

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”

Faculdade de Filosofia e Ciências

Programa de Pós-Graduação em Ciência da Informação

ELIZABETE CRISTINA DE SOUZA DE AGUIAR MONTEIRO

**OPERACIONALIZAÇÃO DE REPOSITÓRIOS DE DADOS: UMA ANÁLISE
SOBRE AS PERSPECTIVAS E ATITUDES DOS PESQUISADORES NAS
QUESTÕES DE AUTORIA E LICENÇA**



Marília
2021

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”

Faculdade de Filosofia e Ciências

Programa de Pós-Graduação em Ciência da Informação

ELIZABETE CRISTINA DE SOUZA DE AGUIAR MONTEIRO

**OPERACIONALIZAÇÃO DE REPOSITÓRIOS DE DADOS: UMA ANÁLISE
SOBRE AS PERSPECTIVAS E ATITUDES DOS PESQUISADORES NAS
QUESTÕES DE AUTORIA E LICENÇA**

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências, da Universidade Estadual Paulista – UNESP – Campus de Marília para obtenção do título de Doutora.

Área de concentração: Informação, Tecnologia e Conhecimento

Linha de Pesquisa: Informação e Tecnologia

Orientador: Prof. Dr. Ricardo César Gonçalves Sant’Ana

Co-orientador: Antonio Hernández-Pérez

Marília
2021

M775o

Monteiro, Elizabete Cristina de Souza de Aguiar

Operacionalização de repositórios de dados : uma análise sobre as perspectivas e atitudes dos pesquisadores nas questões de autoria e licença /

Elizabete Cristina de Souza de Aguiar Monteiro. -- Marília, 2021

268 f. : il.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, Marília

Orientador: Ricardo César Gonçalves Sant'Ana

Coorientador: Antonio Hernández-Pérez

1. Repositórios institucionais. 2. Anomia. 3. Compartilhamento de dados. 4. Direitos autorais. 5. Consórcio Madroño. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Filosofia e Ciências, Marília. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

ELIZABETE CRISTINA DE SOUZA DE AGUIAR MONTEIRO

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação
da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista –
UNESP – Campus de Marília na área de concentração Informação,
Tecnologia e Conhecimento para obtenção do título de Doutora.

Prof. Dr. Ricardo César Gonçalves Sant’Ana (Orientador)
Universidade Estadual Paulista – UNESP

Profa. Dra. Luana Farias Sales Marques
Instituto Brasileiro de Informação em Ciência e Tecnologia Universidade
Federal do Rio de Janeiro

Prof. Dr. José Eduardo Santarém Segundo
Universidade de São Paulo

Profa. Dra. Diana Vilas Boas Souto Aleixo
Universidade Estadual de Londrina

Dra. Flavia Maria Bastos
Universidade Estadual Paulista

Marília
2021

Dedico a minha família e a todos que participaram de cada passo em direção a
conclusão do Doutorado

AGRADECIMENTOS

Agradeço a Deus por toda obra feita em minha vida.

A minha mãe e irmã que me dão suporte e amor.

Ao Leonardo e ao nosso filho Eduardo pelo amor e carinho que me fortalece a cada dia e me incentiva a crescer como pessoa e como profissional.

Ao meu orientador prof. Dr. Ricardo César Gonçalves Sant'Ana por todo aprendizado, confiança e por acreditar na minha pesquisa.

Ao meu co-orientador prof. Dr. Antonio Hernández-Pérez por acreditar na minha pesquisa e me receber com toda dedicação e atenção.

Às professoras Dra. Sarita Albagli e Dra. Luana Farias Sales Marques pela confiança, colaboração e contribuições no exame de qualificação.

Aos docentes participantes da Banca Examinadora, Dra. Luana, Dr. José Eduardo, Dra. Diana e Dra. Flavia que aceitaram o convite e contribuíram com oportunas observações enriquecendo a pesquisa e a tese.

Ao Grupo de Pesquisa sobre Tecnologias de Acesso a Dados (GPTAD) e os amigos da equipe coordenada pelo prof. Ricardo que, juntos aprendemos, compartilhamos e crescemos juntos.

Ao projeto de extensão Competências Digitais para a Agricultura Familiar (CoDAF).

À Faculdade de Filosofia e Ciências (FFC) que me proporciona conhecimentos a cada dia.

A todos os funcionários da Faculdade de Filosofia e Ciências que participam da construção de minha vida acadêmica.

A todos os funcionários da biblioteca da Faculdade de Filosofia e Ciências que contribuíram para que concluísse as atividades do Doutorado.

A todos os docentes do Programa de Pós-graduação em Ciência da Informação da Faculdade de Filosofia e Ciências.

A Universidad Carlos III de Madrid.

Aos amigos que conquistei nessa jornada, que me agradeceram com palavras, gestos, olhares, atitudes, estendendo a mão, caminhando ora junto ora na plateia incentivando. Me refiro a todos pois sabem quem são, mesmo sem mencionar nomes e correr o risco de deixar alguém de fora.

Cada um que passa em nossa vida, passa sozinho, mas não vai só nem nos deixa sós. Leva um pouco de nós, deixa um pouco de si.

(Antoine de Saint-Exupéry)

RESUMO

Repositórios de dados são realidades em muitas instituições internacionais e contribuem para a divulgação e preservação dos dados coletados por sua comunidade. A operacionalização dos repositórios envolve diversas funções e possuem infraestrutura e recursos humanos que estão paulatinamente trabalhando na gestão de dados e nas orientações de pesquisadores sobre a gestão de dados em seu ciclo de vida. Aspectos do direito autoral são preponderantes nas orientações para melhor reuso e citação dos autores e perpassam por diversas temáticas incluindo Ciência aberta, princípios FAIR, agências de fomento, plano de gestão de dados, carreira de pesquisadores, benefícios para os envolvidos e aspectos éticos. O objetivo da tese foi apresentar as convergências e divergências entre as atitudes e percepções dos pesquisadores e a operacionalização de repositórios de dados no fator Direito de Autor visando identificar o grau de aderência dos pesquisadores ao processo de depósito de conjuntos de dados nos repositórios considerando, ainda, a questão da perspectiva para identificar a possível anomia dos pesquisadores em relação aos repositórios. Foi utilizada a pesquisa documental e exploratória para identificação, coleta e análise dos documentos assim como o método Análise de conteúdo para apoiar os procedimentos metodológicos e mapear as orientações fornecidas sobre as questões do direito de autor. A pesquisa foi composta por dois universos: o primeiro composto pelo repositório *e-cienciaDatos* com seus documentos e bibliotecários; o segundo, composto pelos pesquisadores que depositaram seus conjuntos de dados no repositório. A coleta ocorreu por meio da triangulação com a aplicação de questionário, entrevista estruturada e análise documental. Com a análise do primeiro universo foi proposto um framework com treze índices, baseado nas estruturas teóricas, Teoria Institucional de W. Richard Scott, na Teoria do Comportamento Planejado de Icek Ajzen e no modelo de pesquisa de comportamento de compartilhamento de dados em repositório de dados proposto de Youngseek Kim. O framework é composto por categorias e índices que orientaram as perguntas direcionadas aos pesquisadores. Os resultados demonstram que o repositório conta com infraestrutura tecnológica, recursos humanos, documentos e manuais que instruem pesquisadores sobre os direitos de autor favorecendo o compartilhamento de dados, entretanto, há baixa adesão entre os pesquisadores causada pela anomia. A anomia está relacionada a fatores como falta de clareza das leis e normativas, desequilíbrio entre metas institucionais e os caminhos para atingi-las, descumprimento de regras e negação da moral. Conclui-se que o Consórcio Madroño e as bibliotecas membro subscrevem seu compromisso com Ciência Aberta, o repositório *e-cienciaDatos* armazena, gerencia e dissemina os dados resultantes das pesquisas de sua comunidade de forma aberta, os conjuntos de dados são protegidos pela propriedade intelectual e as políticas estão em consonância com as diretrizes da Ciência Aberta e princípios FAIR demonstrando a relevância dos direitos de propriedade intelectual para os atores envolvidos. A conscientização dos pesquisadores sobre os benefícios alcançados com o compartilhamento contribui para diminuir a retenção de seus dados e a anomia.

Palavras-chave: Repositório de dados; anomia; compartilhamento de dados; direitos autorais; Consorcio Madroño.

ABSTRACT

Data repositories are a reality in many international institutions and contribute to the dissemination and preservation of the data collected by their community. The operationalization of repositories involves several functions and they have infrastructure and human resources that are gradually working on data management and researchers' guidance on data management in its life cycle. Aspects of copyright are prominent in the guidelines for better reuse and citation of authors and cut across a range of topics including open science, FAIR principles, funding agencies, data management plan, researchers' careers, benefits for those involved, and ethical aspects. The objective of this thesis was to present the convergences and divergences between researchers' attitudes and perspectives and the operationalization of data repositories in the Copyright factor aiming to identify the degree of researchers' adherence to the process of depositing datasets in the repositories considering, also, the issue of perspective to identify the possible anomie of researchers in relation to the repositories. Documentary and exploratory research was used for the identification, collection and analysis of the documents and the Content Analysis method was used to support the methodological procedures and map the guidelines provided on copyright issues. The research was composed of by two universes: the first composed by the e-cienciaDatos repository with its documents and librarians; the second, composed of researchers who deposited their datasets in the repository. The collection occurred through triangulation with the application of a questionnaire, structured interviews, and document analysis. With the analysis of the first two universes, a framework with thirteen indexes was proposed and based on the theoretical frameworks, Scott's Institutional Theory, Ajzen's Theory of Planned Behavior and the research model of data sharing behavior in data repository proposed by Kim. The framework is composed of categories and indices that guided the questions directed to the researchers. The results show that the repository has technological infrastructure, human resources and documents and manuals that instruct researchers about copyrights favoring data sharing; however, there is low adherence among researchers caused by anomie. Anomie is related to factors such as lack of clarity of laws and regulations, imbalance between institutional goals and paths to achieve them, non-compliance with rules, and denial of morality. It is concluded that The Madroño Consortium and its members subscribe to their commitment to Open Science and the e-cienciaDatos repository stores, manages and disseminates publications and data resulting from the research of its community in an open way, the datasets are protected by intellectual property and policies are in line with the Open Science guidelines and FAIR principles, and e-cienciaDatos demonstrates the relevance of intellectual property rights for the actors involved. Researchers' awareness of the benefits achieved by sharing contributes to decreasing the retention of their data and anomie.

Keywords: Data repositor; anomie; data sharing; copyrights; Consorcio Madroño.

LISTA DE ABREVIATURAS

ASPI	Associação Paulista da Propriedade Intelectual
CI	Ciência da Informação
DPI	Direitos de Propriedade Intelectual
GPL	General Public License
OECD	Organisation for Economic Co-Operation and Development
ONU	Organização das Nações Unidas
PGD	Plano de Gestão de Dados
EU	União Europeia
WIPO	World Intellectual Property Organization

Lista de figuras

Figura 1 - Caminhos para o desenvolvimento da análise.....	34
Figura 2 - Indicação da fase da coleta	36
Figura 3 - Diagrama estrutural da teoria do compartilhamento Planejado	45
Figura 4 - Elementos da análise da percepção e atitude de pesquisadores	46
Figura 5 - Dimensão da análise da percepção e atitude de pesquisadores	47
Figura 6 - Três Pilares da Teoria Institucional.....	48
Figura 7 - Índices para a Teoria Institucional	49
Figura 8 - Dimensão institucional da percepção e atitude de pesquisadores.....	49
Figura 9 - Modelo para o comportamento de compartilhamento de dados	52
Figura 10 - Índices para teoria do comportamento de compartilhamento de dados.....	53
Figura 11 - Dimensão pesquisador da percepção e comportamento de pesquisadores.....	54
Figura 12 – Framework das perspectivas e atitudes do compartilhamento dos dados.....	56
Figura 13 - Composição do Framework final.....	58
Figura 14 - Aspectos da anomia no contexto institucional	66
Figura 15 - Princípios FAIR	76
Figura 16 – Repositórios que compõe o e-cienciaDatos	80
Figura 17 - Interface do e-cienciaDatos	83
Figura 18 - Propriedade intelectual de dados de pesquisa na Espanha	110
Figura 19 - Políticas na Open Science	115
Figura 20 - Aspectos legais direcionados à dados	116
Figura 21 - Princípios norteadores do repositório <i>e-cienciaDatos</i>	136
Figura 22 - Documentos norteadores do repositório <i>e-cienciaDatos</i> para o direito autoral	137
Figura 23 – Orientações da Ciência aberta.....	162
Figura 24 – Orientação sobre cessão dos direitos autorais para editoras.....	163
Figura 25 - Propostas de reformas na Lei de Propriedade Intelectual	194
Figura 26 – Orientações sobre citação de dados	212

Lista de Quadros

Quadro 1 - Domínios possíveis da aplicação da análise de conteúdo	31
Quadro 2 - Domínios aplicados da Análise de Conteúdo.....	32
Quadro 3 - Princípios e subprincípios FAIR	77
Quadro 4 - Campos de busca do <i>escienciaDatos</i> com filtros.....	84
Quadro 5 - Metadados e sua descrição	85
Quadro 6 - Aspectos dos direitos autorais aplicados aos dados.....	94
Quadro 7 - Aplicação do blockchain em dados científicos	96
Quadro 8 - Licenças Creative Commons	105
Quadro 9 - Documentos relacionados à Propriedade Intelectual na Espanha	108
Quadro 10 - Princípios FAIR relacionados aos direito autorias.....	118
Quadro 11 - Documentos relacionados ao direito autoral	134
Quadro 12 - Quadro sinóptico e as respectivas inferências	140
Quadro 13 - Questões e a aderência com o direito autoral.....	147
Quadro 14 - Respostas dos pesquisadores sobre Ciência Aberta.....	165
Quadro 15 - Princípios FAIR no <i>e-cienciaDatos</i>	169
Quadro 16 – Respostas dos pesquisadores sobre Princípios FAIR.....	170
Quadro 17 – Respostas dos pesquisadores sobre PGD.....	174
Quadro 17 – Respostas dos pesquisadores sobre Agências de Fomento.....	177
Quadro 18 – Respostas dos pesquisadores sobre Repositório <i>e-cienciaDatos</i>	183
Quadro 19 – Respostas dos pesquisadores sobre Propriedade Intelectual.....	195
Quadro 20 – Respostas dos pesquisadores sobre Benefícios.....	198
Quadro 21 – Respostas dos pesquisadores sobre Riscos	201
Quadro 22 – Respostas dos pesquisadores sobre Esforços	205
Quadro 23 – Respostas dos pesquisadores sobre Suporte dos Bibliotecários	208
Quadro 24 – Respostas dos pesquisadores sobre Ética.....	213
Quadro 25 – Resultado das hipóteses	219

Sumário

1	INTRODUÇÃO	14
1.1	Problema	18
1.2	Tese	19
1.3	Objetivos	19
1.3.1	Objetivo geral	19
1.3.2	Objetivos específicos	20
1.4	Justificativa.....	20
1.5	Delimitação do tema.....	25
1.6	Estrutura da Tese	26
2	PROCEDIMENTOS METODOLÓGICOS.....	28
2.1	Universo da pesquisa.....	35
2.2	Procedimentos da coleta de dados	36
2.2.1	Coleta de dados dos documentos	36
2.2.2	Coleta de dados com os Bibliotecários	38
2.2.3	Coleta de dados com os pesquisadores	39
2.3	Procedimentos de análise dos dados.....	40
3	MODELO CONCEITUAL DA PESQUISA	43
3.1	Dimensão de análise do pesquisador	44
3.2	Dimensão Institucional para a análise do pesquisador	47
3.3	Dimensão Pesquisador para a análise da atitude do compartilhamento.....	52
4	COMPARTILHAMENTO DE DADOS	60
4.1	Anomia e o compartilhamento de dados	62
5	REPOSITÓRIO DE DADOS E A COMUNICAÇÃO CIENTÍFICA.....	68
5.1	Repositório de dados de pesquisa	71
5.2	Repositório de dados e-cienciaDatos.....	79
6	PROPRIEDADE INTELECTUAL.....	91
6.1	Direitos autorais	93
6.1.1	Direitos Morais	99
6.1.2	Direitos patrimoniais.....	102
6.2	Propriedade intelectual na União Europeia	106
6.3	Propriedade intelectual na Espanha.....	107
7	DIREITOS AUTORAIS NAS DIMENSÕES INSTITUCIONAL E PESQUISADOR.....	112

7.1	Dimensão Institucional	112
7.1.1	Cultura e o direito autoral	112
7.1.2	Ciência aberta e os aspectos do direito autoral.....	113
7.1.3	Princípios FAIR e os aspectos do direito autoral.....	117
7.1.4	Plano de gestão de dados e a indicação do direito autoral	119
7.1.5	Agências de fomento e os requisitos do direito autoral	121
7.1.6	Repositório e a propriedade intelectual	122
7.1.7	Leis e Normas referentes ao direito autoral	123
7.2	Dimensão pesquisador.....	124
7.2.1	Benefícios relacionados ao direito autoral.....	125
7.2.2	Riscos e o direito autoral.....	126
7.2.3	Esforços e o direito autoral.....	127
7.2.4	Suporte dos bibliotecários	128
7.2.5	Benefício para a sociedade	129
7.2.5	Ética e a citação de dados	131
8	RESULTADOS E DISCUSSÕES	133
8.1	Apresentação dos documentos do Consórcio Madroño.....	133
8.2	Apresentação das categorias construídas para a análise de dado	138
8.3	Análise das entrevistas dos pesquisadores	155
8.3.1	Cultura.....	156
8.3.2	Ciência aberta	161
8.3.3	Princípios FAIR	168
8.3.4	Plano de gestão de dados.....	172
8.3.5	Agências de fomento.....	176
8.3.6	Repositório	181
8.3.7	Leis e normas.....	193
8.3.8	Benefícios	197
8.3.9	Riscos	120
8.3.10	Esforços	203
8.3.11	Suporte dos Bibliotecários.....	207
8.3.12	Benefícios para a sociedade	208
8.3.13	Ética	209
9	CONCLUSÕES	228
	REFERÊNCIAS.....	234

ANEXO A – TÉRMINO DE CONSENTIMIENTO LIBRE Y ESCLARECIDO – BIBLIOTECÁRIOS	250
ANEXO B – TÉRMINO DE CONSENTIMIENTO LIBRE Y ESCLARECIDO – PESQUISADORES	251
APÊNDICE A – APRESENTAÇÃO DA PESQUISADORA E DA PESQUISA	252
APÊNDICE B – ROTEIRO DE QUESTIONÁRIO E ENTREVISTA APLICADOS AOS BIBLIOTECÁRIOS	253
APÊNDICE C – ROTEIRO DE QUESTIONÁRIO E ENTREVISTA APLICADOS AOS PESQUISADORES	258
APÊNDICE D – PLANO DE GESTÃO DE DADOS.....	265

1 INTRODUÇÃO

Dados científicos ou dados de pesquisa científica são conceitos que representam dados comumente gerados ou coletados no decorrer das pesquisas científicas, que validam os resultados da pesquisa e são representados por números, registros textuais, documentos, planilhas, cadernos de laboratório ou de campo, questionários, transcrições imagens, áudio, vídeo, imagens, fotografias, slides, sequências de proteínas ou genéticas, amostras, objetos digitais entre outros (DUDZIAKI, 2016; ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2007).

O processo de pesquisa tem nos dados de pesquisa um importante elemento de insumo e sustentação dos resultados publicados e disseminados que implica em um maior investimento e empenho com os procedimentos relacionados a sua análise, gestão, curadoria, reuso e para o seu compartilhamento, principalmente em momentos de transformações tecnológicas abrangentes como as que são presenciadas em que o volume, a velocidade e a variedade de dados aumentam exponencialmente (MONTEIRO; SANT'ANA, 2018).

Os processos de pesquisas dependem progressivamente de financiamento das agências de fomento. As agências de fomento estão, gradativamente, instituindo em suas políticas de financiamento a orientação para a disponibilização dos dados em repositórios de dados de pesquisa ou publicação em periódicos; a elaboração de Plano de gestão de dados (PGD); o estabelecimento de políticas de acesso aberto e a articulação para que a gestão de dados de pesquisa possam atender aos princípios conhecidos como FAIR (*Findable, Accessible, Interoperable, Reusable*) e, assim, favorecer o compartilhamento e reutilização de dados (BORGMAN et al., 2015; COUTO CORRÊA, 2016; FAPESP, 2017; EUROPEAN COMMISSION, 2018).

Para ilustrar algumas das agências de fomento, relaciona-se: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) do Brasil, *National Science Foundation* (NSF), *National Institutes of Health* (NIH), NASA e Fundação Bill e Melinda Gates dos Estados Unidos, *Wellcome Trust* da Inglaterra, *Research Councils* do Reino Unido, *Centre National de la Recherche Scientifique* (CNRS) da França, Fundação para Ciência e Tecnologia de Portugal e *Horizon 2020* na União Europeia.

O compartilhamento de dados está explícito em diversas chamadas para

financiamento de pesquisas de agências de fomento e, também, no planejamento de muitos projetos e nas colaborações científicas internacionais, fortalecendo a reutilização dos dados e proporcionando benefícios à comunidade (FITZGERALD; FITZGERALD; PAPPALARDO, 2011; PINFIELD; COX; SMITH, 2014). O compartilhamento, o acesso aberto e o reuso dos dados são ações fomentadas pelo movimento internacional da Ciência aberta.

De acordo com Albagli (2015, p. 14, grifo da autora) a Ciência Aberta reflete “[...] novos modos de pensar e de exercer a *cientificidade*, com repercussões diretas sobre os compromissos, normas e arcabouços institucionais que interferem diretamente na prática científica e nas suas relações com a sociedade.” A sociedade é beneficiada com o movimento da Ciência aberta, atividade científica praticada de modo aberto, colaborativo e transparente que advoga pela abertura do processo de pesquisa, de seus resultados e dos dados de pesquisa que os embasaram (ALBAGLI, 2015; SANTOS, 2017).

A Ciência Aberta é um conceito holístico baseado em diferentes pilares que inclui o acesso aberto, metodologias abertas, dados abertos, ciência cidadã, revisão por pares aberta, código aberto, blocos de notas aberto, recursos educacionais abertos e redes sociais científicas, proporcionando ampla divulgação à sociedade e potencialização para a reprodutibilidade da ciência (MORA, 2019; ALBAGLI, 2015; SANTOS, 2017, BEZJAK et al., 2018). Adicionalmente, verifica-se outros elementos na agenda da Ciência aberta como: publicação aberta, metadados abertos, impacto e métricas abertas, os princípios FAIR e os repositórios abertos, incluindo repositórios de dados (CONSÓRCIO MADROÑO, 2016, 2017; MONTEIRO; SANT’ANA, 2019).

Os repositórios de dados desempenham papel basilar no apoio aos pesquisadores, na gestão, na publicação e na divulgação dos dados de pesquisa. Cabe evidenciar o número de repositórios de dados registrados no re3data.org, um registro global de repositórios de dados de pesquisa que abrange repositórios de diferentes disciplinas acadêmicas. Ao pesquisar no repositório, identifica-se o termo “*datas providers*” em que são apresentados 2461 repositórios¹.

Os repositórios de dados armazenam os dados de pesquisa, os Planos de Gestão de Dados (PGD) e, em alguns casos, documentação de dados. Os PGDs são documentos formais elaborados no início do projeto, porém atualizáveis, para que seja

¹ <https://www.re3data.org/search>. Acesso em: 20 jan. 2021.

propício descrever todos os aspectos da gestão dos dados observando as instruções das agências de financiamento e as características dos dados. Tais repositórios fornecem ambientes para o armazenamento e gestão dos dados, são implementados, inclusive, em universidades e tem o apoio dos bibliotecários que contribuem na gestão dos dados, na aplicação dos princípios FAIR e dos princípios da Ciência Aberta e no apoio técnico aos pesquisadores.

A demanda por profissionais capacitados para atuarem nos repositórios de dados é expressiva. Os bibliotecários que atuam nas universidades estão assumindo novos papéis com a implementação de repositórios de dados e com os serviços de dados. Tais profissionais estão direcionando os princípios de Biblioteconomia para dar novos serviços e soluções de problemas relacionados aos dados de pesquisa.

Uma pesquisa sobre Serviços de Dados de Pesquisa realizada por um grupo de pesquisadores em colaboração com o *LIBER's Scholarly Communication & Research Infrastructures Committee and DataONE* expõe que em quase todas as bibliotecas acadêmicas da Europa trabalham cooperativamente e oferecem serviços de dados de pesquisa as suas comunidades como, por exemplo, encontrar informações sobre PGD, padrões de metadados e práticas de citação dos dados (LIBEREUROPE, 2016).

O contexto apresentado instiga o compartilhamento dos dados, no entanto, é possível que os detentores dos dados tenham preocupações relacionadas ao reconhecimento da autoria dos seus conjuntos de dados e sobre como os dados podem ser reutilizados devido às questões pessoais, administrativas ou de recursos, assim como relacionadas ao tempo despendido e aos custos na geração ou coleta, no tratamento, na análise e na gestão de dados. Tais preocupações, além da falta de clareza das regras institucionais, podem instaurar um estado de anomia nos pesquisadores inibindo o compartilhamento dos dados.

A anomia ocorre quando os valores culturais, os conjunto de valores normativos que governam a conduta dos membros de uma determinada sociedade ou grupo e as metas institucionais encontram-se em desequilíbrio, pendendo para uma maior relevância dos valores culturais em relação às metas institucionais (MERTON, 1968). Assim, os indivíduos percebem-se em desconformidade com que as instituições requerem e praticam condutas desviantes.

Com o intuito de amenizar possíveis riscos de uso indevido dos dados e não reconhecimento da autoria, os autores tem recorrido às estratégias para proteger a

autoria dos dados, assim como aplicar em seus conjuntos de dados as licenças e demais permissões para seu uso, aspectos que permeiam a propriedade intelectual.

A proteção dos direitos autorais estimula a inovação e a exploração das descobertas científicas, protege o trabalho criativo e o investimento em descobertas, sendo primordiais na investigação científica (ELLIOT, 2005). As leis de direitos autorais podem conceder aos autores ou criadores proteção para seus conjuntos de dados inclusive, quando suas criações são disponibilizadas em repositórios de dados.

Nessa dinâmica, os aspectos éticos e legais relacionados à propriedade intelectual dos conjuntos de dados de pesquisa devem ser considerados. A propriedade intelectual é a área do direito que protege os inventores ou responsáveis por qualquer produção intelectual, seja no domínio industrial, científico, literário ou artístico, facultando aos seus titulares direitos econômicos os quais ditam a forma de comercialização, circulação, utilização e produção (UNIVERSIDAD CARLOS III DE MADRI (UC3M), [201-a]).

Praticamente todas as legislações nacionais integram disposições que definem as várias medidas a aplicar em caso de infração do direito de autor. Via de regra, as disposições encontram-se dispersas nos inúmeros textos legislativos: leis nacionais sobre o direito do autor e códigos penais. O objetivo é impor uma sanção àquele que cometeu infração e indenizar a parte lesada pelo prejuízo sofrido (UNESCO, 1981). Nas práticas da comunicação científica se estima a ética e as boas práticas da citação dos autores quando utilizado seu conteúdo.

A propriedade intelectual pode ser dividida em duas categorias: Propriedade industrial e Direitos autorais. Direito autoral é um dos ramos da ciência jurídica e permeia um conjunto de direitos morais (reconhecimento da autoria explicitado nas citações e o direito à integridade) e os direitos patrimoniais (econômicos e de exploração) sobre as criações do espírito, expressas por quaisquer meios ou fixadas em quaisquer suportes, tangíveis ou intangíveis, que se concede aos criadores de obras intelectuais (UC3M, [201-a]).

A citação de dados faz parte das boas práticas científicas e promove o reconhecimento dos pesquisadores, assim como a citação de suas publicações tradicionais. Os princípios e práticas de citação de dados estão em discussão na comunidade científica juntamente com a necessidade de orientação na reutilização e replicação de dados.

Relevante destacar que nem todos os dados abertos, especialmente dados de pesquisas, são ou devem ser completamente "abertos" ou "livres", mas devem ser, pelo menos, FAIR conforme a redação do artigo 10.1 do *FAIR4Health Consortium* (2019) "*as open as possible, as closed as necessary*". Existem dois motivos principais para manter os dados de pesquisa "fechados": 1) Razões legais e éticas; 2) Razões econômicas (HERNÁNDEZ-PÉREZ; MÉNDEZ RODRÍGUEZ, 2018).

Dessa forma, o tema deste estudo é a o Direito autoral dos conjuntos de dados de pesquisa na operacionalização dos repositórios, considerando as citações e as licenças de uso.

1.1 Problema

Nesta pesquisa, o compartilhamento de dados refere-se a disponibilização de conjuntos de dados brutos ou analisados, que embasaram as publicações científicas, à outros pesquisadores fora da equipe de pesquisa por vários meios porém, com foco nos repositórios de dados.

Para promover a discussão e compreender a dinâmica da interação dos repositórios de dados com os pesquisadores em suas convergências e divergências relativas aos direitos de autor e, eventualmente ajudar a alcançar uma visão central da resistência ou não no compartilhamento de conjuntos de dados, esta pesquisa se concentra em dois pontos principais: 1) no entendimento dos motivos da dubiedade dos pesquisadores em compartilhar seus dados; e 2) na operacionalização dos repositórios de dados na publicação dos dados.

Embora o compartilhamento de dados de pesquisa seja uma prática corrente em muitas disciplinas, em outras é particularmente crítico, dependendo de fatores externos como políticas governamentais, institucionais, de financiadores, além de comportamento individual dos pesquisadores.

A pesquisa se baseia em duas estruturas teóricas, na Teoria institucional de Scott e na Teoria do comportamento planejado de Ajzen e em um modelo de pesquisa sobre comportamento de compartilhamento de dados de Kim. A Teoria do comportamento planejado é uma teoria específica da motivação individual, e está aplicada para examinar até que ponto o entendimento dos pesquisadores sobre as políticas de direitos de autor impacta em suas atitudes em relação ao compartilhamento de dados. A pesquisa pode fornecer *insights* aos profissionais da

informação, em especial aos bibliotecários, para auxiliar os pesquisadores na gestão de dados e no esclarecimento sobre os direitos de autor.

Assim, o problema de pesquisa da Tese está na falta de conhecimento por parte dos pesquisadores sobre a existência de políticas de gestão de direitos do autor nos repositórios, o que impacta na adesão ao processo de compartilhamento de dados e na operacionalização dos repositórios. Tal barreira pode, ainda, ser potencializada pelas exigências geradas pelas práticas de gestão de dados e critérios para sua abertura no contexto da Ciência Aberta.

1.2 Tese

A existência de um estado de anomia dos pesquisadores em relação às políticas de gestão do direito do autor dos repositórios impacta a falta de aderência entre suas perspectivas sobre o processo de compartilhamento de dados e a operacionalização dos repositórios de dados.

A tese divide-se em duas dimensões centrais: a questão da aderência e a questão da anomia. A aderência devido a pouca adesão dos pesquisadores ao repositório explicitado no número de conjuntos de dados depositados. A anomia devido a quantidade de pesquisadores que depositaram seus conjuntos de dados mediante a quantidade de universidades envolvidas.

1.3 Objetivos

1.3.1 Objetivo Geral

Apresentar as convergências e divergências entre as atitudes e percepções dos pesquisadores e a operacionalização de repositórios de dados no fator Direito de Autor. Esse objetivo foi estabelecido com o intuito de identificar o grau de aderência dos pesquisadores ao processo de depósito de conjuntos de dados nos repositórios considerando, ainda, a questão da perspectiva para identificar a possível anomia dos pesquisadores em relação aos repositórios.

1.3.2 Objetivos específicos

- Identificar a operacionalização de repositórios de dados a partir da perspectiva do armazenamento e gestão dos dados no contexto do direito autoral;
- Contextualizar os aspectos envolvidos com o Direito Autoral - Direito moral e Econômico - de conjuntos de dados;
- Elaborar um framework para identificar as percepções e atitudes de pesquisadores com fatores institucionais, individuais e sociais a partir dos índices identificados;
- Elaborar um framework com índices que elucidam as percepções e atitudes de pesquisadores no compartilhamento de dados;
- Relacionar os aspectos do direito autoral com os índices que compõem o framework;
- Categorizar os elementos relacionados ao direito autoral;
- Identificar atitudes e perspectivas dos pesquisadores sobre Direito Autoral de conjuntos de dados;
- Identificar a possível anomia dos pesquisadores em relação aos repositórios;
- Identificar a aderência dos pesquisadores ao processo de depósito de conjuntos de dados no repositório de dados.

1.4 Justificativa

A proposta de estudar a relação dos direitos autorais de conjuntos de dados será apresentado em dois prismas. O primeiro diz respeito aos aspectos teóricos da discussão sobre os temas que permeiam o direito autoral de conjuntos de dados: os objetivos da Ciência Aberta, a gestão de dados, a atitude dos pesquisadores no compartilhamento de seus dados e a possível preocupação dos pesquisadores em não serem reconhecidos e citados pelos seus trabalhos com os dados. O segundo, diz respeito às contribuições práticas da pesquisa no contexto de sua inserção.

A discussão proposta pela pesquisa trata de um fenômeno cada vez mais constante nas discussões e nas boas práticas científicas – o reconhecimento de autoria dos conjuntos de dados. Os pesquisadores podem ter uma série de preocupações em virtude do compartilhamento de seus conjuntos de dados, com

enfoque nas premissas de como serão reconhecidos pela autoria dos dados e como esses dados serão liberados para reutilização. Propiciar a discussão sobre tais aspectos evidencia a relevância do direito autoral, pois demonstra como os atores envolvidos têm promovido a citação de dados e as vantagens da liberação planejada e, conseqüentemente, a escolha sobre qual licença de uso os dados serão disponibilizados, no caso dessa tese, nos repositórios de dados.

A centralidade da descoberta e compartilhamento de dados nas várias áreas do conhecimento, inclusive na Ciência da Informação, configuram um ambiente complexo que justifica a necessidade de estudos e discussões que sustentem um referencial teórico e amparem a nova práxis exigida no ambiente científico.

A pesquisa tem o intuito de contribuir com a área da Ciência da Informação ao discutir o papel dos dados de pesquisa nas práticas da comunicação científica. Os processos atrelados à comunicação científica são objetos de estudo da Ciência da Informação, sendo essa um campo interdisciplinar voltado ao uso de informação nos diferentes contextos, atrelada de maneira indissociável aos dados de pesquisa. Segundo Saracevic (1995) a Ciência da Informação tem três características gerais: é interdisciplinar; está ligada à tecnologia da informação e é participante ativa na evolução da sociedade da informação, com uma forte dimensão social e humana.

Tal transdisciplinaridade transpassa várias disciplinas e inclui novos elementos nas discussões e estudos da Ciência da Informação: o estudo de dados que, conforme sua relevância, estão gerando várias pesquisas em seu contexto.

A evolução da sociedade da informação traz a luz o protagonismo dos dados. Ademais, o *know-how* da Ciência da Informação reúne os requisitos basilares para a discussão dos elementos entorno dos dados no campo científico. Os pesquisadores da Ciência da Informação estão paulatinamente desenvolvendo pesquisas nas temáticas discutidas nessa pesquisa.

A relevância científica na realização dessa pesquisa está em fornecer subsídios teóricos e práticos para a área da Ciência da Informação e afins, com a finalidade de que os profissionais envolvidos na gestão de dados nos repositórios de dados participem das discussões dos direitos autorais desde a elaboração das políticas do repositório até o treinamento aos consumidores de dados contribuindo para diminuir as inquietações e problemas inerentes a possíveis preocupações dos pesquisadores em não serem reconhecidos e citados pelos seus trabalhos com os dados e a cessão

dos dados para reutilização.

O avanço científico, com o aprimoramento do instrumental tecnológico, a grande quantidade de dados gerados ou coletados e a necessidade de gestão dos dados de pesquisa transformam consideravelmente a forma de se fazer pesquisa. A Ciência da informação (CI) contribui com estudos voltados para a temática pois

[...] é um campo dedicado à investigação científica e a prática profissional resolver os problemas de comunicação eficaz de registros de conhecimento e conhecimento entre seres humanos, no âmbito de utilizações sociais, institucionais e/ou individuais e de necessidades para informações. Ao abordar esses problemas de interesse particular está tomando o máximo proveito possível da moderna tecnologia da informação. (SARACEVIC, 1995, p. 3, tradução nossa).

A confiabilidade dos dados de pesquisa são fundamentais no universo acadêmico, principalmente nas universidades públicas em decorrência do dever do pesquisador em prestar contas à sociedade sobre o financiamento recebido para sua pesquisa (GUIMARÃES, 2018).

As agências de financiamento, entre seus requisitos, solicitam a elaboração de Plano de Gestão de Dados junto aos projetos de pesquisas ou após a liberação do financiamento e juntos as relatórios técnicos. Ademais, em muitos casos, recomendam e incentivam a gestão e o compartilhamento dos dados durante ou após a finalização da pesquisa para seu reuso. O compartilhamento pode ser feito via arquivamento em repositórios de dados.

O repositório *e-cienciaDatos*, repositório de dados constituído pelas seis universidades do *Consórcio Madroño: Universidad de Alcalá, Universidad Autónoma de Madrid, Universidad Carlos III de Madrid, Universidad Politécnica de Madrid, Universidad Rey Juan Carlos y UNED* foi selecionado como escopo da pesquisa devido aos seguintes fatores: Primeiro pela sua importância e relevância no contexto científico da Espanha, além da importância da Espanha no contexto da União Europeia.

Segundo, é um repositório de dados multidisciplinar, resultante do trabalho colaborativo das seis universidades públicas da Comunidade de Madri, planejado e articulado desde 2015 por profissionais, inclusive bibliotecários, das universidades, que abriga os conjuntos de dados de pesquisa dos pesquisadores das universidades,

com intuito de dar visibilidade aos dados da sua comunidade, garantir a sua preservação e facilitar seu acesso e reutilização.

Terceiro, os bibliotecários das universidades membros tiveram um plano de formação para responderem as necessidades dos pesquisadores na gestão dos dados. No plano de formação incluíram tópicos como gestão de dados de pesquisa, metadados para os conjuntos de dados, modelos e políticas de dados, aspectos legais e direitos autorais dos dados, conhecimento em gestão de Repositório de dados e Plano de Gestão de Dados (PGD). O curso foi celebrado em 2016 e participaram aproximadamente 100 bibliotecários.

Quarto, para atender a demanda pelo processo de gestão de dados no ciclo de vida dos dados dos pesquisadores e do repositório foram implementada uma infraestrutura com o repositório, o Portal InvestigaM² e o PaGoDa, ferramenta para elaboração do PGD, uma tradução e adaptação para o espanhol da ferramenta DMPOnline, projetado pelo *Digital Curation Centre* do Reino Unido, é um tutorial que tenta responder aos principais tópicos a serem abordados no Plano de Gestão de Dados de acordo com a *FAIR Data Management at a glance: issues to cover in your Horizon 2020 DMP*, incluído nas *Guidelines on FAIR Data Management in Horizon 2020*³.

Quinto, o *Consórcio Madroño* firmou sua *Declaración en apoyo a la Ciencia Abierta a la información académica y científica*, em 2017, na qual se definem as responsabilidades dos pesquisadores e das universidades membros no que diz respeito à publicação aberta de dados de pesquisa. Desde então, vem sendo feita a difusão do repositório, de seus serviços e o trabalho relacionado à Ciência Aberta, bem como documentos, infografias e vídeos relacionados.

Sexto, a *Universidad Carlos III de Madrid*, onde foi feito o Doutorado Sanduíche, tem pesquisadores trabalhando em projetos Europeus envolvidos à Ciência Aberta e princípios FAIR, como os projetos *FAIRSFAIR* e *FAIR4Health: Improving Health Research in EU through FAIR Data* financiado pela *European Commission Research Executive Agency*, sendo este último com período de vigência entre 01/12/2019 à 30/11/2021.

Sétimo, um dos pesquisadores é o professor Tony Hernández-Pérez, que se

² <http://www.consorciumadrono.es/investigam/>

³ https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

tornou co-orientador dessa Tese. É docente do *Departamento de Biblioteconomía y Documentación* da UC3M, já participou e coordenou projetos Europeus relacionados à Ciência aberta, Acesso aberto, Repositórios digitais, Princípios FAIR e, atualmente, participa conjuntamente à docente Eva Maria Mendez Rodriguez do projeto Europeu *FAIR4Health: Improving Health Research in EU through FAIR Data* financiado pela *European Commission Research Executive Agency* com período de vigência entre 01/12/2019 à 30/11/2021. Além dos projetos, o prof. Tony tem diversas publicações nos temas foco da pesquisa, é líder do grupo de pesquisa *Tecnologías Aplicadas a la Información y la Documentación* (TECNODOC) e integrante da *Red Española sobre Datos de Investigación en Abierto* (MAREDATA) composta por grupos de pesquisas das instituições: *CSIC-Instituto de Agroquímica y Tecnología de Alimentos (IATA)*; *CSIC-Instituto de Gestión de la Innovación y el Conocimiento (INGENIO)*-*Universitat de València*; *Universidad de Alicante*; *Universitat de Barcelona*; *Universidad Carlos III de Madrid*; *Universitat Oberta de Catalunya*, KIMO; *Universitat Politècnica de València*, DATASEA.

