, e PARA QUEM serão direcionadas, apresenta-se uma seleção de temas relacionados a dados de pesquisa que podem ser abordados nas atividades de competência em dados:

**a) Coleta de dados de pesquisa:**

- plano de gestão de dados
- escolha de formatos
- direito de propriedade intelectual
- proteção de dados e ética

**b) Organização de dados de pesquisa:**

- Nomear e organizar arquivos
- Documentação e descrição de Metadados
- Gerenciar referências

**c) Preservação de dados de pesquisa:**

- Cópia de segurança
- Armazenamento e preservação de longo prazo
- Seleção - escolher o que manter
- Compartilhamento - o que é, por que e como compartilhar dados
- Repositórios digitais

**d) Compartilhamento e acesso de dados de pesquisa:**

- Quais dados compartilhar
- Por que compartilhar dados
- Como compartilhar dados
- Declarações de acesso a dados
- Uso de licenças

Destaca-se que a relação de temas acima relacionados é uma recomendação com base nas análises realizadas nos documentos selecionados nas universidades internacionais deste estudo. No caso das instituições brasileiras, para alcançar o objetivo proposto, deve-se inicialmente obter um entendimento junto à comunidade, da situação e o nível de conhecimento sobre os aspectos que envolvem a gestão dos dados de pesquisa durante todo seu ciclo de vida, para então planejar e implementar as práticas de competência em dados de maneira que venham realmente a atender às necessidades dos pesquisadores e demais interessados em ampliar suas competências em dados e auxiliar nas atividades diárias de pesquisa.

Buscou-se apresentar uma estratégia para que as bibliotecas das universidades brasileiras possam iniciar suas práticas de competência em dados, considerando os aspectos abordados no estudo e as análises realizadas. A partir das observações durante a coleta dos dados nos *sites* das universidades, considera-se que é possível que as sugestões apresentadas neste estudo sejam colocadas em prática, visando o desenvolvimento de competências dos pesquisadores, docentes e acadêmicos brasileiros que produzem dados de pesquisa.

Outro fator que deve ser levado em consideração são os conhecimentos dos bibliotecários brasileiros enquanto indivíduos competentes em dados para oferecer esses serviços de dados de pesquisa. No país, institutos de pesquisa como o IBICT, a Fiocruz, a Rede Nacional de Ensino e Pesquisa (RNP), e a Federação Brasileira de Associações de Bibliotecários, Cientistas de Informação e Instituições (FEBAB),

por exemplo, têm desenvolvido uma série de iniciativas com o oferecimento de cursos voltados à qualificação dos profissionais bibliotecários para compreenderem e aplicarem práticas eficientes de gestão de dados e reprodutibilidade da pesquisa.

Enquanto profissionais da informação, compreende-se que os bibliotecários devem dominar essa prática uma vez que se trata da ampliação de suas competências habituais em apoio à pesquisa e à informação, para que continuem desempenhando um papel de agente da informação, e auxiliando a comunidade acadêmica, agora a enfrentar os desafios relacionados à gestão de dados, e aproveitar as oportunidades que as pesquisas baseadas em dados disponibilizam.

## 5 CONSIDERAÇÕES FINAIS

Ao longo da jornada acadêmica e profissional, as pesquisas científicas têm envolvido, muito mais nos últimos anos, a coleta, reutilização e compartilhamento de diferentes tipos de dados. Dados esses que, assim como qualquer outro produto de pesquisa, exigem boas práticas que garantam a qualidade do trabalho no final de um projeto. Construir bons hábitos para lidar com os dados desde o início garantirá que a pesquisa seja mais precisa, reproduzível e potencialmente mais impactante ao longo do tempo. No contexto acadêmico, as universidades são as responsáveis por produzir significativo volume de dados por meio dos estudos e pesquisas de professores, alunos e pesquisadores, e com isso, têm desenvolvido iniciativas para apoiar sua comunidade acadêmica no que se refere às competências em dados de pesquisa.

Este estudo teve como objetivo geral *analisar as práticas adotadas para desenvolver competências em dados de pesquisa em universidades internacionais*. Para tanto, inicialmente identificaram-se as 10 primeiras universidades que desenvolvem atividades de competência em dados de pesquisa utilizando o *Webometrics.info*, ranking acadêmico de Instituições de Ensino Superior edição de janeiro de 2023, conforme proposto no objetivo específico a) *identificar universidades internacionais que adotam práticas para desenvolver competências em dados de pesquisa*. Nessas instituições foi possível conhecer de que forma os conteúdos são abordados nas atividades de competência em dados, e quais os assuntos desenvolvidos a partir dos documentos analisados e assim atingir o objetivo específico b) *identificar as práticas adotadas para desenvolver competências em dados de pesquisa nas universidades internacionais selecionadas*.

Observou-se que as instituições abordaram durante suas atividades assuntos referentes ao manuseio dos dados de pesquisa, isto é, destacaram sobre as boas práticas da pesquisa científica voltadas a documentação, preservação e armazenamento e *backup*, e a importância de os pesquisadores conhecerem a respeito dessas ações. Trataram também aspectos sobre dados pessoais, sensíveis e anonimizados, ética na pesquisa científica, citação de dados, compartilhamento, curadoria de dados, o que constitui as boas práticas de gestão de dados, repositórios de dados e as políticas adotadas e os termos de uso.

Cada assunto desenvolvido nas atividades de competência em dados, de um modo geral, era relativo às práticas que garantem que os dados de pesquisa sejam coletados de maneira correta, organizados, tratados, armazenados, preservados e compartilhados, visando a qualidade e integridade, isto é, correspondia às ações de uma boa gestão de dados de pesquisa. Mas, algumas universidades, optam pelo desenvolvimento de conteúdos mais específicos como compartilhamento de dados, descrição de metadados, citação de dados, entre outros.

De um modo geral, as instituições oferecem à comunidade acadêmica conteúdos relacionados à gestão de dados de pesquisa, abordando os motivos pelos quais deve-se investir tempo e esforço em um bom gerenciamento de dados e apresentando as melhores práticas para implantar nos processos de pesquisa. Os desafios da gestão de dados de pesquisa esbarram com a necessidade de desenvolver habilidades e competências em dados para que os pesquisadores possam ter independência de manusear seus dados e praticar ações que os façam durar a longo prazo, visando, também, a reutilização dos dados por outros pesquisadores, e assim, contribuindo com novas pesquisas e estudos.

Ao procurar identificar as práticas adotadas para desenvolver competências em dados de pesquisa nas universidades internacionais selecionadas observou-se que o desenvolvimento das práticas de pesquisa foram efetuados por meio de cursos, oficinas e atendimento individual, mostrando que as instituições internacionais estão criando um ambiente de pessoas com habilidade em dados, pois no contexto das universidades acadêmicas oferecer recursos e serviços para sua comunidade se capacitar e saber manusear os dados que produzem torna-se essencial para o futuro da pesquisa científica.

Ambientes adequados que contribuam para o desenvolvimento, a aplicação e a integração das competências em dados, são caracterizados por vários elementos-chave, incluindo infraestrutura adequada, setores ou departamentos específicos voltados para atividades de gestão de dados que garantam o desenvolvimento dessas competências. Com base nisto, por meio do objetivo c) *verificar, nas universidades selecionadas, quais são as estruturas de apoio às atividades de desenvolvimento de competências em dados de pesquisa*, observou-se que as universidades possuem serviços e desenvolvem atividades que são oferecidas pela Rede de Biblioteca ou por um setor de Serviço de Dados de Pesquisa. Esses

setores específicos elaboram projetos, programas, e consultorias voltados às necessidades e demandas dos indivíduos da instituição. Tanto as bibliotecas como os setores de serviços de dados de pesquisa, são responsáveis por desempenhar as atividades de orientação e suporte para o avanço das práticas de competência em dados nas universidades analisadas.

As bibliotecas universitárias se destacaram no oferecimento das atividades de competência em dados, participando ativamente da organização dos serviços oferecidos, apoiando e colaborando com o conhecimento que envolve coleta, organização, arquivamento, preservação, acesso e compartilhamento de dados, e demais aspectos que envolvem a pesquisa científica. Destaca-se que os profissionais responsáveis por planejar organizar e desenvolver as atividades são, majoritariamente, bibliotecários com conhecimento em gestão de dados, e especialistas que atuam nessa área.

O fato das bibliotecas universitárias terem ampliado sua visibilidade a partir da atuação nos serviços de gestão de dados de pesquisa, indica que o esforço em tornar a comunidade científica qualificada para lidar com a administração dos dados de suas pesquisas, tem tido êxito. Percebeu-se também, que vários elementos-chave têm contribuído para o êxito das atividades, como infraestrutura adequada, redes de colaboração e suporte, setores e departamentos voltados para a análise e gestão de dados que garantem o desenvolvimento dessas competências. O público alvo atividades de competência compreendem pesquisadores, alunos de pós-graduação, alunos de graduação, e em algumas universidades, funcionários e público em geral.

Com base na identificação e análises desses serviços e atividades oferecidas pelas universidades, foi possível, sintetizar informações de forma a relacionar aspectos que podem ser usados para criar serviços de dados de pesquisa em âmbito nacional para que possam auxiliar pesquisadores, alunos, e a comunidade para se tornarem cada vez mais competente em dados de pesquisa, conforme apresentado na seção 4.4.

Como dificuldades das concepções metodológicas desta pesquisa, encontraram-se alguns contratemplos, principalmente em relação à busca exploratória por informações relacionadas às atividades de competência em dados nos *sites* das instituições, e ao uso do *software* NVivo utilizado para análise e

sistematização do conteúdo dos documentos selecionados nas Universidades da amostra.

O acesso às informações nos *sites* das universidades foi realizado manualmente recorrendo à leitura extensiva e observação. O conteúdo sobre as práticas de competência em dados, em algumas universidades, não foi de fácil localização. A recuperação de informações, principalmente, sobre as estruturas de apoio às práticas de competência em dados foi dificultada pelo desconhecimento da estrutura organizacional de cada instituição e em algumas situações não foi possível acesso direto ao *site* das bibliotecas e explorar os serviços de dados que era fornecido, pois nem sempre era suficientemente claro que as bibliotecas disponibilizam as atividades e serviços de desenvolvimento de competência em dados.

Já em relação ao Nvivo, o *software* foi escolhido devido à quantidade de documentos coletados, 146 no total, e necessidade de uma ferramenta que possibilitasse compreender e visualizar graficamente as nuances e contextos dos documentos coletados. A grande vantagem ao utilizar esse *software* foi que ele atendeu às exigências de codificação, classificação, cruzamentos e visualização dos dados, permitindo o retorno constante aos dados da análise, e assim, podendo descrever e observar COMO as atividades de competência em dados ocorrem nas instituições, PARA QUEM são disponibilizados os serviços, e QUEM são os responsáveis pelo planejamento, organização e desenvolvimento.

O tempo dispendido para aprender a manusear o *software* e extrair dele mais gráficos e recursos visuais para a análise, foi necessário para ajudar na organização dos dados e codificação dos textos identificando temas e padrões. Devido à limitação de tempo, não foi possível conhecer mais da ferramenta e aprofundar a utilização dos recursos que ela oferece, mas até onde necessário, ela mostrou-se eficiente.

O desenvolvimento dessa pesquisa mostrou que os bibliotecários podem, e devem, apropriar-se de conhecimento para oferecer serviços de gestão de dados de pesquisa, levando em consideração, principalmente, seu papel de agente da informação que facilita e conduz ao acesso da informação, e no caso dos dados de pesquisa, o acesso e capacitação para que os indivíduos possam aperfeiçoar suas práticas de gestão de dados.

Outrossim, por meio dessa pesquisa, observou-se a participação e envolvimento dos bibliotecários no oferecimento e planejamento de serviços de dados de pesquisa nas universidades internacionais selecionadas, sendo considerados imprescindíveis na geração de conhecimento científico e na orientação durante todo o ciclo de vida das pesquisas. No entanto, têm-se que considerar que, no Brasil, a realidade de alguns profissionais bibliotecários é incompatível com a realidade apresentada nas universidades estudadas, sendo necessário capacitações para que possam adquirir habilidades para cumprir seu papel como agentes da informação nesse contexto. Visando “corrigir” esse déficit, institutos de pesquisa como o IBICT, Fiocruz, FEBAB, e Universidades como a USP, UFRGS, UFSC, têm buscado fortalecer os bibliotecários universitários em suas práticas de gestão de dados por meio de iniciativas auxílio na implementação de repositórios, oferta de cursos para capacitação de profissionais e pesquisadores, e demais atividades desenvolvidas.

Como sugestão para futuros estudos, recomenda-se investigar as atividades desenvolvidas pelos institutos e associações de pesquisa brasileiras que desenvolvem cursos para pesquisadores e profissionais voltados à gestão de dados de pesquisa, como por exemplo, o que são trabalhados, ementa, objetivo, público participante, instituições que esse público pertence, interesse e motivo das capacitações. Além disso, investigar como as bibliotecas universitárias estão se preparando para enfrentar os desafios e ampliar a competência em dados de suas comunidades acadêmicas.

Sugere-se, também, ampliar este estudo e comparar com dados de universidades europeias, como elas atuam ao desenvolverem atividades de competência em dados para sua comunidade acadêmica.