A proposta de estudar a relação dos direitos autorais que podem ser aplicados e que estão permeando os conjuntos de dados de pesquisa se iniciou em 2015 no ingresso da autora da Tese no mestrado juntamente com as temáticas de Plano de Gestão de dados e gestão de dados. Na ocasião, o interesse foi instigado pela participação na pesquisa do orientador dessa Tese, Ricardo Sant'Ana, com o Ciclo de Vida dos Dados, no qual o direito autoral dos dados é um dos fatores; pelas leituras relacionadas aos temas e as poucas discussões envolvidas com a preocupação dos pesquisadores em compartilhar seus dados e não serem reconhecidos pelo seu trabalho, em nível internacional. Ademais, um tema relativamente novo para a área da Ciência da Informação no Brasil, com poucos profissionais da área, vislumbrando o potencial e novo contexto de trabalho e pesquisa para a área. Com a pesquisa no desenvolvimento da dissertação, teve-se contato com o contexto das cem melhores universidades do mundo identificando a atuação dos bibliotecários e dos repositórios de dados. Como bibliotecária atuante em uma universidade pública a 12 anos, a Unesp, com vários *campi* e centenas de docentes e pesquisadores, vislumbrou-se a oportunidade em ser pioneira nas temáticas envolvidas na Dissertação e na Tese para atuação na universidade como bibliotecária de dados e de oferecer conteúdo relevante, treinamentos, produtos e serviços de dados para a comunidade, inclusive, para outras universidades, constatados nas participações com palestras e minicursos.

Em busca de uma infraestrutura com repositório de dados, pois a presente autora atua em uma rede de bibliotecas na Unesp, com uma rede de bibliotecas e atuação de bibliotecários na gestão de dados no repositório e nos treinamentos de pesquisadores, foi identificado o repositório *e-cienciaDatos*, o *Consórcio Madroño* e o prof. Tony, profissional ativo nas temáticas e com potencial de alavancar os conhecimentos da autora, todos com a justificativa pelo estudo já apresentada.

1.5 Delimitação do tema

O tema da Tese está envolto às questões de autoria e licenças de conjuntos de dados, considerando o direito de propriedade intelectual, com a análise sobre os documentos do repositório de dados *e-cienciaDatos* que abordam a temática e os questionários e entrevistas aplicados aos pesquisadores para identificar suas perspectivas e suas atitudes no compartilhamento de seus conjuntos de dados.

O repositório *e-cienciaDatos* considera normativas e a Lei de Propriedade Intelectual em seu artigo 12 - *Colecciones. Bases de datos*, através do chamado direito *sui generis* de bases de dados para elaboração de sua documentação e orientação aos pesquisadores.

A Tese não tem a pretensão de discutir o âmbito da Lei de propriedade intelectual e nem de examinar se as leis e diretivas apresentadas abrangem os dados de pesquisa ou qual o seu alcance. O intuito foi apresentar conteúdos que abordam os direitos autorais sobre dados de pesquisa e como os repositórios operacionalizam seu trabalho baseados nos documentos apresentados.

A anomia, e sua definição mais profunda, não são o foco dessa Tese. A anomia é utilizada como instrumento para identificar a dificuldade da percepção dos pesquisadores em relação às normas e aos procedimentos definidos pelos repositórios bem como as temáticas apresentadas.

Os objetos de estudo da pesquisa foram: o repositório *e-cienciaDatos* e suas bibliotecas, com seus documentos; os bibliotecários, com as entrevistas e questionários aplicados; os pesquisadores, com os questionários e entrevistas aplicados. O escopo da pesquisa foi o repositório *e-cienciaDatos*, um repositório de dados constituído pelas seis universidades do *Consórcio Madroño*: *Universidad de Alcalá*, *Universidad Autónoma de Madrid*, *Universidad Carlos III de Madrid*,

Universidad Politécnica de Madrid, Universidad Rey Juan Carlos y UNED. O repositório organiza os conjuntos de dados de cada uma das universidades cooperantes e oferece o armazenamento e a publicação de conjuntos de dados, atribuindo um identificador de objeto digital DOI e referência normalizada a cada um deles com a finalidade de facilitar a verificação dos dados, sua disseminação, reutilização, impacto e acesso a longo prazo.

1.6 Estrutura da Tese

A presente tese está estruturada nas seguintes seções:

Seção 1 – INTRODUÇÃO é apresentado o contexto da pesquisa com conceitos iniciais sobre as temáticas abordadas. Em suas subseções apresenta-se o problema da pesquisa, o objetivo geral e os objetos específicos, a delimitação do tema e objeto da pesquisa e a justificativa e a metodologia utilizada.

Na seção 2 – PROCEDIMENTOS METODOLÓGICOS são apresentados o universo da pesquisa e os procedimentos de coleta e análise dos dados. Ao final é apresentada a estrutura da Tese.

Na seção 3 – MODELO CONCEITUAL DA PESQUISA são apresentadas as dimensões que embasaram a discussão da pesquisa e o modelo conceitual utilizado para compor as perguntas das entrevistas e questionários aplicados aos pesquisadores com os passos para sua composição e apresentação do framework da pesquisa.

Na seção 4 – COMPARTILHAMENTO DE DADOS discorre sobre o compartilhamento de dados e a anomia com apresentação dos conceitos dos autores e relacionando a anomia com o contexto do compartilhamento dos dados.

Na seção 5 – REPOSITÓRIO DE DADOS DE PESQUISA A COMUNICAÇÃO CIENTÍFICA, traz conceitos acerca dos repositórios de dados, sua origem, características e benefícios; aborda a comunicação científica; perpassa os princípios FAIR e apresenta o repositório *e-cienciaDatos* com suas características, funcionalidades e metadados.

Na seção 6 – PROPRIEDADE INTELECTUAL, há uma contextualização de propriedade intelectual, incluindo a propriedade intelectual no contexto Europeu e Espanhol, as definições de propriedade intelectual, direitos autorais e licenças de uso e seu uso em repositórios de dados. É feita breve apresentação e características do Blockchain.

A seção 7 - DIREITOS AUTORAIS NAS DIMENSÕES INSTITUCIONAL E PESQUISADOR são apresentadas as dimensões do *framework*, suas categorias e índices correlacionando cada índice com o direito autoral.

A seção 8 – RESULTADOS E DISCUSSÕES apresenta o resultado das análises dos documentos, dos questionários e entrevistas correlacionado com o direito autoral e com a anomia. Os resultados são apresentados com os figuras e quadros para melhor visualização dos resultados. Para cada dimensão do framework são apresentados os índices e as categorias, as definições para cada um deles e a aderência dos índices as suas categorias. Ademais são apresentadas as hipóteses levantadas, as questões do questionário e da entrevista dos pesquisadores, a justificativa para cada questão e sua aderência ao tema direito autoral.

Na seção 9 – CONSIDERAÇÕES FINAIS apresenta as reflexões e conclusões sobre os tópicos abordados em toda a pesquisa.

2 PROCEDIMENTOS METODOLÓGICOS

Com o propósito de verificar a tese da presente pesquisa que consistem na existência de um estado de anomia dos pesquisadores em relação às políticas de gestão do direito do autor dos repositórios impacta a falta de aderência entre suas perspectivas sobre o processo de compartilhamento de dados e a operacionalização dos repositórios de dados, e a consecução dos objetivos estabelecidos, definiu-se pela pesquisa bibliográfica de natureza qualitativa, do tipo documental, exploratória e descritiva, de modo a subsidiar um estudo sobre o repositório *e-cienciaDatos* composto pelas universidades da *Comunidad de Madrid* participantes do *Consórcio Madroño*.

A pesquisa bibliográfica abrangeu a bibliografia publicada sobre as temáticas apresentadas na Tese, publicadas em meio impresso e digital (MARCONI; LAKATOS, 2016) justificada por oferecer subsídios para compor a base teórica sobre os temas abordados.

A pesquisa qualitativa é descritiva e tem o ambiente natural como fonte direta de dados, sendo os fenômenos melhor observados e compreendidos no contexto no qual o pesquisador faz parte (nessa Tese são as universidades que compõem o repositório *e-cienciaDatos*), objetivando compreender o fenômeno que está sendo estudado (as convergências e divergências entre a operacionalização do repositório e as perspectivas e atitudes dos pesquisadores) e considerando que todos os dados da realidade são significativos e devem ser examinados para verificar como se manifestam e interagem (GODOY, 1995).

As características da pesquisa qualitativa são: trabalhar com os fenômenos e com a análise hermenêutica dos dados sendo analisados subjetivamente pelo pesquisador, focando indicadores do funcionamento de estruturas e organizações (APPOLINÁRIO, 2011; HAGUETTE, 2003). A pesquisa qualitativa tem como objetos de pesquisa: locais (nessa Tese foi representada pelo repositório *e-cienciaDatos*; pessoas (representada pelos pesquisadores e bibliotecários); fatos (representados pela operacionalização dos repositórios e pelas perspectivas e atitudes dos pesquisadores) (CHIZZOTTI, 2003).

Os procedimentos metodológicos foram aplicados em dois universos: o primeiro composto pelo repositório *e-cienciaDatos* com seus documentos e

bibliotecários; o segundo, composto pelos pesquisadores que depositaram seus conjuntos de dados no repositório.

A pesquisa documental caracteriza-se por ter como fonte para a coleta dos dados os documentos, sendo eles de instituições públicas ou privadas que retratam e fornecem dados sobre o seu contexto (GODOY, 1995; MARCONI; LAKATOS, 2016). Nesse tipo de pesquisa deve-se atentar a três aspectos: 1) a escolha dos documentos não é um processo aleatório, mas se dá em função de alguns propósitos, ideias ou hipóteses levantadas pelo pesquisador; 2) o acesso aos documentos disponibilizados, principalmente em meio eletrônico, está mais acessível ao público e 3) a análise final (GODOY, 1995).

Para a seleção dos documentos foram seguidos os procedimentos relacionados à pesquisa documental indicados por Godoy (1995), em seus três aspectos: 1) os documentos selecionados foram Leis, Normativas, políticas do repositório, manuais, materiais ilustrativos, infografias, audiovisuais disponíveis, questionários e entrevistas; 2) o acesso aos documentos foram os disponibilizados nas páginas do *Consórcio Madroño* e das bibliotecas das universidades bem como os enviados pelos bibliotecários; 3) a análise final resultou na seleção das categorias e dos índices que ampararam a composição de um *framework*. A partir do *framework* foram elaboradas as perguntas que compuseram o questionário e a entrevista dos pesquisadores.

O estudo exploratório objetivou ampliar a compreensão de um fenômeno pouco conhecido ou de um problema de pesquisa não delineado perfeitamente (APOLLINÁRIO, 2011).

Na pesquisa descritiva, o pesquisador descreve e interpreta o fenômeno observado (APPOLINÁRIO, 2011). Para constituição desta tese e em consonância com os tipos de pesquisas apresentadas, foram utilizadas as bibliografias levantadas para o referencial teórico e os documentos para a análise documental.

Os procedimentos metodológicos seguiram três etapas. Na primeira etapa foi realizado o levantamento bibliográfico da literatura nacional e internacional nas bases de dados *Scopus*, *Web of Science* e *Redalyc*, nos catálogos das bibliotecas e repositórios digitais sobre as temáticas centrais dessa pesquisa. O levantamento bibliográfico foi realizado por meio das palavras-chave: propriedade intelectual, direitos autorais, copyright, repositórios de dados, repositórios digitais, dados de pesquisa, dados de pesquisa e compartilhamento de dados pesquisados nos idiomas

português, espanhol e inglês. Os campos utilizados para as buscas foram título, resumo e palavras-chave. Como o tema relacionado aos repositórios de dados e direitos autorais de dados é relativamente recente não foi empregado recorte temporal. Foram recuperados 50 documentos, selecionados para leitura 17 pois os outros mencionavam a temática sem fazer discussões.

Na segunda etapa foi realizada a pesquisa documental para identificação a análise dos documentos selecionadas. Na terceira etapa foi aplicada a pesquisa exploratória para coleta dos documentos para análise e identificação dos pesquisadores que compuseram o universo da pesquisa.

O método Análise de conteúdo foi aplicado para apoiar os procedimentos metodológicos em virtude de ser um método apropriado para reduzir e entender as informações das fontes onde os dados foram coletados. Por meio da aplicação do método Análise de conteúdo foi realizada a análise do conteúdo de tópicos abordados pelos documentos identificados para mapear as orientações fornecidas sobre as questões do direito de autor, sendo ele moral ou patrimonial. Os dois universos da pesquisa foram codificados separadamente para após, cruzar os dados entre eles.

Bardin (2010, p. 44) define Análise de Conteúdo como:

Um conjunto de técnicas de análise das comunicações visando obter por procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens indicadores (quantitativos ou não) que permitam a inferência de conhecimentos relativos às condições de produção/recepção (variáveis inferidas) destas mensagens.

O método Análise de Conteúdo foi aplicado para a coleta e a análise dos dados visto que seu ponto de partida é a mensagem verbal (oral ou escrita), gestual, silenciosa, figurativa, documental ou diretamente provocada (FRANCO, 2008). Os dados coletados para esta tese partiram dos documentos e das entrevistas com os bibliotecários.

Franco (2008, p. 10) considera que

São perfeitamente possíveis e necessários o conhecimento e a utilização da análise de conteúdo, enquanto procedimento de pesquisa, no âmbito de uma abordagem metodológica crítica e epistemologicamente apoiada numa concepção de ciência que reconhece o papel ativo do sujeito na produção do conhecimento.

O método Análise de Conteúdo tem sua história ilustrada pela primeira vez por H. Lasswell o qual fez análise de imprensa e propaganda a partir de 1915 e, na década de 70, o método foi potencializado com a proliferação dos computadores (BARDIN, 2010).

O método se aplica para “escutar” cientificamente e com rigor, assim como para compreender, analisar, sintetizar e descrever as comunicações configuradas em palavras, como um monólogo, um diálogo, entre um grupo restrito ou comunicação de massa (QUADRO 1) (BARDIN, 2010, p. 40).

Quadro 1 - Domínios possíveis da aplicação da análise de conteúdo

Código e Suporte	Número de Pessoas Implicadas na Comunicação			
	Uma pessoa "Monólogo"	Comunicação Dual – "Diálogo"	Grupo Restrito	Comunicação de Massa
Lingüístico Escrito	Agendas, maus pensamentos, conjecturas, Diários Íntimos.	Cartas, respostas a questionários, a teste projetivos, trabalhos escolares.	Ordens de serviço numa empresa, todas as comunicações escritas, trocadas dentro de um grupo.	Jornais, livros, anúncios publicitários, cartazes, literatura, textos jurídicos, panfletos.
Oral	Delírio do doente mental, sonhos.	Entrevistas e conversações de qualquer espécie.	Discussões, entrevistas, conversações de grupo de qualquer natureza.	Exposições, discursos, rádio, televisão, cinema, publicidade, discos.
Ícônico (Sinais, grafismos, imagens, fotografia, filmes etc.).	Rabiscos mais ou menos automáticas, gráficos, sonhos.	Respostas aos testes projetivos, comunicação entre duas pessoas através da imagem.	Toda a comunicação ícônica num pequeno grupo (p. ex.: símbolos icônicos numa sociedade secreta, numa casta...).	Sinais de trânsito, cinema, publicidade, pintura, cartazes, televisão.
Outros Códigos Semióticos				
(i. é, tudo o que não sendo lingüístico, pode ser portador de significações; ex.: música, código olfativo, objetos diversos, comportamentos, espaço, tempo, sinais patológicos etc.).	Manifestações históricas da doença mental, posturas, gestos, tiques, dança, coleções de objetos.	Comunicação não-verbal com destino a outrem (posturas, gestos, distância espacial, sinais olfativos, manifestações emocionais, objetos cotidianos, vestuário, alojamento...), comportamentos diversos, tais como os ritos e as regras de cortesia.		Meio físico e simbólico: sinalização urbana, monumentos, arte...: mitos, estereótipos, instituições, elementos de cultura.

Fonte: Bardin (2010, p. 35).

Considerando os domínios de aplicação da Análise de conteúdo, foi identificado

para esta tese os elementos indicados no quadro 2.

Quadro 2 – Domínios aplicados da Análise de Conteúdo

Código e suporte	Comunicação Dual “Diálogo”	Grupo restrito	Comunicação de massa
Linguístico	Resposta a questionários		Livros, artigos, literatura cinzenta
Escrito			Documentos das bibliotecas e do Consórcio Madroño
Oral	Entrevistas	Entrevista com os Bibliotecários Entrevista com os pesquisadores	Palestras, lives, cursos
Ícônico		Infografia	

Fonte: Elaborado pela autora baseado em Bardin (2010).

A Análise de Conteúdo organiza-se em três fases: 1) Pré-análise; 2) Exploração do material; 3) Tratamento de resultados, inferência e interpretação (BARDIN, 2010; FRANCO, 2008):

- 1) A Pré-análise: corresponde a fase da busca e da organização do material que será analisado com o objetivo de sistematizar as idéias iniciais a fim de conduzir o esquema traçado para realização das fases seguintes. Essa fase contém três incumbências que não se sucedem, obrigatoriamente, segundo uma ordem cronológica: a escolha dos documentos a serem submetidos à análise; a formulação das hipóteses e/ou dos objetivos; e, a elaboração dos índices e dos indicadores que vão fundamentar a interpretação final. A fase da Pré-análise engloba:
 - a) leitura flutuante: o primeiro contato com o material que será potencialmente analisado;
 - b) escolha dos documentos que serão submetidos à análise: pode ser definida *a priori* (pré-determinados) ou *a posteriori* (definidos na pré-análise) e sua escolha deve ser guiada por regras de: a) exaustividade – garantia de que todo o *corpus* que será submetido ao processo analítico seja contemplado; b) representatividade - possibilidade de retirar uma amostra para a realização da análise, desde que a amostra

seja representativa do universo inicial; c) homogeneidade - deve-se obedecer a critérios precisos e idênticos de escolhas; d) pertinência - os documentos que compõem o *corpus* devem ser adequados e corresponder ao objetivo da análise;

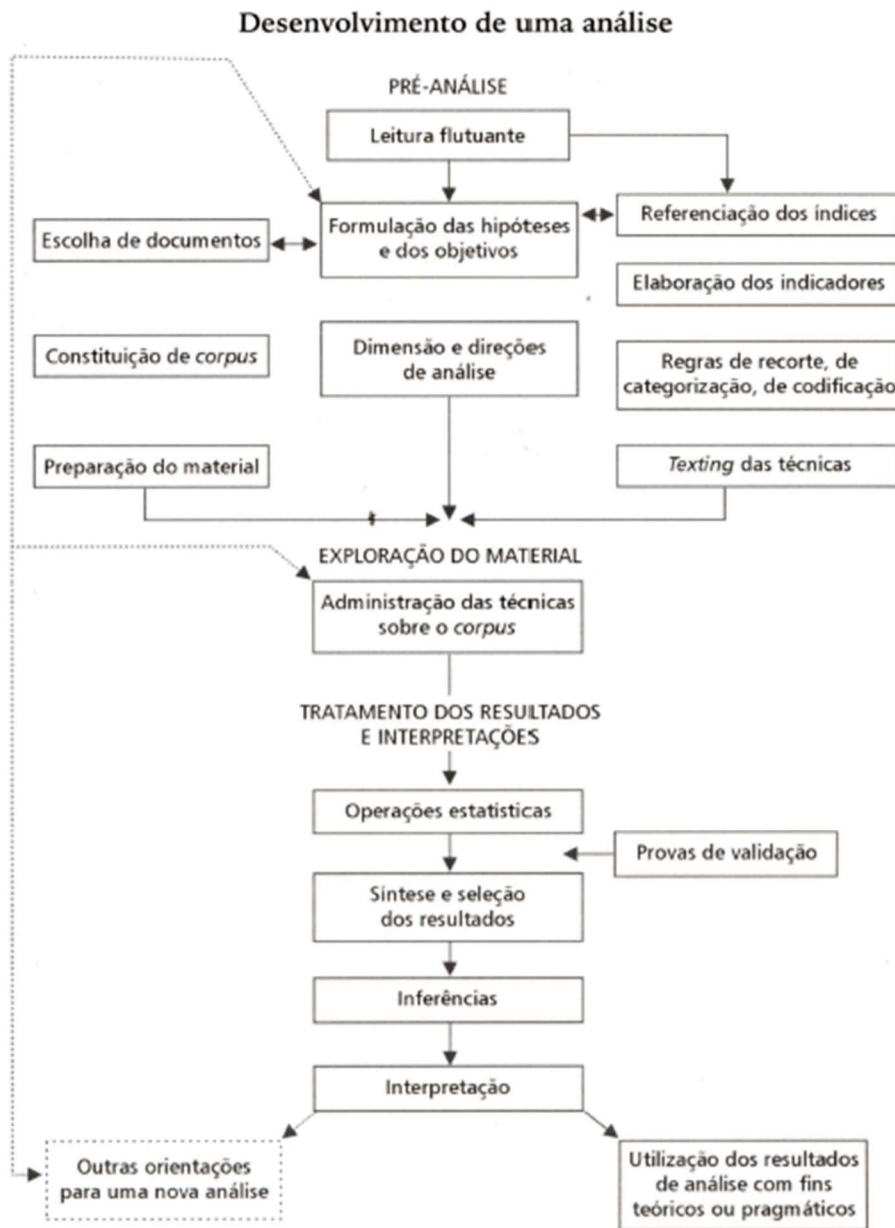
- c) formulação das hipóteses e/ou dos objetivos: levantar a hipótese e objetivos do trabalho tendo em vista as dimensões, a direção e os procedimentos de análise;
- d) A referência aos índices e elaboração de indicadores: Em considerável parte das pesquisas científicas, qualquer que seja o tema (índice) explicitado, o mesmo passa a ter mais importância para a análise dos dados, quanto maior a frequência observada (indicadores) de sua menção (FRANCO, 2008). Para determinação dos indicadores deve-se recorrer a uma análise quantitativa sistemática “[...] para que seja possível identificar a frequência relativa ou absoluta do tema escolhido e a proporcionalidade de sua menção em relação a outros temas igualmente presentes.” (FRANCO, 2008, p. 57-58).

Nessa perspectiva, Bardin (2010, p. 126, grifos do autor) destaca que na referenciação aos índices e elaboração dos indicadores são feitas “[...] determinadas operações de *recorte do texto* em unidades comparáveis de *categorização* para análise temática e de modalidade de *codificação* para o registro dos dados.”

- 2) Exploração do material: aplicação sistemática das decisões tomadas na Pré-análise com a codificação, decomposição ou enumeração e categorização do material.
- 3) Tratamento de resultados, inferência e interpretação: resultados são tratados de modo a se tornarem significativos e válidos permitindo estabelecer os resultados que serão apresentados em quadros, diagramas, figuras e modelos.

As três fases da análise de conteúdo estão sistematizadas na figura 1.

Figura 1 - Caminhos para o desenvolvimento da análise



Fonte: Fonseca Júnior (2014, p. 289).

Na etapa da Pré-análise, a leitura flutuante, nessa Tese, se iniciou com os documentos disponibilizados nas páginas das bibliotecas, do *Consórcio Madroño*, os enviados pelas bibliotecárias; a escolha dos documentos foi feita a *posteriori* seguindo a regra da pertinência; com leitura foram levantadas as hipóteses para o *framework* e os objetivos da Tese; também foi feita a referenciação das categorias e dos índices do *framework*.

As fases foram seguidas para a sistematização, análise e discussão final.

2.1 Universo da Pesquisa

As pesquisas científicas delimitam um universo, também indicado como uma extensão, um corpus ou uma amostra, do que será pesquisado (uma unidade observacional ou experimental, uma parcela conveniente) que podem ser pessoas, ambientes, documentos ou fatos (APPOLINÁRIO, 2012; FRANCO, 2008; MARCONI; LAKATOS, 2016). Nessa perspectiva a presente pesquisa foi composta por dois universos: o primeiro composto pelo repositório *e-cienciaDatos* com seus documentos e bibliotecários; o segundo, composto pelos pesquisadores que depositaram seus conjuntos de dados no repositório.

O primeiro universo foi composto pelo repositório *e-cienciaDatos* com documentos disponibilizados pelo *Consórcio Madroño* e pelas as bibliotecas que o compõem e as entrevistas e os questionários aplicados aos bibliotecários. As entrevistas e os questionários foram aplicados em dezembro de 2018. Com efeito, foram seguidos os procedimentos da Análise de conteúdo na fase da Pré-análise com a leitura flutuante e aplicando a regra da pertinência a fim de atender os objetivos da pesquisa.

O segundo universo constituiu-se pelos pesquisadores que depositaram seus conjuntos de dados no repositório. As entrevistas e os questionários foram aplicados entre junho e agosto de 2020.

Para a escolha do segundo universo dessa pesquisa foi aplicada a amostra não probalística intencional em que “[...] outros critérios que não a teoria da probabilidade são utilizados para a determinação dos sujeitos da amostra.” (APPOLINÁRIO, 2011, p. 3). De acordo com critérios estabelecidos pela pesquisadora para atingir os objetivos propostos a amostra foi composta por pesquisadores que arquivaram os conjuntos de dados resultantes dos projetos de pesquisas que coordenam no repositório. Nessa perspectiva foi seguido as indicações de Franco (2008, p. 54) onde destaca “Para se recorrer à amostragem, é preciso identificar a distribuição das características dos elementos da amostra.” Sendo assim, o delineamento final desse universo se constituiu após averiguação de sua extensão e pela representatividade dos *datasets* no repositório.

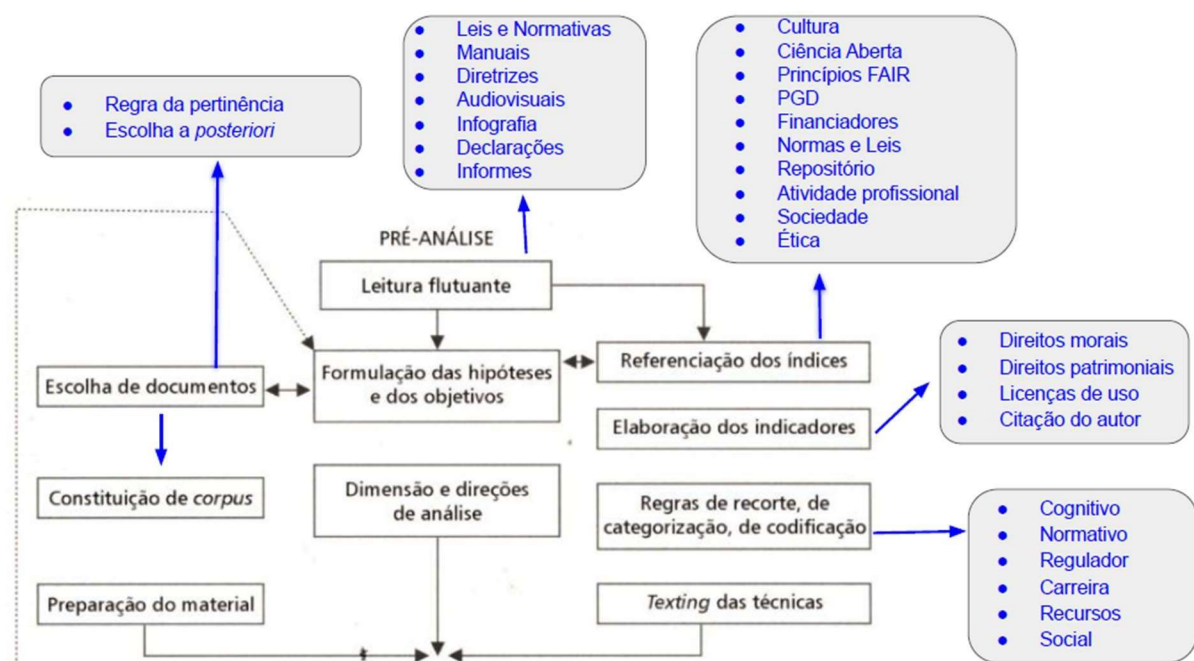
2.2 Procedimentos da Coleta de Dados

A coleta dos dados é a etapa da pesquisa em que se inicia a aplicação dos instrumentos elaborados e das técnicas selecionadas para a coleta dos dados (MARCONI; LAKATOS, 2016). Os instrumentos de coleta de dados foram os documentos disponibilizados pelo *Consórcio Madroño* e pelas bibliotecas que o constitui, bem como os questionários e as entrevistas aplicados aos bibliotecários que atuam nos repositórios e os pesquisadores que arquivaram os conjuntos de dados resultantes dos projetos de pesquisas que coordenam no repositório.

2.2.1 Coleta de dados dos documentos

A coleta de dados foi realizada a partir do universo selecionado para a pesquisa, com a utilização do método Análise de Conteúdo, seguindo os caminhos para o desenvolvimento da análise da figura 2.

Figura 2 - Indicação da fase de coleta



Fonte: Elaborada pela autora com base em Fonseca Júnior (2014).

Os documentos selecionados para a leitura flutuante foram as Leis e Normativas, manuais, diretrizes, audiovisuais, infografias, Declarações e informes. Na leitura flutuante dos documentos selecionados foram definidos os índices: Cultural, Ciência aberta, Princípios FAIR, PGD, Financiadores, Normas e Leis, Repositório, Atividade profissional, Sociedade e Ética (FIGURA 2). Destaca-se que o foco da pesquisa é o Direito de autor e, com o intuito de compreender e evidenciar o contexto em que o direito de autor está inserido, foram identificados os índices apresentados.

Posteriormente a seleção dos índices, foi elaborado um quadro (QUADRO 12) contendo os atributos e características considerados partes integrantes dos mesmos apresentado na seção dos resultados e discussões. O quadro viabilizou o delineamento do questionário e da entrevista aplicados aos pesquisadores.

A continuação da leitura flutuante identificou-se os indicadores: Direitos Morais, Direitos Patrimoniais, Licenças de uso e Citação do autor. Explorando um pouco a ideia de indicadores, Rozados (2005) destaca que os indicadores são medidas qualitativas ou quantitativas, reservadas para a descrição ou representação de um dado evento ou fenômeno servindo para sua avaliação, além de possibilitar comparações internacionais. Para ter indicadores de avaliação confiáveis de mensuração e aferição a busca por indicadores foi realizada em conceitos mencionados nos documentos analisados.

Em seguida, foram selecionadas as categorias identificadas na análise dos documentos. A Análise de Conteúdo dispõe de diferentes técnicas, tais como: Análise Categorical; Análise de Avaliação; Análise da Enunciação; Análise Proposicional do Discurso e Análise de Expressão. O tipo de análise escolhida para a pesquisa foi a Análise Categorical.

A Análise Categorical é a técnica mais antiga e a mais utilizada entre as técnicas da Análise de Conteúdo funcionando por desmembramento do texto em categorias e reagrupamento das unidades baseado em analogias, a partir de critérios definidos (BARDIN, 2010; FONSECA JÚNIOR, 2014; FRANCO, 2008).

A definição de categorias pode ser guiada por dois caminhos: *a priori* - pré-determinados, definidas antes da coleta; *a posteriori* - emergem do conteúdo analisado e implicam constante ida e volta do material de análise à teoria (BARDIN, 2010; FRANCO, 2008). Para essa pesquisa foi definido escolher as categorias *a posteriori* seguindo os critérios apresentados em Bardin (2010) e em Franco (2008)

no que diz respeito à definição de categorias de qualidade:

- A exclusão mútua: um único princípio de classificação deve orientar sua organização em que um dado só deve estar em uma categoria;
- A pertinência: uma categoria deve estar adaptada ao material de análise escolhido e ao quadro teórico definido e deve, também, refletir as intenções da pesquisa, às questões do pesquisador e/ou corresponder às características das mensagens;
- A objetividade e a fidedignidade: as diferentes partes de um mesmo material, ao qual se aplica uma determinada categoria, devem ser codificadas do mesmo modo, mesmo quando submetidas a várias análises;
- A produtividade: um conjunto de categorias é produtivo contanto que concentre a possibilidade de fornecer resultados férteis.

Com a análise categorial e com a regra da pertinência foram identificadas seis categorias: Cognitiva, Normativa, Reguladora, Carreira, Recursos e Social.

A partir da identificação dos índices, dos indicadores e das categorias foram estabelecidas as perguntas que fizeram parte do questionário e entrevista aplicados aos pesquisadores. Os dados coletados subsidiaram o estudo proporcionado elementos para a análise que proporcionou inferências e compreensão do universo pesquisado.

2.2.2 Coleta de dados com os Bibliotecários

Para a coleta de dados com os bibliotecários foi aplicado um questionário composto por 30 perguntas sendo 10 abertas e 20 fechadas sobre as atividades desempenhadas com a gestão dos dados e seu ciclo de vida, assim como as competências e habilidades necessárias para atuarem nos repositórios, tanto em serviços técnicos quanto no auxílio e treinamento dos usuários. O questionário serviu de pré-teste para a identificação dos índices e a elaboração dos indicadores, além de subsidiar as perguntas que compuseram o roteiro da entrevista e do questionário aplicados aos pesquisadores. O Termo de Consentimento livre e Esclarecido

precedeu a pesquisa e foi apresentado aos participantes da pesquisa no início do questionário e da entrevista (ANEXO A).

2.2.3 Coleta de dados com os pesquisadores

A definição de realizar entrevistas com os pesquisadores que compartilharam seus conjuntos de dados resultantes dos projetos coordenados no repositório surge da necessidade de compreender a realidade do compartilhamento de dados via repositório e os pontos de vista dos pesquisadores sobre direitos de autor e as temáticas ao seu redor. Verificou-se a existência de pouca literatura sobre a temática e, por isso, um estudo baseado apenas na literatura não traria elementos suficientes para reflexão proposta nessa Tese. Desta forma, o objetivo do questionário e da entrevista foi identificar percepções e atitudes dos pesquisadores e, assim, confrontar com a operacionalização do repositório, o que comprovariam as hipóteses levantadas. Os detalhes de seleção dos entrevistados e método de análise e sistematização dos resultados foram descritos na seção resultados e discussões.

Os pesquisadores participantes da pesquisa foram selecionados a partir da amostra indicada na seção 2.1. Os entrevistados tiveram a liberdade de se retirar da pesquisa a qualquer momento e não foi necessário responder a nenhuma pergunta para avançar no questionário ou na entrevista. O Termo de Consentimento Livre e Esclarecido precedeu a pesquisa e foi apresentado aos participantes da pesquisa no início do questionário e da entrevista (ANEXO B).

Foi realizado um pré-teste do questionário com pesquisadores do grupo de pesquisa Tecnologias de Acesso a Dados em maio de 2020 visando verificar as ambiguidades, dificuldades de compreensão das perguntas e conhecimento das temáticas, bem como o tempo de resposta. No final do questionário havia uma pergunta para os pesquisadores relatarem as dificuldades encontradas nas respostas, tempo despendido e indicação de sugestões. O pré-teste serviu para calibrar o questionário. A partir do resultado do pré-teste optou-se em aplicar questionário com perguntas fechadas e entrevista com perguntas abertas para compreensão da percepção dos entrevistados sobre as temáticas apresentadas.

O questionário final contou com 19 questões, sendo sete abertas e 12 fechadas. As perguntas abertas serviram para o roteiro da entrevista. Os

pesquisadores ficaram livres para comentarem as perguntas fechadas. A aplicação do questionário e da entrevista foram feitas no mesmo horário agendado com os participantes nos meses de junho, julho e agosto de 2020.

A aplicação do questionário e da entrevista foram feitas via *Skype* e *Hangout* e as mesmas foram gravadas com autorização dos participantes. Considerando que a entrevista foi gravada, os pesquisadores responderam às perguntas que foram transcritas para análise final.

2.3 Procedimentos de análise de dados

Com o plano de trabalho estabelecido *a priori* e seguindo o delineamento do tipo, natureza e método aplicados para a pesquisa, foi iniciada a análise e interpretação dos dados, evitando distorções para assegurar qualidade das inferências apresentadas e discutidas.

O método de pesquisa Análise de Conteúdo foi aplicado para a análise dos dados coletados no primeiro e segundo universos da pesquisa.

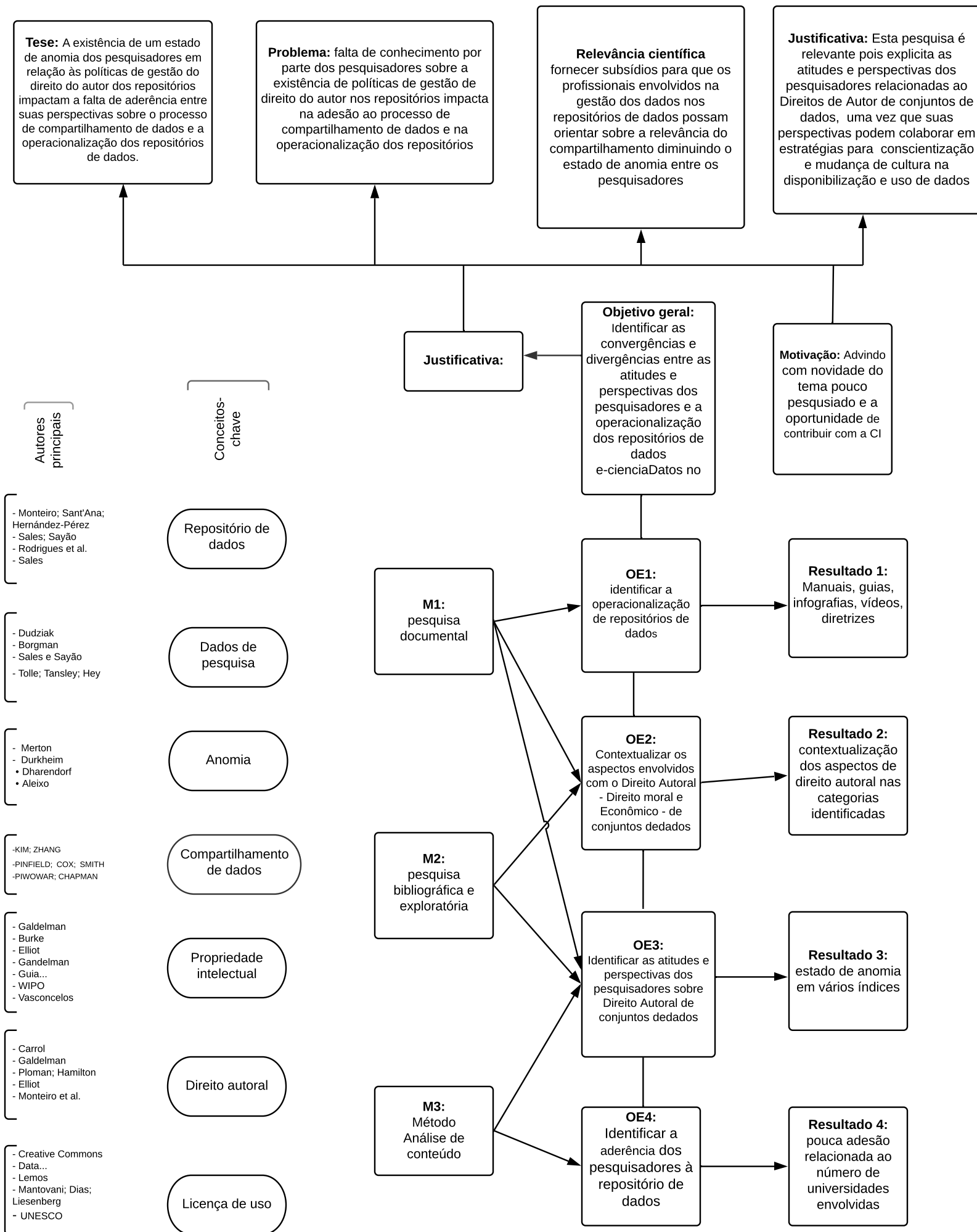
Após a coleta dos dados, se iniciou a etapa de análise e interpretação dos conteúdos seguindo os processos da análise de conteúdo. Os processos são passos “[...] a serem seguidos e, para o efetivo caminhar neste processo, a contextualização deve ser considerada como um dos principais requisitos, e mesmo como o pano de fundo para garantir a relevância dos sentidos atribuídos às mensagens.” (FRANCO, 2008, p. 16). A contextualização foi garantida com a identificação das categorias apresentadas na subseção 2.2.1.

A Análise de Conteúdo dispõe de uma atitude interpretativa da comunicação, porém, sustentada por processos técnicos de validação pois “Qualquer análise objectiva procura fundamentar impressões e juízos intuitivos, através de operações conducentes a resultados de confiança.” (BARDIN, 2010, p. 44).

O pesquisador cria ou tem a sua disposição todo um jogo de procedimentos analíticos que podem ser adaptados à natureza do material analisado e à questão que se procura resolver (BARDIN, 2010).