## REFERÊNCIAS

- ALMEIDA, F. G. **Suporte à gestão de dados de pesquisa**: uma ampliação dos serviços oferecidos pelas bibliotecas. 2019. 249 f. Tese (doutorado em Gestão & Organização do Conhecimento) – Programa de Pós-Graduação em Gestão & Organização do Conhecimento, Escola de Ciência da Informação, Universidade Federal de Minas Gerais, 2019. Disponível em: [https://repositorio.ufmg.br/bitstream/1843/31596/4/Tese\\_PPGGOC\\_Fernanda%20Gomes%20Almeida\\_11out.pdf](https://repositorio.ufmg.br/bitstream/1843/31596/4/Tese_PPGGOC_Fernanda%20Gomes%20Almeida_11out.pdf). Acesso em: 25 fev. 2023.
- AMERICAN LIBRARY ASSOCIATION. **Framework for information literacy for higher education**. Chicago: ALA, 11 jan. 2016. Disponível em: [http://www.ala.org/acrl/sites/ala.org.acrl/files/content/issues/infolit/Framework\\_ILHE.pdf](http://www.ala.org/acrl/sites/ala.org.acrl/files/content/issues/infolit/Framework_ILHE.pdf). Acesso em: 11 maio 2023.
- ANDRADE, S. B. M. **Competências necessárias ao bibliotecário curador de dados de pesquisa**. 2023. 83 f. Dissertação (Mestrado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2023. Disponível em: <https://lume.ufrgs.br/handle/10183/264017>. Acesso em: 30 jul. 2024.
- ARAÚJO, D. G; DIAS, G, A; AUTRAN, M. M. M. Compartilhamento de dados no contexto da ciência brasileira: um estudo integrativo. **Informação & Informação**, Londrina, v. 26, n. 3, p. 100–124, jul./set. 2021. DOI: <https://doi.org/10.5433/1981-8920.2021v26n3p100>. Acesso em: 30 jul. 2024.
- ASHIQ, M; WARRAICH, N. F. A systematized review on data librarianship literature: Current services, challenges, skills, and motivational factors. **Journal of Librarianship and Information Science**: v. 55, n. 2, p. 414-433, 2023. Disponível em: <https://journals.sagepub.com/doi/pdf/10.1177/09610006221083675>. Acesso em: 20 jun. 2024.
- AVENTURIER, P. **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 fev. 2023.
- BALBINOTTI, S. *et al.* Competência em dados: uma necessidade contemporânea para pesquisadores e para a sociedade. **AtoZ: Novas práticas em informação e conhecimento**, [S.l.], v. 11, p. 1 - 11, ago. 2022. ISSN 2237-826X. Disponível em: <https://revistas.ufpr.br/atoz/article/view/83569>. Acesso em: 11 maio 2023.
- BALL, A. **Review of data management lifecycle models**. Bath: University of Bath, 2012. Disponível em: <https://purehost.bath.ac.uk/ws/portalfiles/portal/206543/redm1rep120110ab10.pdf>. Acesso em: 25 fev. 2023.
- BARROS, S. *et al.* O ciclo de vida dos dados no contexto da pesquisa em user experience. **Encontros Bibli**: v. 28, pp. 1-19, 2023. Disponível em: <https://www.scielo.br/j/eb/a/LVRCb7Qr8Ld7GsM8FWZd6SH/#>. Acesso em: 26 jun. 2024.

BARROS, A. J.S.; LEHFELD, N. A.S. Fundamentos de metodologia científica. 3 ed. São Paulo: Pearson Prentice Hall, 2007.

BATISTA, A. F. M. **Alfabetização em dados é competência essencial para qualquer profissão**. In: Insper, 2022. Disponível em: <https://www.insper.edu.br/noticias/alfabetizacao-em-dados-e-competencia-essencial-para-qualquer-profissao/>. Acesso em: 28 jan. 2023.

BELLUZZO, R. C. B.; SANTOS, C. A.; ALMEIDA JÚNIOR, O. F. A Competência em informação e sua avaliação sob a ótica da mediação da informação: reflexões e aproximações teóricas. **Informação & Informação**, Londrina, v. 19, n.2, p. 60 - 77, maio/ago. 2014. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/19995>. Acesso em: 25 maio 2023

BERTIN, P. R. B; VISOLI, M. C.; DRUCKER, D. P. 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://periodicos.ufba.br/index.php/revistaici/article/view/21449/15200>. Acesso em: 24 fev. 2023.

BONETTI, L. G.; ARAKAKI, A. C. S. Desafios para a implementação de dados FAIR. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 29., 2022, *online*. **Anais** [...]. *Online*, 2022, p. 1-20. Disponível em: <https://portal.febab.org.br/cbbd2022/article/view/2455>. Acesso em: 25 maio 2023.

BORGMAN, C. L. The Conundrum of Sharing Research Data. **Journal of the American Society for Information Science and Technology**, Hoboken, v. 63, n. 6, p. 1059-1078, June, 2012. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.22634>. Acesso em: 22 jan. 2023.

BRANDÃO, T.; MOREIRA, A.; TANQUEIRO, S. R. As políticas de acesso aberto: história, promessas e tensões. **Ler História**, Portugal, n. 78, p. 253- 276, 2021. Disponível em: <https://journals.openedition.org/lerhistoria/8560>. Acesso em: 12 jan. 2023.

BRESNAHAN, M.; JOHNSON, A. Assessing scholarly communication and research data training needs. **Reference Services Review**: v. 4, n. 3, pp. 413-433, 2013. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/RSR-01-2013-0003/full/pdf?title=assessing-scholarly-communication-and-research-data-training-needs>. Acesso em: 16 jul. 2024.

BROCHU, L; BURNS, J. Librarians and research data management—a literature review: Commentary from a senior professional and a new professional librarian. **New Review of Academic Librarianship**, Londres v. 25, n. 1, pp. 49-58. 2019.

BUDAPEST Open Access Initiative. Portuguese (Brazilian) Translation. **Dez anos da Iniciativa de Budapeste em Acesso Aberto**: a abertura como caminho a seguir. Disponível em: <https://www.budapestopenaccessinitiative.org/boai-10-translations/portuguese-brazilian-translation>. Acesso em: 12 jan. 2023.

BURNES, T.; MANN, E.; NEVILLE, T. Exploring data literacy via a librarian-faculty learning community: a case study. **The Journal of Academic Librarianship**, Amsterdã, v. 45, jan. 2020. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0099133319304069?via%3Dihub#s0090>. Acesso em: 27 set. 2022.

BERKELEY UNIVERSITY OF CALIFORNIA. **About Berkeley**. 2024. Disponível em: <https://www.berkeley.edu/about/>. Acesso em: 13 maio 2024.

CALISKAN, A.; DANGWAL, S; DANDEKAR, T. Metadata integrity in bioinformatics: Bridging the gap between data and knowledge. **Computational and Structural Biotechnology Journal**, Amsterdã v. 21, p. 4895-4913, 2023. DOI: <https://doi.org/10.1016/j.csbj.2023.10.006>. Acesso em: 10 jul. 2024.

CAMPELLO, B. S. O movimento da Competência Informacional: uma perspectiva para o letramento informacional. **Ciência da Informação**, Brasília, v. 32, n. 3, p. 28-37, set./dez. 2003. Disponível em: <https://revista.ibict.br/ciinf/article/view/986/1027>. Acesso em: 11 maio 2023.

CAREGNATO, S. E. O desenvolvimento de habilidades informacionais: o papel das bibliotecas universitárias no contexto da informação digital em rede. **Revista de Biblioteconomia & Comunicação**, Porto Alegre, v. 8, p. 47-55. Disponível em: <https://cedap.ufrgs.br/xmlui/bitstream/handle/20.500.11959/137/v8a3.pdf?sequence=4>. Acesso em: 11 maio 2023.

CARMI, E.; YATES, S.J; LOCKLEY, E. Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. **Internet Policy Review**, Berlim, v. 9, n. 2, mai. 2020. Disponível em: <https://policyreview.info/pdf/policyreview-2020-2-1481.pdf>. Acesso em: 2 fev. 2023.

CARVALHO, É. R. S. **Diferenças na produção, compartilhamento e (re)uso de dados**: a percepção de pesquisadores da química, antropologia e educação. 2018. 173 p. Dissertação (Mestrado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, 2018. Disponível em: [https://repositorio.unb.br/bitstream/10482/32559/1/2018\\_%C3%89rikaRayanneSilvadeCarvalho.pdf](https://repositorio.unb.br/bitstream/10482/32559/1/2018_%C3%89rikaRayanneSilvadeCarvalho.pdf). Acesso em: 27 jan. 2023.

CARLSON, J. *et al.* The International Journal of Digital Curation. **International Journal of Digital Curation**, Edimburgo, v. 8, n. 1, p. 204–217, 2013. Disponível em: <https://ijdc.net/index.php/ijdc/article/view/8.1.204/306>. Acesso em: 23 jun. 2024.

CHOUDHURY, S. *et al.* **Research Data Curation A Framework for an Institution-Wide Services Approach**. Educause Working Group Paper: Louisville, CO: ECAR, May 2018. Disponível em: <https://library.educause.edu/-/media/files/library/2018/5/ewg1803pdf.pdf>. Acesso em: 18 jun. 2024.

COLUMBIA UNIVERSITY. **History of Columbia University**. 2024. Disponível em: <https://www.columbia.edu/content/university-history>. Acesso em: 13 maio 2024.

CONCEIÇÃO, G. C.; NHACUONGUE, J. A. A privacidade e os dados pessoais na Ciência da Informação: uma análise de domínio. **Liinc em Revista**, Rio de Janeiro, v. 19, n. 2, p. 1-22, nov./2023. Disponível em: <https://revista.ibict.br/liinc/article/view/6471/6389>. Acesso em: 30 jul. 2024.

CONRAD, L. Y.; URBERG, M. **The Experience of Good Metadata: Linking Metadata to Research Impacts**. The Scholarly Kitchen, 2021. Disponível em: <https://scholarlykitchen.sspnet.org/2021/09/30/the-experience-of-good-metadata-linking-metadata-to-research-impacts/>. Acesso em: 15 jul. 2024.

CORNELL UNIVERSITY. **About Cornell**. 2024. Disponível em: <https://www.cornell.edu/about/>. Acesso em: 1 maio 2024.

COSTA, M. P.; LEITE, F. C. L. Acesso Aberto no mundo e na América Latina: uma revisão a partir da BOAI. **Transinformação**, Campinas, v. 28, n.1, p. 33-46, 2016. Disponível em: [https://repositorio.unb.br/bitstream/10482/19433/2/AcessoAbertoMundo\\_%202016.pdf](https://repositorio.unb.br/bitstream/10482/19433/2/AcessoAbertoMundo_%202016.pdf). Acesso em: 12 jan. 2023.

COSTA, S. M S. Filosofia aberta, modelos de negócios e agências de fomento: elementos essenciais a uma discussão sobre o acesso aberto à informação científica. **Ciência da Informação**, Brasília, v. 35, n. 2, p. 39-50, maio/ago. 2006. Disponível em: <https://www.scielo.br/j/ci/a/XsgXnnC7xWHNR7gXrP9Hw3M/?format=pdf&lang=pt>. Acesso em: 12 jan. 2023.

COSTAL, M. **Competência em dados: mapeamento e análise das ações de apoio à pesquisa em bibliotecas universitárias dos Países Baixos e Reino Unido**. 2021. 155 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro. Disponível em: [https://ridi.ibict.br/bitstream/123456789/1191/1/PPGCI\\_IBICT\\_UFRJ\\_Dissertacao\\_CostalMarcelle\\_2021.pdf](https://ridi.ibict.br/bitstream/123456789/1191/1/PPGCI_IBICT_UFRJ_Dissertacao_CostalMarcelle_2021.pdf). Acesso em: 10 maio 2024.

COSTAL, M.; SALES, L.; ZATTAR, M. Proposta de classificação para promoção da competência em dados de pesquisa. *In*: TOGNOLI, Natália Bolfarini; ALBUQUERQUE, A. C.; CERVANTES, Brígida Maria Nogueira (Org.). **Organização e representação do conhecimento em diferentes contextos: desafios e perspectivas na era da datificação**. Londrina, ISKO Brasil: PPGCI-UDEL, 2023. p. 481-493. Disponível em: [https://isko.org.br/wp-content/uploads/2023/06/livro-isko-Brasil\\_23.pdf](https://isko.org.br/wp-content/uploads/2023/06/livro-isko-Brasil_23.pdf). Acesso em: 04/ jun. 2024.

COSTAL, M.; SALES, L.; ZATTAR, M. Competência em dados: habilidades na atuação e formação do bibliotecário. **BIBLOS**, Florianópolis, v. 34, n. 2, 2020. DOI: 10.14295/biblos.v34i2.11809. Disponível em: <https://periodicos.furg.br/biblos/article/view/11809>. Acesso em: 25 abr. 2023.

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: 24 fev. 2023.

DATA OBSERVATION NETWORK FOR EARTH – DataONE. Best practices. Disponível em: [https://repository.oceanbestpractices.org/bitstream/handle/11329/502/DataONE\\_BP\\_Primer\\_020212.pdf?sequence=3&isAllowed=y](https://repository.oceanbestpractices.org/bitstream/handle/11329/502/DataONE_BP_Primer_020212.pdf?sequence=3&isAllowed=y). Acesso em: 25 maio 2023.

DIGITAL CURATION CENTRE (DCC) Plan management Plan. Disponível em: <https://www.dcc.ac.uk/dmps>. Acesso em: 18 ago. 2024.

DDI-Lifecycle. 2020. Disponível em: <https://ddialliance.org/Specification/DDI-Lifecycle/#understanding>. Acesso em: 25 fev. 2023.

DEJA, M. *et al.* The Impact of Basic Data Literacy Skills on Work-Related Empowerment: The Alumni Perspective. **College & Research Libraries**, Chicago, v. 82, n. 5, 2021, p. 708-729. Disponível em: <https://crl.acrl.org/index.php/crl/article/view/25016/32878>. Acesso em: 27 set. 2023.

DUDZIAK, E. A. **A Information literacy e o papel educacional das bibliotecas**. 2001. 173f. Dissertação (Mestrado em Ciências da Comunicação). Escola de Comunicação e Artes. Universidade de São Paulo, São Paulo, 2001. Disponível em: <https://www.teses.usp.br/teses/disponiveis/27/27143/tde-30112004-151029/pt-br.ph>. Acesso em: 25 maio 2023.

EDWARDS, F; COWIE, B; TRASK, S. Using colleague coaching to develop teacher data literacy. **Professional development in education**, London, p. 1-14, 2022. DOI: <https://doi.org/10.1080/19415257.2022.2081247>. Acesso em: 30 jul. 2024.

EYNDEN, V. V. D. **Data life cycle & data management planning**. In: Looking after and managing your research data (going digital and ESRC ATN events) UK Data Archive, Colchester. University of Essex: [s.n.], 2013. Disponível em: <https://dam.ukdataservice.ac.uk/media/187718/dmplanningdm24apr2013.pdf>. Acesso em: 25 fev. 2023.

EYNDEN, V. V. D *et al.* **Managing and data sharing: best practice for researchers**. Colchester: University of Essex, 2011. Disponível em: <https://dam.ukdataservice.ac.uk/media/622417/managingsharing.pdf>. Acesso em: 25 fev. 2023.

FARMEL, S.; SHIRI, A. Metadata for Research Data: current practices and trends. **Int'l Conf. on Dublin Core and Metadata Applications: Canadá**, [s.l.], p. 1-9, 2014. Disponível em: <https://dcpapers.dublincore.org/files/articles/952136534/dcmi-952136534.pdf>. Acesso em: 2024.

FERREIRA, A. F.; SILVA, V. B. produção científica: Conceitos, iniciativas e fatores complicadores. In: CADERNO DE RESUMOS ELETRÔNICO DO ENEBD, 34., 2012, GT4 Temática livre. Manaus. **Resumos** [...]. Manaus: Universidade Federal do Amazonas, 2012. Disponível em: 64  
<https://www.periodicos.ufam.edu.br/index.php/enebd2011/article/view/45>. Acesso em: 12 jan. 2023.

FLEURY, A.; FLEURY, M. T. L. **Estratégias empresariais e formação de competências**: um quebra-cabeça caleidoscópico da indústria brasileira. São Paulo: Atlas, 2000. p. 17-33

FLICK, U. **Introdução à pesquisa qualitativa**. 3 ed. São Paulo: Artmed, 2008.