Para ter índices e categorias confiáveis de mensuração e aferição, a busca pelos mesmos foi realizada em temas, conceitos e características do primeiro universo da pesquisa. A seleção e delimitação dos índices e categorias foram amparadas pela Teoria Institucional de W. Richard Scott, pela Teoria do Comportamento Planejado de

Icek Ajzen e no Modelo de pesquisa de comportamento de compartilhamento de dados de Youngseek Kim. Os procedimentos seguidos serviram para composição de um *framework*, tópico apresentado na próxima seção.



3 MODELO CONCEITUAL DA PESQUISA

Para entender a dinâmica que abarca as convergências e divergências entre repositórios de dados e pesquisadores relacionados aos direitos de propriedade intelectual e a disponibilização dos conjuntos de dados, é apropriado considerar as perspectivas e atitudes dos pesquisadores. A pesquisa pressupõe que a dubiedade dos pesquisadores no que se refere aos direitos de propriedade intelectual de conjuntos de dados diminui a aderência ao compartilhamento de dados, agravadas pelas boas práticas de gestão de dados no contexto da Ciência Aberta e dos princípios FAIR.

Assim, delinea-se um modelo teórico para identificar e elucidar as percepções e atitudes de pesquisadores vinculados às Universidades da Comunidade de Madri que integram o *Consórcio Madroño* em relação aos direitos de propriedade intelectual ao compartilhar conjuntos de dados. Para ter um instrumento robusto, o modelo teórico foi baseado nas estruturas teóricas, Teoria Institucional (TI) (SCOTT, 1995), Teoria do Comportamento Planejado (TCP) (AJZEN, 1991) e no modelo de pesquisa de comportamento de compartilhamento de dados em repositório de dados proposto por Kim (2017) sendo o último estruturado para fatores que influenciam o comportamento dos cientistas que compartilham dados por meio de repositórios de dados. Kim (2017, p. 879) afirma que “[...] as pressões institucionais, motivações individuais e recursos afetam os diversos comportamentos de compartilhamento de dados dos cientistas [...]”

O modelo teórico da Tese tem o intuito de elucidar como os cientistas tomam a decisão para disponibilizar seus dados com o arquivamento no repositório de dados e em que medida os fatores institucionais, individuais e sociais os influenciam. Para cada elemento do modelo proposto foram levantadas hipóteses, indicadas com a letra “H” adicionado um número para cada uma. Nas hipóteses há a indicação se influencia “positivamente” ou “negativamente” o compartilhamento dos dados via repositório. A indicação de “positivamente” e “negativamente” é embasada na análise das respostas dos pesquisador. Com o resultado dos questionários e entrevistas, tais hipóteses foram confirmadas ou refutadas.

A utilização desses modelos se baseou na identificação e aplicação por outras pesquisas como tendo relevância no domínio do compartilhamento de dados: no

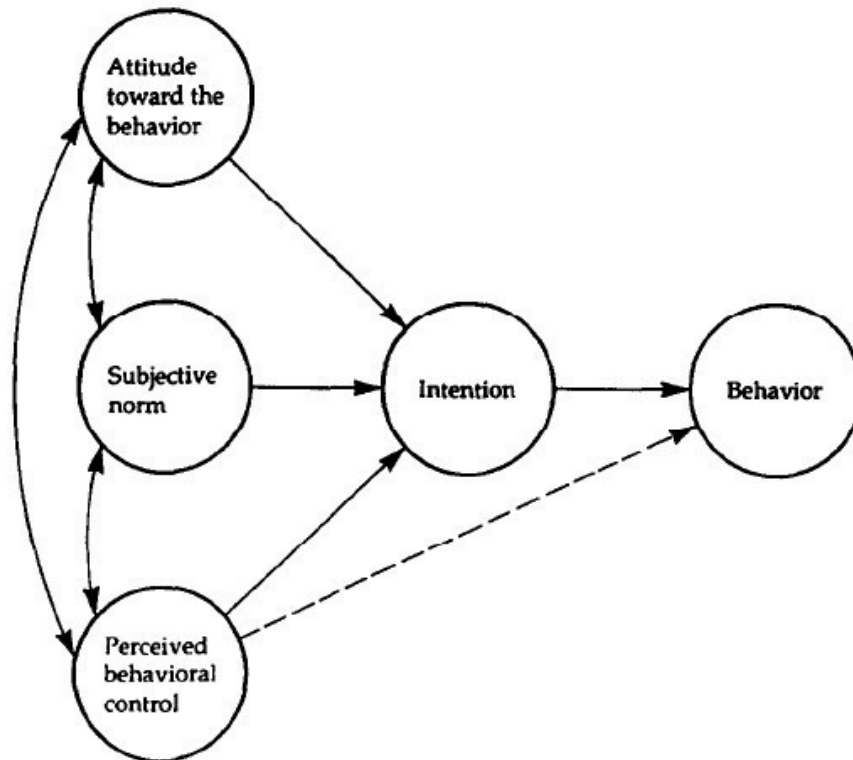
contexto de dados geoespaciais (LUBIDA et al., 2014); nas áreas de ciência, tecnologia, engenharia e matemática (KIM; ZHANG, 2016) e ciências sociais (KLIM; ADLER, 2015) e disponibilizados em repositórios (KIM, 2017). A Teoria do Comportamento Planejado (TCP) é uma teoria da psicologia social bem estabelecida que descreve como as crenças pessoais influenciam suas intenções e seus comportamentos subsequentes e explica como as pessoas tomam suas decisões com base em suas próprias motivações (KIM; ZHANG, 2016). A Teoria Institucional propicia percepções significativas sobre a importância dos ambientes institucionais com suas regras, normas e cultura nas ações e comportamentos das pessoas e como podem moldar as crenças das pessoas e seus comportamentos não racionais (KIM; STANTON, 2012, 2016).

Para facilitar a apresentação e compreensão do modelo teórico proposto nessa Tese, os modelos de Scott, Ajzen e Kim são apresentados e correlacionados com cada etapa da elaboração do *framework* da Tese.

3.1 Dimensão de análise do pesquisador

A Teoria do Comportamento Planejado (TCP) é “[...] uma teoria projetada para prever e explicar o comportamento humano em contextos específicos [...]” com fator central onde há “[...] a intenção do indivíduo em executar um determinado comportamento.” (AJZEN, 1991, p. 181, tradução nossa). A TCP explica a racionalidade dos comportamentos dos sujeitos com base em suas motivações atitudinais e condições facilitadoras de recursos que promovam, inclusive, seus comportamentos em compartilhar dados (KIM, 2017). Essa teoria ajuda a explicar as motivações subjacentes e a disponibilidade de recursos, facilitando o compartilhamento de dados dos pesquisadores no contexto institucional (AJZEN, 1991; KLIM; ADLER, 2015).

Figura 3 - Diagrama estrutural da Teoria do comportamento planejado



Fonte: Ajzen (1991, p. 182).

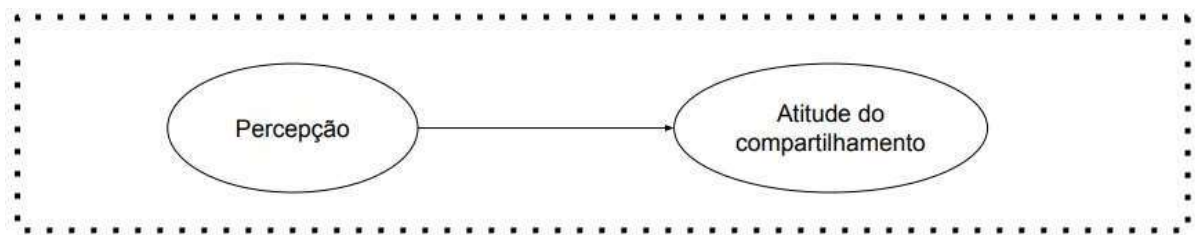
A teoria do comportamento planejado apresenta uma relação clara de como as atitudes de um indivíduo (*Attitude toward the behavior*), as normas subjetivas (*Subjective norm*) e os controles comportamentais percebidos (*Perceived behavioral control*) influenciam seus comportamentos, mediados por intenções comportamentais (AJZEN, 1991; KIM; ZHANG, 2015). Um fator central na Teoria do comportamento planejado é a intenção do indivíduo em executar um determinado comportamento (AJZEN, 1991). Relevante destacar que o desempenho de um comportamento é uma função conjunta de intenções e controle comportamental percebido.

Segundo Kim e Zhang (2015, p. 190, tradução nossa)

Como teoria motivacional, a teoria do comportamento planejado pode explicar como os indivíduos tomam suas decisões baseadas em suas próprias motivações. Essa teoria é apropriada para examinar os comportamentos de compartilhamento de dados de pesquisadores individuais.

Os aspectos da TCP aplicados para embasar o modelo teórico proposto resultaram nos dois componentes com foco na propriedade intelectual (FIGURA 4): Os dois componentes estão na camada central do modelo teórico proposto em que são baseadas as hipóteses levantadas segundo a percepção, perspectiva e atitude dos pesquisadores.

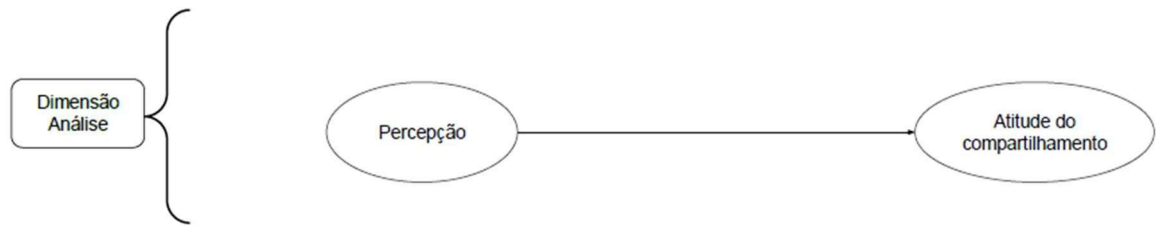
Figura 4 - Elementos da análise da percepção e atitude de pesquisadores



Fonte: Elaborado pela autora.

- Percepção: correspondente ao *Perceived behavior control* da Teoria do comportamento planejado. A concepção pode ser entendida como os recursos e oportunidades disponíveis para o pesquisador que impactam suas intenções e ações e vão ditar a probabilidade de realização do compartilhamento dos conjuntos de dados; entendida também como a percepção das pessoas sobre a facilidade ou dificuldade em executar o comportamento de interesse. Ademais, corresponde a norma subjetiva pois refere-se à percepção da pressão social que vai influenciar na execução ou não do compartilhamento de dados. Suas intenções, juntamente com controle do comportamento percebido, são responsáveis por variações consideráveis no comportamento real;
- Atitude do compartilhamento de dados: corresponde ao comportamento, à ação de compartilhar seus conjuntos de dados. Após a avaliação e percepção das vantagens e desvantagens que o cenário possibilita e das pressões sociais, o pesquisador decide entre as alternativas disponíveis.

Figura 5 - Dimensão da análise da percepção e atitude de pesquisadores



Fonte: Elaborado pela autora.

O arcabouço da Teoria do Comportamento Planejado apoiou a construção da Dimensão de Análise com a percepção e atitude do compartilhamento de conjuntos de dados dos pesquisadores. A Dimensão de Análise está relacionada com a propriedade intelectual e fundamentou todas as hipóteses levantadas apresentadas no framework final.

A teoria do comportamento planejado é uma estrutura teórica profícua para explicar as percepções e as atitudes do comportamento de compartilhamento de dados dos pesquisadores das universidades do *Consórcio Madroño*.

3.2 Dimensão Institucional para a análise do pesquisador

A Teoria institucional advém de estudos sociológicos e organizacionais e auxilia a conexão entre as forças entrelaçadas de instituições, infraestrutura e pessoas além de examinar como os atores sociais executam comportamentos aceitáveis para serem legitimados diante das pressões institucionais (SCOTT, 1995; KIM, 2017). O contexto em que os cientistas estão atuando fornece *insights* sobre como os atores sociais são influenciados pelas pressões institucionais dentro do ambiente institucional (SCOTT, 1995; KLIM; ADLER, 2015). O compartilhamento de dados pelos pesquisadores está associado em seus respectivos contextos institucionais, incluindo universidades ou institutos de pesquisa, agências de fomento, editores de periódicos e áreas de conhecimento fazendo-se necessário examinar as influências institucionais nos comportamentos de compartilhamento de dados dos pesquisadores (KLIM; ADLER, 2015).

A Teoria Institucional tem três pilares de apoio institucional (SCOTT, 1995) apresentado na figura 6.

Figura 6 – Três pilares da Teoria Institucional

	Regulatório	Normativo	Cognitivo
Bases da submissão	Experiência Conveniência	Obrigaç�o social	Entendimento compartilhado
Indicadores	Regras Leis Sanç�es	Certifica��o Acredita��o Pertin�ncia	Cren�as comuns
Bases da Legitimidade	Sancionada legalmente	Governada moralmente	Reconhec�vel Apoiada culturalmente

Fonte: Elaborado pela autora baseado em Scott (1995).

- **Pilar Regulador:** pilar mais destacado na TI devido ao fato de ser o pilar que estabelece regras, as inspecionam e estabelecem recompensas ou puni  es na tentativa de influenciar futuros comportamentos (SCOTT, 1995). Em uma perspectiva da realidade institucional, a Institui  o assume que seus atores t m interesse natural motivado em fazer suas escolhas com a l gica custo/benef cio (SCOTT, 1995).
- **Pilar Normativo:** introdu  o da dimens o prescritiva, avaliativa e obrigat ria na vida social em que especificam como seus elementos e bases devem ser feitos, definindo significados legitimados aos valores e regras adotados (SCOTT, 1995);
- **Pilar Cognitivo:** presen a e intera  o entre os atores, entendimento internalizado de cada ator, a partir de sua interpreta  o da realidade social em que participa, juntamente com o sistema cultural atuante (SCOTT, 1995). Os atores tendem a imitar aqueles que tiveram sucesso (SCOTT, 1995).

Os três pilares de Scott embasaram a definição de três categorias apresentadas na figura 6 sendo elas Cognitiva, Normativa e Reguladora. No contexto dos três pilares foram atribuídos os índices: Cultural, Ciência aberta, Princípios FAIR, PGD, Agência de fomento, Repositório, Normas e Leis (FIGURA 7). Destaca-se que os índices em vermelho foram acrescentados, embasados no conceito e característica dos três pilares da TI.

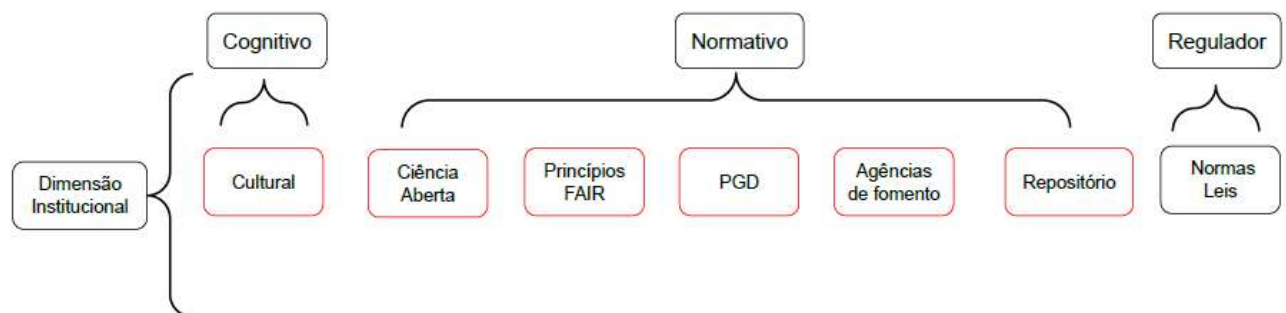
Figura 7 - Índices para a Teoria Institucional



Fonte: Elaborado pela autora.

Os três pilares da TI aplicados para embasar o modelo teórico proposto juntamente aos índices resultaram nos componentes da figura 8.

Figura 8 - Dimensão Institucional da percepção e atitude de pesquisadores



Fonte: Elaborado pela autora.

O arcabouço dos três pilares da TI embasaram a aderência com os elementos da figura 8. Toda a Dimensão Institucional com seus índices e categorias estão relacionados com a propriedade intelectual. Junto aos elementos são indicadas as hipóteses:

- Cultural: relacionado ao Pilar cognitivo, esse componente aponta a relação e interação entre os atores em nível institucional, nacional e internacional destacando o sistema organizacional e a vida social com os direitos, regras e responsabilidades. A propriedade intelectual foi trabalhada com os atores envolvidos perpassando as crenças e como entendem o aspecto da propriedade intelectual dentro do contexto em que atuam.

H1.: A cultura influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório.

- Ciência aberta: integrando o pilar normativo engloba as práticas científicas abertas e a mudança do comportamento dos pesquisadores relacionadas, entre outros aspectos, ao compartilhamento de dados de forma aberta, orientando sobre o uso de licenças e a importância da citação dos detentores dos dados;

H2.: Ciência Aberta influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório.

- Princípios FAIR: dispõe aspectos do pilar normativo com seus princípios norteadores em que estabelecem e regulamentam os procedimentos para que os meta(dados) estejam acessíveis e reutilizáveis com os direitos associados e termos de uso, além de demandar mudanças em termos de cultura e prática de pesquisa; nos princípios *Findable* e *Reusable* são recomendados os metadados dos detentores dos dados e das licenças de uso respectivamente;

H3.: Princípios FAIR influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório.

- Plano de Gestão de Dados (PGD): condizente com o pilar normativo, o PGD é um documento que articula os aspectos da gestão dos dados pelos pesquisadores e contribuem para a gestão no repositório; algumas ferramentas para o desenvolvimento do plano de gestão de dados norteiam sua elaboração

com as diretrizes das agências. O aspecto da propriedade intelectual está previsto nas orientações das agências de fomento e do repositório, inclusive nas orientações das ferramentas e deve ser sinalizada pelos autores qual licença será aplicada ao seu conjunto de dados.

H4.: Plano de Gestão de Dados influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório.

- Agências de fomento: perpassando o pilar normativo, suas diretrizes vão sendo articuladas de acordo com as necessidades institucionais e sociais, envolvendo o estabelecimento de regras e como as mesmas devem ser aplicadas; entre seus requerimentos, a propriedade intelectual e outros direitos relevantes são reconhecidos e adequadamente indicados;

H5.: As agências de fomento influenciam positivamente a percepção dos pesquisadores fomentando o compartilhamento dos dados via repositório.

- Repositório: consiste no pilar normativo e, como instituição, estabelece e implementa políticas e diretrizes, conduzindo o comportamento de compartilhamento dos dados e articulando a demanda dos outros elementos do modelo teórico proposto.

H6.: Repositório influencia positivamente o compartilhamento dos dados.

H7.: Repositório influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados.

- Leis e normas: Relacionado ao pilar regulador, esse componente discorre sobre o entendimento dos pesquisadores relacionado a legislação e normativas da União Europeia, da Espanha e das universidades relativos à propriedade intelectual das quais o repositório de dados utilizou para desenvolver sua política;

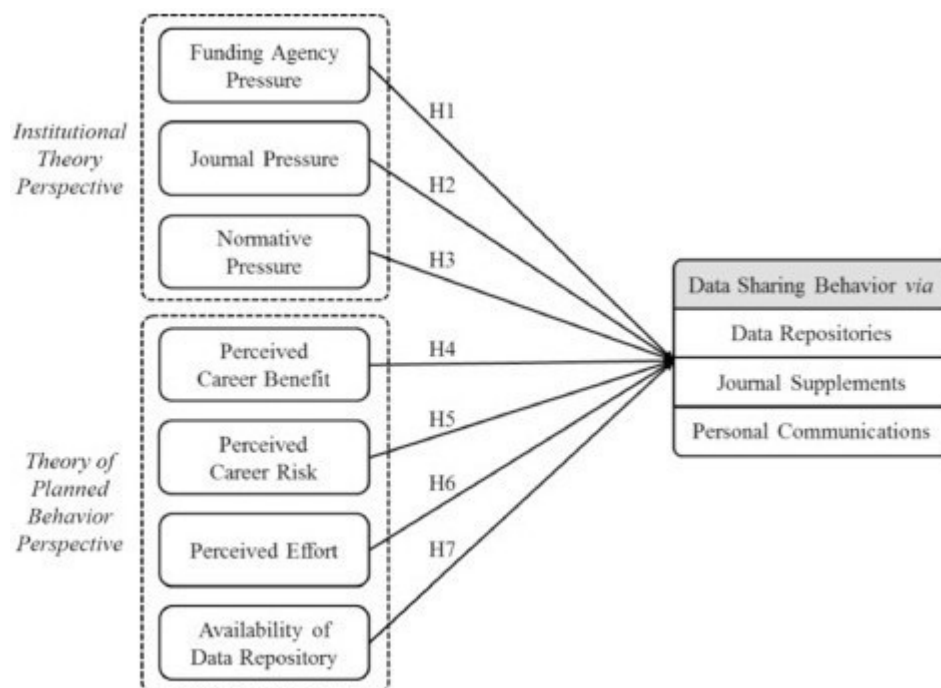
H8.: Normas e Leis influenciam positivamente o compartilhamento dos dados via repositório.

O modelo conceitual da pesquisa é completado com a adição da Dimensão Pesquisador apresentada a seguir.

3.3 Dimensão Pesquisador para análise da atitude do compartilhamento

O modelo para comportamento de compartilhamento de dados de cientistas através de repositórios de dados é um modelo baseado na Teoria Institucional e na Teoria do Comportamento Planejado (FIGURA 9).

Figura 9 - Modelo para o comportamento de compartilhamento de dados



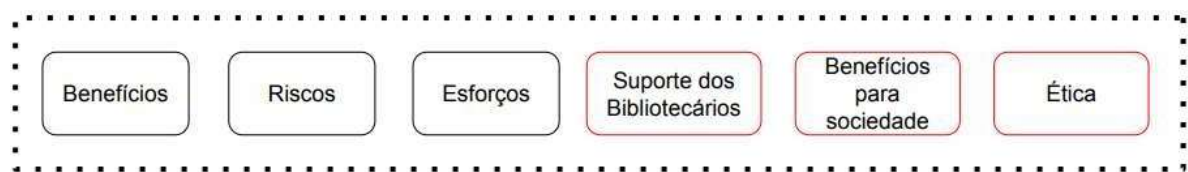
Fonte: Kim (2017, p. 826)

O modelo de pesquisa de Kim (2017) fornece um mapa dos comportamentos de compartilhamento de dados dos cientistas, com base na combinação da Teoria Institucional e da Teoria do Comportamento Planejado. O modelo elucida de que forma os pesquisadores tomam suas próprias decisões para o uso de métodos

diversificados para compartilhar dados. Sustentado na TI, o autor considera como influência as pressões regulatórias de agências de financiamento e editores de periódicos e a pressão normativa das expectativas de seus colegas. Da mesma forma, o autor considera a TCP para embasar o comportamento de compartilhar dados seguindo as influências na sua carreira.

Os aspectos do modelo de Kim (2017) também serviram de suporte para embasar o modelo teórico proposto nessa Tese e, por meio da identificação dos índices selecionados na análise dos documentos (FIGURA 2), determinou-se os índices da figura 10 com foco na propriedade intelectual. O índice Atividade profissional resultou nos índices Benefícios, Riscos e Esforços; o índice Repositório resultou no índice Suporte dos bibliotecários; o índice Sociedade resultou no índice Benefícios para a sociedade e o índice Ética não teve alteração. O *framework* corresponde a Dimensão do Pesquisador. Relevante destacar que os índices em vermelho foram acrescentados, embasados no modelo de Kim (2017).

Figura 10 - Índices para a Teoria do Comportamento de compartilhamento de dados



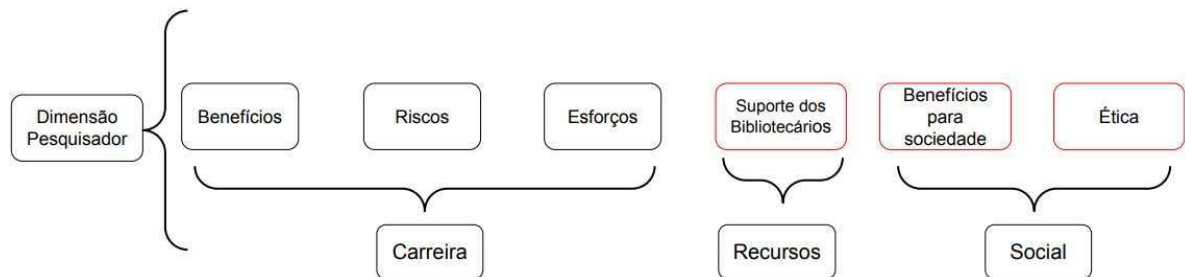
Fonte: Elaborado pela autora.

Embasado na Teoria do Comportamento de Compartilhamento de Kim (2017) e nos pilares da Teoria Institucional de Scott (1995), foram acrescentados três pilares no *framework* destacados em vermelho: Carreira; Recursos; Social:

- Pilar Carreira: grau em que um cientista pensa que o compartilhamento de dados de pesquisa causaria resultados que influenciam negativamente e positivamente em sua carreira incluindo o tempo e energia exigidos com a preparação dos dados para o compartilhamento (KIM; STANTON, 2012, 2016);
- Pilar Recursos: recursos disponíveis para o pesquisador que impactam suas intenções e ações, além de ditar a probabilidade de realização do compartilhamento dos conjuntos de dados (KIM, 2017).

- **Pilar Social:** aumenta a transparência, dá mais credibilidade às descobertas da pesquisa, fornecendo evidências para apoiar estruturas e decisões políticas e uma fonte para os pesquisadores consultarem quando considerarem como aproveitar os estudos existentes (KIM; ADLER, 2015).

Figura 11 - Dimensão Pesquisador da percepção e comportamento de pesquisadores



Fonte: Elaborado pela autora.

Os pilares seguem os respectivos índices:

- **Benefícios na carreira:** são os benefícios percebidos e esperados com o compartilhamento de dados decorrentes do trabalho do pesquisador;

H9.: Benefícios na carreira influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório.

- **Riscos na carreira:** designa os riscos percebidos no processo de compartilhar seus dados;

H10.: Riscos na carreira influencia negativamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório.

- **Esforços na carreira:** são os trabalhos decorrentes e energia despendida com a gestão dos dados e com sua preparação para o compartilhamento.

H11.: Esforços na carreira influenciam negativamente no compartilhamento dos dados via repositório.

- Suporte dos Bibliotecários: equivale ao suporte institucional percebido, que nessa pesquisa foi acrescentado pois corresponde ao papel estratégico dos bibliotecários na gestão dos dados, no trabalho do repositório e nos treinamentos dos pesquisadores;

H12.: Suporte dos Bibliotecários influencia positivamente no compartilhamento dos dados via repositório.

- Benefícios para a sociedade: elemento incluído no *framework* pois está alinhado as necessidades de disponibilização dos dados para a agilidade em respaldar determinada demanda da sociedade, rapidez na análise situacional e toma de decisão;

H13.: Benefícios para a sociedade influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório.

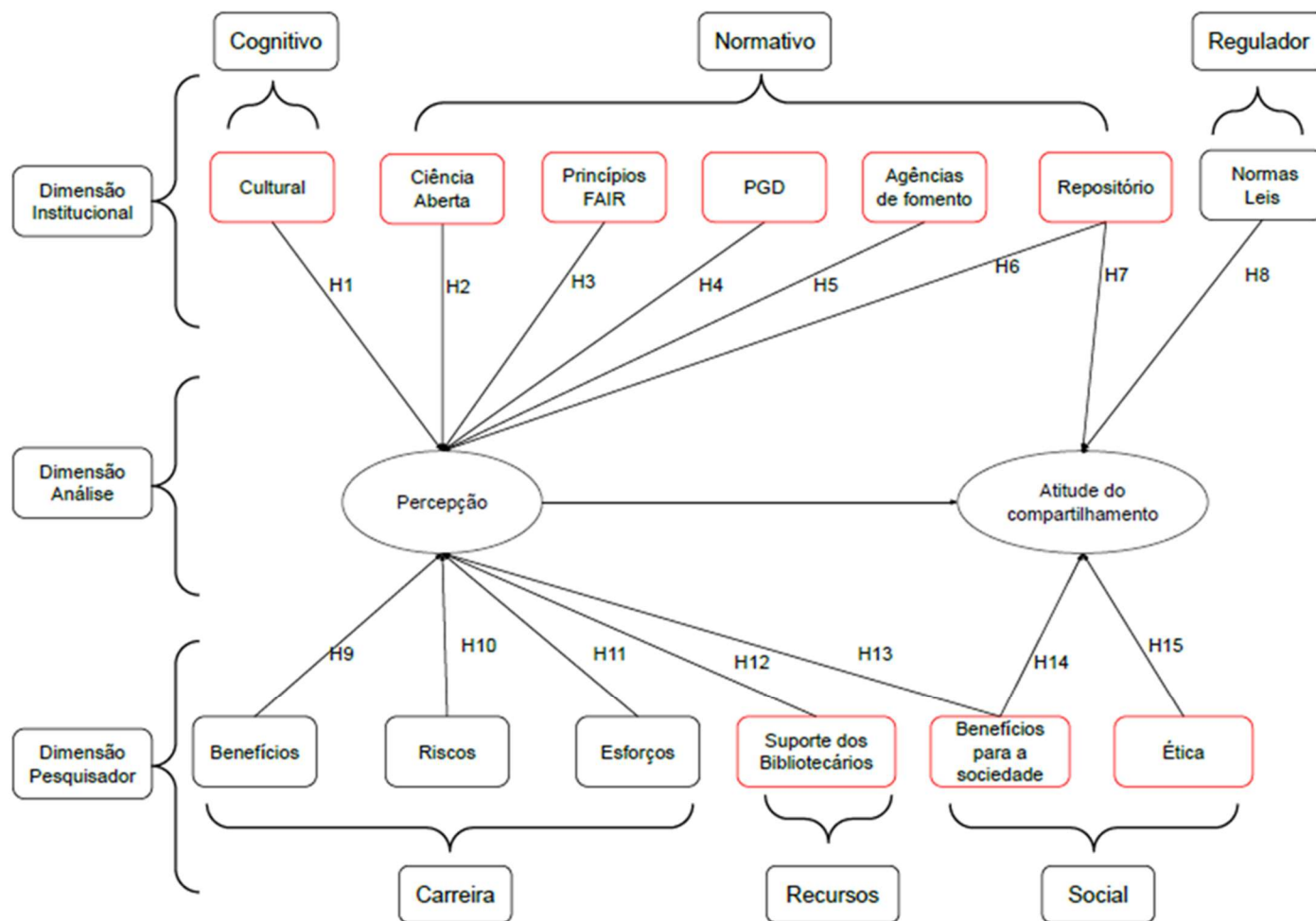
H14.: Benefícios para a sociedade influencia positivamente no compartilhamento dos dados via repositório.

- Ética: elemento acrescentado ao *framework* indicando o conjunto de valores de uma área do conhecimento e seu comportamento apropriado ou impróprio juntamente aos deveres, os princípios e as normas que são compatíveis aos grupos dos pesquisadores em seu contexto;

H15.: Ética influencia positivamente o compartilhamento dos dados via repositório.

Os itens de medição dessa pesquisa foram baseados na TI de Scott (2001), na TCP de Ajzen (2001) e no modelo de Kim (2017) (FIGURA 12), incluindo alguns dos itens de medição criados especificamente para esta pesquisa destacados em vermelho.

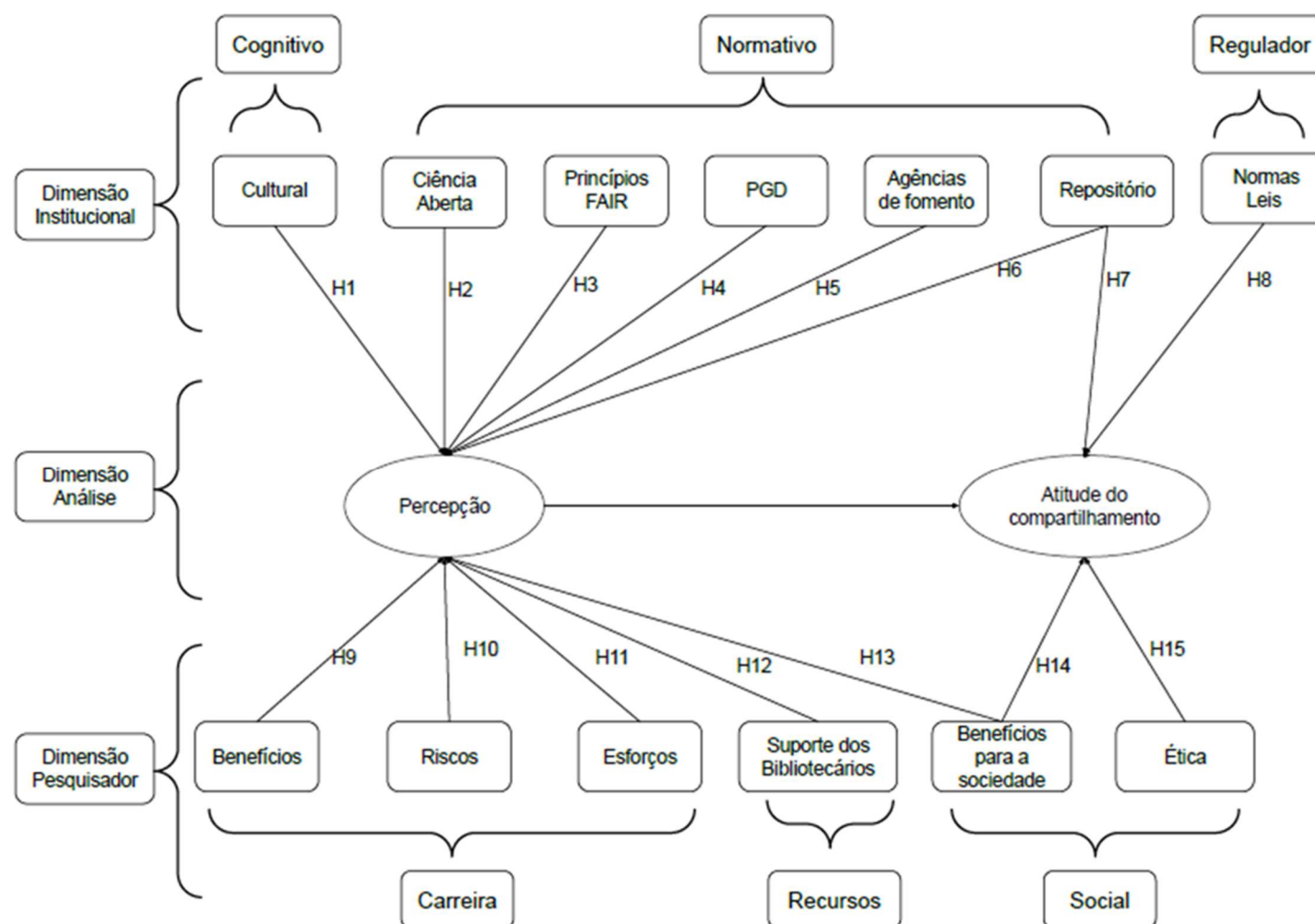
Figura 12 - Framework das perspectivas e atitudes do compartilhamento dos dados



Fonte: Elaborado pela autora.

O framework final foi composto baseado nos índices, indicadores e categorias identificados na Pré-análise proposta no método Análise de conteúdo. O framework serviu de base para a elaboração e organização das perguntas que fizeram parte do questionário e da entrevista aplicados aos pesquisadores.

Figura 13 - Composição do Framework final



Fonte: Elaborado pela autora.

Esta pesquisa validou empiricamente o modelo proposto e as hipóteses de pesquisa empregando dados coletados com os pesquisadores das universidades (FIGURA 13). O modelo auxiliou nas construções e a avaliação dos relacionamentos hipotéticos. Também produziu resultados sobre as percepções e atitudes de compartilhamento de dados dos pesquisadores em diferentes disciplinas.

No próximo capítulo é apresentado o compartilhamento de dados e o estado de anomia que pode causar retenção de dados pelos pesquisadores, com exposição de quatro autores, filósofos e sociólogos, alinhando a anomia com o compartilhamento de dados.

4 COMPARTILHAMENTO DE DADOS

A introdução, em larga escala, de dados na sociedade está redefinindo o domínio de ação e interação tanto entre humanos, quanto humanos-computador, em todas as esferas, públicas e privadas, redefinindo, inclusive, as atividades de pesquisa científica e os processos da comunicação científica, gerando novas atuações. Geração, compartilhamento e reuso de dados de pesquisa estão presentes nas pesquisas científicas.

O compartilhamento de dados entre pesquisadores pode ser essencial em diversificadas áreas e contribuir para acelerar a colaboração científica, conduzir estudos comparativos, permitir pesquisas em larga escala, verificar e validar resultados de pesquisas, aumentar a eficiência e a qualidade da pesquisa, contribuir em diversos campos de atuação e facilitar a reutilização de dados por terceiros (KIM; ZHANG, 2015; PINFIELD; COX; SMITH, 2014; PIWOWAR; CHAPMAN, 2010).

Os dados podem ser definidos de formas diferentes, dependendo do setor observado e, no contexto das pesquisas, correspondem aos *inputs* que servem para validar seus resultados. No contexto científico, tais dados podem ser planilhas, cadernos de laboratório ou de campo, questionários, transcrições, áudio, vídeo, protocolos, imagens, fotografias, slides, sequências de proteínas ou genéticas, amostras, verificações e respostas do teste, algoritmos, *scripts*, *software* de simulação, código de *software*, observações digitais, monitoramento científico, dados de sensores, metadados, desenvolvimento de modelos, modelo de saída, cenários, dados comportamentais qualitativos ou observados, visualizações e dados estatísticos coletados para fins administrativos ou comerciais, leitura de temperatura de sensores, medição da salinidade da água, dados geográficos, de recenseamento, geoespaciais, genômicos, observações, números e medidas tradicionais armazenados em planilhas e bancos de dados, fósseis, árvores filogenéticas e amostras de herbários, em meio digital, entre outros (BORGAM et al. 2011, 2015; DUDZIAK, 2016; GREENBERG et al., 2009; TOLLE; TANSLEY; HEY, 2011; UNIVERSITY OF MELBOURN, 2009). Esta lista de exemplos não é exclusiva nem absoluta.

Os diversos autores citados ao longo da Tese utilizam os termos dados científicos, dados de pesquisa e dados para a pesquisa. Foi adotado nesta pesquisa

o termo dados de pesquisa para os dados gerados pelos pesquisadores ao longo de seus projetos de pesquisa científica.

Em uma perspectiva conceitual, os dados, em um nível fundamental, são definidos como símbolos não interpretados que representam as propriedades de objetos e eventos (ACKOFF, 1999). Em relação ao conceito anterior, informações “[...] são dados com significado adicional”; e conhecimento “[...] é a capacidade de atribuir significado aos dados para obter novas informações[...]” que consistem em dados processados, o processamento direcionado para aumentar sua utilidade (ACKOFF, 1999, p. 170; STEPANOV, 2019, p. 67). Dentre essa ordem de ideias, é habitual que os dados sejam definidos em sua relação com a informação e o conhecimento (STEPANOV, 2019).

No entorno da comunidade científica, há esforços significativos no encorajamento e conscientização para compartilhar e reutilizar dados de pesquisa com o intuito de melhorar a reprodutibilidade e avançar nas descobertas científicas, mas com a segurança que os dados não serão copiados, modificados ou distribuídos ilegalmente. O compartilhamento de dados oriundos de pesquisas científicas com financiamento de agências de fomento, em muitos casos, é obrigatório e os repositórios de dados, entre outros canais de comunicação, são vias para a publicação. As questões intrínsecas à propriedade intelectual de dados estão evidenciadas em diversos documentos desses canais, inclusive de repositórios de dados.

Embora os dados estejam em evidência na comunicação científica, o compartilhamento de dados nem sempre é uma prática comum, nem compreendido ou bem implantado como uma prática de pesquisa na comunidade científica (KIM; ZHANG, 2015; PIWOWAR; CHAPMAN, 2010).

Apesar dos repositórios de dados disponibilizarem informações sobre aos aspectos da propriedade intelectual, pode haver baixa aderência entre as perspectivas relacionadas ao depósito e compartilhamento de dados pelos pesquisadores devido à ausência ou incompreensão (anomia) das informações e das políticas dos repositórios. Pesquisadores podem ter uma série de dúvidas e preocupações ao compartilharem seus dados de pesquisa, destacando as premissas de como serão reconhecidos pela autoria dos dados e como esses dados serão liberados para reutilização. Esse estado de anomia pode impactar negativamente o compartilhamento de conjuntos de dados, tópico apresentado na próxima subseção.

4.1 Anomia e o compartilhamento de dados

A palavra anomia tem origem do grego *anomos*, em que “a” representa ausência, inexistência ou privação de; e “nomos” significa lei, norma (ALEIXO, 2020). Ao longo da história, escritores sofistas, filósofos, sociólogos, entre outros, empregaram ou definiram anomia nos contextos social e psicológico do ser humano.

Um dos escritores foi o sociólogo e filósofo polonês Bauman que apresenta anomia como: “A ausência, ou a mera falta de clareza das normas - anomia - é o pior que pode acontecer às pessoas em sua luta para dar conta dos afazeres da vida. As normas capacitam tanto quanto incapacitam [...]” (BAUMAN, 2001, p. 28).

Bauman (2001), ao abordar a modernidade da sociedade, destaca que a sociedade avança em diversos sentidos e discute como a dicotomia do líquido e do sólido a delineia retratando a mudança da sociedade sólida para a sociedade líquida. A liquidez apresentada em seu texto está no fato de os líquidos não terem forma e se moldarem conforme o recipiente em que são inseridos, diferentemente dos sólidos, que são rígidos e precisam sofrer uma força para moldar-se. Destaca-se que a análise de Bauman foi realizada no contexto do “bug do milênio”, em que computadores e aplicativos estavam preparados para executar até 1999, exigindo adaptações para não ocorrer o caos tecnológico.

Observa-se as “adaptações” na atualidade pela mudança nas práticas científicas advogadas pela Ciência Aberta. As áreas do conhecimento ou pesquisadores individuais, que não estavam acostumados ou não eram adeptos a disponibilização pública dos dados, deparam-se com a necessidade de adaptações na gestão de dados e sua disponibilização, que podem seguir nas duas direções: fluir como o líquido se ajustando as necessidades de abertura e compartilhamento dos dados; ou precisar de uma tensão de força para o mesmo acontecer.

As adaptações nas atuais práticas científicas podem gerar tensões ou enfraquecimento das relações na sociedade ou na ordem estabelecida entre as comunidades científicas. Essa abordagem tem como princípio a complexidade que a sociedade está atingindo, fator observado e discutido por Émile Durkheim (1977, 1999) em seus livros *Da divisão do trabalho social* e *O suicídio*. Durkheim é o sociólogo da moral e ressalta a importância da consciência coletiva, do compartilhar formas de pensar e de agir identificando males da sociedade, definidas por ele como “anomia”,

ou seja, negação da moral, falta de clareza e descumprimento das regras existentes.

No livro *Da divisão do trabalho social*, Durkheim explanou sobre as consequências da divisão do trabalho na vida dos indivíduos e da sociedade considerando que, sob a tensão de mudanças repentinas na sociedade, nas normas e nas regras sociais já estabelecidas, as mesmas tornam-se obsoletas e “[...] falham em manter a uniformidade com atitudes e expectativas, consequentemente, quando estabelecidas de forma inapropriada resultam no desprezo por todas as outras regras.” (QUEIROZ, 2015). Nesse aspecto, Durkheim apresenta o estado de anomia que a sociedade se encontra. Para Durkheim, a anomia está associada a: negação da moral; falta de clareza; perda das referências normativas e descumprimento das regras existentes; condição em que as normas sociais e morais são confundidas, pouco esclarecidas ou simplesmente ausentes (ALEIXO, 2020; QUEIROZ, 2015; RIBEIRO, 2012).

O sociólogo Merton, pesquisador que deu continuidade ao debate acerca da Teoria da anomia iniciado por Durkheim elucida que

Conforme foi elaborado por Durkheim, o conceito de anomia se referia a uma condição relativa [de] normalidade numa sociedade ou grupo. Durkheim deixou claro que este conceito se referia a uma propriedade da estrutura social e cultural, e não a uma propriedade dos indivíduos que confrontavam tal estrutura. (MERTON, c1968, p. 235).

Merton, relevante teórico para a burocracia e para a sociologia da ciência, elucidou a conduta de cientistas e o que os motivam: a recompensa e a intimidação. Para Merton (1973), de acordo com as normas tradicionais da ciência, os cientistas devem compartilhar suas descobertas científicas e as informações a elas relacionadas. Nesse aspecto, os dados de pesquisa também devem ser compartilhados. Entretanto, “[...] tradições disciplinares, barreiras institucionais, [...] preocupações com a propriedade intelectual e percepções individuais impedem os cientistas de compartilhar seus dados (KIM; STANTON, 2016, p. 777, tradução nossa).

Em sua teoria funcional, o autor apontou como essa teoria tenta determinar o modo pelo qual a estrutura cultural (conjunto de valores normativos que regulam o comportamento em uma determinada sociedade ou grupo) e a estrutura social (relações sociais em que os membros estão inseridos) pressionam para comportamentos socialmente divergentes em pessoas situadas em posições

diferentes nessas estruturas (MERTON, c1968).

A anomia na estrutura cultural de Merton

[...] é então concebida como uma ruptura na estrutura cultural, ocorrendo especialmente quando houver uma aguda disjunção entre, de um lado, as normas e os objetivos culturais e, de outro, as capacidades socialmente estruturadas dos membros do grupo em agirem de acordo com as primeiras. Conforme esta concepção, os valores culturais podem ajudar a produzir um comportamento que esteja em oposição aos mandatos dos próprios valores. (MERTON, c1968, p. 236-237).

No entendimento de Merton, a estrutura cultural deve ser analisada separadamente. As discussões do conceito sociológico de Merton (c1968, p. 236) ao longo do seu trabalho,

[...] pressupõe que o ambiente mais destacado dos indivíduos possa ser concebido de maneira útil, como envolvendo a estrutura cultural, de um lado, e a estrutura social, de outro. Tal conceito pressupõe que, mesmo sendo tais conceitos intimamente ligados, devem ser conservados separados para finalidades de análise, antes que sejam novamente reunidos.

Partindo da análise de Merton, observa-se aspectos na ruptura cultural da comunidade científica: culturalmente, após a análise dos dados e publicação dos resultados, os dados eram guardados ou descartados pelos pesquisadores ou grupos de pesquisa, com algumas exceções, fato observado em trabalhos publicados. Com a evolução da ciência, do instrumental tecnológico e a incorporação dos elementos adicionados nas práticas da Ciência Aberta, nota-se um intenso fomento para o compartilhamento dos dados culminando em seu reuso e com a universalização do seu acesso.

Merton destaca que o conceito da anomia apresentado por Durkheim faz-se compreensível para entender as diversas formas de comportamento, e destaca que o conceito psicológico, apresentado em suas discussões, “[...] é uma contrapartida do conceito sociológico da anomia, e não um substituto para ele.” (MERTON, c1968, p. 236).

Os estudos de Merton foram apresentados e discutidos pelo filósofo e sociólogo Dahrendorf. O autor complementa as discussões de Merton e destaca as sanções como forma de punição. Para ele, a anomia não é um estado de espírito, mas um estado da sociedade, uma condição social, que pode fazer brotar vários tipos de

comportamentos (DAHRENDORF, 1987). Na obra *A lei e a Ordem* Dahrendorf (1987, p. 31) conceitua anomia como:

Uma condição social onde as normas reguladoras do comportamento das pessoas perderam sua validade. Uma garantia dessa validade consiste na força presente e clara das sanções. Onde prevalece a impunidade, a eficácia das normas está em perigo. Nesse sentido, a anomia descreve um estado de coisas onde as violações de normas não são punidas.

Aleixo apresenta anomia no contexto dos dados governamentais. Em seu trabalho, a autora resgata os conceitos de anomia dos autores apresentados nessa Tese e autores que empregaram o conceito no contexto da Ciência da Informação. Para a autora “[...] o estado de anomia ocorre quando o elemento que se faz ausente existe ou já existiu, ou seja, se as Normas, Regras, Leis, Metas, Regulações e Sanções se fazem presente, porém não são colocadas em prática.” (ALEIXO, 2020).

O termo anomia é empregado pelos autores apresentados com variadas definições, envoltas ao convívio em sociedade e suas relações e atuações. O emprego do termo pelos autores apresentados caracteriza a aspecto Cultural, a Institucional e a Social apresentadas na figura 14.

Figura 14 – Aspectos da anomia no contexto institucional



Fonte: Elaborado pela autora.

Os autores apresentam a anomia de forma que podem ser relacionadas com as Dimensões Institucional e Pesquisador apresentadas no *framework*. A Tese manifesta aderência com o conceito de anomia de Durkheim e a anomia institucionalizada de Merton.

Considera-se, a partir do exposto, como estado de anomia a baixa aderência ou não observância dos pesquisadores das normas estabelecidas e a percepção dos pesquisadores sobre os benefícios, riscos, esforços e suas ações. Parte-se do pressuposto que existem normas e instruções, porém não são seguidas em conformidade com o estabelecido.

Nessa perspectiva, ao associar a anomia com o compartilhamento de dados, pode ser entendida como a ruptura entre quem impõe as regras de depósito e compartilhamento dos dados e os pesquisadores, que não percebem seus benefícios

ou não sabem cumpri-las, assim como, pode-se inferir que se os pesquisadores não conhecem ou não tem clareza sobre os diversos aspectos norteadores do contexto apresentado é possível ocasionar a anomia e, conseqüentemente, possível retenção dos dados.

5 REPOSITÓRIO DE DADOS E A COMUNICAÇÃO CIENTÍFICA

Gray (2009) destacou a importância da Tecnologia da Informação e Comunicação na forma de se fazer ciência. Para esse autor, os desafios tecnológicos incluem a necessidade de êxito ao capturar, analisar, modelar, visualizar e preservar os dados e as informações científicas, tornando os sistemas de computação vitais para o moderno ambiente de pesquisa (GRAY, 2009). O planejamento das tecnologias aplicadas nesse contexto é essencial. A coleta de dados é comumente realizada empregando alguma forma de tecnologia digital ou os dados coletados ou gerados manualmente, em grande maioria, são convertidos em uma forma digital para sua análise (GREENBERG, 2009).

Observa-se que as universidades, centros de pesquisas e laboratórios são locais que propiciam a coleta, a produção, o processamento, a transferência e o uso de dados. Tais ambientes incentivam as comunidades científicas para produção, divulgação e publicação dos resultados das pesquisas e das descobertas de novos fenômenos, incluindo o compartilhamento dos dados que embasaram as pesquisas contribuindo, assim, para a evolução do conhecimento científico.

Morin (2007, p. 5) destaca que o conhecimento científico “[...] também foi durante muito tempo e com frequência ainda continua sendo concebido como tendo por missão dissipar a aparente complexidade dos fenômenos a fim de revelar a ordem simples a que eles obedecem.”. A comunicação do conhecimento científico colabora com a viabilidade de reprodução de resultados das pesquisas e experimentos, além de oferecer evidências para a qualidade do trabalho científico e sua validação (LYNCH, 2011). A comunicação do conhecimento científico está sendo feita, inclusive, com a liberação de dados de pesquisa brutos, com certa camada de análise ou analisados via periódicos e repositórios de dados, entre outros canais.

Meadows (1999) em suas discussões sobre a comunicação científica destacou que comunicação está situada no coração da ciência. A comunicação científica consiste em assegurar o intercâmbio de [dados] e informações sobre as pesquisas em andamento ou já finalizadas e viabilizar o contato entre pesquisadores e suas pesquisas sendo possível o trabalho colaborativo (LE COADIC, 2004).

O trabalho colaborativo é uma das premissas do compartilhamento de dados. Rebouillat (2019) destaca que os dados fazem parte de um ciclo de credibilidade que

lhes confere um valor de uso e reuso para a produção de novas publicações e um valor como moeda de troca no contexto de colaborações com parceiros da mesma instituição ou instituições diferentes.

Para Mueller (2000, p. 25) os resultados das pesquisas publicados pelos pesquisadores “[...] são freqüentemente retomados por outros cientistas, teóricos ou aplicadores, que dão continuidade ao estudo, fazendo avançar a ciência ou produzindo tecnologias ou produtos neles baseado.”. Nessa dinâmica, a publicação dos dados também tem como um de seus objetivos a reutilização por outros pesquisadores fazerem novas análises, validar os resultados apresentados ou dar continuidade às análises dos autores dos dados, propiciando, inclusive, o remix por outros pesquisadores.

Acerca da retomada dos resultados de pesquisa e a recombinação de trabalhos publicados de outros pesquisadores, Albagli (2015) argumenta que na ciência essa prática é frequente, o que a autora denomina de remix. A autora considera o remix sob a perspectiva da Ciência aberta em sua dimensão ética na qual o trabalho científico e seus resultados devem estar públicos para a reutilização por outros. Nas boas práticas da Ciência aberta relacionadas aos dados de pesquisa, o remix está presente ao considerar que os conjuntos de dados devem ser compartilhados para a reutilização desses ou combinação com outros dados.

Esse argumento é fundamentado com a apresentação do trabalho dos pesquisadores Tolle, Tansley e Hey (2011) na discussão dos dados na *e-Science*. Os autores apresentam uma pirâmide com os dados de pesquisa em sua base, devido a sua grande quantidade, com os dados derivados e/ou recombinações no meio e a literatura no topo, resultado final da análise dos dados. Do ponto de vista dos autores, os dados coletados e armazenados em repositórios podem ser integrados e podem “[...] unificar todos os dados de pesquisa e toda a literatura para criar um mundo em que os dados e a literatura possam interagir.” (TOLLE; TANSLEY; HEY, 2011, p. 24). Os dados podem ser combinados e recombinações para novas análises, unificados e novamente publicados, interoperados em rede e usados por pesquisadores de áreas multidisciplinares. Os pesquisadores tem a possibilidade de analisar e utilizar dados de terceiros em suas pesquisas, publicar esses resultados junto com os seus dados proporcionando a outros o acesso a todo o conteúdo disponibilizado. Dessa forma, há a possibilidade de integrar os dados e a literatura, aumentando a velocidade e produtividade da comunicação científica. Todo esse processo necessita uma gestão de dados eficaz.

A gestão de dados de pesquisa é complexa e, tornar esses dados úteis e compreensíveis para outros pesquisadores, que não fizeram parte do processo de coleta e análise ou que são de domínios diversos, para sua replicação para fins não previstos, é difícil (BORGMAN et al., 2015). Os dados de pesquisa necessitam de gestão apropriada para melhor recuperação e reuso (SALES; SAYÃO, 2012). Essa gestão integra o arquivamento, a preservação, os meios para acrescentar valor aos dados e maximizar a capacidade de reuso (SALES; SAYÃO 2012).

Um elemento relevante na gestão de dados de pesquisa é assegurar que os mesmos possam ser compreendidos e interpretados por outros pesquisadores ao longo do tempo e em pesquisas divergentes das que fizeram a coleta.

Considerando a grande quantidade de dados, sua heterogeneidade e a complexidade em seu tratamento, cada etapa na manipulação dos dados requer o conhecimento dos passos que foram dados anteriormente, pois os detalhes documentados ajudam na interpretação, na manipulação e na reutilização dos dados por outros pesquisadores (BORGMAN et al., 2015).

Para tal, é essencial uma descrição clara e detalhada dos dados, as anotações adicionais e as informações que contextualizam os dados e, assim, que possibilitem transmitir informação e conhecimento no tempo e no espaço (SAYÃO; SALES, 2017). A documentação dos dados é o instrumento que apoia e efetiva essas ações (SAYÃO; SALES, 2017). A documentação compreende informações sobre a proveniência dos dados que inclui: como, quando e onde os dados foram coletados, com qual instrumento, calibragem, medidas, entre outros fatores.

Os dados de pesquisa, informações digitais, documentação de dados e Planos de gestão de dados da comunidade científica são armazenados e gerenciados em repositórios de dados de diversas universidades e propiciam aos usuários terem acesso à conteúdo dos quais não participaram da coleta ou análise, diminuindo as dificuldades impostas por limitações orçamentárias ou barreiras geográficas.

Os repositórios de dados viabilizam o armazenamento e a gestão de dados visando à otimização da recuperação, o que amplia as potencialidades de reuso dos dados entre os pesquisadores. Desta forma, contribui para agilizar os processos da investigação científica e, conseqüentemente, o avanço na ciência.

Esse panorama viabiliza aos pesquisadores terem acesso e analisar dados coletados por outros pesquisadores e armazenados nos repositórios de dados, o que torna exequível a produtividade científica, por meio do aumento da velocidade com

que os dados e as informações podem ser tratados e produzidos (LYNCH, 2011) além de reduzir os custos e investimentos aplicados pelas agências de fomento entre outros, evitando que esforços na coleta de dados sejam duplicados.

Repositórios de dados de pesquisa tem contribuição relevante na gestão de dados e serão apresentados na próxima subseção.

5.1 Repositório de dados de pesquisa

Desde a década de 2000, o acesso aberto aos dados vem gradativamente ocupando um lugar cada vez mais estratégico nas agendas de financiadores de pesquisa, nas políticas de pesquisa e nas práticas científicas. Essas questões foram incorporadas por profissionais e universidades que desenvolveram serviços e produtos destinados a apoiar os pesquisadores na gestão, abertura e publicação dos dados.

Sendo assim, os anos subsequentes vem sendo marcado por práticas e descobertas científicas que enfatizam a gestão de dados de pesquisa e a conscientização de pesquisadores sobre o valor contido nesses dados e dos riscos de perde-los ao longo do tempo (SESARTIC; TOWE, 2016). A inevitabilidade de infraestrutura para a gestão dos dados alterou a configuração de serviços necessários ao contexto de repositórios digitais.

Os repositórios digitais estão presentes nas comunidades científicas e são ambientes que contribuem no processo da comunicação científica, que armazenam objetos digitais, que gerenciam e preservam esses objetos por longo tempo, que promovem o acesso e o uso dos objetos e impulsionam a visibilidade da produção científica e minimiza custos de publicações (MÁRDERO ARELLANO, 2008; CAMARGO; VIDOTTI, 2009).

Os repositórios digitais de acesso aberto é uma realidade em diversos países, são 5.586 no total, desses, 170 estão na Espanha e 381 contém *datasets* (OPENDOAR, 2020). Entre os repositórios digitais, há os repositórios de dados que, nos últimos anos vislumbra-se um desenvolvimento quantitativo expressivo. São registrados no re3data.org 2461. Desses, 35 estão na Espanha. Uma parte considerável de tais repositórios foram implementados por instituições, comunidades científicas, financiadores de pesquisas, universidades, entre outros para armazenar,

preservar e disponibilizar os conjuntos de dados, PGDs e documentação de dados. Seus serviços são direcionados a sua comunidade e à terceiros.

Os repositórios de dados configuram-se como uma ampliação da proposta dos repositórios temáticos e institucionais, sendo esses últimos utilizados, entre outros fins, para o armazenamento, a gestão, a divulgação, a preservação e a visibilidade de teses, dissertações e publicações científicas.

A origem dos repositórios temáticos data da década de 1990 com a implementação do repositório arXiv⁴ para publicação de *pre-prints* (MOLFINO; GONZÁLES, 2012). Os repositórios temáticos armazenam e disponibilizam um conjunto de trabalhos de pesquisa, *pre-prints* ou não, de uma determinada área do conhecimento (CAFÉ et al., 2003).

Os Repositórios Institucionais (RI) são implementados por órgãos governamentais, instituições públicas e acadêmicas. No contexto acadêmico são ambientes digitais com materiais digitais criados pela instituição e pelos membros de sua comunidade e com serviços que uma universidade oferece a sua comunidade como armazenamento, gestão, preservação, divulgação e acesso à produção intelectual (LYNCH, 2003; RODRIGUES, 2004). A quantidade desses repositórios vem aumentando desde do final dos anos 1980 e “[...] tiveram um crescimento muito orgânico, o que significa que as instituições podem definir pré-requisitos para atender às suas necessidades individuais em vez de se conformar aos padrões estabelecidos.” (STEVENSON; ZHANG, 2015, tradução nossa).

Os repositórios de dados têm sua gênese na necessidade de gestão dos dados de pesquisa, contribuem para assegurar que os dados sejam publicados e disponibilizados para a comunidade científica com o menor número possível de restrições (MONTEIRO; SANT’ANA, HERNÁNDEZ-PÉREZ, 2019).

Os repositórios de dados formam um conjunto de sistemas e serviços que viabilizam o armazenamento, a gestão, a disponibilização e a curadoria de dados de pesquisa, promovendo o acesso aberto aos dados, difundindo e projetando seu conteúdo junto a outras comunidades universitárias, científicas e sociedade em geral, visando à otimização da recuperação e a maximização de reuso dos dados entre os pesquisadores. Desta forma, contribui para agilizar os processos de investigação e, consequentemente, o avanço na ciência.

⁴ <https://arxiv.org/>

Ademais, destaca o valor científico, cultural, econômico e social da comunidade, podendo maximizar a captação de recursos financeiros para as atividades de pesquisas. Gerenciar, disseminar e preservar os conjuntos de dados produzidos nas universidades facilitando o reuso é uma conduta que incentiva a retroalimentação da pesquisa além de reconhecer e valorizar sua comunidade.

A quantidade significativa de instituições em diversos países que aderiu aos repositórios de dados, principalmente as universidades, evidencia tanto a relevância dos dados e da produção intelectual de sua comunidade quanto o papel fundamental da disseminação dos dados para dar visibilidade à produção científica, ao pesquisador e à instituição a qual está vinculado.

Observa-se, também, um crescente interesse das instituições pelo estudo da gestão de dados impulsionado pela obrigatoriedade de Plano de gestão de dados pelas agências de fomento. Sayão e Sales (2017) consideram que o planejamento na gestão de dados se inicia antes da implementação do repositório, aspecto crucial no delineamento do projeto e de sua implementação e, também, para o pesquisador, no momento do delineamento da pesquisa. A gestão deve considerar como os dados serão gerenciados durante o desenvolvimento do projeto do pesquisador, como serão compartilhados, onde serão publicados e sob quais condições serão reutilizados (SAYÃO; SALES, 2017).

A propósito dos benefícios da implementação de repositórios de dados, Sales (2014) destaca vários aspectos. Sobre a visibilidade, a autora destaca que “Amplia a visibilidade dos resultados de pesquisa, posto que, via de regra, só é formalmente disseminada a fração que está registrada nas publicações acadêmicas.” (SALES, 2014, p. 63). Pesquisadores depositam seus conjuntos de dados nos repositórios não só para cumprir a exigência das agências de fomento, mas, também, para divulgação e visibilidade de sua pesquisa.

Em segundo lugar, a autora destaca a preservação que “Proporciona mecanismos de preservação de longo prazo, em termos de preservação digital, de arquivamento seguro e de curadoria digital.” (SALES, 2014, p. 63). Para cumprir a finalidade de reuso dos dados a longo prazo, os repositórios estão estruturando sua capacidade técnica para preservação a longo prazo. Dependendo dos tipos dos dados e dos propósitos dos pesquisadores, dados podem ser descartados após um período determinado.

Em terceiro lugar, a autora destaca que “Permite que o material depositado esteja disponível on-line continuamente para ser consultado e citado mais frequentemente.” (SALES, 2014, p. 63). Aspecto com forte aderência à temática principal dessa Tese pois os dados consultados e reutilizados devem ser citados dando créditos aos autores. Além disso, está relacionado ao benefício anterior pois, para ser continuamente disponível, os dados precisam ser preservados.

Em quarto lugar a autora menciona, o que pode ser associado ao remix discutido anteriormente sobre a ótica da autora Albagli (2015). Sales menciona que “É um instrumento chave para os processos de reformatação e recriação de dados proporcionados pela curadoria digital.” (SALES, 2014, p. 63).

Em sequência, Sales (2014, p. 64) destaca “[...] a possibilidade de criação de novos serviços de informação para pesquisadores e gestores a partir da análise dos dados arquivados e a integração de dados e publicações acadêmicas.” Nesse aspecto, vislumbra-se diversos serviços oferecidos aos pesquisadores, inclusive a integração apresentada por Tolle, Tansley e Hey (2001) discutida anteriormente.

A continuação, a autora mencionam a interoperabilidade: “Permite a criação de redes de repositórios interoperáveis.” (SALES, 2014, p. 64). A interoperabilidade de dados está na agenda dos princípios FAIR e nas discussões relacionadas aos repositórios e, também, sendo apresentadas por pesquisadores demonstrando as possibilidades e tecnologias.

Em seguida, Sales (2014, p. 64) enfatiza que “Aumenta o grau de reuso dos dados minimizando a duplicação de esforços e otimizando os investimentos na geração de dados.”. Tópico enfatizado pelas agências de fomento e estimulado pelos repositórios e pela comunidade científica. O reuso de dados também é discutido nessa Tese com os resultados das entrevistas e aplicação de questionários apresentados nos resultados e discussões.

Os repositórios de dados podem ter dados brutos ou estruturados, armazenados de forma mais simples ou com nível detalhado de representação. “Os conteúdos digitais estruturados possibilitam uma grande variedade de mecanismos de recuperação e de tratamento, variedade esta ampliada pelos próprios recursos disponibilizados pelas tecnologias da informação.” (SANT’ANA, 2008, p. 148).

Relevante destacar que a representação de dados com maior nível de detalhamento é um pré-requisito para a eficácia no seu compartilhamento,

recuperação e reuso entre uma comunidade científica específica e comunidades que não fazem parte do grupo alvo, contudo podem se beneficiar desses dados, principalmente, dados que só podem ser coletados uma única vez.

A crescente preocupação com a disponibilização e reuso de dados de pesquisa resultou em uma série de iniciativas que promovem o arquivamento e a curadoria de conjuntos de dados (SALES, 2014; SICILIA; GARCÍA-BARRIOCANAL; SANCHEZ-ALONSO, 2017). A representação descritiva dos conjuntos de dados nos repositórios apoiada em padrões ou esquemas de metadados é um fator preponderante para a descoberta e consumo dos dados. Descrever e organizar dados e fazer os relacionamentos entre eles pode ter "[...] efeitos profundos na capacidade de descobrir, trocar e armazenar dados" (BORGMAN, 2015, p. 65). Os padrões de metadados aumentam essas habilidades. Os repositórios de dados utilizam padrões de metadados reconhecidos e utilizados mundialmente para descreverem seus conjuntos de dados. O uso de padrões e esquemas de metadados para a descrição dos dados e a interoperabilidade entre os repositórios se configuram de forma proeminente nesse contexto (GREENBERG et al., 2009).

O termo metadados possui diferentes definições de acordo com a área e o contexto de aplicação e ou análise. "Metadados entendidos como informações sobre dados, têm como finalidade documentar e organizar, de forma estruturada, conjuntos de dados [...]" padronizando-os e, dessa forma, minimizando retrabalho e facilitando a manutenção destes dados (SIQUEIRA; COSTA SANTOS, 2004, p. 96).

A estruturação de dados pode ser obtida por meio de conjuntos de elementos previamente definidos, identificados por etiquetas (tag), e seus respectivos atributos. A informação assim estruturada e com uso de padrão reconhecido em âmbito internacional propicia uma maior confiabilidade dos dados (ROSETTO, 2003).

A representação descritiva dos metadados precisa ser o mais completa possível a fim de facilitar a recuperação dos dados. Os metadados apresentados nos resultados de busca fornecem informações para que os consumidores de dados conheçam os aspectos relacionados a autoria, ao conteúdo, às licenças de uso, aos direitos de acesso, aos formatos entre outros. Os metadados estão presentes nos princípios FAIR.

Os repositórios estão alinhando suas atividades para atender os princípios *Findable, Accessible, Interoperable, Reusable* (FAIR) (DE SMEDT et al., 2018; DE

JONG et al., 2018). A concepção dos repositórios de dados está em conformidade com a visão subjacente às políticas europeias emergentes e aos paradigmas da Ciência Aberta, interconectando pesquisadores através das fronteiras disciplinares e geográficas, proporcionando a gestão e acesso contínuo à dados de pesquisa e à serviços *online*. A figura 15 mostra seus princípios.

Figura 15 – Princípios FAIR



Fonte: Australian National Data Service ([201-]).

Os princípios FAIR oferecem subprincípios para aprimorar a utilidade dos (meta)dados, e marca um refinamento essencial dos conceitos necessários para dar maior valor aos dados e maximizar sua propensão para reutilização, por seres humanos e por máquinas, apoiando na interação entre quem deseja usar os dados e quem os fornece (KALINAUSKAITĖ; 2017; EUROPEAN COMMISSION, 2018).

Quadro 3 - Princípios e subprincípios FAIR

Princípio	Descrição
Findable	F1 – (meta)dados são atribuídos a um identificador globalmente exclusivo e persistente
	F2 – os dados são descritos com metadados ricos (redefinidos por R1 no princípio Reusable)
	F3 - metadados de forma clara e explícita incluem o identificador dos dados que descreve
	F4 - (meta) dados são registrados ou indexados em um recurso pesquisável
Accessible	A1 - (meta) dados são recuperáveis pelo seu identificador usando um protocolo de comunicação padronizado
	A1.1 - o protocolo é aberto, gratuito e universalmente implementável
	A1.2 - o protocolo permite um procedimento de autenticação e autorização, quando necessário
	A2 - os metadados estão acessíveis, mesmo quando os dados não estão mais disponíveis
Interoperable	I1 - Os (meta) dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento.
	I2 - (meta) usam vocabulários que seguem os princípios do FAIR
	I3 - Os (meta) dados incluem referências qualificadas a outros (meta) dados
Reusable	R1 meta (data) são ricamente descritos com uma pluralidade de atributos precisos e relevantes
	R1-1 – (meta) dados são liberados com uma licença de uso de dados clara acessível
	R1.2 - (meta) dados estão associados à proveniência detalhada
	R1.3 - (meta) dados atendem aos padrões da comunidade relevantes ao domínio

Fonte: Adaptado de European Commission (2018, tradução nossa).

O FAIR exige grandes mudanças em termos de cultura e práticas de pesquisa, bem como a implementação de um ecossistema com serviços e componentes de dados como Políticas, Plano de Gestão de Dados, Identificadores, Padrões e Repositórios, sendo esse último, essencial no ecossistema de dados FAIR pois são necessários para executar a função de oferecer dados e metadados acessíveis e reutilizáveis para usuários e por máquinas (EUROPEAN COMMISSION, 2018).

O PGD é um documento, mandatário ou recomendado pelos financiadores, elaborado no início do projeto de pesquisa, mas com viabilidade de atualizações, no qual são descritas as diretrizes do ciclo de vida dos dados (CVD) (MONTEIRO, 2017, MONTEIRO; SANT'ANA, 2018). As diretrizes orientam todos os envolvidos de alguma forma com a gestão dos dados. Os PGDs fornecem um índice dinâmico que articula as informações relativas a um projeto e as ligações com os vários componentes FAIR e, os repositórios, oferecem bancos de dados e serviços de dados devendo ser certificados para garantir a sua confiabilidade (EUROPEAN COMMISSION, 2018). O princípio FAIR “F1 – (meta)dados são atribuídos a um identificador globalmente exclusivo e persistente” destaca a relevância de utilização de identificadores

persistentes (QUADRO 3).

A vantagem em agregar um identificador persistente aos conjuntos de dados consiste no fato de que esse identificador aponta sempre para o conjunto de dados, mesmo que tenha sido alterada sua localização. Existem vários sistemas de identificação persistente baseados em padrões abertos como *Digital Object Identifiers* (DOI), *Handle System* e PURL (SAYÃO, 2007).

Outro aspecto previsto nos subprincípios FAIR está relacionado ao “R1-1 – (meta) dados são liberados com uma licença de uso de dados clara e acessível” que orienta os envolvidos no compartilhamento e reuso de dados armazenados nos repositórios de dados. Os repositórios estão se resguardando de possíveis conflitos, violações e problemas éticos e jurídicos pelo uso ou apropriação indevidos dos conjuntos de dados neles armazenados. Suas ações estão direcionadas aos depositantes e aos consumidores de dados. Aos depositantes, são assinados contratos que definem os direitos e obrigações. Para os consumidores, estão explicitadas a indicação de quais licenças são aplicadas aos conjuntos de dados disponibilizados evidenciando as condições de uso, bem como a indicação da autoria, a referência dos conjuntos de dados e a indicação de citação.

A missão, os objetivos e as áreas de conhecimento das universidades são elementos que influenciam sobre quais licenças os conjuntos de dados são liberados. Repositórios multidisciplinares tendem a ampliar o leque de opções para os autores liberarem seus dados devido ao fato da diversidade e características das áreas de conhecimento e dos conjuntos de dados. A comunidade da universidade influencia a elaboração das diretrizes dos repositórios relacionadas ao direito autoral.

O acesso e uso do conteúdo depositado nos repositórios de dados é definido pela política do repositório, resguardando a legislação do seu país em relação aos direitos autorais. Os repositórios cumprem a legislação sobre direitos autorais e orientam os usuários na forma de liberação e de uso do seu conteúdo.

Um conjunto de dados, antes de ser disponibilizado ou consumido por outros usuários (pesquisador), percorre um caminho que começa com a definição legal da licença de uso (transferência de direito patrimonial), prossegue com a escolha de como será publicado (físico ou digital), continua pelos agentes distribuidores/divulgadores (periódicos científicos, periódicos de dados, repositórios institucionais ou repositórios de dados) que, como detentores dos conjuntos de dados,

se tornam responsáveis em eventuais pleitos jurídicos⁵.

Em vista do exposto, os Repositórios de dados regularizam os aspectos do direito autoral pois orientam quem vai depositar os conjuntos de dados, o qual deve deixar explícito qual licença será aplicada sobre cada conjunto de dados para que o repositório o disponibilize, o torne público, além de explicitar quem é o autor dos conjuntos de dados e como deve-se referenciar esses conjuntos de dados com uma norma de citação.

A decisão sobre a aplicação de licenças nos conjuntos de dados é acordado entre os repositórios e os autores que podem ser regidos pelo uso de uma licença como a *Creative Commons* (CC) que são aplicados sobre diversos tipos de materiais ou pelas licenças ODL direcionadas a dados e/ou bases de dados. As licenças flexibilizam o uso dos dados, porém não são garantias que o mau uso e o plágio não aconteçam. Os direitos morais, que configuram a vinculação entre o nome do autor e sua obra de criação não podem ser transferidos a terceiros, porém seu direito patrimonial pode ser cedido sob determinada licença a título oneroso ou gratuito.

O chamado licenciamento aplicado sobre os dados é geralmente um 'contrato de edição' no qual o autor vincula aos seus conjuntos de dados. O licenciamento não é oneroso para os usuários de repositório de dados visto que esses seguem a máxima que nasceu com o conceito dos repositórios: disponibilizar o conteúdo em acesso aberto no meio digital. A citação dos autores também é advertida evidenciando que os autores devem ter os créditos pelo trabalho da coleta e organização dos conjuntos de dados. Modelos de citação dos dados são disponibilizados nos repositórios.

Na próxima seção é apresentado o repositório *e-cienciaDatos*.

5.2 Repositório de dados e-cienciaDatos

O *e-cienciaDatos* é um repositório de dados resultante do acordo de cooperação entre bibliotecas da Comunidade de Madrid composto por seis universidades: *Universidad de Alcalá, Universidad Autónoma de Madrid, Universidad Carlos III de Madrid, Universidad Politécnica de Madrid, Universidad Rey Juan Carlos e UNED*, membros do *Consórcio Madroño*.

⁵ Parafraseando Gandelman (2007) - sobre criação intelectual

Figura 16 - Repositórios que compõe o e-cienciaDatos



Fonte: e-cienciaDatos (c2021).

O repositório é implementado com o *software Dataverse*, desenvolvido pelo *Institute for Quantitative Social Science* da Universidade de Harvard. O *Dataverse* é uma aplicação de código aberto para implementação de repositórios de dados de pesquisa que permitem armazenar, gerenciar, representar, preservar e disponibilizar dados de pesquisa, facilitando a recuperação dos dados nele armazenados para que outros pesquisadores tenham a possibilidade de reutilizá-los e recriá-los em suas pesquisas, inclusive, oferece a citação e a referência padronizada dos dados permitindo dar créditos a todos os responsáveis pelo conjunto de dados utilizados como os autores, distribuidores e as instituições afiliadas (THE DATAVERSE PROJECT, 2015a).

O principal objetivo da Rede *Dataverse* é resolver os problemas de compartilhamento de construção de tecnologias que viabilizem instituições reduzir a carga de dados para os pesquisadores e editores de dados, bem como incentiva-los no compartilhamento de seus dados (CROSAS, 2011). *Dataverse* disponibiliza uma infraestrutura com suporte para serviços arquivísticos, assim como o trabalho de representação, metadados, *backup*, identificadores persistentes e conversão de arquivos (CROSAS, 2011).

Dessa forma, o *software* fornece uma estrutura de publicação de dados que possibilita o reconhecimento ao autor, citação persistente, descoberta de dados, gerenciamento, descrição com uso de metadados e preservação dos dados (CROSAS, 2011).

O repositório *Dataverse* dispõe de características intrínsecas (The DATAVERSE PROJECT, c2015b):

- Citação de dados - o modelo de referência e dados para a citação é gerado automaticamente pelo sistema. Ao clicar no conjunto de dados, o sistema exibe a referência normalizada do conjunto de dados, indicando a autoria para a citação;
- Integração com o *DataCite* - quando os conjuntos de dados são publicados,

seus metadados são enviados para o *DataCite*;

- Login via ORCID, Google ou GitHub;
- Suporte para os princípios FAIR;
- Vários fluxos de trabalho de publicação – Incluem conjunto de dados em rascunho, em revisão e depois publicado;
- Termos de Uso + Livro de Visitas – indicação de utilização da licença *Creative Commons*, termos de uso personalizados e métricas de *download* para cada conjunto de dados;
- Notificações de conta + dados - Acesso às funções que são concedidas pelo repositório;
- Pesquisa facetada - Campos de metadados baseados em facetas. O repositório disponibiliza vários campos de metadados para o refinamento da pesquisa permitindo montar vários tipos de estratégias de busca;
- Três Níveis de Metadados - Descrição/citação, campos específicos do domínio ou campos personalizados, metadados de arquivos que permitem a descrição dos dados com maior nível de detalhes. A interface de busca avançada disponibiliza três níveis de campos específicos facilitando buscas mais específicas;
- Extrair metadados de *datasets* pré-estabelecidos a partir dos metadados do *Astronomy* (FITS);
- APIs para interoperabilidade - API de pesquisa, API de depósito de dados. Pode ser aplicado conjunto de instruções de programação;
- *Shibboleth* - Assinar usando as credenciais da instituição a qual pertence. Pode ser criado *login* de usuário institucional;
- Suporte ao controle de acesso - Funções predefinidas e personalizadas;
- Arquivos Restritos + Capacidade de solicitar acesso a arquivos restritos - permiti que alguém, determinadas pessoas ou ninguém tenha acesso para fazer o *download* de arquivos. Configurado para dar acesso, por exemplo, apenas para quem fizer cadastro no sistema;
- Personalização de *dataverses* - *Branding*, facetas baseadas em metadados, sub-dataverses, dataverses; além de ser possível configurar o sistema para outros níveis conforme a necessidade;
- Re-formatar, estatísticas de resumo e análise de arquivos tabulares - Integração com o *TwoRavens*, sistema de ferramentas estatísticas interligadas

integrada ao *Dataverse* e que permite aos usuários executar modelos estatísticos;

- Mapeamento de arquivos geoespaciais - Integração com o *WorldMap* para indicar e mostrar a localização geográfica).

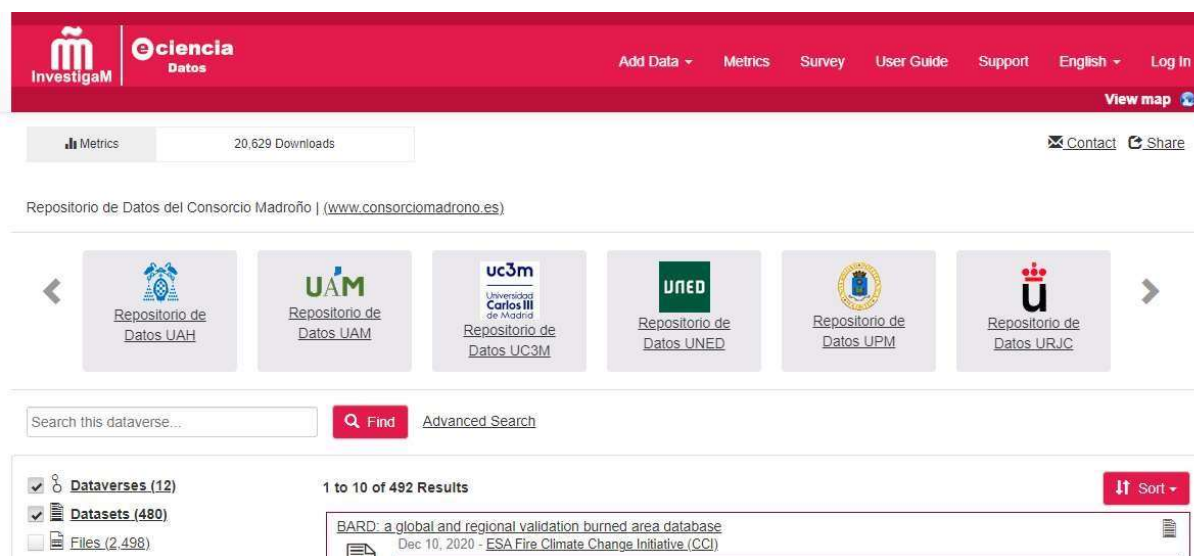
O *Dataverse* permite a customização da interface inicial com a disponibilização das informações sobre o repositório e os conjuntos de dados. As características do *Dataverse* facilitam o uso pelos pesquisadores que vão depositar seus conjuntos de dados e pelos usuários. O sistema disponibiliza elementos para a descrição dos dados com utilização de metadados, formas de referenciar atuando na citação dos dados e fornecem informações e dados aos administradores com análise de uso e estatísticas.