GABRIEL JÚNIOR, R. F. *et al.* Acesso aberto a dados de pesquisa no Brasil: mapeamento de repositórios, práticas e percepções dos pesquisadores e tecnologias. **Ci.Inf.**, Brasília, DF, v.48 n.3 (Supl.), p.87-101, set./dez. 2019. Disponível em: <https://lume.ufrgs.br/handle/10183/212266>. Acesso em: 26 fev. 2023.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GONZÁLEZ, Y. M.; RODRÍGUEZ, A. I. Alfabetización en datos: Diseño de un nuevo escenario formativo para el contexto universitario. **Revista Ibero-americana de Ciência da Informação**, Brasília, v. 14, n. 1, p. 318 -330, jan./abril. 2021. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/35521/28759>. Acesso em: 26 jan. 2023.

GOODMAN, Alyssa *et al.* Ten Simple Rules for the Care and Feeding of Scientific Data. **PLoS Comput Biol**, San Francisco, v. 10, n.4, 2014. Disponível em: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003542>. Acesso em: 31 maio 2024.

GUANAES, P. C. V. **Modelos de gestão de periódicos científicos eletrônicos em acesso livre**: estudo para um modelo de gestão sustentável na área de saúde pública. 2011. 120 f. Dissertação (Mestrado em Informação e Comunicação em Saúde) - Programa de Pós-Graduação em Informação e Comunicação em Saúde, Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, 2011. Disponível em: <https://www.arca.fiocruz.br/bitstream/icict/6185/1/Disserta%C3%A7%C3%A3o%20Paulo%20Guaaes.pdf>. Acesso em: 12 jan. 2023.

GUANDALINI, C. A.; FURNIVAL, A. C. M.; ARAKAKI, A. C. S. Boas práticas científicas na elaboração de planos de gestão de dados. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 17, p. 1-20, 2019. Disponível em: <https://brapci.inf.br/index.php/res/download/125743>. Acesso em: 25 maio 2023.

GUIMEIRO, K. A.; COSTA, S. M. S. O uso de modelos de negócios por editoras de periódicos científicos eletrônicos de acesso aberto. **Perspectivas em Ciência da Informação**, Belo Horizonte, MG, v.17, n.4, p.100- 122, out./dez. 2012. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1489/1069>. Acesso em: 12 jan. 2023.

HAENDEL, Melissa A.; VASILEVSKY, Nicole A.; WIRZ, Jacqueline A. Dealing with Data: A Case Study on Information and Data Management Literacy. **PLoS Biol**, San Francisco, v. 10, n. 5, 2012, 5 p. Acesso em: 10 jun. 2024. Disponível em: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001339>.

HENDERSON, Jessa; CORRY, Michael. Data literacy training and use for educational professionals. **Journal of Research in Innovative Teaching & Learning**, Bingley, v. 14, n. 2, 2021, p. 232-244. Acesso em: 10 jun. 2024. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/JRIT-11-2019-0074/full/pdf?title=data-literacy-training-and-use-for-educational-professionals>.

HARVARD UNIVERSITY. **History**. 2024a. Disponível em: <https://www.harvard.edu/about/history/timeline/#1600s>. Acesso em: 01 maio 2024.

HARVARD UNIVERSITY. **Research Data Management**. 2024b. Disponível em: <https://research.harvard.edu/research-policies-compliance/research-data-management/#federal>. Acesso em: 01 maio 2024.

HATSCHBACH, M. H. L. **Information Literacy**: aspectos conceituais e iniciativas em ambiente digital para o estudante de nível superior. 2002. 109 f. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2002. Disponível em: <https://ridi.ibict.br/bitstream/123456789/722/1/mariahelena2002.pdf>. Acesso em: 25 maio 2023.

HENNING, P. C.; RIBEIRO, C. J. S.; SALES, L. F.; MOREIRA, J. L. R.; SANTOS, L. O. B. Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. **Pesq. Bras. em Ci. da Inf. e Bib.**, João Pessoa, v. 14, n. 3, p. 175-192, 2019a. Disponível em: <https://periodicos.ufpb.br/index.php/pbcib/article/view/46969>. Acesso em: 27 jan. 2023.

HENNING, P. C. *et al.* GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. **Em Questão**, Porto Alegre, v. 25, n. 2, p. 389-412, maio/ago. 2019b. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/84753>. Acesso em: 27 jan. 2023.

HIGGINS, S. The DCC Curation Lifecycle Model. **The International Journal of Digital Curation**, Edimburgo, v. 3, n. 1, p. 134-140, 2008. Disponível em: <http://www.ijdc.net/article/view/69/48>. Acesso em: 24 fev. de 2023.

HJØRLAND, B. Data: with *Big Data* and database semantics. *In: ISKO Encyclopedia of Knowledge Organization*. Disponível em: <http://www.isko.org/cyclo/data>. Acesso em: 04 jun. 2023.

HUMPHREY, C. **e-Science and the Life Cycle of Research**. Edmonton: [s.n.], 2006. Disponível em: <https://era.library.ualberta.ca/items/3334684b-fa6a-4c9d-a74b-559fecd42f9f>. Acesso em: 24 fev. 2023.

JACOBSEN, P. S.; MILETTO, E. M.; LOUREIRO, C. B. Pesquisa científica na pós-graduação: elementos da competência informacional em formato de objeto de aprendizagem. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.27, n. 1, p. 3-26, jan/mar 2022. Disponível em: <https://www.scielo.br/j/pci/a/trjV4QbLMgXRnQgSFS764Rp/?format=pdf&lang=pt>. Acesso em: 25 maio 2023.

KJELVIK, M. K.; SCHULTHEIS, E. H. Getting Messy with Authentic Data: Exploring the Potential of Using Data from Scientific Research to Support Student Data Literacy. **CBE-Life Sciences Education**, Bethesda, v.18, n. 2, p. 1–8, 2019. Disponível em: <https://www.lifescied.org/doi/pdf/10.1187/cbe.18-02-0023>. Acesso em: 27 set. 2022.

KOLTAY, T. Data literacy for researchers and data librarians. **Journal of Librarianship and Information Science**, London, v. 49 n. 1, p. 3-14. 2017a. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/0961000615616450>. Acesso em: 22 jan. 2023.

KOLTAY, T. The bright side of information: ways of mitigating information overload. *Journal of Documentation*, **Bingley**, West Yorkshire, v. 73, n. 4, p. 767-775, 2017b. DOI: <https://doi.org/10.1108/JD-09-2016-0107>. Acesso em: 11 maio 2023.

KOLTAY, T. Data literacy: in search of a name and identity. *Journal of Documentation*, **Bingley**, West Yorkshire, v. 71, n. 2, p. 401-415, 2015. DOI: <https://doi.org/10.1108/JD-02-2014-0026>. Acesso em: 11 maio 2023.

KOLTAY, T. Data governance, data literacy and the management of data quality. **IFLA Journal**, Haia, Holanda, v. 42. v.4, p. 303-312, 2016. Disponível em: <https://journals.sagepub.com/doi/epub/10.1177/0340035216672238>. Acesso em: 13 mar. 2023.

LIMA J. S; PINTO V. B; FARIAS, M. G. O bibliotecário na gestão de dados de pesquisa: uma revisão sistemática. **Em Questão**, Porto Alegre, v. 26, n. 3, p. 43-69. Disponível em: <https://www.redalyc.org/articulo.oa?id=465664724003>. Acesso em: 3 fev. 2023.

LYON, L. **Dealing with data: role, rigths, responsibilities and relationships** consultancy report. p. 1-65, jun. 2007. Disponível em: [https://purehost.bath.ac.uk/ws/portalfiles/portal/419529/dealing\\_with\\_data\\_report-final.pdf](https://purehost.bath.ac.uk/ws/portalfiles/portal/419529/dealing_with_data_report-final.pdf). Acesso em: 22 jan. 2023.

MACHADO, J. A. S. Difusão do conhecimento e inovação - o Acesso Aberto a publicações científicas. In: Baumgarten, M. (org). **Conhecimento e Redes - Sociedade Política e Inovação**. Porto Alegre: Ed. UFRGS, 2005. Disponível em: [http://each.usp.br/machado/art04-05/jorge\\_machado-acesso\\_aberto.html](http://each.usp.br/machado/art04-05/jorge_machado-acesso_aberto.html). Acesso em: 12 jan. 2023.

MATA, M. L. **A inserção da Competência Informacional nos currículos dos cursos de Biblioteconomia no Brasil e nos cursos de Informação e documentação na Espanha**. 2014. 197 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista - Campus de Marília. Disponível em: <https://repositorio.unesp.br/bitstream/handle/11449/110393/000792902.pdf?seq>. Acesso em: 25 maio 2023.

MAYBEE, C.; ZILINSKI, L. Data informed learning: a next phase data literacy framework for higher education. **Association for Information Science and Technology**, Bethesda, v. 52, n. 1, pp. 1-4, 2015. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/pra2.2015.1450520100108>. Acesso em: 26 jun. 2024.

MIT FACTS. **Origins**. 2024. Disponível em: <https://facts.mit.edu/origins/>. Acesso em: 01 maio 2024.

MIT LIBRARIES. **Mit Research Data Principles**. 2024. Disponível em: <https://libraries.mit.edu/data-management/mit-research-data-principles/>. Acesso em: 01 maio 2024.

MONTEIRO, E. C. S. A. **Direitos autorais nos repositórios de dados científicos: análise sobre os planos de gerenciamento dos dados**. 2017. 115 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de filosofia e Ciências, Universidade Estadual Paulista, Marília, 2017. Disponível em: <http://hdl.handle.net/11449/149748>. Acesso em: 24 fev. 2023.

MONTEIRO, E. C. S. A.; SANT'ANA, R. C. G. Plano de gerenciamento de dados em repositórios de dados de universidades. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, [S. l.], v. 23, n. 53, p. 160–173, 2018. DOI: <https://doi.org/10.5007/1518-2924.2018v23n53p160>. Acesso em: 27 jan. 2023.

MOTA, F. R. L. *et al.* Características da produção e comunicação científica dos docentes da Universidade Federal de Alagoas – UFAL. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador, Bahia. **Anais** [...] Salvador, Bahia. GT 7 – Produção e Comunicação da Informação em CT&I, Disponível em: <http://www.enancib.ppgci.ufba.br/artigos/GT7--229.pdf>. Acesso em: 12 jan. 2023.

MOTA, M. **Oficina, Workshop e Curso: qual a diferença?** LinkedIn. Publicado em 21. jul. 2017. Disponível em: <https://pt.linkedin.com/pulse/oficina-workshop-e-curso-qual-diferen%C3%A7a-mika-mota>. Acesso em: 18 jun. 2024.

MOZZATO, A. R.; GRZYBOVSKI, D.; TEIXEIRA, A. N. Análises qualitativas nos estudos organizacionais: as vantagens no uso do software Nvivo®. **Revista Alcance**, Araraquara, v. 23, n. 4., out./dez. 2016. Disponível em: <https://www.redalyc.org/journal/4777/477749961009/477749961009.pdf>. Acesso em: 15 abr. 2024

National Science Board. **Long-Lived Digital Data Collections: enabling research and education in the 21st century.** (2005). Disponível em: <https://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>. Acesso em: 27 jan. 2023.

NAZAROVETS, S.; KULYK, Y. Library 4.0: next generation services and technologies. **Bibliotečnij visnik**, Kiev, v. 5, p. 3-14, jan. 2017.  
NGUYEN, D. Mediatisation and datafication in the global COVID-19 pandemic: on the urgency of data literacy. **Media International Australia**, London, v.178, n. 1, p. 210-214. Disponível em: <https://journals.sagepub.com/doi/pdf/10.1177/1329878X20947563>. Acesso em: 27 set. 2022.

Organização para a Cooperação e Desenvolvimento Econômico. **OECD Principles and Guidelines for Access to Research Data from Public Funding.** OECD, 2007. DOI 10.1787/9789264034020-en-fr. Disponível em: <https://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 22 jan. 2023.

**OHAJI**, I.K.; **CHAWNER**, B.; **YOONG**, P. The role of a data librarian in academic and research libraries. **Information Research**, Suécia, v. 24, n. 4, dez./2019. Disponível em: <https://informationr.net/ir/24-4/paper844.html#swa08>. Acesso em: 30 jul. 2024.

OLIVEIRA, C. C . *et al.* A teoria da citação de dados: uma revisão da produção científica na América Latina. **Transinformação**, Araraquara, v. 34, p. 1- 18, 2022. DOI: <http://doi.org/10.1590/2318-0889202234e210062>. Acesso em: 30 jul. 2024.

OYELUDE, A. A. Trending issues in advancing blockchain technology in libraries, archives and museums. **Library Hi Tech News**, Bingley, v. 39, n. 6, p. 6-7, 2022. DOI: <https://doi.org/10.1108/LHTN-06-2021-0040>. Acesso em: 30 jul. 2024.

PENN LIBRARIES. **Research Data & Digital Scholarship.** 2024. Disponível em: <https://www.library.upenn.edu/rdds>. Acesso em: 14 maio 2024.

PRADO, J. C.; MARZAL, M. Á. Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. **Libri**, Berlim, v. 63, n. 2, p. 123-131. 2013. Disponível em: <https://core.ac.uk/download/pdf/288499712.pdf>. Acesso em: 22 jan. 2023.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do trabalho científico.** Novo Hamburgo: Feevale, 2013.

RAFFAGHELLI, J. E; MANCA, S. Exploring the social activity of open research data on ResearchGate: implications for the data literacy of researchers. **Online Information Review**, Bingley, v. 47, n.1, p. 197-217, 2023. DOI: <https://doi.org/10.1108/OIR-05-2021-0255>. Acesso em: 30 jul. 2024.

REIS, M. J. **Ciência da Informação e Ciência de Dados: guia para alfabetização de dados para bibliotecários**. 2019. Dissertação (Mestrado Profissional em Gestão da Informação e do Conhecimento) – Universidade Federal de Sergipe - UFS. 143 p. 2019. Disponível em: [https://ri.ufs.br/bitstream/riufs/12667/2/MAKSON\\_%20JESUS\\_REIS.pdf](https://ri.ufs.br/bitstream/riufs/12667/2/MAKSON_%20JESUS_REIS.pdf). Acesso em: 22 jan. 2023.

REIS, M. J; SENA, N. C. S. Biblioteconomia de dados e ciência de dados no contexto da e-science. **Revista fontes documentais: Aracajú**, v. 4, 2021, Edição especial: I EBBUP, p. 51-64. Disponível em: <https://cip.brapci.inf.br/download/193856>. Acesso em: 30 jul. 2024

RESEARCH UNIVERSITY MICHIGAN. **About the Research Data Stewardship Initiative**. 2024. Disponível em: <https://research.umich.edu/research-data-stewardship/about-research-data-stewardship/>. Acesso em: 01 maio 2024.