O repositório *e-cienciaDatos* visa organizar os conjuntos de dados e os PGDs dos pesquisadores vinculados às universidades cooperantes e oferece a gestão, a organização, o armazenamento, a preservação, a publicação e divulgação dos conjuntos de dados de cada uma das universidades atribuindo um identificador de objeto digital (DOI) e referência normalizada à cada um deles com a finalidade de facilitar a verificação dos dados, sua disseminação, reutilização, impacto e acesso a longo prazo.

O repositório *e-cienciaDatos* possui atualmente 13 *Dataverses* (nome das coleções), 534 *datasets* e 3.106 arquivos. Todo o material do repositório está associado a uma publicação acadêmica e uma parte significativa dos dados arquivados no repositório está associada a artigos revisados por pares e publicados.

O *e-cienciaDatos* armazena conjuntos de dados, PGD e documentação de dados, das diversas disciplinas das universidades membros do *Consórcio Madroño*. O repositório permite que as universidades agreguem os dados de sua comunidade no nível institucional e, também, agregando os repositórios de cada uma das universidades, além de oferecem espaço de armazenamento para colaboração, permitindo que o corpo docente trabalhe em arquivos antes da publicação (ENIS, 2013). Na página inicial é possível selecionar a coleção de cada uma das universidades (FIGURA 17).

Figura 17 – Interface do e-cienciaDatos



Fonte: e-cienciaDatos (c2021).

Outro aspecto relacionado ao repositório é a descrição com o máximo de representatividade para que reflita em uma melhor recuperação. A opção de incluir padrão de metadados para descrição que melhor represente as características dos dados possibilita uma descrição personalizada, pois os conjuntos de dados são heterogêneos exigindo metadados específicos.

No *e-cienciaDatos* os conjuntos de dados são descritos de forma padronizada, de acordo com um ou mais padrões de metadados, como *Dublin Core* e *DataCite* e, assim, garante a compatibilidade com a infraestrutura de pesquisa europeia *OpenAIRE* (MALO DE MOLINA, 2019). O *DataCite Metadata Schema* é uma lista das principais propriedades de metadados escolhidas para uma identificação consistente de um recurso, geralmente um conjunto de dados, para fins de citação e recuperação, juntamente com instruções como definições e usos de metadados (DATA CITE METADATA WORKING GROUP, 2017).

Os metadados constituem-se em elementos fundamentais no processo de tratamento descritivo, pois refletem a conjunção entre aspectos tecnológicos e de representação necessários aos novos tipos de recursos e ambientes informacionais, contribuindo para eficiência dos processos de recuperação dos repositórios (ALVES, 2010).

O detalhamento na representação dos dados permite uma recuperação no campo de busca simples e avançado com os filtros apresentados no quadro 4:

Quadro 4 - Campos de busca do e-cienciaDatos com filtros

CAMPOS DE BUSCA	FILTRO	
Dataverses	Name Sector Description Subject	
Conjunto de dados: Citation metadata	Title Author name Author affiliation Description text Subject Keyword term Topic Classification Term Related Publication Citation Related Publication ID Type Related Publication ID Number Language Producer Name	Contributor Type Contributor Name Distributor Name Distribution Date Time Period Time Period Covered Start Time Period Covered End Kind of Data Series Name Publication data ID persistent from the data packet
Arquivos	Name Description File type	Variable name Variable label

Fonte: Elaborado pela autora.

Observa-se no quadro 4 que o repositório possui vários filtros para especificar a busca tornando-a mais refinada. Pelos filtros é possível perceber a possibilidade de uma descrição mais detalhada dos conjuntos de dados, aspecto que contribui para atender o princípio F2 do FAIR “F2 - os dados são descritos com metadados ricos (definidos por R1 no princípio *Reusable*”. O repositório tem a estrutura de descrição que prevê diversos metadados e a descrição de cada conjunto de dados varia conforme suas características.

No quadro 5 são apresentados os metadados mais comuns localizados e coletados durante a análise de repositórios com suas respectivas descrições para compreensão de seus significados conforme aparecem nos resultados de busca. O *Dataverse* permite a customização dos metadados.

Quadro 5 - Metadados e sua descrição

(continua)

METADADO	DESCRIÇÃO
Access righth	Se o dataset está open; closed; embargoed; restricted
Altmetrics	Métrica alternativa baseada em redes sociais
Autor	Autor do dataset
Available in	Disponível
Categories	Área do conhecimento
Citation	Quantidade de citação do dataset
Citation / referencia	Modelo de citação do dataset
Cite / Cite as	O modelo de citação do dataset
Cite data	Modelo de citação
Contact	e-mail de contato de quem fez o depósito
Data	Data de publicação
Data / data e hora	Data de adesão
Data available	Data de disponibilização para coleta
Data e hora	Data e hora
Data issued	Data de envio ao repositório
DC.Coverage.spatial	Área geográfica
DC.Coverage.temporal	Período de abrangência dos dados
DC.Format.extent	Medida
DC.Identifier.URI	Endereço que remete ao dataset no repositório DRYAD
DC.Type.embargo	Se o dataset está embargado
Deposit date	Data de depósito
Depositor	Pessoa que fez o depósito

(continua)

Description	Sumário descrevendo o propósito, a natureza e o escopo do dataset
Distribution Date	Data em que os dados foram disponibilizados para distribuição
Distributor	Organização indicada pelo autor para produzir cópias
DOI	Endereço de localização no repositório
Download	Quantidade de downloads
Download	Quantidade de downloads / link para baixar o dataset /
DWC.Scificname	Nome científico dos dados do dataset
Embed	
Export	
Funding	Financiador da pesquisa
Guestbook	Livro de visitas
Indexed in	Base em que os dados estão indexados
Keywords	Termos que descrevem os principais aspectos do dataset
Kind of Data	Tipo de dados inclusos
Licence	Licenças de uso estabelecida, forma pela qual o dataset pode ser utilizado
ORCID	Número identificador do autor – dá o link para o ORCID do autor no sítio do ORCID
Part of	Logotipo com o link para a comunidade a qual faz o acesso a outros tipos de publicações
Producer	Pessoa ou instituição com responsabilidade financeira ou administrativa sobre o dataset
Production data	Data quando o dataset foi produzido
Publication data	Data de publicação

(conclusão)

Read the peer-reviewed publication	Referência do documento ao qual o dataset está vinculado
Related identifier	Link para o documento/periódico ao qual o dataset está vinculado
Related publication-material	Publicação que usa os dados do dataset
Share	Rede social para compartilhar o dataset
Subject	Categorias de assunto especializadas do domínio que são relevantes para o dataset
Title	Título do dataset
Time Period Covered	Período de tempo para o qual o dado se refere
URL/DOI	Endereço de localização no repositório
Version	Informações sobre as versões do dataset e indicação de como citar
Versions	Todas as versões do dataset
View	Quantidade de visualizações
Waiver	Renúncia dos direitos - como os dados podem ser usados

Fonte: Elaborado pela autora.

Os metadados *Title* e *Author* são metadados mínimos para identificação do conjunto de dados e, juntamente com as *Keywords*, o repositório tem metadados mínimos para a recuperação dos *datasets*, tanto no repositório como em buscadores como o *datasetsearch.research*⁶. A licença aplicada sobre os conjuntos de dados determina como os mesmos poderão ser utilizados e já instrui o pesquisador no uso ou não dos dados. Desse modo atende o princípio FAIR “R1.1 - (meta) dados são liberados com uma licença de uso de dados clara e acessível”. O elemento descrito *Share* contribui para o pesquisador compartilhar os conjuntos de dados e sua localização com outros pesquisadores.

O *Digital object identifier* (DOI) é disponibilizado pelos repositórios, o que contribui para a identificação persistente do conjunto de dados. Esse metadado indica

⁶ <https://datasetsearch.research.google.com/>

a exata localização do conjunto de dados sendo relevante sua utilização. O DOI atende o princípio FAIR “F1- (meta) dados são atribuídos a um identificador globalmente exclusivo e persistente”.

Um ID Autor é um identificador exclusivo para os autores e podem ser usados para distinguir o autor de outros com o mesmo nome ou nome semelhante. O autor tem a possibilidade de assegurar que todas as suas publicações estejam associadas ao seu perfil em bancos de dados ao utilizar esse tipo de ID. Inscrever-se para ID Autor é uma prática propícia aos pesquisadores para tornar suas publicações mais fáceis de serem localizadas e identificadas. O uso do *Open Researcher & Contributor ID* (ORCID), identificador digital persistente para autor, contribui com esse aspecto (THOMAS; CHEN; CLEMENT, 2015). Na descrição dos *datasets* do repositório o metadado ORCID é preenchido.

O protocolo OAI-PMH é utilizado, aspecto que atende o princípio FAIR “A1 - (meta)dados são recuperáveis pelo seu identificador usando um protocolo de comunicação padronizado” e o princípio FAIR “A1.2 - o protocolo permite um procedimento de autenticação e autorização, quando necessário”.

Os metadados *Citation* e *Citation/reference* contribuem com a citação e referenciação dos dados, assim como os itens documentários como artigo e livro. Sua relevância consiste em dar créditos aos autores além de padronizar a sua referência na lista de referências dos trabalhos que utilizaram os dados e está em consonância com as boas práticas da Ciência aberta.

Tanto a interface para representação descritiva quanto a interface de busca do repositório disponibilizam diversos metadados, o que pode configurar a aplicação de metadados com requisitos e metadados de alto nível, apoiada com o perfil de aplicação baseado no padrão *Dublin Core* e no *DataCite* que o repositório utiliza.

O repositório disponibiliza informações sobre as métricas de visualizações, citações e *downloads*. O repositório não disponibiliza informações sobre os serviços *altmetrics* prestados por empresas como a *Plum Analytics* e *ImpactStory* (ENIS, 2013).

Os Altmetrics são indicadores alternativos baseados em mídias sociais e estão emergindo como um instrumento útil para avaliar o impacto dos dados da pesquisa (PETERS et al., 2016) e tem preeminente potencial para pesquisadores que pesquisam sobre métricas. As métricas de dados estão previstas nos princípios FAIR e a utilização pelos outros repositórios vai contribuir para a visibilidade e fomentar a

citação dos dados neles armazenados.

A indexação do conteúdo de repositórios ajuda os usuários na sua localização. A indexação do *e-cienciaDatos* é feita pelos bibliotecários utilizando as palavras-chave fornecidas pelos pesquisadores no PGD e na documentação dos dados. Não são utilizados vocabulários controlados. Apesar da importância da indexação para melhor recuperação dos *datasets*, não é utilizado um vocabulário controlado e sim uma lista de termos e as palavras-chave indicadas pelos autores.

Os dados do repositório *e-cienciaDatos* estão em acesso aberto, a menos que tenham razões legais, éticas ou de transferência de tecnologia que o impeçam. Há opção de aplicar as licenças livres CC0, CC-BY, CC-BY-SA e *Open Data Commons* PDDL aos conjuntos de dados e licenças para *software* aprovado pela *Open Source Initiative* (OSI), *GNU General Public License Versions* e licença MIT (MALO DE MOLINA, 2019). Caso um pesquisador necessite de outra licença não especificada pelo repositório ele deve solicitá-la aos administradores (MALO DE MOLINA, 2019). A indicação para atribuição de licença sobre os dados de maior destaque no repositório é CC-BY, permitindo ao usuário fazer maior uso dos conjuntos de dados.

Destaca-se que *Consórcio Madroño* e suas universidades membros têm compromisso com a Ciência aberta e o Acesso aberto e, em suas políticas, os aspectos relacionados à propriedade intelectual estão explicitados (CONSÓRCIO MADRONO, 2017).

O *Consórcio Madroño* e suas bibliotecas elaboraram os documentos com as instruções que regulamentam vários aspectos de sua infraestrutura, organização e gestão dos dados. Entre essas instruções, estão enfatizadas as diretrizes sobre o direito autoral, a citação dos dados, modelos de referência para os dados e como os dados estão disponíveis para uso sendo aplicadas as licenças apresentadas.

Os documentos que regulamentam e instrui os autores e os usuários são preponderantes no contexto dos repositórios de dados em que, em sua maioria, são implementados por universidades e instituições de pesquisa, as quais definem políticas que abrangem formas de disponibilização e uso do conteúdo, sendo identificados os aspectos dos direitos autorais. Relevante destacar a necessidade de o repositório incluir em sua agenda a adoção de políticas e diretrizes de autoavaliação para ajustes dos aspectos deficitários.

As diretrizes do repositório orientam os autores para a liberação de seus conjuntos de dados e os usuários no seu uso, assessorando os envolvidos sobre as

questões éticas e jurídicas. Essas diretrizes precisam ser claras para não gerar dúvidas e incertezas. É recomendável que esteja de fácil localização e visualização na página do repositório e nas páginas das bibliotecas.

A próxima seção explana sobre a propriedade intelectual na União europeia e na Espanha explicitando o direito moral e patrimonial.

6 PROPRIEDADE INTELECTUAL

A propriedade intelectual surgiu de uma morosa evolução no contexto histórico. Sua legitimidade é reconhecida mundialmente. Leis e tratados internacionais foram elaborados como dispositivos legais para dar proteção legal às criações humanas e regularizar e intermediar essas ações e os interesses dos consumidores. Atualmente, as instituições que direta e indiretamente trabalham com publicação de dados de pesquisa se baseiam em leis e normativas sobre propriedade intelectual para elaboração de suas políticas e manuais. Nos documentos elaborados, são explicitados os dispositivos jurídicos que regulamentam o processo, além das boas práticas científicas.

A propriedade intelectual é o conjunto de direitos que protege os autores, os inventores e demais proprietários no que diz respeito a qualquer produção intelectual, resultantes da sua criação, seja no domínio industrial, científico, literário ou artístico, facultando aos seus titulares os direitos econômicos (direitos patrimoniais) os quais ditam a forma de produção, comercialização, circulação, divulgação e utilização dos ativos intelectuais ou dos produtos e serviços que incorporam tais criações (BRASIL, 2007; ASSOCIAÇÃO PAULISTA DA PROPRIEDADE INTELECTUAL (ASPI), c2016; WORLD INTELLECTUAL PROPERTY ORGANIZATION (WIPO), [20--]).

Direitos de propriedade intelectual concedem aos criadores ou proprietários das patentes, das marcas comerciais ou dos direitos autorais trabalharem para benefício de sua própria atividade ou dos investimentos em sua criação (WIPO, [20--]).

A fonte de todos os direitos de propriedade intelectual é a legislação nacional (CARROL, 2015). Cada país elaborou sua legislação baseada em seu contexto e no modo mais propício em conciliar os interesses do público que usam a produção intelectual disponibilizada e os autores que dão expressão a elas.

Além das legislações nacionais, os Tratados Internacionais, como os resultantes da Convenção de Paris (1883)⁷ e da Convenção de Berna (1886)⁸, harmonizam os direitos do detentor da propriedade intelectual, porém, permitem que os direitos dos usuários variem de acordo com seu país (CARROL, 2015). Gandelman

⁷ Proteção da Propriedade Industrial

⁸ Proteção das Obras Literárias e Obras artísticas

(2007, p. 31) enfatiza que os tratados internacionais têm o intuito de “[...] dar aos autores e titulares dos países aderentes aos convênios a mesma proteção legal que cada país dá a seu autor ou titular nacional. É o princípio de reciprocidade no tratamento jurídico da autoria.”

A importância da propriedade intelectual foi reconhecida pela primeira vez na Convenção de Paris para a Proteção da Propriedade Industrial e da Convenção de Berna para a Proteção das Obras Literárias e Obras artísticas nos anos de 1883 1886 respectivamente. Ambos os tratados são administrados pela Organização Mundial da Propriedade Intelectual (OMPI) agência especializada da Organização das Nações Unidas (ONU) (WIPO, [20--]).

A Convenção de Berna teve revisões em seu teor conforme apresentado pela WIPO ([20--]): completada em Paris em 4 de maio de 1896; completada em Berlim em 13 de novembro de 1908; com emenda em Berna em 20 de março de 1914 e revisada em Roma em 2 de junho de 1928; revisada em Bruxelas em 26 de junho de 1948; revisada em Estocolmo em 14 julho de 1967; revisada em Paris em 24 julho de 1971 e com emenda em 28 de setembro de 1979.

Normas internacionais importantes são: o Tratado da União de Berna, assinado em 1886 e revisado várias vezes, e os Tratados da OMPI de 20 de dezembro de 1996, um sobre direitos autorais e outro sobre direitos conexos (UC3M, [201-a]).

A Convenção de Berna, pacto internacional que estruturou os princípios básicos dos direitos autorais, tornou o instrumento internacional de referência para proteção de direitos autorais, com adesão de diversos países. A Convenção foi realizada em 1886, atualizada em 1971, para dar proteção adicional a artistas, músicos, fotógrafos, designers, entre outros e, assim garantir proteção dos trabalhos literários e artísticos (OMPI, 1996). Obras intelectuais, no campo científico, estão estabelecidas no artigo 2, n. 1 e, segundo a Convenção, essas obras abrange “[...]toda produção no domínio literário, científico e artístico, qualquer que seja o modo ou a forma de sua expressão, como livros, panfletos e outros escritos [...]” (OMPI, 1996, p. 22).

Para a *World Intellectual Property Organization* ([20--]) a expressão “propriedade intelectual” engloba duas vertentes:

- Propriedade Industrial: inclui patentes, marcas, desenhos industriais,

indicações geográficas e obtenções vegetais.

- Direitos autorais: abrangem obras literárias como: romances, poemas e peças; filmes, música, trabalhos artísticos como, por exemplo, desenhos, pinturas, fotografias e esculturas e projeto arquitetônico.

A *World Intellectual Property Organization* também apresenta uma definição para *copyright*

Copyright (ou direitos de autor) é um termo legal usado para descrever os direitos que os criadores têm sobre suas obras literárias e artísticas. As obras protegidas por direitos autorais variam de livros, música, pinturas, esculturas e filmes a programas de computador, bancos de dados, anúncios, mapas e desenhos técnicos. WIPO ([20--], tradução nossa).

De acordo com a WIPO ([20--]) o *copyright* (ou direito do autor) tem duas vertentes:

- direitos econômicos: permitem ao detentor dos direitos obter recompensa financeira pelo uso de suas obras por terceiros;
 - direitos morais: protegem os interesses não econômicos do autor.
- Essas duas vertentes são apresentadas nas próximas subseções.

6.1 Direitos autorais

O direito autoral apresenta fundamentalmente dois aspectos: o moral, que garante ao criador (pessoa física ou jurídica) o controle à menção de seu nome no uso, reprodução ou divulgação de sua obra e o respeito a sua integridade; e o patrimonial, que visa regular as relações jurídicas da utilização econômica das obras intelectuais (GANDELMAN, 2007).

Os direitos autorais protegem o trabalho criativo e o investimento despendido da mesma forma que estimulam a inovação e a exploração ordenada das descobertas científicas (ELLIOT, 2005). Direitos autorais são aplicados como um mecanismo legal para a ordenação da vida cultural, científica e social além de promover a maneira mais propícia em conciliar os interesses do público que usam conteúdos compartilhados e os autores que dão expressão às ideias (PLOMAN; HAMILTON, 1980).

Considerando as discussões apresentadas na comunidade científica sobre a disponibilização, o compartilhamento e a reutilização dos dados de pesquisa de forma aberta, gratuita e com menos restrições de uso possíveis, as orientações de propriedade intelectual sobre os dados de pesquisa podem seguir a mesma lógica das obras exemplificadas pela WIPO ([20--]). No quadro 6 estão relacionadas algumas características dos direitos autorais de dados de pesquisa adotadas nas boas práticas científicas e explicitadas em repositórios de dados.

Quadro 6 - Aspectos dos direitos autorais aplicados aos dados

Registro	Declaratório, optativo e gratuito
Titularidade	Desde sua coleta, geração ou registro
Requisitos básicos	Originalidade da forma de expressão
Prazo de proteção	Variável, conforme determinação do detentor do direito autoral ou do publicador
Cessão de direitos	Os direitos patrimoniais podem ser total ou parcialmente cedidos
Integridade	Controle exercido através dos direitos morais
Divulgação	<i>Datapapers</i> , repositórios de dados, periódicos científicos
Relação instituição X autor	Titularidade patrimonial estabelecida em contrato
Beneficiário do uso	Autor(es) ou consumidores
Proteção internacional	Automática e independente de registro; segundo reciprocidade dos tratados ou das comunidades científicas

Fonte: adaptado de Gandelman (2007, p. 38).

Conforme mostrado no quadro 6, os direitos patrimoniais podem ser total ou parcialmente cedidos e os direitos morais controlam a integridade dos titulares. Destaca-se que os titulares dos conjuntos de dados precisam ficar atentos a cessão dos direitos patrimoniais. Ao enviar seu artigo para publicação em periódicos comerciais, podem ser solicitados os dados que embasaram as discussões do artigo.

Caso o titular tenha requerido financiamento para sua pesquisa, em muitos casos, as agências condicionam o recurso à disponibilização aberta e gratuita dos dados e, com o envio para os periódicos comerciais, pode ocorrer choque de interesses.

Na literatura há discussões e descrições sobre ações e opções para o registro de dados e, assim, garantir a propriedade intelectual. Autores destacam a necessidade de registro em virtude do trabalho realizado com os dados exigir: tempo e recursos consumidos nos processos de coleta e análise; organização dos dados e de sua documentação; competências específicas no ciclo de vida dos dados, o que requer, em muitos casos, a participação de especialistas; utilização de ferramentas, de *hardware* e de *software*; equipamentos e infraestruturas especiais relacionadas aos ambientes de coleta desses dados, entre outros, que são específicos de cada área do conhecimento (ABUBAHIA; COCEA, 2015; ZOPE-CHAUDHARI; VENKATACHALAM; BUDDHIRAJU, 2017). Os autores realçam, ainda, que partes destes requisitos apresentam alto custo e que parte dos processos de coleta não podem ser replicados por se tratarem de dados únicos (ABUBAHIA; COCEA, 2015; ZOPE-CHAUDHARI; VENKATACHALAM; BUDDHIRAJU, 2017).

O registro dos dados não é obrigatório, é um meio de proteção e de prova. O local e forma de registro pode variar dependendo do país. Relevante destacar que devido ao tempo despendido e os custos na geração ou coleta, no tratamento e na gestão de dados serem relativamente dispendiosos, os detentores dos dados têm recorrido a estratégias para proteger a autoria dos dados. Um recurso utilizado para o registro da proveniência e garantia da autoridade de dados, incluindo dados de pesquisa, é a tecnologia *blockchain* (MONTEIRO et al., 2018).

Na operacionalização do *blockchain*, quando um conjunto de dados é publicado, uma versão do contrato deve ser implantada e a primeira função deve ser chamada para registrar os termos sob os quais os dados podem ser reutilizados. Quando um usuário tem o intuito de acessar o conjunto de dados, as duas partes (autor e reutilizador) devem chegar a um acordo sobre seguir os termos e registrá-los no *blockchain*. A prova de que os termos foram realmente seguidos será registrada. Estudos feitos pelos autores Elizabete Monteiro, Jackelin Camperos Reyes, Elaine Affonso e Ricardo Sant'Ana (MONTEIRO et al., 2018, 2020) demonstram como o *blockchain* está sendo utilizado para registro dos dados, inclusive dados de pesquisa e relacionados aos usos no contexto da Ciência aberta.

Quadro 7 – Aplicação do blockchain em dados de pesquisa

Autores	Título do artigo	Tecnologia proposta	Propósito do uso da Tecnologia	Contexto de uso	Eixo Ciência Aberta
GIPP et al.	CryptSubmit: Introducing Securely Timestamped Manuscript Submission	CryptSubmit, plataforma para submissão de documentos e dados de pesquisa para a publicação acadêmica	Incrementar confiança nos sistemas de envio e avaliação por pares, no contexto da publicação de documentos científicos e permitir	Autores e o público verificam independentemente [...] que um conjunto de dados [...] já existia em um formato preciso no	Open Science Evaluation - Open Peer Review
	and Peer Review Feedback Using the Blockchain	a qual gera uma impressão digital única, do manuscrito de pesquisa e dos dados acompanhantes	aos pesquisadores verificarem com segurança a existência de ideias, dados ou resultados no momento do envio para revisão	momento da submissão a uma conferência ou revista	Open Science Tools - Open Repositories
PĂNESCU; MANTA	Smart Contracts for Research Data Rights Management over the Ethereum Blockchain Network	Solidity, Smart contracts. Geração de registro do acordo entre os autores dos dados de pesquisa e os usuários que vão reutilizar, permitindo aos autores supervisionar o acesso e as condições de reutilização	Ajudar a rastrear todo e qualquer acesso e reutilização de um conjunto de dados compartilhado e contribuir para aumentar a granularidade na indicação do licenciamento dos dados	Publicação dos dados, implantação do contrato, registro dos termos para reutilização e acesso mediante acordo entre o autor e reutilizador com os termos registrados no blockchain	Open Access - Open Access Use and Reuse

Fonte: Monteiro et al. (2019).

Segundo Kim e Laskowski (2018) *blockchain* é uma tecnologia de registro público baseado em uma arquitetura distribuída, uniformemente replicado em todos os seus nós, com a descentralização como medida de segurança, à prova de adulteração, que fornece um método para estabelecer a existência de uma transação em um momento específico, contendo todas as transações que já foram executadas no sistema e que podem ser verificadas por qualquer parte interessada, incluindo o rastreamento da proveniência dos dados.

No contexto da gestão de dados de pesquisa, *blockchain* possibilita aos pesquisadores interagirem com o conteúdo em qualquer etapa do ciclo de vida dos dados, gerir esses dados de forma descentralizada, com a possibilidade da auto-regulação por meio de uma infraestrutura compartilhada para transações ou fluxos dos dados, bem como definir a granularidade das condições de reutilização dos dados (MONTEIRO et al., 2018, 2019).

O *blockchain* visa assegurar o registro e, conseqüentemente, a proveniência dos dados. Esse aspecto auxilia na rastreabilidade, na verificação de acessos e de

usos, na autenticidade e na integridade dos dados. Com isso, identifica as alterações do conteúdo dos conjuntos de dados propiciando a menção do autor quando usados, estando em consonância com os aspectos do acesso aberto que garante a disseminação, acesso livre e, em alguns casos, visualização dos dados com alguns direitos de utilização adicionados.

De certo, o registro é fator significativo, uma vez que confere a autenticidade do protocolo original e a segurança jurídica ao autor, facilita a comprovação da criação dos dados, assegura o direito autoral na publicação e a obtenção de um certificado de registro (MONTEIRO et al., 2018, 2020). De acordo com Karafiloski (2017) quando se tem o registro de data e hora das transações, é possível obter prova no tribunal, no caso de uma disputa de propriedade.

A legislação sobre o direito do autor protege as obras literárias, artísticas e científicas de diversas formas, define os beneficiários da proteção e determina o alcance e a duração dessa proteção (UNESCO, 1981) e podem servir de orientação para elaboração de diretrizes para uso de dados de pesquisa. Os detentores de direitos autorais podem recorrer à justiça na ocasião em que o resultado de sua atividade intelectual ou científica for utilizado sem a sua devida autorização (UNESCO, 1981). Da mesma forma, na ocasião de utilização sem a devida citação.

Em uma perspectiva geral, um projeto pode conter dois elementos: os dados gerados ou coletados e o banco de dados. Fazer a distinção entre esses dois elementos em relação a propriedade intelectual nos ajuda a compreender a dinâmica nos repositórios de dados.

Dados factuais, geralmente, não tem proteção de direitos autorais sob as leis de alguns países por não ser possível o registro de direitos autorais de fatos. Além disso, os metadados que os descrevem também são considerados dados factuais e, portanto, sem proteção de direitos autorais (RESEARCH DATA MANAGEMENT SERVICE GROUP (RDMSG), [201-]). Cabe considerar que, apesar de não haver legislação que conceda direitos de propriedade sobre os dados em si, ressalva para alguns tipos de dados que já estão amparados pelas leis, diversos campos do direito abordam questões relacionadas aos dados e banco de dados como propriedade (STEPANOV, 2019).

Por outro lado, um banco de dados pode ter uma camada fina de proteção de direitos autorais. Decidir quais dados serão incluídos no banco de dados, como os dados serão organizados e como relacionar diferentes elementos de dados são todas

decisões criativas e, dessa forma, podem receber proteção de direitos autorais (RDMSG, [201-]).

Devido a disparidade dos *status* de direitos autorais de bancos de dados e conteúdo de dados, diferentes mecanismos são aplicados para gerenciar cada um. Os direitos autorais podem reger o uso de bancos de dados e algum conteúdo de dados (o que é original), mas a lei contratual, as marcas registradas e outros mecanismos são necessários para regular os dados factuais (RDMSG, [201-]).

Vale destacar que o repositório de dados, considerado como banco de dados ou recurso de dados, deve ser resguardado, tendo o conhecimento de quem é o proprietário dos conjuntos de dados ou dos recursos depositados e, também, como outras pessoas podem acessá-los e utilizá-los.

O responsável pelo repositório organiza e disponibiliza os conjuntos de dados recebidos por terceiros e deve se certificar de que todas as obrigações éticas, legais e profissionais que alguém possa ter para com o repositório sejam cumpridas.

Os repositórios podem, por exemplo, indicar no contrato ou nos termos de depósito que os detentores concedam permissão de uso com mínimo de restrição possível ou abdicuem de quaisquer direitos de propriedade intelectual que possam ter sobre os dados (CARROL, 2015). Além disso, deve-se ter a explicitação das licenças aplicadas nos termos de uso do repositório, a indicação da licença de cada conjunto de dados, a orientação da devida citação e os modelos da citação e referência dos conjuntos de dados.

No artigo de Carrol (2015), na seção sobre as recomendações para aumentar o compartilhamento e a abertura de dados, são apresentados os mecanismos pelos quais os repositórios de dados podem tornar mais fácil para os depositantes sinalizarem o escopo da permissão que eles concedem aos consumidores dos dados. O autor apresenta os contratos, as licenças e os termos de uso, itens que os repositórios de dados utilizam em sua operacionalização. Considerando os contratos e as licenças, o autor parte do pressuposto que, quando um ou mais direitos de propriedade intelectual se aplicam aos dados de pesquisa, o proprietário desses direitos pode conceder permissão para sua reutilização por meio de uma licença. Carrol (2015) evidencia que em termos legais, uma concessão de permissão é uma licença não exclusiva. Essa concessão é vislumbrada nos repositórios de dados através, por exemplo, de licenças como a *Creative Commons* e *Open data Commons*.

O autor diferencia uma licença não exclusiva de uma licença exclusiva. A

segunda é aquela em que o proprietário dos direitos concorda em renunciar quaisquer direitos de uso da propriedade intelectual (patrimonial), geralmente em troca de alguma forma de compensação (CARROL, 2015). Esse último aspecto pode ser observado quando o detentor dos direitos os concede aos periódicos comerciais que publicam os dados juntamente ao artigo ou indicam o repositório de dados para a sua publicação.

A cerca dos termos de uso, é fortemente recomendado ao titular dos direitos sobre os dados deixar claro para os usuários potenciais se há quaisquer direitos de propriedade intelectual associados aos dados e quem os possui, inclusive indicar qual é a licença aplicada sobre o conjunto de dados ou as permissões associadas.

Cabe a comunidade acadêmica discutir e indicar se é ético ou apropriado condicionar o acesso aos dados ao acordo de um contrato que impõe restrições de uso dos conjuntos dos dados (CARROL, 2015). Os repositórios de dados disponibilizam aos autores opções de liberação dos dados. A liberação total ou condicional dos conjuntos de dados pode variar de acordo com a cultura de determinada área do conhecimento, de comunidades ou de práticas adotadas por pesquisadores individuais.

Tanto os direitos morais quanto os direitos patrimoniais estão descritos na Declaração Universal dos Direitos do Homem, em seu artigo 27: “Todo ser humano tem direito à proteção dos interesses morais e materiais decorrentes de qualquer produção científica literária ou artística da qual seja autor.” (NAÇÕES UNIDAS, 2009). Partindo do pressuposto apresentado, esses direitos pertencem, inicialmente, ao autor. O direito autoral é dividido em direito moral e direito patrimonial abordados nas próximas subseções.

As definições de propriedade intelectual foram apresentadas para contextualizar a temática discutida ao longo da Tese. Em virtude de haver uma pequena diferença entre a definição da propriedade intelectual na Espanha em relação a definição da WIPO, nesta Tese, ao abordar os “direitos autorais”, está se referindo ao direito de autor, na vertente direito moral e na patrimonial.

6.1.1 Direitos Morais

As diversas atividades da comunidade científica avançaram na coleta e

produção de dados, gerando colaboração internacional, conhecimento compartilhado e progresso na ciência (MONTEIRO et al., 2020). O domínio da comunicação científica é complexo e um de seus desafios é a atribuição da autoria nos fluxos das pesquisas, observado em contextos digitais e perante o dinamismo e a produtividade presente nesse ambiente (VAN ROSSUM, 2017). No domínio dos dados de pesquisa, organizações têm trabalhado para incentivar e conscientizar os diversos atores do contexto científico da necessidade e relevância da atribuição de autoria, explicitada na citação dos dados.

O Direito moral é um direito fundamental do autor, inalienável e imprescritível, um direito ligado a pessoa ou instituição jurídica que criou uma obra (UNESCO, 1981). O direito de paternidade é o qual o autor tem o reconhecimento de uma obra como sua e a ela associar ou não o seu nome (UNESCO, 1981). O autor pode almejar que seu nome seja mencionado todas as vezes que a sua obra for utilizada (UNESCO, 1981).

O Direito moral é o reconhecimento pela autoria com sua explicitação nas citações. A *Joint Declaration of Data Citation Principles – Final* declaração do Grupo de trabalho *The Future of Research Communications and e-Scholarship* (FORCE11) evidencia três aspectos primordiais: primeiro que os dados devem ser considerados produtos de pesquisa legítimos e citáveis; segundo que e a citação de dados se configura em boas práticas de pesquisa; terceiro que os dados fazem parte do ecossistema acadêmico que apoia a reutilização de dados (MARTONE, 2014).

A referida declaração possui oito princípios que são a síntese do trabalho de profissionais vinculados a várias instituições⁹ e cobrem o propósito, a função e os atributos das citações, com o intuito de orientar e incentivar as comunidades a desenvolverem práticas e ferramentas que incorporem esses princípios. Seu escopo fornece recomendações de citações de dados, não inclui especificações detalhadas para implementação e nem direciona para tecnologias, ferramentas ou repositórios de dados de pesquisa, porém se estende a todos os tipos de dados, inclusive os dados de pesquisa (ALTMAN et al., 2015). Os princípios não seguem ordem definida por critérios e são apresentados como:

1. Importância

Os dados devem ser considerados produtos de pesquisa legítimos. As citações de dados devem ter a mesma importância no registro

⁹ <https://www.force11.org/datacitation/workinggroup>

acadêmico que as citações de outros objetos de pesquisa, tais como publicações.

2. Crédito e Atribuição

As citações de dados devem facilitar a concessão de crédito acadêmico e atribuição normativa e legal a todos os contribuintes dos dados, reconhecendo que um único estilo ou mecanismo de atribuição pode não ser aplicável a todos os dados.

3. Evidência

Na literatura acadêmica, sempre e onde quer que uma declaração se baseie em dados, os dados correspondentes devem ser citados.

4. Identificação única

Uma citação de dados deve incluir um método persistente de identificação que seja acionável por máquina, globalmente único e amplamente utilizado por uma comunidade.

5. Acesso

As citações de dados devem facilitar o acesso aos próprios dados e aos metadados associados, documentação, código e outros materiais, conforme necessário para que tanto humanos quanto máquinas façam uso informado dos dados referenciados.

6. Persistência

Os identificadores e metadados únicos que descrevem os dados, e sua disposição, devem persistir - mesmo além da vida útil dos dados que eles descrevem.

7. Especificidade e Verificabilidade

As citações de dados devem facilitar a identificação, o acesso e a verificação dos dados específicos que suportam uma reivindicação. Citações ou metadados de citação devem incluir informações sobre proveniência e fixação suficientes para facilitar a verificação de que o tempo específico, versão e/ou porção granular dos dados recuperados subsequente são os mesmos que foi originalmente citado.

8. Interoperabilidade e Flexibilidade

Os métodos de citação de dados devem ser suficientemente flexíveis para acomodar as práticas variantes entre as comunidades, mas não devem diferir tanto que comprometam a interoperabilidade das práticas de citação de dados entre as comunidades. (MARTONE, 2014, tradução nossa).

Dentre seus princípios destaca-se o 2. Crédito e Atribuição, “As citações de dados devem facilitar a concessão de crédito acadêmico e a atribuição normativa e

legal a todos os autores dos dados.” (MARTONE, 2014, tradução nossa).

Nos países em que a legislação relacionada ao direito de autor reconhece expressamente os direitos morais, os autores conservam esses direitos mesmo que tenham cedido os seus direitos patrimoniais (UNESCO, 1981). Os direitos morais, que configuram a vinculação entre o autor e sua obra de criação, não podem ser transferidos a terceiros, porém seu direito patrimonial pode ser cedido sob determinada licença a título oneroso ou gratuito.

O avanço na ciência e o aumento vultoso de dados de pesquisa proporcionam novas possibilidades com a sua reutilização, o que requer um equilíbrio jurídico entre duas partes: o autor detentor dos direitos autorais e o usuário final (MONTEIRO, 2017). As licenças de uso contribuem nesta direção.

6.1.2 Direitos Patrimoniais

Os direitos patrimoniais regulamentam as relações de uso com aplicação das licenças de uso, as quais são utilizadas com o intuito de publicar, reutilizar e replicar os dados. As licenças *Creative Commons* e *Open Data Commos* para os conjuntos de dados e licenças como *General Public Licence* para *software* são indicadas pelos repositórios de dados para a reutilização e replicação dos dados. Uma licença evidencia todas as liberações e restrições no compartilhamento e uso dos dados armazenados nos repositórios de dados e, também, a importância do reconhecimento do autor ou detentor dos dados.

As licenças de uso são “[...] instrumentos legais que, normalmente, objetivam garantir o reconhecimento da autoria e promover a distribuição livre e aberta do conhecimento construído coletivamente.” (MANTOVANI; DIAS; LIESENBERG, 2006, p. 273). As licenças de uso asseguram a publicação de dados sem ferir aspectos dos direitos autorais, transfere o direito patrimonial a outrem, cria clareza quanto a utilização do conteúdo e constitui uma segurança para os usuários no sentido da utilização dos conjuntos de dados (MONTEIRO, 2017).

As licenças *Creative Commons* oferecem uma maneira fácil de gerenciar os termos de direitos autorais que se ligam automaticamente a todo o material criativo sob *copyright* (CREATIVE COMMONS, 2016). *Creative Commons* é uma organização global, não governamental, sem fins lucrativos localizada em *Mountain View*, Califórnia,

fundada em 2001 por Lawrence Lessig e autoriza a partilha e reutilização da criatividade e conhecimento através do fornecimento de instrumentos legais gratuitos (CREATIVE COMMONS, 2016).

Lessig, advogado americano, é conhecido por suas obras que analisam o direito do autor assim como opositor do *software* proprietário, além de defender o Código Livre para que todos os usuários tenham acesso à informação e a estrutura construtiva da informação (JORENTE; SANTOS, 2008).

A *Creative Commons* possui seis licenças ao total que permitem:

- 1 - **Creative Commons Atribuição (CC-BY)** - distribuir, remixar, adaptar e criar a partir do trabalho, mesmo para fins comerciais, desde que lhe atribua o devido crédito pela criação original;
- 2 - **Creative Commons Atribuição-Compartilhual (CC BY-SA)** - remixar, adaptar e criar a partir do trabalho original, mesmo para fins comerciais, desde que lhe atribuam o devido crédito e que licenciem o trabalho derivado sob termos idênticos;
- 3- **Creative Commons Atribuição-SemDerivação (CC BY-ND)** - redistribuir, comercial e não comercial, desde que o trabalho seja distribuído inalterado e no seu todo, com crédito atribuído ao autor;
- 4 - **Creative Commons Atribuição-NãoComercial (CC BY-NC)** - remixar, adaptar e criar a partir do trabalho original para fins não comerciais;
- 5 - **Creative Commons Atribuição-NãoComercial-Compartilhual (CC BY-NC-SA)** - remixar, adaptar e criar a partir do trabalho original para fins não comerciais, desde que atribuam o devido crédito ao autor e que licenciem as novas criações sob termos idênticos;
- 6 - **Creative Commons Atribuição-SemDerivações-SemDerivados (CC BY-NC-ND)** – fazer o *download* dos trabalhos e os compartilharem desde que atribuam crédito ao autor, mas sem que possam alterá-los de nenhuma forma ou utilizá-los para fins comerciais (CREATIVE COMMONS, [2016]).