RESENDE, L. C. **A curadoria de dados científicos na Ciência da Informação: levantamento do cenário nacional**. 2019. 134 f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) - Programa de Pós-Graduação em Gestão e Organização do Conhecimento, Universidade Federal de Minas Gerais, Belo Horizonte, 2019. Disponível em: <https://repositorio.ufmg.br/bitstream/1843/32413/1/Disserta%C3%A7%C3%A3o%20CD-ECI-UFMG%20-%20Liliane%20CResende.pdf>. Acesso em: 30 jul. 2024.

RICE, R.; SOUTHALL, J. **The data librarian's handbook**. London: Facet Publishing, 2016. *In: RIDSDALE, Chantel et al.(orgs.)*. Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report. Dalhousie University, Halifax, NS, 2015.

ROBAINA, J. V. L. *et al.* **Fundamentos teóricos e metodológicos da pesquisa em educação em ciências**. Curitiba: Bagai, 2021.

ROCHA, L. L.; SALES, L. F.; SAYÃO, L. F. Descrever para preservar: metadados como ferramenta para gestão de dados de pesquisa. *In: PINHO, F. J.; GUIMARÃES, A. C. (Orgs.)*. **Memória, tecnologia e cultura na organização do conhecimento**. Recife, PE: Ed. UFPE, 2017. Disponível em: <https://cip.brapci.inf.br/download/121928>. Acesso em: 10 jul. 2024.

RDPdata: Tesouro de Dados de Pesquisa. Universidade Federal do Rio Grande do Sul – UFRGS. 2024. Disponível em: <https://www.ufrgs.br/tesauros/index.php/thesa/terms/374>. Acesso em: 18 ago. 2024.

SALES, L. F. **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, 268 p. 2014. Disponível em: <https://ridi.ibict.br/bitstream/123456789/874/1/LUANA%20SALES%20D.pdf>. Acesso em: 22 jan. 2023.

SALES, L. F.; SAYÃO, L. F. 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/view/26337>. Acesso em 22 jan. 2023.

SANTA ANNA, J.; DIAS, C. C.; MACULAN, B. C. 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/19169/16262>. Acesso em: 22 jan. 2023.

SANTOS, A. S.; TOLFO, S. R. Competências demandadas dos bibliotecários frente às novas tecnologias de informação em bibliotecas universitárias 10.5007/1518-2924.2006v11n21p69. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, v. 11, n. 21, p. 69-84, 2006. DOI: 10.5007/1518-2924.2006v11n21p69. Acesso em: 20 mar. 2023.

SANTOS, M.C. C. **Competência em Dados**: mapeamento e análise das ações de apoio à pesquisa em bibliotecas universitárias dos Países Baixos e Reino Unido. 2021. 155 f. Dissertação (Mestrado em Ciência da Informação) - Pós-Graduação em Ciência da Informação, Universidade Federal do Rio de Janeiro, Rio de Janeiro. Disponível em: [http://ridi.ibict.br/bitstream/123456789/1191/1/PPGCI\\_IBICT\\_UFRJ\\_Dissertacao\\_CostalMarcelle\\_2021.pdf](http://ridi.ibict.br/bitstream/123456789/1191/1/PPGCI_IBICT_UFRJ_Dissertacao_CostalMarcelle_2021.pdf). Acesso em: 25 maio 2023.

SAYÃO, L. F.; SALES, L. Curadoria digital: um novo patamar para a preservação de dados digitais de pesquisa. **Inf. & Soc.:Est.**, João Pessoa, v.22, n.3, p. 179-191, set./dez. 2012. Disponível em: <https://www.marilia.unesp.br/Home/Instituicao/Docentes/EdbertoFerneda/curadoria-digital---sayao.pdf>. Acesso em: 24 fev. 2023.

SAYÃO, L. F.; SALES, L. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. Disponível em: <http://www.icb.usp.br/~sbibicb/images/guia%20gestaoPDF/Guia%20de%20gestao%20dados%20de%20pesquisa.pdf>. Acesso em: 24 fev. 2023.

SAYÃO, L. F.; SALES, L. Afinal, o que é dado de pesquisa? **Biblos**: Rio Grande, v. 34, n. 02, p. 32-51, jul./dez. 2020. DOI 10.14295/biblos.v34i2.11875. Disponível em: <https://periodicos.furg.br/biblos/article/view/11875>. Acesso em: 22 jan. 2023.

SEMELEER, A. R. **Ciência da Informação em contextos de E-Science: bibliotecários de dados em tempos de *Data Science***. 2017. Tese (Doutorado) - Universidade federal de Santa Catarina, Programa de Pós-Graduação em Ciência da Informação. 2017. Disponível em:

<https://repositorio.ufsc.br/bitstream/handle/123456789/185593/PCIN0168-T.pdf?sequence=-1&isAllowed=y>. Acesso em: 3 fev. 2023.

SEMELEER, A. R.; PINTO, A. L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. **Ciência da Informação**, Brasília, DF, v. 48, n. 1, 2019. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/111525>. Acesso em: 22 jan. 2023.

SEMIDÃO, R. A. M. **Dados, informação e conhecimento enquanto elementos de compreensão do universo conceitual da ciência da informação: contribuições teóricas**. 2014. 199 f. Dissertação (Mestrado em Ciência da Informação) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências de Marília, 2014. Disponível em: <https://repositorio.unesp.br/handle/11449/110783>. Acesso em: 25 maio 2023.

SESARTIC, A.; TOWE, M. Research Data Services at ETH-Bibliothek. **IFLA Journal**, London, v. 42, n. 4. p. 284-291, mai/2016. Disponível em: <https://journals.sagepub.com/doi/epub/10.1177/0340035216674971>. Acesso em: 24 fev. 2023.

SHINTAKU, M.; APPEL, A. L.; OLIVEIRA, A. F. **Tecnologias para gestão de dados de pesquisa segundo preceitos FAIR**. In: SALES, L. F.; VEIGA, V. S. S. O.; HENNING, P.; SAYÃO, L. F. (orgs.). **Princípios FAIR aplicados à gestão de dados de pesquisa**. Rio de Janeiro: IBICT, 2021. Disponível em: <https://ridi.ibict.br/handle/123456789/1182>. Acesso em: 27 jan. 2023.

SHORISH, Y. Data Information Literacy and Undergraduates: a critical Competency. **College and Undergraduate Libraries**, v. 22, n. 1, p. 97–106, 2015. Disponível em: [doi:10.1080/10691316.2015.1001246](https://doi.org/10.1080/10691316.2015.1001246). Acesso em: 27 jan. 2023.

SILVA, A. R. *et al.* A curadoria de dados de pesquisa sensíveis por profissionais da informação. In: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO, 8., 2019, Paraná. **Anais [...]**. Paraná: UEL, 2019, p. 77-92. Disponível em: <https://www.uel.br/eventos/cinf/index.php/secin2019/secin2019/paper/viewFile/559/368>. Acesso em: 30 jul. 2024.

SILVA, F.C. C.; SILVEIRA, L. O ecossistema da Ciência Aberta. **TransInformação**, [S. l.], v. 31, n. 1, 2022. Disponível em: <https://periodicos.puc-campinas.edu.br/transinfo/article/view/5926>. Acesso em: 26 fev. 2023.

SILVEIRA, J. I. **Acesso aberto a dados de pesquisa em repositórios universitários internacionais: um estudo sobre políticas de depósito, acesso e uso**. 2021. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal do Rio Grande do Sul. 145 p. Disponível em: <https://lume.ufrgs.br/handle/10183/222095>. Acesso em: 22 jan. 2023.