A *Creative Commons* fornece também instrumentos que operam no espaço "todos os direitos concedidos", do domínio público. A licença *No rights reserved* (CC Zero) ou (CC0) = CC Zero permite aos criadores e proprietários de conteúdo renunciar todos seus direitos sobre suas obras e, assim, colocá-los em domínio público, para que outros possam construir livremente sobre a obra original, melhora-las e reutilizá-las para quaisquer fins sem restrições de direito autoral. Destaca-se que as licenças da versão CC 4.0 abordam as características das bases de dados.

Essas licenças podem ser remixadas e combinadas para a liberação do conteúdo. Lessig afirma que a cultura do remix é a peça central do processo criativo contemporâneo e sempre defendeu que essa cultura “[...] é aquela que vai derrubar as barreiras entre criador e ‘consumidor’ da cultura. [...] Entre Estado e cidadão.” (LEMOS, 2005, p. 18).

A *Creative Commons* tem em seu âmbito a iniciativa *Science Commons* que tem o intuito de quebrar barreiras e desenvolver ferramentas para facilitar a reutilização dos dados resultantes de projetos de pesquisa. Nessa mesma direção, o *Science Commons Open Access Data Protocol* estabelece uma metodologia e boas práticas para a criação de ferramentas que permitem a integração de bancos de dados de pesquisa entre si e a disponibilização em domínio público (GRUPO DE TRABALHO..., 2012).

A licença *Creative Commons* atua como um dispositivo para autorizar previamente a utilização da obra, comercialmente ou não, com alteração ou não do conteúdo original dependendo de qual a chancela CC escolhida (LEMOS, 2005). As licenças estão apresentadas no quadro 8 com suas respectivas permissões.

Quadro 8 – Licenças Creative Commons

SÍMBOLO		
	Atribuição	distribuir, remixar, adaptar e criar a partir do seu trabalho, mesmo para fins comerciais, desde que lhe atribua o devido crédito pela criação original.
	Atribuição-Compartilha Igual	remixar, adaptar e criar a partir do trabalho original, mesmo para fins comerciais, desde que lhe atribuam o devido crédito e que licenciem o trabalho derivado sobre termos idênticos.
	Atribuição-Sem Derivações	redistribuir, comercial e não comercial, desde que o trabalho seja distribuído inalterado e no seu todo, com crédito atribuído ao autor.
	Atribuição-Não Comercial	remixar, adaptar e criar a partir do trabalho original para fins não comerciais e os usuários não precisam licenciar os trabalhos derivados sob os mesmos termos.
	Atribuição-Não Comercial-Compartilha Igual	remixar, adaptar e criar a partir do trabalho original para fins não comerciais, desde que atribuam o devido crédito ao autor e que licenciem as novas criações sob termos idênticos.
	Atribuição-Sem Derivações-Sem Derivados	Fazer <i>download</i> dos trabalhos e os compartilhem desde que atribuam crédito ao autor, mas sem que possam alterá-los de nenhuma forma ou utilizá-los para fins comerciais.
	No <i>rights reserved</i>	renunciar os direitos sobre suas obras e, assim, colocá-los tão completamente quanto possível no domínio público, para que outros possam construir livremente sobre a obra original, melhora-las e reutiliza-las para quaisquer fins sem restrições de direito autoral.

Fonte: Elaborado pela autora a partir de Creative Commons ([20--]).

O *Open Data Commons* (ODC), projeto da *Open Knowledge Foundation*, gerido pelo Conselho Consultivo, lançou sua primeira licença de dados abertos em março de 2008 a *Public Domain Dedication and License* (PDDL). Além das licenças CC, a *Open Data Commons* (ODC) oferece três opções de licenças criadas especificamente para dados / bancos de dados:

1 - **Public Domain Dedication and License (PDDL)**: os usuários são livres para copiar, distribuir e usar o trabalho; produzir trabalhos a partir dos dados / bancos de dados armazenados e para modificar, transformar e construir sobre os dados / banco de dados.

2 - **Attribution License (ODC-By)**: permite que outras pessoas copiem, distribuam e usem o trabalho e produzam trabalhos a partir dos dados armazenados além de

modificar, transformar e construir sobre os dados / banco de dados, desde que eles forneçam a atribuição adequada.

3 - Open Database License (ODC-ODbL): permite que outras pessoas copiem, distribuam e utilizem o trabalho e produzam trabalhos a partir dos dados / banco de dados e modifiquem, transformem e construam sobre os dados / banco de dados, desde que eles forneçam a atribuição adequada e que todos os trabalhos derivados sejam disponibilizados sob esta mesma licença (OPEN KNOWLEDGE FOUNDATION, [201-]).

Todas as licenças possibilitam a utilização de materiais sem ferir os direitos autorais. Carrol (2015) enfatiza que os direitos autorais são muito mais complexos do que o necessário. Sendo assim, apesar de haver leis de proteção de camadas de banco de dados ou tipos de dados e tratados e harmonizações internacionais, a comunidade científica está conciliando a disponibilização e reutilização de dados conforme as necessidades e orientações da ciência aberta.

6.2 Propriedade intelectual na União Europeia

A gestão de direitos de propriedade intelectual (DPI) é um item indispensável em qualquer programa de gestão de dados. Os dados são resultados da pesquisa e, assim como os textos publicados, são passíveis de aplicação de propriedade intelectual. A União Europeia (UE) criou a *Intellectual Property Rights* (IPR) *sui generis* com o intuito de incentivar a criação de bases de dados ao oferecer direitos de propriedade aos seus fabricantes, indicando a proteção sobre o investimento nos dados obtidos, verificados ou apresentados (STEPANOV, 2019). De acordo com a Directiva, base de dados “[...] significa um conjunto de obras, dados ou outros materiais independentes, organizado de forma sistemática ou metódica e acessível individualmente por meios eletrônicos ou outros.” (STEPANOV, 2019, p. 70).

Na União Europeia os direitos de propriedade que podem ser aplicados aos dados de pesquisa são: segredos comerciais (informações confidenciais), direitos autorais e direitos especiais de banco de dados (CARROL, 2015). Dois tipos de Direitos de Propriedade Intelectual pertinentes para a contextualização e discussão da Tese são: a proteção de banco de dados e direitos autorais. Os dados podem, também, ser objeto de proteção de segredos comerciais, porém, essa área não

será abordada.

O artigo de Carrol (2015) discute a legislação de propriedade intelectual no contexto Europeu e identifica os direitos que podem ser aplicados em dados de pesquisa. No artigo é destacado a Diretiva de Bancos de Dados da União Europeia. De acordo com a Diretiva, esses direitos se aplicam à base de dados, sob qualquer forma, em que tenha sido empregado um “investimento substancial” para ser montado ou mantido. O autor destaca a interpretação do Tribunal de Justiça da União Europeia (*Court of Justice*) sobre esse direito de bases de dados. Segundo a interpretação do Tribunal, o direito se limita às bases de dados que requerem investimentos na obtenção de dados e não na criação dos dados subjacentes.

As discussões apresentadas por Carrol (2015) incluem o fato de que os direitos de bases de dados *Sui Generis* protegem contra a extração ou a reutilização de partes substanciais de uma base de dados protegida, bem como a extração frequente de partes não substanciais de uma base de dados protegida. O autor evidencia uma apreensão com relação ao compartilhamento de dados de pesquisa. Em sua argumentação, destaca-se que os direitos de bases de dados *Sui Generis* seria uma barreira para o compartilhamento de dados de pesquisa, caso não estivesse sujeita a uma limitação para pesquisa não comercial. O autor enfatiza que uma quantidade significativa de dados de pesquisa atende ao requisito limite de “investimento substancial” de recursos financeiros e mão de obra devido à sua ampla definição, mas uma quantidade substancial de uso de tais dados em universidades ou hospitais sem fins lucrativos provavelmente se qualificaria para tal limitação. Carrol (2015) afirma que a Diretiva limita consideravelmente a possibilidade de um proprietário de base de dados usar termos de uso ou outras formas de acordo contratual para adicionar restrições de uso que excedam as previstas na Diretiva.

6.3 Propriedade intelectual na Espanha

A principal norma nacional que regulamenta a propriedade intelectual na Espanha é o *Texto Refundido de la Ley de Propiedad Intelectual, aprobado por Real Decreto Legislativo 1/1996 de 12 de abril*, que sofreu várias modificações, dentre elas, pela Lei 23/2006 de 07 de julho, a fim de adaptar os regulamentos espanhóis às novas circunstâncias criadas pela sociedade da informação (UC3M, [201-a]). Ademais, são consideradas as *Directivas Comunitarias* sobre o assunto, especialmente a *Directiva*

2001/29/CE sobre direitos autorais e direitos conexos na Sociedade da Informação. São consideradas, também, as normas internacionais: Tratado de Berna e os Tratados da OMPI de 20 de dezembro de 1996, sobre direitos de autor e direitos conexos. Ao longo dos anos, foram sendo implementadas resoluções e outros instrumentos para atualização da Lei e para atender as necessidades da sociedade da informação, inclusive relacionada aos dados de pesquisa (QUADRO 9).

Quadro 9 – Documentos relacionados à Propriedade Intelectual na Espanha
(continua)

DATA	DOCUMENTO	SUMA
1996 12/04	Ley de Propiedad Intelectual Española (Vigente y Actualizada) ¹⁰	Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el Texto Refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia.
2001	Derechos de autor y derechos afines a los derechos de autor en la sociedad de la información ¹¹ Directiva 2001/29/CE	relativa a la armonización de determinados aspectos de los derechos de autor y derechos afines a los derechos de autor en la sociedad de la información
2006	Ley 23/2006, de 7 de julio, por la que se modifica el texto refundido de la Ley de Propiedad Intelectual ¹²	aprobado por el Real Decreto Legislativo 1/1996, de 12 de abril
2012	La conservación y reutilización de los datos científicos en España ¹³	Informe elaborado por el Grupo de Trabajo de “Depósito y Gestión de datos en Acceso Abierto” del proyecto RECOLECTA – Informe del Grupo de trabajo de buenas prácticas
2015 09/12	Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on ensuring the cross-border portability of online content services in the internal market ¹⁴	Directiva que armoniza los derechos de autor en la sociedad de la información data del año 2001.

¹⁰ <https://www.leydepropiedadintelectual.es/>

¹¹ <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=LEGISSUM:I26053>

¹² <https://www.boe.es/buscar/doc.php?id=BOE-A-2006-12308>

¹³ https://recolecta.fecyt.es/sites/default/files/contenido/documentos/informe_datos_cientificos_en_esp.pdf

¹⁴ https://ec.europa.eu/commission/presscorner/detail/es/IP_15_6261

(conclusão)		
2016 19/01	Resolución del Parlamento Europeo, de 19 de enero de 2016, sobre la iniciativa «Hacia un Acta del Mercado Único Digital» (2015/2147(INI)) ¹⁵	El Informe del Parlamento Europeo sobre la reforma del copyright.
2016 14/09	Propuesta de DIRECTIVA DEL PARLAMENTO EUROPEO Y DEL CONSEJO sobre los derechos de autor en el mercado único digital ¹⁶	La Comisión Europea propone una Directiva para reformar el copyright en la UE. Presentada el 14.09.2016.
2019 17/04	Directiva (UE) 2019/790 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, sobre los derechos de autor y derechos afines en el mercado único digital y por la que se modifican las Directivas 96/9/CE y 2001/29/CE. ¹⁷	Directiva Copyright de la Unión Europea.

Fonte: Elaborado pela autora.

Além da *Ley de Propiedad Intelectual*, são consideradas outras normas como as que regulam a prestação de serviços na sociedade da informação, sendo adaptada para ambientes digitais (UC3M, [201-a]). Em 2019, foi publicada a *Directiva (UE) 2019/790 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, sobre los derechos de autor y derechos afines en el mercado único digital y por la que se modifican las Directivas 96/9/CE y 2001/29/CE*.

O Informe elaborado pelo *Grupo de Trabajo de “Depósito y Gestión de datos en Acceso Abierto” del proyecto RECOLECTA* (2012) ao abordar os aspectos da produção, da gestão e da reutilização dos dados de pesquisa destaca que as coleções de dados e bancos de dados são protegidas pela Lei de Propriedade Intelectual do *Texto Refundido de la Ley de Propiedad Intelectual* segundo seu artigo 12 - *Colecciones. Bases de datos*, através do chamado direito *sui generis*, na medida em que constituem criações intelectuais. O mesmo Informe traz em seu teor que

As coleções de dados e os bancos de dados são protegidos por propriedade intelectual, de acordo com o art. 12 da TRLPI¹⁸ mencionados por meio do chamado direito *sui generis*, na medida em que constituem criações intelectuais. A proteção se refere apenas à sua estrutura como uma forma de expressão da seleção ou organização de conteúdo e não aos dados em si. Os direitos autorais pertencem a seus criadores, desde que sejam obras originais.”

¹⁵ https://www.europarl.europa.eu/doceo/document/TA-8-2016-0009_ES.html?redirect

¹⁶ <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A52016PC0593>

¹⁷ <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32019L0790&from=ES>

¹⁸ Decreto Legislativo Real 1/1996, de 12 de abril, que aprova o texto consolidado da Lei de Propriedade Intelectual, regularizando, esclarecendo e harmonizando as disposições legais vigentes sobre o assunto.

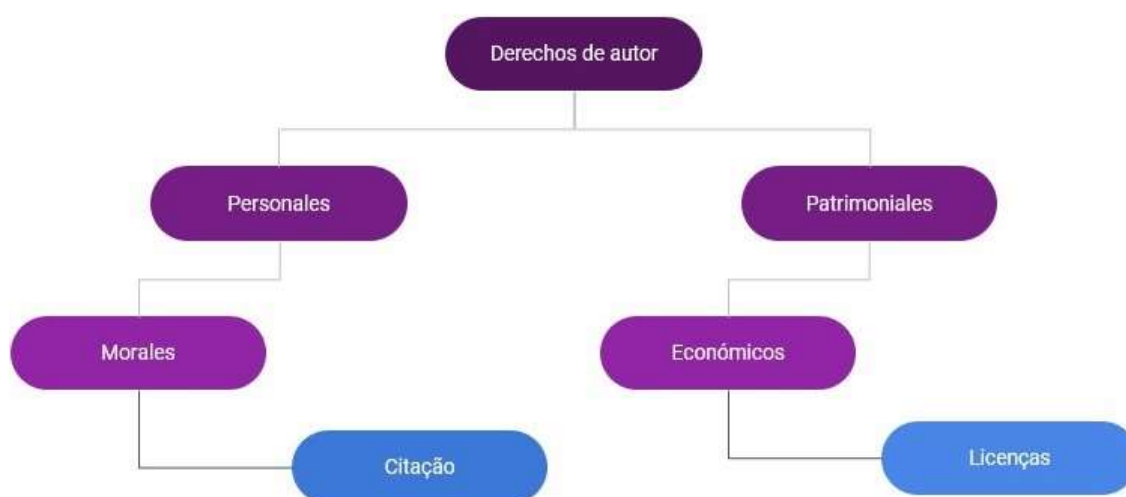
(GRUPO DE TRABAJO..., 2012, p. 21, tradução nossa).

Assim, as coleções de dados e bases de dados são protegidas pela propriedade intelectual e se distinguem entre:

- Direitos autorais: pertencem aos criadores, desde que sejam obras originais.
- Direitos morais: direitos pessoais, pertencem exclusivamente aos autores e não podem ser renunciados.
- Direitos de exploração ou *copyright*: direitos patrimoniais e transferíveis (GRUPO DE TRABAJO..., 2012).

A propriedade intelectual aplicada sobre os dados de pesquisa, no contexto dos repositórios de dados e da Ciência aberta, podem ser explicitados através das citações e da aplicação de licenças de uso (FIGURA 19)

Figura 18 - Propriedade intelectual de dados de pesquisa na Espanha



Fonte: Elaborada pela autora.

Os repositórios de dados, contemplados na *Ley de Propiedad Intelectual* da Espanha como as bases de dados, podem aplicar contratos e termos de uso para a reutilização dos dados nele armazenados. Carrol (2015) destaca que os contratos se

sobrepõem sobre todos os direitos apresentados e podem ser usados com o intuito de fornecer permissão para reutilização dos dados por meio do licenciamento de direitos subjacentes ou, por outro lado, restringi-los.

O contrato é o instrumento jurídico que regulamenta o acordo entre o titular dos direitos e o usuário. Assim, estão definidos os direitos e as obrigações das partes envolvidas em relação ao uso e exploração dos conjuntos de dados visando sempre o melhor desenvolvimento e aproveitamento dos mesmos (MONTEIRO, 2017).

Os repositórios de dados, geralmente, dispõem de termos de depósito ou contratos para os detentores dos dados e solicitam aos autores a aplicação de licenças de uso tanto a *Creative Commons* quanto a *Open Data Commons*. Além disso, os repositórios podem sugerir que os autores cedam quaisquer direitos de propriedade intelectual que possam ter sobre os dados, nesse caso aplicando a licença CC0. Ao depositarem seus conjuntos de dados, os autores concordam com os termos de depósito ou com o contrato do repositório. Da mesma forma, quando o consumidor de dados vai utilizar os conjuntos de dados, ele concorda com os termos de uso do repositório.

A legislação sobre o direito do autor é um recurso infraestrutural das comunicações e da educação, chamada a desempenhar um papel gradualmente importante em vista da progressão e da diversificação dos modos de utilização das obras protegidas pelo direito do autor (UNESCO, 1981). Hodiernamente, as organizações que direta e indiretamente transferem dados requerem modificações do dispositivo jurídico que regulamenta suas transferências. Por conseguinte, a Espanha foi propondo, publicando e aderindo as alterações nos seus dispositivos legais conforme demonstrado no quadro 9 e explicitado ao longo da subseção. Os aspectos dos direitos autorais perpassam os pilares com suas categorias e índices do *framework* proposto na Tese e são apresentados e relacionados com cada um dos índices na próxima subseção.

7 DIREITOS AUTORAIS NAS DIMENSÕES INSTITUCIONAL E PESQUISADOR

Os aspectos relacionados aos direitos autorais de dados perpassam diferentes vertentes. Neste capítulo são apresentados os índices do framework e sua relação com os direitos de autor.

As orientações sobre direitos de propriedade intelectual contribuem para direcionar como os dados podem ser disponibilizados, processados, divulgados e reutilizados.

7.1 Dimensão Institucional

Com base no Modelo para o comportamento de compartilhamento de dados, na Dimensão Institucional são apresentados as três categorias e os sete índices com potencial de estimular ou inibir o compartilhamento dos dados.

7.1.1 Cultura e o direito autoral

Dados de pesquisa fazem parte do cotidiano do processo de pesquisa e, a necessidade de acesso e compartilhamento de dados é reconhecida e evidenciada em muitos projetos e nas colaborações científicas internacionais, e a sua reutilização proporciona benefícios à comunidade científica (FITZGERALD; FITZGERALD; PAPPALARDO, 2011; PINFIELD; COX; SMITH, 2014).

A gestão de dados é um pré-requisito para a eficácia no compartilhamento de dados entre as diversas comunidades científica e contribui, inclusive, com comunidades que não integram o grupo alvo da pesquisa, mas que podem ter benefícios com a reutilização dos dados (MONTEIRO; SANT'ANA, 2018).

Para Mueller (2000, p. 25) os resultados das pesquisas alcançados por determinados pesquisadores “[...] são freqüentemente retomados por outros cientistas, teóricos ou aplicadores, que dão continuidade ao estudo, fazendo avançar a ciência ou produzindo tecnologias ou produtos neles baseado”. O que foi pontuado por Muller está fortemente recomendado para os dados de pesquisa. Percebe-se que muitos dos resultados de pesquisas são enviados para a publicação em periódicos científicos para que sejam legitimados por seus pares e utilizados em novas pesquisas,

uma vez que a publicação traz êxito e prestígio para o pesquisador e para a instituição que ele atua. O mesmo é observado para os dados de pesquisa, em que os pesquisadores tem a opção de enviar seus conjuntos de dados para publicações ampliadas, *datapapers* ou repositórios de dados.

Culturalmente, a comunidade acadêmica reconhece os autores das publicações com suas respectivas citações. Os pesquisadores também precisam expandir sua cultura, até então focada no reconhecimento de autoria de itens documentários, para essa cultura baseada em dados e, assim, reconhecer os autores dos dados e citá-los quando os dados forem reutilizados ou consultados. Essa perspectiva vai ao encontro das observações de Scott (1995) ao considerar a cultura como um entendimento internalizado de cada ator a partir de sua interpretação da realidade social em que participa juntamente com o sistema cultural atuante. A cultura dos atores do contexto científico é impactada pelas demandas de mudanças nas práticas de pesquisas direcionando para práticas abertas.

O tópico Cultura está articulado com os outros tópicos apresentados nas subseções.

7.1.2 Ciência Aberta e os aspectos do direito autoral

Os dados de pesquisa são heterogêneos e diferentes nas diversas disciplinas. Os dados podem ser efêmeros e multifacetados, estar em estado bruto ou com certo nível de análise, ser numéricos, descritivos ou visuais, ser fatos, observações ou experiências nos quais o argumento, teoria ou prova é baseado e devem, portanto, ser considerados como o conjunto de possibilidades emergentes, materialidades e ideias em constante mudança e constantemente reavaliados que encontram incorporação na obra de arte final (UNIVERSITY OF MELBOURNE, 2009). Os dados de pesquisa são os elementos chave no contexto atual da Ciência Aberta (ABADAL et al., 2018).

A Ciência Aberta é um conceito complexo que exige adoção de novas práticas científicas incluindo a transformação no *modus operandi* da pesquisa, além de mudança no comportamento dos pesquisadores tanto na realização de pesquisas quanto no compartilhamento de seus resultados, incluindo os dados (PONTIKA et al.,

2015). A Ciência Aberta pressupõe a abertura de todo o processo científico maximizando sua transparência e colaboração. Além disso, inclui o conceito de Ciência Cidadã, onde o engajamento da sociedade no processo científico é valorizado (ALBAGLI, 2015). Sendo assim, pressupõe uma série de mudanças de paradigmas que buscam fazer com que a ciência seja um bem comum à sociedade (ALBAGLI, 2015; SANTOS, 2017).

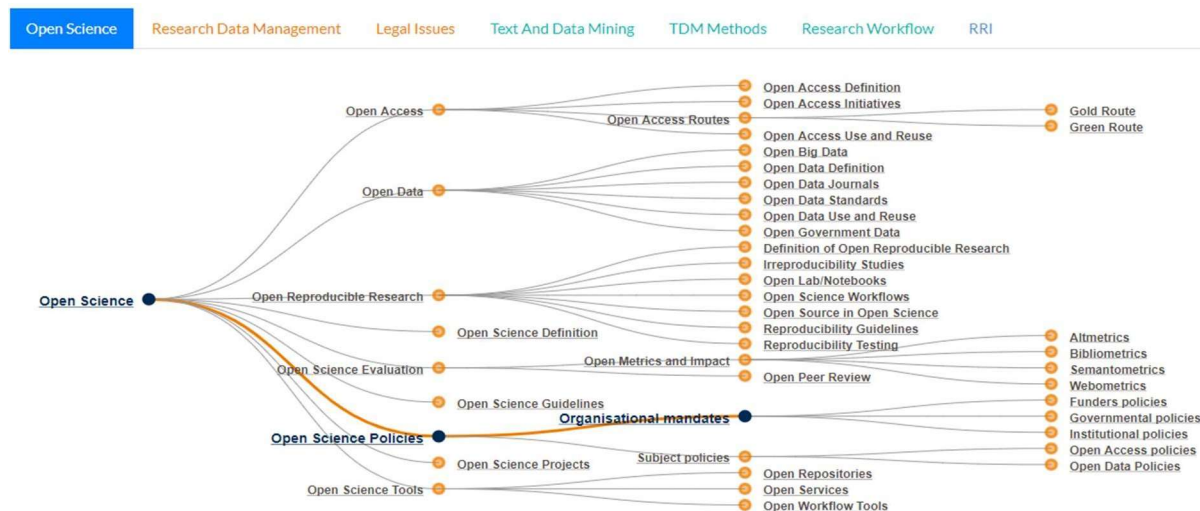
Para a Comissão Europeia, "*Open Science*" é a transformação, a abertura e a democratização da ciência, da pesquisa e da inovação, com o objetivo de tornar a ciência mais eficiente, transparente e interdisciplinar, de mudar a interação entre ciência e sociedade e de possibilitar a ampliação do impacto social e da inovação (RAMJOUÉ, 2015).

Fostering the practical implementation of Open Science in Horizon 2020 and beyond (FOSTER)² (2018, tradução nossa) complementa o entendimento sobre Ciência aberta e afirma que

[...] a prática da ciência de tal forma que outros possam colaborar e contribuir, onde dados de pesquisa, notas de laboratório e outros processos de pesquisa estão disponíveis gratuitamente, em termos que permitam a reutilização, redistribuição e reprodução da pesquisa e seus dados e métodos subjacentes.

A Ciência Aberta engloba uma série de práticas e estratégias abertas que envolvem conceitos abertos como acesso aberto às publicações e aos dados, aos *softwares* e às ferramentas de código aberto, aos fluxos de trabalho, à ciência cidadã, aos recursos educacionais, aos métodos abertos e alternativos para apoiar a pesquisa como a revisão por pares de forma aberta, entre outros (PONTIKA et al., 2015) conforme ilustrado na figura 19 e seguindo a *Open Science Taxonomy*.

Figura 19 - Políticas na Open Science

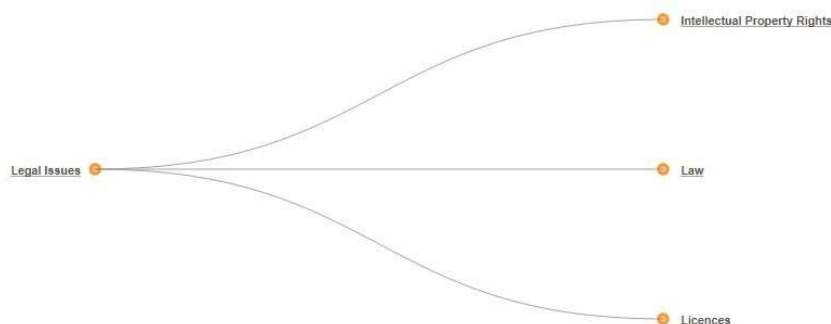


Fonte: FOSTER, [201-].

Evidencia-se na figura 19 que são indicadas as políticas da Ciência Aberta as quais englobam as políticas dos órgãos financiadores, as políticas governamentais e as políticas institucionais que, nesse contexto, tem forte contribuição. As políticas são implementadas de acordo com a realidade de cada comunidade. Cabe destacar na figura 19 o impacto das métricas abertas que estão sendo fortemente recomendadas para os dados de pesquisa.

Oportuno enfatizar que na *Open Science Taxonomy*, na *prática Open Data*, está relacionado o *Open Data use and Reuse* (FIGURA 19). A prática “openness” não implica que os dados publicados abertamente não devam ser citados, não ter seus autores reconhecidos ou liberados sem indicações de licenças de uso. As questões legais também são apresentadas na *Open Science Taxonomy* enfatizando a propriedade intelectual dos dados (FIGURA 20). Um imprescindível componente do sistema operacional está relacionado com as questões legais vinculado ao compartilhamento, reutilização e reprodução de texto, dados, metodologias de pesquisa, práticas e experimentos (PONTIKA et al., 2015).

Figura 20 - Aspectos legais direcionados à dados



Fonte: FOSTER, [201-].

Os principais elementos da Ciência Aberta são apresentados e discutidos no trabalho de Pontika et al. (2015, grifo nosso), com trabalhos de autores que destacam o direito autoral de dados no contexto da Ciência Aberta:

- **Open access:** resultados das pesquisas acadêmicas avaliados pelos pares, acessíveis, sem restrições para a leitura, reutilizáveis, **com direitos autorais e licenças** limitados ou nulos (SUBER, 2012);
- **Open data:** publicação *on-line* disponível para acesso e reutilização dos dados da pesquisa coletados ou gerados ao longo do projeto (MURRAY-RUST, 2008);
- **Open source:** *software* com acesso gratuito, **com uma licença de código fonte** que permite seu uso, criação de derivados e distribuição (ALTENHÖREN, 2005);
- **Open reproducible research:** a prática da Ciência Aberta que permite a reprodutibilidade dos resultados da pesquisa (STODDEN, 2009).

O Grupo de Trabalho da Ciência Aberta da Fiocruz desenvolveu um estudo em que foram sistematizados um mapeamento e uma análise do processo de implantação da Ciência Aberta em oito países e na União Europeia com ênfase nas políticas e na infraestrutura de dados abertos. Neste estudo são apresentados os desafios no processo de abertura e compartilhamento dos dados, apresentado nessa tese com destaque **as disputas por prioridade da descoberta**; a preservação de dados para futuras pesquisas; **a proteção da propriedade intelectual**; a proteção de dados

sensíveis ou confidenciais, as cláusulas contratuais e regulações específicas dos países (SANTOS, 2017, grifo nosso). Mesmo com a abertura e compartilhamento dos dados o fator relacionado aos direitos autorais perpassa as discussões.

Em setembro de 2018, a comunidade científica europeia presenciou uma mudança potencialmente transformadora na política de pesquisa. O *Plano S*, lançado pela *Science Europe*¹⁹, destinou-se a aumentar o acesso aberto (OA) à publicação acadêmica e aos dados de pesquisa resultantes de pesquisas produzidas com financiamento público até 01 de janeiro de 2020 (QUADERI, 2018). Após consulta à comunidade de pesquisa e de revisão de suas diretrizes, o prazo foi estendido até janeiro de 2021 (PELLS, 2019). O *Plano S* requer que os resultados de pesquisas financiadas por organizações signatárias sejam publicados em repositórios abertos ou em periódicos de acesso aberto para sua gestão e disponibilização (QUADERI, 2018).

Interessante destacar que o incentivo a disponibilização e a reutilização de dados de pesquisa está atrelado à indicação do direito autoral. Oportuno destacar que nem todos os dados abertos, especialmente dados de pesquisa, são ou devem ser completamente "abertos" ou "gratuitos", mas devem pelo menos ser FAIR (EUROPEAN COMMISSION, 2018), tópico abordado na próxima subseção.

7.1.3 Princípios FAIR e os aspectos do direito autoral

A gestão de dados de pesquisa está sendo articulada para atender aos princípios conhecidos como FAIR (*Findable, Accessible, Interoperable, Reusable*) divulgados nas comunidades científicas em todo o mundo. O relatório final e plano de ação da Comissão Europeia o *Final Report and Action Plan from the European Commission Expert Group on FAIR Data* apresenta esses princípios com uma visão holística dos requisitos necessários para sua implementação (QUADRO 10).

¹⁹ <https://www.scienceeurope.org/>

Quadro 10 - Princípios FAIR relacionados aos direitos autorais

Princípio	Descrição
Findable	F1- (meta) dados são atribuídos a um identificador globalmente exclusivo e persistente
	F2 - os dados são descritos com metadados ricos (definidos por R1 no princípio Reusable)
	F3 - metadados de forma clara e explícita incluem o identificador dos dados que descreve
	F4 - (meta) dados são registrados ou indexados em um recurso pesquisável
Accessible	A1 -(meta) dados são recuperáveis pelo seu identificador usando um protocolo de comunicação padronizado
	A1.1 - o protocolo é aberto, gratuito e universalmente implementável
	A1.2 - o protocolo permite um procedimento de autenticação e autorização, quando necessário
	A2 - os metadados estão acessíveis, mesmo quando os dados não estão mais disponíveis
Interoperable	I1 -Os (meta) dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento.
	I2 - (meta) usam vocabulários que seguem os princípios do FAIR
	I3 - Os (meta) dados incluem referências qualificadas a outros (meta) dados
Reusable	R1 meta (data) são ricamente descritos com uma pluralidade de atributos precisos e relevantes
	R1.1 - (meta) dados são liberados com uma licença de uso de dados clara e acessível
	R1.2 - (meta) dados estão associados à proveniência detalhada
	R1.3 - (meta) dados atendem aos padrões da comunidade relevantes ao domínio

Fonte: Adaptado de European Commission (2018).

No princípio *Findable*, destaca-se o F1 e F2. Como descrito no relatório acima citado, os dados devem ser registrados com descrição detalhada de metadados para que sejam facilmente localizados, como também “[...] os identificadores persistentes devem ser atribuídos de forma que os dados possam ser referenciados e citados inequivocamente em comunicações de pesquisa.” (EUROPEAN COMMISSION, 2018, p. 19, tradução nossa). Em relação ao princípio *Reusable*, destaca-se R1.1, a reutilização requer que os dados sejam liberados com uma licença de uso de dados clara e acessível (EUROPEAN COMMISSION, 2018).

Os princípios destacados do quadro 10 vão ao encontro dos benefícios considerados no relatório da Comissão Europeia sobre o compartilhamento dos dados via depósito ou publicação e que os autores devem ser reconhecidos e recompensados. No relatório é enfatizada a relevância das métricas de citação, reutilização e impacto dos dados

Métricas e indicadores para contribuições de pesquisa precisam ser reconsiderados e enriquecidos para garantir que funcionem como incentivos convincentes para Open Science e FAIR. Reconhecimento e recompensas eficazes são vitais para a mudança de cultura. (EUROPEAN COMMISSION, 2018, p. 8, tradução nossa).

Os princípios FAIR oferecem um conjunto de princípios e subprincípios para aprimorar a utilidade dos dados; marca um refinamento importante dos conceitos necessários para dar maior valor aos dados e aumentar sua propensão para a reutilização, tanto para seres humanos quanto por máquinas, apoiando a interação entre quem deseja usar os dados e aqueles que os fornecem (KALINAUSKAITĖ, 2017; EUROPEAN COMMISSION, 2018).

O FAIR exige mudanças significativas na cultura dos atores envolvidos na prática de pesquisa além da implementação de um ecossistema com Políticas institucionais, Plano de gestão de dados (PGD), identificadores persistentes, padrões, incluindo padrões de metadados e repositórios de dados, sendo esse último, essencial no ecossistema de dados FAIR pois são necessários para executar a função de oferecer dados e metadados acessíveis e reutilizáveis para usuários e por máquinas (EUROPEAN COMMISSION, 2018).

Os Metadados podem ser definidos como dados sobre dados. Alves (2010, p. 47) define metadados como “[...] dados que descrevem outros dados em um sistema de informação, com o intuito de identificar de forma única uma entidade (recurso informacional) para sua posterior recuperação.” Metadados são dados ou informações descritivas ou contextuais que se referem ou estão associados a outro recurso ou objeto (DIGITAL CURATION CENTRE (DCC), c2004-2015).

Esses metadados, formando um conjunto estruturado de elementos chamado de padrão de metadados, descrevem o recurso e auxilia em sua identificação, localização e recuperação pelos usuários de um sistema de informação, além de formalizar a estrutura do recurso para garantir que os objetivos de uma comunidade de usuários possam ser cumpridos (DCC, c2004-2015).

O PGD, um dos elementos do ecossistema do FAIR, será apresentado na próxima subseção.

7.1.4 Plano de gestão de dados e a indicação do direito autoral

O Plano de Gestão de Dados (PGD) é um documento com informações sobre a gestão do ciclo de vida dos dados e orienta àqueles que estão envolvidos de alguma forma nessa gestão. É recomendável que o PGD seja elaborado no início do projeto de pesquisa e, por ser um documento dinâmico, é possível ser atualizado ao longo do

desenvolvimento do projeto. Sua elaboração envolve diferentes graus de formalidade, extensão, periodicidade, metas e objetivos e seu conteúdo deve considerar as particularidades de cada área do conhecimento. Nele são descritos, essencialmente, quais dados serão coletados ou gerados durante o projeto e qual será seu destino após o término da pesquisa.

No PGD são descritos diferentes atividades e processos relacionados ao ciclo de vida dos dados e envolve “[...] a concepção e criação de dados, armazenamento, segurança, preservação, recuperação, partilha e reutilização, todos tendo em conta as capacidades técnicas, considerações éticas, questões legais e estruturas de governança.” (COX; PINFIELD, 2014, tradução nossa).

O PGD adequadamente elaborado inclui as decisões tomadas no planejamento da pesquisa, contribui com o pesquisador para a indicação de quem será responsável pelos conjuntos de dados e quem terá os direitos de gestão e divulgação e apoia na orientação dos pesquisadores para replicar os dados (DATA..., 2016).

Na elaboração do PGD é imprescindível analisar e planejar o seu conteúdo observando os aspectos como a identificação da titularidade e as licenças aplicadas aos dados, o que contribui para evitar problemas éticos e legais e conflitos resultantes de violações de propriedade intelectual no futuro.

Monteiro e Sant’Ana (2018) destacam que os PGDs relacionam e descrevem os aspectos e as diretrizes que envolvem a gestão dos dados de pesquisa, os quais estão em consonância com as exigências das agências de fomento. Os PGDs não precisam seguir um modelo padronizado, cada um pode seguir um tipo de estrutura e forma de apresentação e, também, seguir a estrutura de uma ferramenta que fornece os *templates*.

A título de exemplo, há ferramentas para elaboração de PGD que orientam os pesquisadores com instruções, incluindo os requisitos necessários para tal como *DMPTool*²⁰, *DMPonline*²¹ e *Argos*²². Nas orientações sobre os requisitos sugeridos apresenta-se questões para serem respondidas e, entre elas, sobre o compartilhamento de dados, ética e conformidade legal, onde destaca-se as questões relacionadas ao direito autoral:

- São necessárias restrições ao compartilhamento de dados?

²⁰ <https://dmptool.org/>

²¹ <https://dmponline.dcc.ac.uk/>

²² <https://argos.openaire.eu/splash/>

- Com quem você compartilhará os dados e em que condições?
- Será necessário um acordo de compartilhamento de dados (ou equivalente)?
- Como os dados serão licenciados para reutilização?
- Como você gerenciará as questões de direitos autorais e direitos de propriedade intelectual (IP / DPI)?

Nas instruções do preenchimento do *template* do PGD é evidenciada a necessidade de declarar quem é o proprietário dos direitos autorais de quaisquer dados e a indicação de (s) licença (s) para reutilização.

Os financiadores de pesquisas, conscientes da necessidade de incentivar boas práticas científicas e alcançar maior transparência e valor para as pesquisas que eles patrocinam, pressionam os pesquisadores a elaborar um Plano de Gestão de Dados para disponibilização de recursos (PINFIELD; COX; SMITH, 2014).

Esse é um documento solicitado e muitas vezes mandatório das agências de fomento, tópico apresentado na próxima subseção.

7.1.5 Agências de fomento e requisitos do direito autoral

Gestão de dados é um aspecto primordial no processo de investigação científica que está gradativamente inserida na rotina dos pesquisadores tanto para atender às exigências de agências de financiamento quanto às revistas científicas para liberação de dados de pesquisa no momento da publicação do artigo (BORGMAN et al., 2015).

As agências de financiamento podem ter requisitos específicos para o compartilhamento dos dados para cada área do conhecimento. Dependendo das orientações do edital, os pesquisadores podem disponibilizar seus dados com aplicação de licenças. Conforme apresentado na subseção do PGD, algumas das orientações dos financiadores englobam os aspectos do direito autoral.

As agências de financiamento, entre elas, *National Science Foundation* (NSF), *National Institutes of Health* (NIH) e NASA dos Estados Unidos, Horizon2020 da Europa, exigem a disponibilização dos dados de pesquisa para financiamento dos projetos de pesquisas, avançando para a realização de ideais de disponibilização pública dos dados, concretizado muitas vezes com o depósito de dados em repositórios de dados, além de solicitar o envio do PGD (BORGMAN et al., 2015;

COUTO CORRÊA, 2016).

Os pesquisadores que vão disponibilizar seus dados e aqueles que vão reutiliza-los demandam orientações sobre as questões de direitos autorais e as licenças de uso. As orientações facilitam aos pesquisadores citarem os dados e, assim, evitar plágio e indicam como os dados podem ser reutilizados. Proteção da propriedade intelectual se apresenta como instrumental importante para o desenvolvimento científico e tecnológico e as agências de fomento estão alinhados a esse princípio.

A exigência das agências de fomento para o compartilhamento dos dados gerados ou coletados no processo de pesquisa é um coadjuvante para a implementação de repositórios de dados de pesquisa e para a gestão dos dados. Tal gestão integra o Ciclo de Vida dos Dados considerando fases como Coleta, Armazenamento, Recuperação e Descarte, perpassando por fatores: Preservação, Disseminação, Direitos Autorais, Qualidade, Integração e Privacidade (SANT'ANA, 2016).