SMALE, N. *et al.* The history, advocacy and efficacy of data management plans. **Cold Spring Harbor Laboratory**, 2018, Australia, 30 p. Disponível em: <https://www.biorxiv.org/content/10.1101/443499v1.full>. Acesso em: 27 jul. 2024.

SOBRE nós. **RDP Brasil - Rede de Dados de Pesquisa**, 2022. Disponível em: <https://www.ufrgs.br/rdpbrasil/wp/sobre-nos/>. Acesso em: 26 fev. 2023.

SOUZA, E. C.; NASCIMENTO, A. P. S.; SANTOS, W. C. A relevância da informação para o conhecimento a partir do esclarecimento de conceitos e do sentido presentes na informação. *In*: COLÓQUIO EM ORGANIZAÇÃO, ACESSO E APROPRIAÇÃO DA INFORMAÇÃO E DO CONHECIMENTO, 4., 2019., Londrina. **Anais [...]**, Londrina, 2019, p. 1-16.

STANFORD LIBRARIES. **Data support and services**. Disponível em: <https://library.stanford.edu/research-support/data-support-and-services>. Acesso em: 01 maio 2024.

STANFORD UNIVERSITY. **About**. 2024. Disponível em: <https://www.stanford.edu/about/>. Acesso em: 01 maio 2024.

STEINEROVÁ, J.; ONDRIŠOVÁ, M. Research Data Literacy Perception and Practices in the Information Environment. p. 545-555. *In*: KURBANOĞLU, Serap *et al.* (orgs.). **Information Literacy in Everyday Life**. Finlândia: Springer, 2019. 630 p. Disponível em: <https://link.springer.com/content/pdf/10.1007/978-3-030-13472-3.pdf>. Acesso em: 27 jan. 2023.

SURKIS, A.; READ, K. Research data management. **J. Med. Libr. Assoc**, Chicago, v. 103, n. 3, p. 154–156, jul./2015. DOI: <http://dx.doi.org/10.3163/1536-5050.103.3.011>. Acesso em: 30 jul. 2024.

TEAL, T. K. *et al.* Data Carpentry: Workshops to Increase Data Literacy for Researchers. **International Journal of Digital Curation**, Edimburgo, v. 10, n. 1, p. 135-143, 2015. DOI: <https://doi.org/10.2218/ijdc.v10i1.351>. Acesso em: 30 jul. 2024.

TENOPIR, C. *et al.* Data Sharing by Scientists: Practices and Perceptions. **PLoS ONE**, San Francisco, v. 6, n. 6, p. 1-21, jun./2011. DOI: <https://doi.org/10.1371/journal.pone.0021101>. Acesso em: 30 jul. 2024.

TENOPIR, C.; BIRCH, B.; ALLARD, S. **Academic libraries and research data services**: Current practices and plans for the future. Association of College and Research Libraries. An ACRL White Paper. Chicago, IL. 2012. 56 p. Disponível em: [https://trace.tennessee.edu/utk\\_dataone/20/](https://trace.tennessee.edu/utk_dataone/20/). Acesso em: 19 jun. 2024.

THIRYCHERQUES, H. R. Saturação em pesquisa qualitativa: estimativa empírica de dimensionamento. **PMKT – Revista Brasileira de Pesquisas de Marketing, Opinião e Mídia**, São Paulo, v. 2, n. 3, p. 20-27, set., 2009. Disponível em: [https://revistapmkt.com.br/wp-content/uploads/2009/03/SATURACAO\\_EM\\_PESQUISA\\_QUALITATIVA\\_ESTIMATIVA\\_EMPIRICA\\_DE\\_DIMENSIONAMENTO.pdf](https://revistapmkt.com.br/wp-content/uploads/2009/03/SATURACAO_EM_PESQUISA_QUALITATIVA_ESTIMATIVA_EMPIRICA_DE_DIMENSIONAMENTO.pdf). Acesso em: 13 fev. 2023.

UNIVERSITY OF OXFORD. **Facts and figures**. 2024a. Disponível em: <https://www.ox.ac.uk/about/facts-and-figures/full-version-facts-and-figures>. Acesso em: 09 maio 2024.

UNIVERSITY OF OXFORD. **History**. 2024b. Disponível em: <https://www.ox.ac.uk/about/organisation/history>. Acesso em: 09 maio 2024.

UNIVERSITY OF OXFORD. **Organization**. 2024c. Disponível em: <https://www.ox.ac.uk/about/organisation>. Acesso em: 09 maio 2024.

UNIVERSITY OF PENNSYLVANIA. **About**. 2024. Disponível em: <https://www.upenn.edu/about>. Acesso em: 14 maio 2024.

UNIVERSITY OF WASHINGTON. About the U.W. 2024. Disponível em: <https://www.washington.edu/about/>. Acesso em: 10 maio 2024.

UK DATA ARCHIVE. Research data lifecycle.2012. Disponível em: <https://www.ukdataservice.ac.uk/manage-data/lifecycle.aspx> . Acesso em 25 maio 2020.

VAN DER LAAN, Regina Helena; SATTAMINI, Glória Isabel Ferreira. Tesouros e terminologia. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 19, 2000, Porto Alegre, RS. **Anais** [...]. Porto Alegre, 2000. Disponível em: <https://lume.ufrgs.br/bitstream/handle/10183/10208/000294943.pdf>. Acesso em: 10 abr. 2024.

VEIGA, V. *et al.* Plano de gestão de dados FAIR: uma proposta para a Fiocruz. **Liinc em Revista**, Rio de Janeiro, v.15, n.2, p. 275-286, nov., 2019. Disponível em: <http://revista.ibict.br/liinc/article/view/5030>. Acesso em: 3 fev. 2023.

VILAR, P.; ZABUKOVEC, V. Research data management and research data literacy in Slovenian science. **Journal of Documentation**, Bingley, v. 75 n. 1, p. 24-43, 2019. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/JD-03-2018-0042/full/pdf?title=research-data-management-and-research-data-literacy-in-slovenian-science>. Acesso em: 13 mar. 2023.

VITORINO, E. V. A formação contínua do profissional da informação: princípios epistemológicos à competência informacional. In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 9., 2008. São Paulo. **Anais** [...]. São Paulo: USP, 2008. p. 1-15. Disponível em: <https://brapci.inf.br/index.php/res/download/181654>. Acesso em: 25 maio 2023.

WHITMIRE, A. L.; BOOCK, M.; SUTTON, S. C. Variability in academic research data management practices: Implications for data services development from a faculty survey. **Program: electronic library and information systems**, Bingley, v. 49, n. 4, p. 382-407, 2015. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/PROG-02-2015-0017/full/html>. Acesso em: 26 fev. 2023.

WHITMIRE, A. L. Implementing a Graduate-Level Research Data Management Course: Approach, Outcomes, and Lessons Learned. **Journal of Librarianship and Scholarly Communication**, Ann Arbor, v. 3, n. 2, 2015, p. 1-22. Acesso em: 10 jun. 2024. Disponível em:  
<https://www.iastatedigitalpress.com/jlsc/article/12760/galley/12446/view/>.

YANG, N.; LI, T. How stakeholders' data literacy contributes to student success in higher education: a goal-oriented analysis. **International Journal of Educational Technology in Higher Education**, Berlim, v. 17, n. 41, 2020. Disponível em:  
<https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-020-00220-3#citeas>. Acesso em: 27 set. 2022.