A disponibilização de dados oriundos de pesquisas com financiamento de agências de fomento está indicada em muitos editais e os repositórios de dados estão incumbidos de publica-los. A proteção autoral de dados está evidenciada em diversos repositórios de dados tópico apresentado na próxima subseção.

7.1.6 Repositório e a propriedade intelectual

O avanço científico e a necessidade de arquivamento de dados de pesquisa em repositório de dados mudam consideravelmente a relação de dados e pesquisas. As universidades propiciam meios de disseminar o conjunto de dados de sua comunidade com o apoio dos repositórios de dados que potencializam a disseminação e acesso aos dados e a visibilidade e reconhecimento de seus pesquisadores e da instituição.

Com uma infraestrutura implementada por repositórios de dados, os pesquisadores têm aporte propício para depósito de seus conjuntos de dados e para busca e recuperação de dados já coletados e depositados por outros pesquisadores (MONTEIRO, 2017). As políticas e instruções dos repositórios tem o intuito, entre outras coisas, de amenizar a preocupação relacionada aos direitos autorais.

No entorno da comunidade científica há esforços significativos direcionados ao encorajamento para compartilhar e reutilizar dados de pesquisa, com o objetivo de maximizar a reprodutibilidade e avançar nas descobertas científicas contudo, deve-se ter a segurança que os dados não serão copiados, modificados ou distribuídos ilegalmente. Esse aspecto vai ao encontro dos achados de Sayão e Sales (2016, p. 70, grifo nosso) ao discorrerem sobre dados de pesquisa e curadoria digital onde evidencia que “[...] existe uma preocupação forte com questões éticas, de privacidade e de **propriedade intelectual**. Isso significa que teremos pela frente debates interessantes ao redor dessa questão.”.

A propriedade intelectual é um componente expressivo no planejamento de gestão de dados. Repositórios de dados que orientam sobre direito autoral cooperam para que todas as informações sobre obrigações legais e éticas sejam apresentadas em relação aos conjuntos de dados nele armazenados e favorecem o compartilhamento de conjuntos de dados visto que este aspecto é persuasivo e documentado (MONTEIRO, 2017).

As orientações dos repositórios de dados estão alinhadas com as Leis e normas institucionais, tópico da próxima subseção.

7.1.7 Leis e Normas referentes ao direito autoral

O domínio da comunicação científica é complexo e, nos últimos anos, uma discussão adicional na comunidade científica é o reconhecimento e a atribuição da autoria sobre os dados que estão entorno dos fluxos das pesquisas.

Com efeito, é notório a importância de Leis e normas que orientem e regularizem os aspectos da propriedade intelectual para que a comunidade científica disponha de um sistema que gere confiança no compartilhamento e reuso dos dados e, assim, contribuir para assegurar a fluidez nos processos da comunicação científica, a fim de minimizar o receio pela perda da originalidade nos seus estudos, possibilidades de plágio e uso indevido dos dados. Destaca-se que “A segurança proporcionada pelas leis de Propriedade intelectual estimula a inovação e o desenvolvimento de novas soluções para os problemas existentes. “ (GUIA..., [20--]).

Nas discussões de Scott (1995, p. 35, tradução nossa) é apresentado que as instituições restringem e regularizam o comportamento, e menciona que

“[...] os processos reguladores envolvem a capacidade de estabelecer regras, inspecionar ou revisar a conformidade dos outros a elas e, conforme necessário, manipular sanções - recompensas ou punições - na tentativa de influenciar o comportamento futuro.

As Leis, normas e regras são conscientemente elaboradas para dar suporte às ações instrumentalmente racionais; constituem a natureza da realidade e as estruturas por meio das quais o sentido é feito e especificam como as ações devem ser feitas (SCOTT, 1995).

Os repositórios de dados representam um elemento com forte potencial de contribuição na gestão de dados. Os diversos requisitos funcionais desses repositórios podem ajudar na gestão do ciclo de vida dos dados de pesquisa e no apoio técnico aos pesquisadores. Os repositórios disponibilizam produtos e serviços embasados nas leis e normativas nos quais mencionam, também, instruções sobre direito autoral dos dados orientando os pesquisadores sobre como e quais licenças poderão ser aplicadas, e para disponibilizar seus conjuntos de dados para reutilização por outros pesquisadores.

A atribuição pode ser ampla na qual permite ao usuário fazer maior uso de seus dados sem restrições ou uma licença mais restrita que limita o uso. A escolha de uma licença pode depender do tipo de dados disponibilizados. A indicação que os autores serão citados, com demonstração de formas padronizadas de referenciar os dados também está presentes nas instruções aos pesquisadores.

Os pesquisadores não podem esperar compreender o comportamento social sem considerar os significados que medeiam a ação social; eles não respondem mecanicamente aos estímulos; eles primeiro os interpretam e, em seguida, moldam sua resposta (SCOTT, 1995). Os pesquisadores atribuem significados aos eventos em sua volta, mas não respondem mecanicamente aos estímulos ou fatores inibidores, primeiro eles os interpretam e, em seguida, moldam sua ação e comportamentos. Esses eventos serão apresentados e discutidos nas subseções a seguir.

7.2 Dimensão Pesquisador

Com base no Modelo para o comportamento de compartilhamento de dados, na dimensão do pesquisador são apresentadas as três categorias e os seis índices com

potencial de estimular ou inibir o compartilhamento dos dados.

7.2.1 Benefícios relacionados ao direito autoral

A gestão de conjuntos de dados em seu ciclo de vida é um processo ativo pelo qual é possível torna-los detectáveis, acessíveis, interoperáveis e reutilizáveis, com o potencial de agregação de valor à pesquisa sendo estimulado o seu reuso entre os pesquisadores. Na ambiência da investigação científica, tal processo de agregação de valor é um contribuinte significativo para a realização tão desejada de impacto (PRYOR, 2012).

Os benefícios percebidos na carreira do pesquisador são concernentes ao impacto da qualidade do trabalho apresentado, obtendo mais citações e créditos. Para o pesquisador que tem em seu trabalho a rotineira avaliação da qualidade de desempenho, é frequente sua busca em melhorar seu *status* profissional e seu reconhecimento perante seus pares, explicitado nos agradecimentos e citações, pois o compartilhamento de seus dados propicia oportunidades para ser reconhecido como o autor, ser citado por membros mais proeminentes de seu campo de atuação e receber o reconhecimento formal (KIM; STANTON, 2013, 2016). Da mesma forma, a divulgação de dados fora do contexto científico impacta positivamente nas carreiras dos profissionais propiciando visibilidade inclusive para sua instituição (MONTEIRO et al., 2018).

Com a replicação dos dados por terceiros é possível constituir parcerias e gerar novas publicações e, conseqüentemente, aumentar o número de citações e das métricas. Esse aspecto está relacionado ao direito moral e patrimonial no tocante a aplicação das licenças e citação da autoria pois prevê o direito do autor sobre sua criação.

O compartilhamento ou publicação de dados oferecem outras vantagens para os cientistas pois, quando os pesquisadores publicam seus dados, o seu trabalho pode ser compartilhado entre as fronteiras regionais, nacionais e internacionais (TENOPIR et al., 2015).

Como os pesquisadores consideram o reconhecimento e a reputação valiosos para suas carreiras, suas percepções sobre os benefícios do compartilhamento de dados aumentariam suas atitudes positivas em relação ao compartilhamento de dados (KIM; ZHANG, 2015).

7.2.2 Riscos e o direito autoral

Os riscos percebidos na carreira são definidos como a crença de um pesquisador sobre a possibilidade de receber uma consequência indesejável com o compartilhamento de dados (KIM; ZHANG, 2015).

Ao contrário do índice *Benefício*, o pesquisador pode considerar que ao compartilhar seus conjuntos de dados não será citado ou reconhecido, perdendo o reconhecimento sobre o resultado do seu trabalho, assim como terceiros poderiam reutilizar os conjuntos de dados como sendo seus e assinar a autoria sobre os mesmos.

Em segundo lugar, poderia considerar o compartilhamento como perda de oportunidades de novas publicações (outros publicarem resultados embasados em seus dados antes, e ele ter que citar ao invés de ser citado). Ademais, o compartilhamento aberto dos dados pode comprometer novas publicações em periódicos comerciais que solicitam os dados juntamente com o artigo.

À continuação, o índice também está relacionado ao direito patrimonial, pois são aplicadas licenças de uso. Nesse aspecto, o pesquisador poderia ter em mente que, apesar de serem aplicadas licenças de uso aos conjuntos de dados, essas podem não ser considerados para o reuso dos dados.

O conjunto de dados é resultado da atividade do pesquisador, assim como um texto científico. Em várias situações, a coleta despense mais trabalho do que o próprio texto científico. A coleta pode envolver simulação, observação, testes, pesquisa de campo e anotações que demandam uma série de variáveis como:

- escolha ou preparação do ambiente;
- espera pelo momento preciso da coleta;
- a articulação dos atores envolvidos (sujeitos humanos ou não humanos);
- a articulação dos fatores envolvidos (instituição, profissionais, equipamentos);
- utilização de equipamentos com alto valor agregado, que podem ser de instituições ou laboratórios.

Um pesquisador e sua equipe que estejam, por exemplo, em um navio em alto mar e utiliza um robô para coleta dos dados no fundo do oceano e, como os dono dos equipamentos são os “detentores” desses dados, esses pesquisadores podem pleitear serem reconhecidos como titulares desses dados e receberem os méritos e o

reconhecimento (autoria). Esse aspecto está relacionado ao direito moral pois prevê o direito do autor sobre sua criação, assim como o pesquisador pode não ser considerado como tal ou, ainda, o autor poderia não ser citado, perdendo o reconhecimento sobre o resultado do seu trabalho.

Da mesma forma, ao disponibilizar seus dados, o autor pode determinar como os dados serão disponibilizados e consumidos. Relacionado ao direito patrimonial, aos dados são aplicadas licenças de uso que podem não ser consideradas ao serem reutilizados, assim como o autor pode perder oportunidades de novas publicações já que resultados embasados em seus dados podem ser publicados antes do trabalho do autor, perdendo a originalidade da descoberta, aspecto associado ao direito moral pois trabalhos publicados com maior rapidez, o autor deve citar ao invés de ser citado.

Portanto, se os pesquisadores consideram o compartilhamento de dados como um risco para suas carreiras, é mais provável que tenham uma atitude negativa em relação ao compartilhamento de dados (KIM; ZHANG, 2015).

7.2.3 Esforços e o direito autoral

A gestão de dados é um conjunto de atividades que envolvem a coleta, armazenamento, gestão, disseminação, compartilhamento e preservação de conjuntos de dados provenientes de pesquisa e preparação da documentação de dados. É uma parte essencial de todos os esforços de pesquisa pois viabiliza o compartilhamento dos dados oferecendo à pesquisadores:

- a possibilidade de replicação e de reprodução dos dados;
- a avaliação das afirmações científicas;
- a verificação dos resultados publicados;
- a possibilidade de realização de análises alternativas, entre outros.

Assim, a coleta, a análise e organização necessitam de diversas variáveis, que demandam tempo e esforços. O resultado do trabalho do pesquisador pode gerar recompensas tanto de reconhecimento com as citações quanto lucrativas, vinculadas ao direito patrimonial.

Apesar de todas as recompensas que um pesquisador possa ter, os mesmos podem não ter habilidades necessárias para a preparação dos dados ao ponto de

compartilhá-los, de modo que outros possam compreendê-los e reutilizá-los, sendo necessários, em muitos casos, treinamentos específicos.

Evidentemente, as tecnologias digitais, cada vez mais robustas, somadas ao crescimento da produção, coleta e uso de dados no âmbito científico, são componentes importantes na forma de se fazer ciência e contribuem para o trabalho do pesquisador, porém, trazem desafios como a necessidade de melhorias nos processos de gestão dos dados com sua captura, análise, modelagem, visualização, preservação e curadoria (GRAY, 2009; MAYER-SCHONBERGER; CUKIER, 2013).

Adicionalmente, os pesquisadores necessitam aprimorar as habilidades inerentes a gestão dos dados o que inclui a documentação para o entendimento e interpretação dos dados. A gestão e a documentação favorecem para uma melhor interpretação e reutilização dos dados por terceiros, bem como novas publicações e citações.

Se os pesquisadores acreditarem que a gestão dos dados e seu compartilhamento exigem seu trabalho e tempo adicionais, é mais provável que tenham atitudes negativas em relação ao compartilhamento de dados (KIM; ZHANG, 2015). Profissionais, como os bibliotecários, podem auxiliá-los nesse aspecto, tópico apresentado na próxima subseção.

7.2.4 Suporte dos bibliotecários

Desde a década de 2000, o acesso aberto aos dados vem ocupando um lugar cada vez mais estratégico nas políticas de pesquisa. Essas questões foram retransmitidas por profissionais intermediários que desenvolveram serviços destinados a apoiar os pesquisadores com recomendações, treinamentos e instruções para a gestão e abertura dos dados (REBOUILLAT, 2019).

Ressalta-se que para os investigadores serem competitivos no Espaço Europeu de Investigação e na utilização de dados de pesquisa deve-se adquirir novas competências e, conseqüentemente, os bibliotecários necessitam de novas habilidades para auxiliar os primeiros nos processos de pesquisas, especialmente na gestão de dados e nos princípios FAIR (HERNÁNDEZ-PÉREZ; MÉNDEZ RODRÍGUEZ, 2018).

Os bibliotecários há tempos atuam nos repositórios institucionais, já estão

acostumados com a dinâmica da organização e gestão dos documentos nos repositórios e sua disponibilização. Os repositórios de dados têm no fluxo de trabalho, as particularidades dos dados e os bibliotecários das universidades estão, gradativamente, sendo preparados para atuarem em tais ambientes.

O bibliotecário “[...] tem o potencial de auxiliar a sociedade “[...] a dispor de infraestrutura intelectual capaz de promover a distribuição do saber acumulado de forma mais equitativa.” (SOUZA, 2004, p. 90). A capacidade operativa do papel social do bibliotecário em atender essas missões, cumulativas e simultaneamente, que faz com que os bibliotecários sejam reconhecidos como prestadores de serviços relevantes. Afirmar diariamente suas habilidades na execução de suas tarefas significa oferecer de forma consciente o melhor resultado que cada usuário pode obter (SOUZA, 2004).

O papel dos bibliotecários é influenciado pelo contexto ao qual está inserido e, no campo científico, observa-se uma metamorfose na maneira como a pesquisa científica é conduzida suscitada pelo avanço tecnológico, fortalecendo o foco e o alcance dos dados de pesquisa, criando desafios para a comunidade de pesquisa na gestão e curadoria dos dados (TENOPIR; BIRCH; ALLARD, 2012; SHAKERI; GRACY, 2014).

7.2.5 Benefícios para a sociedade

O crescimento exponencial da produção e do uso de dados tem promovido debates entre os atores do âmbito científico sobre sua importância, usos, potencialidades, desafios e impactos na sociedade, alterando as práticas dos pesquisadores que atuam paulatinamente em várias disciplinas com pesquisas colaborativas, coletando e analisando grandes quantidades de dados, aplicando ou utilizando métodos computacionais sofisticados e computação de alto desempenho (APPEL, 2014; SANTOS, 2017).

As necessidades da sociedade científica estão sendo alteradas e novos elementos estão sendo adicionados nas práticas científicas. Se antes estávamos na era da informação e do conhecimento, atualmente estamos na era dos dados. Se a informação era reconhecida como recurso, agora os dados também o são. Quem tinha a informação tinha poder, atualmente quem tem, também os dados, tem o poder.

Castells em 1999 abriu debates na comunidade científica com sua afirmação de que a revolução da tecnologia da informação juntamente com a reestruturação do

capitalismo introduziu uma nova forma de sociedade denominada sociedade em rede (CASTELLS, 1999) que evoluiu para sociedade baseada em dados. Aquela tecnologia já redefinia perfis, fazeres, hábitos, formas de interação e novas identidades da sociedade civil e científica e propiciaram mudanças nos ambientes, nas práticas profissionais e na identidade social e profissional (MORIGI; PAVAN, 2003/2004).

Como destacado no índice Suporte dos bibliotecários, desde a década de 2000, o acesso aberto aos dados vem ocupando um lugar cada vez mais estratégico nas políticas de pesquisa e, inclusive com destaque nas emergências sanitárias.

O cenário apresentado, e em especial os dados, contribuem para que se fortaleça a tendência de que a disponibilização de dados de forma aberta contribui na agilidade em respaldar demandas da sociedade com maior rapidez na análise situacional e tomada de decisões.

Momentos de urgências sanitárias, como as pandemias, evidencia a relevância e a contribuição do compartilhamento e disponibilização de dados de pesquisa de forma aberta. Além disso, surgem iniciativas que promovem o arquivamento, a disponibilização e a descoberta de materiais e conjuntos de dados sobre o assunto em repositórios de dados.

Ao considerar a rápida disseminação da doença, o compartilhamento de dados é essencial para a realização de diagnósticos, como também melhorar a análise epidemiológica, o tratamento, a modelagem de pandemia transfronteiriça e desenvolvimento coordenado de políticas internas e entre países (JORGE, 2018). Compartilhar dados torna-se extremamente necessário para cientistas que coletam dados relacionados à vírus ou bactérias, por exemplo, a fim de maximizar os benefícios dos dados, visto que, no cenário de emergência sanitária é um imperativo preservar os dados e disponibilizá-los da maneira mais aberta possível para cada caso.

O surto da doença do Coronavírus (COVID-19) detectada em dezembro de 2019 na China está exigindo esforços de colaboração da comunidade científica internacional que precisam agir com rapidez e compartilhar resultados de suas descobertas em diversos meio de comunicação. Com o intuito de contribuir e propiciar melhores resultados no compartilhamento dos resultados de pesquisa relacionados com a doença de Coronavírus (COVID-19) e SARS-CoV-2 foi criada, a título de exemplo, a comunidade *Coronavirus Disease Research Community - COVID-19* reunindo conteúdos que possam ser relevantes para a comunidade científica que

trabalha com a doença mencionada (ZENODO, 2020).

Advoga-se que a ciência aberta promove o aumento dos estoques de conhecimento público, propiciando não apenas a ampliação dos índices gerais de produtividade científica e de inovação, como também a das taxas de retornos sociais dos investimentos em ciência e tecnologia. Tem-se demonstrado que, historicamente, é no compartilhamento e na abertura à produção coletiva e não individual que melhor se desenvolvem a criatividade e a inovação (ALBAGLI, 2015).

A complexidade dos desafios científicos e a urgência das questões sociais e ambientais que se apresentam às ciências impõem, por sua vez, facilitar a colaboração e o compartilhamento de dados, informações e descobertas.

Nesse contexto, a liberação e abertura dos dados com maior agilidade é necessário para respaldar as respostas para sanar as necessidades apresentadas, porém, pesquisadores e equipes podem ter o interesse de registrar a antecedência da descoberta para serem citados nos resultados publicados quando embasados em seus dados (direito moral). Por outro lado, o intuito e antecedência da descoberta, que pode gerar patentes, pode contribuir para a não liberação e abertura dos dados (direito patrimonial).

A sociedade, em muitos casos, discute ou indica o que é ético em determinados contextos. A ética é apresentada na próxima subseção.

7.2.6 Ética e a citação de dados

Os aspectos éticos da pesquisa estão atrelados aos esforços, aos benefícios e aos riscos na carreira do pesquisador, aos benefícios para a sociedade assim como aos fatores institucionais. O movimento da Ciência aberta advoga pela abertura dos dados, porém, pesquisadores podem não ter o intuito de compartilhar seus dados por motivos variados como disputa pela antecedência de descobertas científicas ou competitividade. Nesse aspecto, destacam-se as afirmações de Albagli em relação à dimensão ética da Ciência aberta. A autora afirma que a ética “[...] diz respeito ao compromisso ético de tornar o trabalho de pesquisa e seus resultados imediatamente disponíveis para utilização e *remix* de outros [...]” (ALBAGLI, 2015, p. 17).

A complexidade dos desafios científicos e a urgência das questões sociais e ambientais que se apresentam às ciências impõem, por sua vez, facilitar a

colaboração e o compartilhamento de dados, informações e descobertas (ALBAGLI, 2015).

A título de exemplo, pesquisadores também podem não compartilhar seus conjuntos de dados por não perceber benefício em sua carreira (citação pelos seus dados), não sentir nenhuma obrigação ética em disponibilizar os dados para outros pesquisadores, perceber que outros terão mais benefícios com seus dados do que ele (mais publicações) ou não atender as normas institucionais.

A ética está inserida nas boas práticas científicas, entre outras coisas, no compartilhamento dos dados, na citação dos autores sempre que os dados forem utilizados, bem como no uso respeitando a licença aplicada.

Uma vez que a prática é legal e aplicável, deve ser um tópico de discussão da comunidade se é ético ou apropriado condicionar o acesso aos dados à concordância de um contrato que impõe restrições de uso de dados que, de outra forma, estão livres de quaisquer direitos de propriedade intelectual (CARROL, 2015).

As informações sobre os direitos autorais no contexto das categorias e índices do framework contextualizaram as temáticas e os resultados das análises dos dados que seguem na próxima seção e são amparados pelas discussões apresentadas.

8 RESULTADOS E DISCUSSÃO

Nesta seção são apresentados, analisados e discutidos os resultados obtidos que retratam as convergências e divergências entre a operacionalização do Repositório *e-cienciaDatos* e as perspectivas e atitudes dos pesquisadores relacionados ao compartilhamento de seus conjuntos de dados com enfoque às questões concernentes ao direito autoral: citação e licenças. A pesquisa abrange três grupos de documentos: 1) as leis, normativas, manuais, diretrizes, audiovisuais, infografias, declarações e informes disponibilizados pelo Consórcio Madroño e por suas bibliotecas; 2) os questionários e as entrevistas aplicados aos bibliotecários que trabalham com os repositórios; 3) questionários e entrevistas aplicados aos pesquisadores que são coordenadores de projetos que depositaram seus conjuntos de dados no repositório.

Os resultados foram obtidos com base na aplicação do método Análise de Conteúdo, com a utilização da técnica Análise Categorical, por meio de Categorias, índices e indicadores definidos a partir dos documentos analisados e demarcados no percurso investigativo.

As subseções seguintes dedicam-se a apresentar os documentos que integraram a Pré-análise e, em seguida, discorrer sobre análise propriamente dita.

8.1 Apresentação dos documentos do Consórcio Madroño

O *Consórcio Madroño* e suas bibliotecas sistematizaram suas atividades, seus produtos e serviços para atender os fatores envolvidos com a Ciência Aberta e para assessorar seus pesquisadores na publicação aberta de seus *datasets*, assim como na elaboração de planos de gestão de dados (PGD). Mediante a aplicação da Pré-análise indicada na figura 2 a escolha dos documentos foi feita a *posteriori* seguindo a regra da pertinência. A análise dos documentos proporcionou a verificação dos princípios relacionados ao Direito de autor norteadores do repositório apresentados no quadro 11. Os documentos analisados revelam um panorama sobre o Direito de Autor e suas diretrizes aos pesquisadores.

Quadro 11 - Documentos relacionados aos aspectos do direito autoral
(continua)

Documentos
23 cosas bibliotecas para los datos de investigacion
10 pasos para elaborar um PGD 2016_ Infografías_-5
2016_ Infografías_-4
2016_ Infografías_-6 cita tus datos
Acesso Abierto
Beneficios de publicar en los repositorios institucionales
ciencia abierta la investigacion y los datos cientificos accesibles y abiertos a todos los ciudadanos - 2016_ Infografías_-4
Como depositar sus datos
como cumplir con los mandatos de acceso abierto
como cumplir con los mandatos sobre gestion y publicacion de datos en horizonte 2022 - 2016_ Infografías_- _2
Como depositar sus datos de investigacion en e-cienciaDatos-guia_gestion_datos_investigacion_2019-04-05.pub_
Cómo investigador debo tener un Plan de Gestión de Datos?
comunicacion de la comision al parlamento europeo
Comunicación De La Comisión Al Parlamento Europeo, Al Consejo, Al Comité Económico Y Social Europeo Y Al Comité De Las Regiones
conserva tus derechos de autor frente a las editoriales
Checklist for a Data Management Plan
Datos de investigación
declaracion del Consorcio Madrono en apoyo a la ciencia abierta a la información académica y científica
Declaración institucional de la Universidad Rey Juan Carlos
declaracion institucional de la URJC sobre ciencia educacion y publicacion abierta DCEPA
Difunde tus datos de investigacion en e-cienciaDatos
Directrices para la Gestión de Datos en Horizonte 2020
Directrices para la Gestión de Datos en Horizonte 2020
directrices_gestion_datos_horizon_2020_es
Directiva 2001/29/CE del Parlamento Europeo y del Consejo
Directiva 2006/115/CE
Directiva (UE) 2019/790 del Parlamento Europeo Y del Consejo
el ciclo de los datos cientificos - 2016_ Infografías_-3
El Informe del Parlamento Europeo sobre la reforma del copyright. Aprobado en Sesión del 19.01.2016 ²³
El Plan de gestión de datos
Expert tour guide on Data Management
Horizon 2020
Infografías sobre Acceso Abierto y datos de investigación
La Comisión Europea propone una Directiva para reformar el copyright en la UE. Presentada el 14.09.2016 ²⁴

²³ <https://www.uc3m.es/biblioteca/es/propiedad-intelectual-industrial/propiedad-intelectual>

²⁴ <https://www.uc3m.es/biblioteca/es/propiedad-intelectual-industrial/propiedad-intelectual>

(conclusão)

La Comisión expone su plan para modernizar las normas sobre derechos de autor de la UE. Comunicación 9.12.2015 ²⁵
La conservación y reutilización de los datos científicos en España: Informe del grupo de trabajo de buenas prácticas
Ley de Propiedad Intelectual de 1996
licença de deposito
Licença de uso
Medicion_acceso_abierto_Consorcio_Madrono_2014_2018
Medición del Acceso Abierto en el Consorcio Madroño (2015-2019)
Normativa Vigente y Principales Recomendaciones
Nueva Directiva sobre derechos de autor (copyright) en la era digital ²⁶
OpenAIRE y Datos de Investigación
Presentaciones y vídeos del Curso “Publicación de datos Científicos” organizado por el Consorcio Madroño y celebrado los días 29 y 30 de noviembre de 2016
PAGODA: CREAR SU PGD
Piloto de datos de investigación en abierto
Policy recommendations for open access to research data
Política para la Gestión de datos de Investigación
Por qué Acceso Abierto a los datos en Horizonte 2020?
Practica_open_science
Propiedad intelectual
Propiedad intelectual: cuestiones generales ²⁷
Propiedad intelectual y datos de investigación ²⁸
Propiedad intelectual y edición comercial ²⁹
Propiedad intelectual y nuevas tecnologías ³⁰
¿Qué cuestiones éticas y legales hay sobre los datos?
¿Qué es el Acceso Abierto?
¿Qué son los datos FAIR? ³¹
RDA-CODATA Legal Interoperability Interest Group. (2016). Legal Interoperability of Research Data: Principles and Implementation Guidelines. Zenodo
Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia ³²
The research data group 2012 v7 final

Fonte: Elaborado pela autora.

O repositório *e-cienciaDatos* promove a adoção de práticas de pesquisa abertas e publicação de dados abertos corroborando com suas Políticas, Declaração da

²⁵ <https://www.uc3m.es/biblioteca/es/propiedad-intelectual-industrial/propiedad-intelectual>

²⁶ <https://www.uc3m.es/biblioteca/es/propiedad-intelectual-industrial/propiedad-intelectual>

²⁷ <https://www.uc3m.es/ss/Satellite/Biblioteca/es/TextoDosColumnas/1371213999139/>

²⁸ <https://www.uc3m.es/biblioteca/es/propiedad-intelectual-industrial/propiedad-intelectual>

²⁹ <https://www.uc3m.es/ss/Satellite/Biblioteca/es/TextoDosColumnas/1371214018427/>

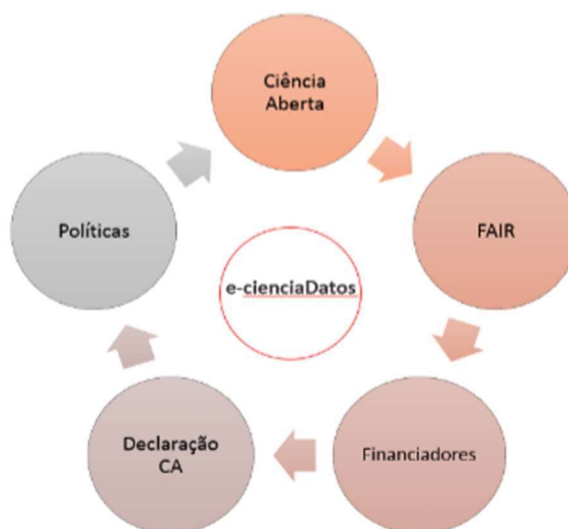
³⁰ <https://www.uc3m.es/ss/Satellite/Biblioteca/es/TextoDosColumnas/1371213998525/>

³¹ <http://www.consorcioamadrono.es/investigam/faq/>

³² https://noticias.juridicas.com/base_datos/Admin/rdleg1-1996.html

Ciência Aberta, Documentos sobre os Financiadores de pesquisa, diretrizes da Ciência aberta e dos princípios FAIR (FIGURA 21).

Figura 21 - Princípios norteadores do repositório *e-cienciaDatos*



Fonte: Elaborada pela autora.

O repositório dispõe de vários documentos que regulamentam o Ciclo de Vida dos Dados (CVD) e que indicam as políticas governamentais e das agências de fomento (FIGURA 21). Os documentos representados na figura 22 são norteadores para elaboração dos outros documentos e manuais: Políticas - instruções gerais sobre os dados depositados; *Declaración del Consorcio Madroño en apoyo a la Ciencia Abierta a la información académica y científica*; *Política para la Gestión de datos de Investigación*; *Normativa Vigente y Principales Recomendaciones*; *Licencia de uso*; *Licencia de depósito*.

Figura 22 - Documentos norteadores do repositório *e-cienciaDados* para o direito autoral



Fonte: Elaborada pela autora.

Na política “Licença de depósito” está explícito que os dados estão em acesso aberto, a menos que haja razões legais, éticas ou de transferência de tecnologia que o impeça. A política “Licença de uso” esclarece que ela é um instrumento legal através do qual o proprietário dos direitos de propriedade intelectual outorga que terceiros utilizem os conjuntos de dados sem infringir esses direitos. As Políticas indicam as licenças para cessão e reutilização dos conjuntos de dados.

Em relação aos direitos morais o repositório indica o *link* das informações sobre padrões de citação dos conjuntos dos dados e disponibiliza a referência normalizada para cada conjunto de dados deixando claro a importância do reconhecimento de sua autoria.

Nas práticas da Ciência aberta estão evidentes que os dados devam ser unicamente identificáveis e listados na seção de referência de um documento. Tanto o *DataCite*³³ como o ORCID³⁴ suportam isso. O *DataCite* tornou possível o uso generalizado de (DOI), está ligado a uma descrição de metadados do objeto e a um local digital, como um URL, onde todos os detalhes sobre o objeto estão acessíveis, além de ligados aos artigos publicados (TAYLOR et al., 2017). O ORCID identifica

³³ www.datacite.org

³⁴ www.orcid.org

exclusivamente os pesquisadores e os ligam às suas publicações, garantindo que seus trabalhos sejam identificados corretamente (TAYLOR et al., 2017). É possível ligar esses identificadores aos artigos científicos, aos resultados de pesquisa, incluindo equipamentos, experimentos, patentes e conjuntos de dados (TAYLOR et al., 2017).

Tanto o *DataCite* como o ORCID viabilizam o desenvolvimento de métricas de citação. Este acompanhamento claro do trabalho (DOIs) com pesquisadores (ORCIDs) significa que os índices de citação são mantidos, possibilitando a cienciometria (por exemplo, h-index) e altmetria (baseado em mídias sociais) e os autores recebem crédito pelo seu trabalho por meio de citações ou “menções” (TAYLOR et al., 2017).

A gestão e publicação dos conjuntos de dados dos repositórios seguem as normas, diretivas e Lei referentes à propriedade intelectual. Nos documentos analisados constatou-se que os conjuntos de dados dispõem da proteção autoral prevista na lei. Nos treinamentos dos bibliotecários aos usuários a questão dos direitos autorais é apresentada expondo as possibilidades da cessão, da disponibilização e da reutilização dos dados. Interessante destacar que a cessão dos direitos autorais não é exclusiva ao repositório.

8.2 Apresentação das categorias construídas para a análise de dados

Com aplicação do método Análise de Conteúdo e da técnica Análise Categorical foram identificadas seis categorias: Cognitiva, Normativa, Reguladora, Carreira, Recursos e Social. Após, para cada categoria foram definidos índices em um total de treze: Cultural, Ciência aberta, Princípios FAIR, PGD, Agências de Fomento, Repositório, Normas e Leis, Benefícios, Riscos, Esforços, Suporte dos Bibliotecários, Benefícios para sociedade e Ética. Em seguida foram determinados quatro indicadores: Direitos Morais, Direitos Patrimoniais, Licenças de uso e Citação do autor que compuseram o “Framework das perspectivas e atitudes do compartilhamento dos dados” (FIGURA 13).

A definição das categorias, dos índices e dos indicadores foi a *posteriori* (Quadro 12), a partir da análise documental, tendo como base os temas abordados nos documentos analisados e no referencial teórico relacionados ao Direito de autor e

que propiciaram o desenvolvimento do roteiro das perguntas do questionário e da entrevista aplicados aos pesquisadores, com análises enfocando o Direito de Autor. A seleção das categorias e dos índices foram baseados na Teoria Institucional (SCOTT, 1995), Teoria do Comportamento Planejado (AJZEN, 1991) e no Modelo para comportamento de compartilhamento de dados (KIM, 2017).

Quadro 12 - Quadro sinóptico e as respectivas inferências

(continua)

Dimensão	Categoria	Definição para a categoria	Índice	Definição para o índice	Aderência do Índice à Categoria
Institucional	Cognitivo	Presença e interação entre os atores, entendimento internalizado de cada ator, a partir de sua interpretação da realidade social em que participa juntamente com o sistema cultural atuante. Os atores tendem a imitar aqueles que tiveram sucesso (SCOTT, 1995).	Cultural	Aponta a relação e interação entre os atores em nível institucional, nacional e internacional destacando o sistema organizacional e a vida social com direitos, regras e responsabilidades.	A Ciência aberta e os princípios FAIR demandam mudanças na cultura e nas práticas de pesquisa relacionadas, entre outras coisas, ao compartilhamento dos dados de forma aberta e citação dos dados.
Institucional	Normativo	Introdução da dimensão prescritiva, avaliativa e obrigatória na vida social, especificam como seus indicadores e bases devem ser feitos, definindo significados legitimados aos valores e regras adotados (SCOTT, 1995).	Plano de Gestão de Dados (PGD)	Documento solicitado pelas agências de fomento que articula os aspectos da gestão dos dados tanto pelos pesquisadores quanto pelo repositório.	Nos <i>Templates</i> para o seu desenvolvimento há diretrizes sobre os Direitos autorais.

Dimensão	Categoria	Definição para a categoria	Índice	Definição para o índice	Aderência do Índice à Categoria
Institucional	Normativo	Introdução da dimensão prescritiva, avaliativa e obrigatória na vida social, especificam como seus indicadores e bases devem ser feitos, definindo significados e regras adotados (SCOTT, 1995).	Ciência Aberta	Movimento internacional de prática científica com a disponibilização de dados científicos de maneira aberta (MORA, 2009; BEZJAK, 2018).	Engloba a práticas científicas abertas e a mudança do comportamento dos pesquisadores relacionadas, entre outros aspectos, ao compartilhamento de dados de forma aberta.
Institucional	Normativo	Introdução da dimensão prescritiva, avaliativa e obrigatória na vida social, especificam como seus indicadores e bases devem ser feitos, definindo significados legitimados aos valores e regras adotados (SCOTT, 1995).	FAIR	Princípios norteadores em que estabelecem e regulamentam os procedimentos para que os meta(dados) estejam acessíveis e reutilizáveis com os direitos associados e termos de uso.	Em seus princípios são indicados o uso de identificadores para autores indicando elementos para a citação e indicação de licenças de uso indicando as formas para compartilhamento e reutilização.
Institucional	Normativo	Introdução da dimensão prescritiva, avaliativa e obrigatória na vida social, especificam como seus indicadores e bases devem ser feitos, definindo significados legitimados aos valores e regras adotados (SCOTT, 1995).	Agências de fomento	Suas diretrizes vão sendo articuladas de acordo com as necessidades institucionais e sociais, envolvendo o estabelecimento de regras e como devem ser aplicadas; entre seus requisitos, a propriedade intelectual e outros direitos relevantes são reconhecidos e adequadamente indicados.	Em suas políticas há orientação sobre o compartilhamento dos dados incentivando uso de licenças abertas.

Dimensão	Categoria	Definição para a categoria	Índice	Definição para o índice	Aderência do Índice à Categoria
Institucional	Normativo	Introdução da dimensão prescritiva, avaliativa e obrigatória na vida social, especificam como seus indicadores e bases devem ser feitos, definindo significados legitimados aos valores e regras adotados (SCOTT, 1995).	Repositório	Estabelece e implementa políticas e diretrizes relacionadas a gestão dos dados, explicitando os aspectos do direito moral (indicação e modelo de citações) e patrimonial com indicação de uso das licenças.	Estabelece quais e como as licenças devem ser aplicadas aos dados e orienta sobre citação dos autores.
Institucional	Recursos	Recursos disponíveis para os pesquisadores que impactam suas intenções e ações, além de ditar a probabilidade de realização do compartilhamento dos conjuntos de dados (KIM, 2017).	Suporte dos bibliotecários	Papel estratégico dos bibliotecários na gestão dos dados, no trabalho do repositório e nos treinamentos dos pesquisadores.	Bibliotecários oferecem treinamentos e a questão dos direitos autorais é apresentada expondo a relevância da citação e as possibilidades da cessão, da disponibilização e da reutilização dos dados.

Dimensão	Categoria	Definição para a categoria	Índice	Definição para o índice	Aderência do Índice à Categoria
Institucional	Regulatório	Estabelece regras, as inspecionam e estabelecem recompensas ou punições na tentativa de influenciar futuros comportamentos. Em uma perspectiva da realidade institucional, a Instituição assume que seus atores têm interesse natural motivado em fazer suas escolhas com a lógica custo/benefício (SCOTT,1995).	Normas Leis	O entendimento dos pesquisadores relacionado a legislação e normativas da União Europeia e da Espanha relativos à propriedade intelectual das quais o repositório de dados utilizou para desenvolver suas políticas e manuais.	Estabelecem e disciplinam a relação entre os interesses dos autores e dos usuários.
Pesquisador	Carreira	Grau em que um cientista pensa que o compartilhamento de dados de pesquisa causaria resultados que influenciam negativamente ou positivamente em sua carreira incluindo o tempo e energia exigidos com a preparação dos dados para o compartilhamento (KIM; STANTON, 2016).	Riscos	Os riscos percebidos no processo de compartilhamento dos seus dados.	Relacionado ao direito moral pois prevê o direito do autor sobre sua criação que pode não ser considerado, ou autor poderia não ser citado, perdendo o reconhecimento sobre o resultado do seu trabalho; perder oportunidades de novas publicações (publicaram antes e ele ter que citar ao invés de ser citado); relacionado ao direito patrimonial - aos dados são aplicadas licenças de uso que podem não ser consideradas ao serem reutilizados.

Dimensão	Categoria	Definição para a categoria	Índice	Definição para o índice	Aderência do Índice à Categoria
Pesquisador	Carreira	Grau em que um cientista pensa que o compartilhamento de dados de pesquisa causaria resultados que influenciam negativamente ou positivamente em sua carreira incluindo o tempo e energia exigidos com a preparação dos dados para o compartilhamento (KIM; STANTON, 2016).	Benefícios	São os benefícios percebidos e esperados com o compartilhamento decorrentes do trabalho do pesquisador.	Relacionado ao direito moral pois prevê o direito do autor sobre sua criação gerando novas publicações, mais citação, aumento das métricas e parceria em novas pesquisas e publicações.
Pesquisador	Carreira	Grau em que um cientista pensa que o compartilhamento de dados de pesquisa causaria resultados que influenciam negativamente e positivamente em sua carreira incluindo o tempo e energia exigidos com a preparação dos dados para o compartilhamento (KIM; STANTON, 2016).	Esforços	Trabalho e energia despendidos com a gestão e preparação dos dados para o compartilhamento.	Esforço intelectual é reconhecido entre os pares a partir da autoria das publicações.

Dimensão	Categoria	Definição para a categoria	Índice	Definição para o índice	Aderência do Índice à Categoria
Pesquisador	Sociais	Aumenta transparência, dá mais credibilidade às descobertas da pesquisa, fornecendo evidências para apoiar estruturas e decisões políticas e uma fonte para os pesquisadores consultarem quando considerarem como aproveitar os estudos existentes (KIM; ADLER, 2015).	Benefícios para a sociedade	As necessidades de disponibilização dos dados de forma aberta para a agilidade em respaldar determinada demanda da sociedade, rapidez na análise situacional e tomada de decisão.	Os benefícios implicam em abertura dos dados aumentando a transparência da pesquisa, agilizando a reprodutibilidade das pesquisas, qualidade dos produtos e serviços (ativos).
Pesquisador	Sociais	Aumenta a transparência, dá mais credibilidade às descobertas da pesquisa, fornecendo evidências para apoiar estruturas e decisões políticas e uma fonte para os pesquisadores consultarem quando considerarem como aproveitar os estudos existentes. (KIM; ADLER, 2015).	Ética	O conjunto de valores de uma área do conhecimento e seu comportamento apropriado ou impróprio juntamente aos deveres, os princípios e as normas que são compatíveis aos grupos dos pesquisadores em seu contexto.	Compartilhamento deve seguir a ética respeitando as licenças aplicadas aos dados assegurando os interesses da ciência ou da sociedade além da citação dos autores; Cumprimento da indicação da licença para a reutilização

Fonte: Elaborado pela autora

O Quadro 12 apresenta todas as categorias e índices de análise definidas para a presente pesquisa. Conforme ilustrado no quadro 12 seus elementos estão articulando cada dimensão com suas categorias e índices. Para cada Categoria e Índice há uma definição contextualizando e a aderência do Índice à Categoria. As Dimensões Institucional e Pesquisador são as dimensões que sustentam a construção dos índices e dos indicadores conforme apresentado na figura 13.

Após a indicação das categorias, índices e indicadores, foram levantadas questões apresentadas em um quadro (QUADRO 13) que, ao final, resultaram no questionário e na entrevista aplicados aos pesquisadores.

Quadro 13 – Questões e a aderência com o direito autoral

(continua)

Índice	Hipótese	Questão	Justificativa para a questão	Aderência da questão com o tema Direito Autoral
Cultural	A cultura influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório	1. Os autores/detentores dos dados tem a preocupação de serem citados	Elementos culturais influenciam o indivíduo a sempre citar os dados	Citação é o ponto chave do direito moral
		2. Os dados são disponibilizados publicamente sem restrições de licenças de uso	porque existem áreas do conhecimento ou urgências sanitárias que impulsionam a cooperação com trocas de dados	na indicação de dados abertos há indicação de licença de domínio público
		3. tem como prática o compartilhamento visando melhores resultados, colaborações e publicações com autoria coletiva	Porque existem disciplinas com forte cultura de compartilhamento e pesquisa coletiva	mais publicações influenciam nas métricas de citação
PGD	Plano de Gestão de Dados influencia positivamente a relação entre a percepção dos pesquisadores e o	1. Em meu PGD indico qual licença é atribuída ao meu conjunto de dados	Porque ainda tem pesquisadores que não sabem quais informações colocar no PGD	PGD fornece informações sobre como os dados serão reutilizados
		2. É imprescindível colocar a informação dos autores no meu PGD para que outros investigadores possam	Porque o PGD está associado aos conjuntos dos dados quando depositado no repositório	PGD fornece informações sobre quem são os detentores dos dados

Índice	Hipótese	Questão	Justificativa para a questão	Aderência da questão com o tema Direito Autoral
	compartilhamento dos dados via repositório	citar os dados		
Ciência Aberta	Ciência Aberta influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório.	1. Conheço suas diretrizes sobre compartilhamento de dados sem restrições de uso	Para identificar se o pesquisador conhece as políticas da ciência aberta	Licença CC0 está relacionado aos direitos patrimoniais
		2. Sou favor de compartilhar dados aplicando a licença CC0 - Declaração de Domínio Público - cedendo todos os direitos de autor	No contexto da Ciência Aberta é indicado que os dados sejam licenciados recorrendo à licença CC0 de domínio público	Licença CC0 está relacionado aos direitos patrimoniais
FAIR	Princípios FAIR influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados	1. conheço seus princípios	Porque está fortemente recomendado que os meta(dados) atendam aos princípios	os dados do FAIR precisam ter uma licença clara, de preferência legível por máquina
		2. Deixa claro como compartilhar dados	porque são reconhecidos por instituições como uma estrutura útil para pensar o compartilhamento e a abertura de dados	
Agências de fomento	As agências de fomento influenciam	1. à agência de fomento indicou o compartilhamento dos	Porque o financiador pode especificar que os dados devem ter uma licença,	Os financiadores garantir que os resultados de projetos com financiamento público sejam

Índice	Hipótese	Questão	Justificativa para a questão	Aderência da questão com o tema Direito Autoral
	positivamente a perspectiva dos pesquisadores fomentando o compartilhamento dos dados via repositório	dados como condição para financiamento? se sim, sobre quais condições?		disponibilizados para reutilização e verificação pela sociedade em geral
		2. Qual foi a indicação do financiador para o compartilhamento de dois dados do projeto?		
		3. concorda com as exigências do financiador?	Porque o pesquisador pode compartilhar para cumprir as exigências do financiador	Algumas políticas de financiadores são mandatórias para o compartilhamento sem restrição de licenças
		4. as agências devem utilizar as métricas de citação de dados	As agências podem analisar o ranque dos pesquisadores e podem ranquear as métricas de citação de dados	Métricas de citação estão relacionadas ao direito moral auxiliando na visibilidade do pesquisador
Repositório	Repositório influencia positivamente o compartilhamento dos dados	1. conheço as suas diretrizes sobre às licenças de uso	1. Porque precisa seguir as diretrizes sobre as licenças aplicadas aos dados do repositório para depositar	As diretrizes norteiam os pesquisadores em como e quais licenças podem aplicar sobre os dados
		2. conheço as suas diretrizes sobre citação	Porque o pesquisador pode ter receio de depositar no repositório e não ter indicações para os usuários dos dados sobre quem, como e porque citar	As diretrizes norteiam os pesquisadores em como e porque citar os dados

Índice	Hipótese	Questão	Justificativa para a questão	Aderência da questão com o tema Direito Autoral
	Repositório influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados	3. Os pesquisadores são incentivados a depositar	Porque o repositório está direccionado a sua comunidade	incentivos favorecem o compartilhamento de maior quantidade de dados propiciando maior reutilização e citação
		4. já deposei meus dados	Porque tem áreas do conhecimento que preferem depósito em repositórios temáticos	O depósito disponibiliza os dados com licenças e indica as formas de citação
		5. Vou depositar meus dados	Porque o pesquisador pode ter receio de não ser citado	Relacionado ao direito moral
		6. me preocupo com a violação dos direitos autorais	O depósito no repositório e mesmo a aplicação de uma licença não garante que os pesquisadores não façam mau uso dos dados ou que os autores serão citados	Relacionado ao direito moral e patrimonial
		7. não deposito porque tem pouca visibilidade	Porque tem áreas do conhecimento que preferem depósito em repositórios temáticos pois é o local com maior probabilidade de os pares depositarem e buscarem dados	Maior visibilidade favorece maior impacto e citação
Suporte do bibliotecário	Suporte dos Bibliotecários influencia positivamente no compartilhamento dos dados via repositório	1. oferecem treinamentos ou orientações sobre gestão de dados .	Porque os bibliotecários fazem a gestão dos dados no repositório, são disseminadores de dados e dão as orientações aos pesquisadores	Promove o entendimento e conscientização sobre aspecto de propriedade intelectual
		2. participei de treinamentos ou orientações que abordaram direito autoral de dados	Porque os treinamentos auxiliam na conscientização	Promove o entendimento e conscientização sobre aspecto de propriedade intelectual

Índice	Hipótese	Questão	Justificativa para a questão	Aderência da questão com o tema Direito Autoral
Normas Leis	Normas e Leis influenciam positivamente o compartilhamento dos dados via repositório	1. conheço toda a legislação e diretrizes nacionais e institucionais	O licenciamento é a forma de estipular quando e como os direitos de proteção nos dados científicos podem ser utilizados por outros	A regulamentação institucional é definida como a nível de pressão coercitiva que cada pesquisador percebe em seus estudos acadêmicos
		2. me sinto seguro pois resguardam os direitos dos detentores dos dados	O direito de autor e o direito Sui Generis das bases de dados (SGDR) surgem automaticamente nos trabalhos qualificados na União Europeia	Proteção do direito moral e patrimonial
		3. resguardam o sigilo dos dados e informações	Durante a gestão dos dados ocorre manuseio de dados sigilosos com interesses econômicos ou de publicações de alto impacto	O sigilo é resguardado com as normas do repositório ou os termos assinados entre pesquisadores e repositório, protegendo o direito moral e patrimonial
		4. as leis e normas viabilizam o desenvolvimento científico e tecnológico	Porque pode haver disputa e competitividade entre pesquisadores ou instituições resultando em novos produtos	Desenvolvimento científico e tecnológico tem resultados de novos produtos com proteção de novas tecnologias recompensados com o direito o patrimonial
Risco	Riscos na carreira: designa os riscos percebidos no processo de compartilhar seus dados	1. não ser citado	Porque há fragilidade da garantia de citação de fontes de dados.	Direito moral prevê a citação dos autores/detentores dos dados
		2. perder oportunidades de novas publicações ou produção de novos produtos	Porque podem ter novas publicações baseados nos dados e impactar negativamente as métricas dos autores	Direito moral prevê a citação das publicações anteriores
		3. fazerem uso dos dados para fins não previstos na licença	Podem fazer mal uso dos dados, usar os dados para fins não previstos e indicados nas licenças	Direito patrimonial - futuras publicações devem seguir a indicação das licenças aplicada aos dados

Índice	Hipótese	Questão	Justificativa para a questão	Aderência da questão com o tema Direito Autoral
		aplicada		
		4. Perder oportunidade de comercialização (patente)	Porque pode haver disputa e competitividade entre pesquisadores ou instituições	Liberação dos dados com ou sem licenças
Benefício	Benefícios na carreira influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório	1. Aumenta minhas citações e métricas	Se beneficiam de crédito acadêmico e progressão na carreira	Métricas de citação estão relacionadas ao direito moral auxiliando na visibilidade do pesquisado
		2. Melhora meu reconhecimento entre os pares	porque a publicação é a melhor forma de obter citação e os pesquisadores tem o intuito de ser reconhecido entre os pares	citação estão relacionadas ao direito moral auxiliando na visibilidade do pesquisado
		3. possibilita colaboração e publicação coletiva	Porque há áreas do conhecimento em que a colaboração é prática visando resultados melhores e publicações robustas	Mais publicações resultando em mais citação
Esforços	Esforços na carreira influenciam negativamente no compartilhamento dos dados via repositório	1. Não tenho habilidades e conhecimentos necessários para gerir e disponibilizar meus dados	As habilidade inerentes a gestão dos dados auxiliam em uma, melhor documentação para o entendimento e interpretação correta	A gestão e documentação favorece melhor interpretação e reutilização dos dados gerando novas publicações e citações
		2. Terei muito trabalho e outros pesquisadores podem ter mais	Pesquisadores podem analisar rapidamente os dados e publicar novos resultados em periódicos com	Melhores resultados podem ter mais citações

Índice	Hipótese	Questão	Justificativa para a questão	Aderência da questão com o tema Direito Autoral
		benefícios como publicação em periódicos com maior fator de impacto	maior visibilidade	
Benefícios para a sociedade	Benefícios na carreira influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório	1. os dados foram disponibilizados durante ou após ter concluído o projeto? Porque?	Em alguns casos, os recursos precisam ser protegidos por senha, apenas para respeitar as restrições legais, de privacidade e éticas.	o compartilhamento sem restrição de licença tem impacto positivo na ciência pela abrangência de futuros usos
		2. Os dados compartilhados podem ser usados para fins comerciais	Porque há áreas do conhecimento marcadas por um mercado lucrativo e com grandes investimentos em pesquisa com foco em patentes	Patentes é um ramo da propriedade intelectual
		3. os detentores de direitos autorais liberem os dados embargados	dados embargados ficam restritos ao compartilhamento entre os detentores dos dados e o compartilhamento entre toda comunidade diminui tempo de análise, diagnóstico, tratamento e tomada de decisão	com o embargo os dados podem estar sujeitos a restrições de leis ou normas da propriedade intelectual
		Justificativa da resposta	Com o embargo os dados não são disponibilizados e podem estar sujeitos a acordos ou direitos estabelecidos	Com o embargo os dados podem estar sujeitos as restrições de contratos
Ética	Ética influencia positivamente no compartilhamento dos dados	1. tenho o dever ético de compartilhar mesmo sem ser citado	Entre os requisitos éticos que se aplicam a gestão de dados científicos está a citação dos dados	Relacionado ao direito moral

Índice	Hipótese	Questão	Justificativa para a questão	Aderência da questão com o tema Direito Autoral
	via repositório			
		2. só compartilho porque é exigido pelo financiador ou pela universidade	Porque a pressão da comunidade científica pode levar os cientistas a compartilharem seus dados com outros pesquisadores	Exigência do financiador pode incluir que os dados estejam com licença CC0 relacionada ao direito patrimonial
		3. só compartilho meus dados se perceber que terei benefícios na carreira	Porque pode haver pesquisador que não percebe nenhum benefício pessoal ou não sentem nenhuma obrigação ética em disponibilizar os dados para outros pesquisadores	Impactos negativos podem surgir a partir da reutilização com publicação de maior impacto gerando a outros pesquisadores mais citações ou recursos financeiros
		4. é ético ou adequado restringir o acesso aos dados que estão livres de direitos de autor é dever ético dos pesquisadores compartilhar os dados em acesso aberto.	Porque há áreas que restringem o acesso devido a disputa e competitividade	Restringir o acesso utilizando embargo ou licença restritiva

Fonte: Elaborado pela autora.

Na próxima subseção são apresentados os resultados das análises das entrevistas com os pesquisadores.

8.3 Análise das entrevistas dos pesquisadores

Nesta seção apresentam-se os resultados obtidos por meio das respostas dos questionários respondidos e das entrevistas realizadas com o objetivo de identificar a percepção e atitude dos pesquisadores sobre compartilhamento de conjuntos de dados. Os pesquisadores fazem parte de Universidades e áreas do conhecimento diferentes fornecendo uma visão holística do contexto analisado.

As entrevistas foram realizadas com os pesquisadores selecionados a partir dos critérios estabelecidos. Os entrevistados receberam um código próprio para preservar suas identidades sendo identificados nesta pesquisa com a letra P seguido de número sequencial, sendo P1, P2, P3, P4. Os pesquisadores são das áreas da Economia, Filologia, Geografia e Linguística. Os projetos coordenados pelos pesquisadores tiveram seus conjuntos de dados depositados no repositório.

Os trechos das falas dos pesquisadores que poderiam revelar suas identidades foram supridos com o recuso [...] ou com intervenção da autora dessa tese para explicação com o recurso [] para não perder o sentido das falas. Em algumas passagens, trechos das falas e explicações, os pesquisadores não foram identificados com P e o seu número pois poderiam ser reveladas suas identidades. Nesses casos foram identificados como “Um dos pesquisadores”.

Os questionários e as entrevistas respondidos pelos pesquisadores obtinham 14 questões fechadas e cinco abertas com foco no direito de autor. As questões estão relacionadas aos assuntos apresentados na tese e no *framework*. Para cada índice foi levantada uma ou duas hipóteses que foram confirmadas ou refutadas de acordo com a análise dos resultados. As justificativas e a aderência das questões com o direito de autor foram apresentadas conforme o quadro 13. Os documentos analisados foram apresentados ao longo dos resultados.

Nas próximas subseções são apresentados os resultados da Dimensão Pesquisador com seus sete índices e Dimensão Pesquisador com seus seis índices.

8.3.1 Cultura

Relacionado ao *Pilar Cognitivo* da Teoria institucional, este índice aponta as relações e interações entre os atores em nível institucional, nacional e internacional, com destaque ao sistema organizacional e a vida social com os direitos, as regras, as responsabilidades, os esforços, os estímulos, as recompensas e as punições (SCOTT, 1995). Destaca-se que a Cultura perpassa por todos os outros índices do *Pilar Normativo* e do *Pilar Carreira*. A análise dos dados indica que o nível institucional está relacionado às universidades que os pesquisadores estão vinculados e ao repositório; o nível nacional está relacionado às outras universidades, aos institutos de pesquisas, ao governo, as Leis e normativas; o nível internacional está relacionado às agências de fomento, às comunidades acadêmicas, científicas e instituições ligadas à pesquisa.

Scott (1995) destaca que em ambientes sociais é considerado natural as pessoas seguirem interesses idealistas ou coletivistas. Para sustentar sua argumentação, o autor discute a visão de sociólogos e antropólogos na vida individual e organizacional. Nesse aspecto, os dados relevam que interesses coletivistas ou idealistas dependem de cada área do conhecimento ou de determinada comunidade. No interesse coletivista, é possível identificar que um pesquisador siga ou, ao menos reconheça, os interesses percebidos em sua comunidade. O aspecto é respaldado pela resposta do P2, quando indagado a dar sua opinião sobre o compartilhamento de dados ao enfatizar as características da sua área de atuação:

Eu acho que para entender um pouco a postura do pesquisador ou a postura coletiva que temos do projeto, tem que esclarecer que viemos de [...] **um campo** de pesquisa que tradicionalmente tem sido **caracterizado por ser bastante proprietário e conservador em relação aos dados**, ou seja, era muito comum coletar os dados e que os dados estivessem disponíveis apenas para quem os tinha coletado, e era muito difícil acessá-los. (P2, grifo nosso).

A influência cultural de uma área do conhecimento denotou ter uma relação significativa com a percepção dos pesquisadores, entretanto, é possível que o pesquisador siga os interesses da comunidade científica em geral, que influencia sua percepção e suas atitudes. Nesse aspecto, P2 menciona que os pesquisadores de seu projeto optaram em seguir o lado oposto praticado pela sua área do conhecimento devido ao fato de o projeto ter sido financiado com recursos públicos. Ademais, o pesquisador chama a atenção para o fato de ser funcionário público e a sociedade

ser, de certa forma, detentora dos dados. No entendimento dos pesquisadores os dados devem estar disponíveis à comunidade:

Nós, quase como uma reação a que optamos por **manter uma posição que será o verdadeiro caminho quase óbvio para o outro lado**, argumentar que se nossa investigação está sendo sustentada por fundos [públicos] e acreditamos que esse é o elemento fundamental em relação aos direitos de autor, somos de certa forma alguns intermediários, mas somos funcionários públicos que também coletamos com a estrutura de um projeto e, portanto, o depositário disso deve ser quem financia a pesquisa, que é o sociedade como um todo. (P2, grifo nosso).

O pesquisador complementa que a postura tradicional de retenção dos dados deve ser excluída das práticas científicas e declara:

[...] é claro que sob nenhuma circunstância devemos estar em aquela situação a qual era tão comum que os dados permanecessem na instituição e deixado ao livre arbítrio do pesquisador responsável concordar em quem acessa [os dados]. **Eu acho que deveríamos excluí-lo totalmente**. Eu acho que nós devemos disponibilizar em aberto. (P2, grifo nosso).

O aspecto vai ao encontro das afirmações de Scott em que “Indivíduos constroem e negociam constantemente a realidade social, mas o faz no contexto de um amplo sistema cultural pré-existente: estruturas simbólicas, percebidas como objetivas e externas, que fornecem orientação e diretrizes.” (SCOTT, 1995, p. 41, tradução nossa).

Oportuno mencionar que a Ciência aberta, os princípios FAIR e as agências de fomento demandam mudanças na cultura e nas práticas de pesquisa relacionadas, entre outras coisas, ao compartilhamento dos dados de forma aberta e ao reuso, incluindo tanto a citação dos dados quanto a aplicação de uma licença de uso.

Nesse sentido, os resultados vão ao encontro do Relatório Final e Plano de Ação Europeu ao ponderar que é preciso formar uma cultura de dados com profissionais capacitados para todo o ciclo de vida dos dados para torna-los FAIR, além de incorporar práticas e habilidades a fim de fornecer e manter serviços de dados (EUROPEAN COMMISSION, 2018). O relatório apresenta um Plano de Ação FAIR com 27 recomendações, sendo 15 recomendações prioritárias. As recomendações se relacionam com os principais conceitos de *Objetos Digitais FAIR* e o ecossistema FAIR, que são implementados por meio de estruturas de interoperabilidade e

mudanças na pesquisa, na cultura, nas habilidades e nas tecnologias, além de incorporação de métricas, incentivos e investimentos para incorporar e sustentar as mudanças (EUROPEAN COMMISSION, 2018).

Os pesquisadores são influenciados pelo seu contexto institucional, nacional e internacional, sendo esse último marcado por iniciativas como *Nube Europea de la Ciencia Abierta* que “[...] visa dar à Europa liderança mundial em infraestruturas de dados científicos, para que os cientistas europeus possam explorar plenamente os benefícios da ciência baseada em dados.” (COMISIÓN EUROPEA, 2016, p. 7, tradução nossa). Para tanto, os pesquisadores necessitam adquirir novas competências e habilidades nos processos de pesquisas, especialmente na gestão de dados e em seu preparo para atender os princípios FAIR (HERNÁNDEZ-PÉREZ; MÉNDEZ RODRÍGUEZ, 2018).

Dentro do marco, os resultados vão de encontro com o estudo de Hedoux (1992) ao discutir sobre construção de identidades sociais e profissionais. Com a associação da identidade com habilidades, o autor pondera que a identidade é construída e reconstruída mediante dois momentos: a) o da atribuição de identidade por instituições e agentes que interagem com o indivíduo, observado nesta Tese, os agentes representados pelas universidades, por instituições governamentais, de financiamento e de pesquisa, por grupos de pesquisadores e por profissionais que direta e indiretamente atuam com os dados; b) a incorporação da identidade por indivíduos, a partir das suas trajetórias sociais que podem ser marcadas pelas características da comunidade científica ao qual o pesquisador faz parte (HEDOUX, 1992).

Tais processos identitários, com novas competências e habilidades, não existem fora de um contexto. De acordo com Michel Agier (2001) é uma relação consigo mesmo e com os outros. O meio urbano multiplica o encontro das pessoas que carregam consigo “[...] seus pertencimentos étnicos, suas origens regionais ou suas redes de relações familiares ou extrafamiliares.” (AGIER, 2001, p. 9). Todos os aspectos tem o potencial de influenciar a cultura dos pesquisadores.

Geralmente o pesquisador se sente pertencente à comunidade na qual se relaciona. Nessa perspectiva, o índice *Cultura* vai ao encontro das observações de Scott (1995) ao considerar a cultura como um entendimento internalizado de cada ator a partir de sua interação com outros atores e sua interpretação da realidade social em que participa juntamente com o sistema cultural atuante.

A cultura atuante em cada área do conhecimento releva as possíveis práticas do compartilhamento ou retenção dos conjuntos de dados antes da finalização das publicações dos resultados das pesquisas, como, também, da prática da citação de dados. Para compreender a cultura do contexto dos pesquisadores a esse respeito, eles foram indagados a dar sua opinião sobre a temática.

Em suas respostas, foram apresentadas que dados com financiamento público devem ser compartilhados em acesso aberto.

O pesquisador P4 responde, sem nenhuma ressalva, que

[...] os dados de pesquisa, eu acho que eles deveriam ser compartilhados abertamente.

O pesquisador P1 concorda com a publicação em acesso aberto, porém com ressalvas demonstrando resistência ao compartilhamento antes da publicação dos resultados das pesquisas, assim como P2 e P3, conforme apresentado:

Eu creio que os dados científicos, principalmente os financiados com recursos públicos, **devem estar em acesso aberto, uma vez publicados os resultados da pesquisa e somente depois**, porque são dados que implica o esforço pessoal do investigador e não pode deixar que outras pessoas utilizem os dados antes do pesquisador e do seu grupo de pesquisa que tiveram seus esforços. (P1, grifo nosso).

Com tal fala, percebe-se que o índice Cultura está alinhado ao índice Esforços apresentado e discutido na seção 8.3.10, assim como a resposta de P3 apresentada abaixo. Ademais, a fala se alinha a do pesquisador P3 que destaca a necessidade de compartilhar os dados de pesquisas com financiamentos públicos e, também, com a ressalva de compartilhar após a publicação dos resultados da análise dos dados. Interessante destacar que em sua fala o pesquisador menciona os riscos do compartilhamento antes da publicação, que se alinha ao índice Riscos, da seção 8.3.9.

[...] bem eu creio que quando a pesquisa é financiada por fundos públicos deveria poder compartilhar para que não seja necessário repetir o trabalho que já foi feito, então em princípio eu estou bastante aberto ao compartilhamento de informações. Naturalmente, precisa haver um determinado período em que essa informação possa ser processada pela pessoa que gerou e que se possa publicar as análises que são necessárias para dar a ele, digamos, algum valor agregado a esses dados.

Compartilhar eu acho importante, mas teremos que esperar um certo tempo até que esta informação seja publicada, porque senão outros poderiam se beneficiar do esforço que você fez ou um grupo. (P3, grifo nosso).

Referente à cultura de citação dos dados de pesquisa, os pesquisadores P1, P2 e P3 estão totalmente de acordo e P4 moderadamente de acordo de que o compartilhamento dos dados permite a colaboração e publicação coletiva, aspecto que pode resultar em aumento de citações e métricas.

Acerca da citação dos conjuntos de dados, os pesquisadores P1 e P4 demonstram pouca preocupação de não serem citados quando outros utilizarem seus conjuntos de dados, aspecto que chama a atenção pelo fato de declarações anteriores ser mencionado que os dados devem ser disponibilizados após a publicação dos resultados. Embora a preocupação da citação sobre o uso dos dados por outros pesquisadores não estejam evidentes, não implica que os autores não devam ser reconhecidos ou citados.

Apesar dos pesquisadores concordarem que os dados devam ser disponibilizados de forma aberta, os mesmos têm a ressalva da liberação ocorrer após o autor fazer todas as análises e publicar os resultados.

Ainda que diversos atores em nível institucional, nacional e internacional advoguem pelo compartilhamento, reuso e acesso aberto aos dados e o *Consórcio Madroño* e suas bibliotecas atuem como disseminadores das boas práticas científicas do compartilhamento dos dados, de fornecer instruções sobre os aspectos da citação de dados, reuso e direitos de autor, nota-se que a cultura ainda não está totalmente favorável ao contexto apresentado. Percebe-se a aderência dos pesquisadores sobre os benefícios propiciados com o compartilhamento dos dados, com destaque para depois das publicações. Ademais, as respostas dos pesquisadores nos outros índices apresentam motivos variados que divergem com a hipótese levantada.

Dessa forma a hipótese: **H1.**: A cultura influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório é refutada.

8.3.2 Ciência aberta

Indicador que integra o pilar normativo. A Ciência Aberta é um conceito complexo que exige adoção de novas práticas e mudança no comportamento dos atores envolvidos no contexto científico. Os resultados obtidos em relação à Ciência aberta demonstram ser o principal índice com forte aderência com os outros índices. Os documentos analisados enfatizam a importância da ciência aberta no contexto europeu e, conseqüentemente, no contexto dos repositórios de dados, de maneira que o *Consórcio Madroño* subscreveu uma declaração em apoio a Ciência aberta e a informação acadêmica e científica. Mediante tal declaração é afirmado que a mesma tem o intuito de impulsionar a adoção de uma série de medidas que facilitem a transição das universidades ao paradigma do acesso aberto. Da mesma forma, outras instituições também subscreveram esse tipo de declaração.

No documento *Declaración del Consorcio Madroño en apoyo a la Ciencia Abierta a la información académica y científica* são enfatizadas as recomendações para a adoção dos princípios da Ciência Aberta, a aplicação das boas práticas seguindo os princípios FAIR considerando as indicações sobre a cessão de direitos autorais às revistas e relevância da citação dos autores (CONSÓRCIO MADROÑO, 2017). Relevante destacar que diversos periódicos solicitam os dados juntamente com o artigo na submissão.

A declaração, em seu item 1: *recomendações aos investigadores*, recomenda realizar boa gestão dos dados de pesquisa e elaborar o PGD, considerando a política de gestão de dados que está anexo ao documento. Nessa política, é informado o valor dos dados de pesquisa e o potencial uso na sociedade, com sua disponibilização em aberto, com menor número possível de restrições (CONSÓRCIO MADROÑO, 2017).

Em relação à propriedade intelectual, a Declaração esclarece que os direitos de propriedade intelectual se definem nos acordos de concessão dos projetos. Ademais, é informado que aos dados se aplicarão licenças abertas como as *Creative Commons*, e que os dados e metadados devem estar em aberto de acordo com as Leis de propriedade intelectual (CONSÓRCIO MADROÑO, 2017).

A infografia da figura 23 da *Red de Bibliotecas Universitarias Españolas* (REBUIN) orienta sobre a Ciência aberta, o acesso aberto, os repositórios abertos e

os dados abertos. Ressalta-se a indicação das métricas e o impacto dos resultados.

Figura 23 – Orientações da Ciência aberta



Fonte: REBUIN (2016a).

Em relação a cessão de direitos autorais às revistas, a infografia abaixo (FIGURA 24) orienta o que o pesquisador deve ter em mente ao enviar seu trabalho para publicação assim como as possibilidades em manter os direitos autorais e de manter uma cópia do documento em aberto.

Figura 24 – Orientação sobre cessão dos direitos autorais para editoras



Fonte: REBUIN (2015).

No atual cenário das publicações em que periódicos solicitam os dados que embasaram o artigo ou para publicação junto ao artigo ou para envio à repositórios de dados, é relevante conhecer os aspectos apresentados, pois podem ocorrer choque de interesses ou de objetivos. Ademais, em muitos casos o pesquisador tem que pagar para a publicação em periódicos comerciais ou não recebem financeiramente pela publicação, inclusive a universidade paga as assinaturas e, quando não paga, a comunidade acadêmica não tem acesso ao artigo. A fala de um dos pesquisadores entrevistados releva a preocupação do acesso não estar aberto para todos mesmo quando os dados estão em tais publicações:

Nas publicações e publicações científicas quando são acompanhados de apêndices de dados nem sempre são acessíveis em aberto e, em muitos casos, foram outros que obrigaram para entrar na própria publicação, o que significa que você tem que pagar ou tua universidade tem que ser assinante daquela revista, que é dizer pagar. Isso é algo que está em forte discussão, o livre acesso. (P1, grifo nosso).

Acerca da citação dos dados, é esclarecido que as normas de citação, assim como os requisitos de publicação, permitirão rastrear as fontes de dados que foram utilizadas e reconhecer a fonte original (CONSÓRCIO MADROÑO, 2017).

A análise dos documentos propiciou pontuar os principais aspectos da Ciência Aberta (CA) relacionados aos direitos patrimoniais apontados no trabalho de Pontika et al. (2015):

- **Open data:** *datasets* estão disponíveis *on-line* para acesso e reutilização. Cada conjunto de dados tem informações sobre a licença aplicada. Ressalta-se que na política do repositório é explicitado que são aplicadas somente licenças livres e, caso o autor requeira aplicar outra licença ou uma combinação de licenças, deve-se entrar em contato com os administradores do *Consórcio Madroño*. A opção de embargo também será aplicada, caso necessário, e com a indicação de prazo;
- **Open source:** é recomendado pelo repositório o uso de formatos abertos, *software* livre e aplicação de licenças livres para *software* aprovadas por *Open Source Initiative* (OSI), *GNU General Public License*, licença MIT entre outras além do uso de recursos educativos abertos. Esses fatores refletem na visualização ou reuso dos dados;
- **Open reproducible research:** nos documentos *Declaración del Consorcio Madroño en apoyo a la Ciencia Abierta a la información académica y científica, Política para la Gestión de datos de Investigación, Normativa Vigente y Principales Recomendaciones* e Políticas do *Consórcio Madroño* estão discriminadas as práticas para a reprodutibilidade aberta da pesquisa.

Após a avaliação dos documentos, é constatado que os vários aspectos da citação dos dados, das licenças de uso e da propriedade intelectual são esclarecidos e divulgados.

Referente ao indicador Ciência Aberta, os pesquisadores foram solicitados a responder a questão: “O quanto você concorda com cada afirmação sobre CIÊNCIA ABERTA:

- (a) Eu conheço suas diretrizes para compartilhar dados sem restrições de uso
- (b) Estou disposto a compartilhar dados aplicando a licença CC0, atribuindo todos os direitos autorais

Quadro 14 – Respostas dos pesquisadores sobre Ciência Aberta

	Pergunta	P1	P2	P3	P4
(a)	Eu conheço suas diretrizes para compartilhar dados sem restrições de uso	Totalmente de acordo	Totalmente de acordo	Moderadamente de acordo	Totalmente de acordo
(b)	Estou disposto a compartilhar dados aplicando a licença CC0, atribuindo todos os direitos autorais	Totalmente de acordo	Totalmente de acordo	Totalmente em desacordo	Totalmente em desacordo

Fonte: Elaborado pela autora.

Apesar das respostas apresentadas no quadro 14, os pesquisadores demonstraram não conhecer a filosofia da Ciência aberta e suas diretrizes de compartilhamento de dados, inclusive o compartilhamento de dados sem restrições, colocado em domínio público. Após apresentar a questão e verificar que os pesquisadores não conheciam a temática, a autora dessa tese fez uma explanação sobre os principais aspectos e explicação do significavam as perguntas. Como resposta foi obtido que:

Eu mentiria se dissesse que eu conheço suas diretrizes, li um par de algumas declarações [...] que eu também tenho uma cabeça muito ruim para nomes, que um é o manifesto, eu não sei agora como é chamado [...] então eu o li em geral, acho, ou a declaração da ciência aberta. (P2, grifo nosso).

Podemos conversar para explicar que é a Ciência aberta, **porque a verdade é que não sei muito bem a que esse termo se refere**. Eu não conheço. (P3, grifo nosso).

É que eu não as conheço. (P4, grifo nosso).

Destaca-se que após a explicação pela autora dessa tese sobre a temática o pesquisador menciona:

Sim, bom sim sabia que existe esse requisito porque os projetos europeus em que eu estou pedem que todos sejam publicados em abertos. **Não conhecia a expressão concretamente**, mas há mais disso, o conteúdo sim, mas já estou acostumado é porque a União Europeia já trabalha há alguns anos. (P3, grifo nosso).

Ainda referente a questão a) Eu conheço suas diretrizes para compartilhar dados sem restrições de uso, o pesquisador responde:

Eu concordo totalmente [em compartilhar dados sem restrições de uso], desde que a propriedade intelectual seja reconhecida claro, a propriedade intelectual tem que ser reconhecida. (P4).

Ao serem indagados se estão dispostos a compartilhar seus dados, aplicando a licença CC0 e cedendo todos seus direitos de autor, P1 e P2 estão totalmente de acordo entretanto, P3 e P4 estão totalmente em desacordo. Destaca-se que P3 não sabia o que abrange todos os direitos de autor, sendo necessário a autora dessa tese explicar a temática:

Não sei muito bem a que se refere todos os direitos autorais. É que não conheço os detalhes do que significa todos os direitos autorais. (P3, grifo nosso).

Após a explanação da autora, o pesquisador responde que não está de acordo e enfatiza a falta de clareza a respeito

Eu não entendo bem o que é isso [...] Claro que acredito nisso [que se deva citar os autores], não concordo, não existe razão [para ceder todos os direitos]. (P3, grifo nosso).

Seguindo as respostas dos outros pesquisadores, P2 faz uma relevante observação sobre a restrição da licença não comercial e até que ponto a Lei espanhola considera comercial:

Acho que sim, que eu estaria disposto de fato. Se não o fizermos, principalmente devido a essa coisa da citação, digamos [...] eu diria apenas que **há [...] muitos cientistas que mantém a restrição, que mantém é a restrição do não comercial**, mas porque parece, em certo modo, como se estivéssemos “contaminados” ou com o dinheiro, no sentido de que nos parece que temos que caricaturá-lo como pecaminoso quanto alguém possa ganhar dinheiro com o que nós geramos para nós. Em particular, devolvemos pouco, me dá um pouco o mesmo, é que se ganhe ou não, e eu o vejo e aí voltamos para a lei espanhola, muitas vezes as pessoas não são conscientes de até que ponto a lei espanhola considera comerciais coisas que para um cientista não o seria, ou seja, se eu produzo algo é apenas uma cópia de uma página web que tem anúncios no google, de acordo com a lei Espanhol essa página web tem

um objeto de lucro que estaria ganhando com meu conteúdo indiretamente e, por tanto, não estaria de acordo com a licença comercial não, então nós não temos nada de problemas com isso e eu acho que se a sociedade científica como um todo apoiasse mais fortemente a liberação para o domínio público nós seríamos os primeiros a ir por essa via. (P2, grifo nosso).

Isso é o que eu tenho na minha área [...] [indicou a sua temática] (P4).

Nesse ponto, a autora da Tese explicou para P4 o significado de se aplicar a licença CC0, enfatizando que os direitos sobre os dados são cedidos, inclusive abdicando de ser citado, pois não há a necessidade de citar os autores dos dados. Com a explicação, P4 corrigiu sua resposta e ficou com dúvidas sob qual licença foi aplicada aos dados depositados no repositório e foi conferir.

Eu não estou de acordo com isso, não concordaria [em aplicar a licença CC0] e não me lembro se fui informado a respeito [pela bibliotecária].

Destaca-se que P4 fez uma pequena confusão com a indicação da CC0, não sabia o que era domínio público. Quando viu a pergunta com CC0 estava de acordo pois não sabia que sedia todos os direitos de autor. Isso pode levar o pesquisador ao erro pois pensa que está apenas deixando os dados livre para o uso. Isso demonstra certa confusão, falta de clareza e pouco esclarecimento sobre as normas do repositório e sobre os aspectos do direito autoral, na aplicação das licenças.

Tendo em vista a operacionalização do repositório, através do *Consórcio Madroño* e suas bibliotecas, alinhando sua documentação, a organização dos trabalhos dos atores institucionais e a divulgação dos aspectos da Ciência aberta, além da operacionalização da União Europeia no assunto, pressupõe-se que os pesquisadores das universidades membro do Consórcio estivessem mais esclarecidos sobre os tópicos da Ciência aberta. Porém, após a declaração dos pesquisadores entrevistados constata-se que há pesquisadores sem o conhecimento prévio, com pouco conhecimento ou esclarecimento.

De forma geral, os pesquisadores se posicionaram a favor da publicação de seus conjuntos de dados de forma aberta, fato comprovado pelo depósito de seus conjuntos de dados no repositório, entretanto, são contrários colocar os conjuntos de dados em domínio público, destacando que, mesmo concordando com a liberação os dados de forma aberta, podem ser aplicadas as licenças indicando a forma de reuso, inclusive

que terceiros devem citar os autores.

Sendo assim, a hipótese **H2.**: Ciência Aberta influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório é refutada.

8.3.3 Princípios FAIR

Dispõe sobre aspectos do pilar normativo com princípios norteadores em que estabelecem e regulamentam os procedimentos para que os meta(dados) estejam encontráveis, acessíveis, interoperáveis e reutilizáveis (FAIR).

Os resultados obtidos mostram que esses princípios estão diretamente relacionados ao indicador *Cultura* pois requerem mudanças consideráveis em termos de cultura e prática de pesquisa entre todos os atores envolvidos. Mediante a análise dos documentos, foi identificadas instruções a respeito dos princípios FAIR.

No que tange a gestão dos dados, o *Plano de ação e relatório final da European Commission Expert Group on FAIR Data* evidencia que, garantir a gestão de dados com o intuito de que se torne uma parte central de todas as práticas de pesquisa é mais um dos elementos críticos da mudança cultural necessária para que os dados se tornem FAIR (EUROPEAN COMMISSION, 2018).

No prisma, é coerente destacar o aperfeiçoamento das habilidades dos pesquisadores para a gestão de dados de pesquisa e, em especial, habilidades em relação ao FAIR. Em virtude do exposto, destaca-se com coerência as considerações de Souza (2004) quando discute habilidades de “profissional”. Na discussão o autor aponta que um profissional possui habilidades intrínsecas a execução de uma determinada função, em uma comunidade e no espaço temporal, sendo reescrita ou definida por razões que vão desde descobertas científicas à novas aplicações tecnológicas, incluindo decisões econômicas e políticas (SOUZA, 2004).

Um aspecto de discussão entre os atores envolvidos é a replicação dos dados pois, como destaca o mesmo relatório citado, uma das maneiras pelas quais grandes investimentos em pesquisa, principalmente em escala global, podem aumentar seu impacto é abordando a questão de como os dados gerados no programa que teve financiamento serão administrados e usados no futuro para replicação, reanálise e integração com novos dados (EUROPEAN COMMISSION, 2018).

No documento *Declaración del Consorcio Madroño en apoyo a la Ciencia Abierta a la información académica y científica* é recomendado às universidades potencializar o uso de boas práticas no manejo dos dados de pesquisa, trazendo como exemplo tornar os dados FAIR (CONSÓRCIO MADROÑO, 2017). Na recomendação há menção para o documento FORCE11³⁵.

Entre os princípios e subprincípios FAIR, ressalta-se os relacionados ao direito de autor. A título de ilustração, apresenta-se no quadro 15 os princípios *Findable* e *Reusable*, princípios aplicados aos conjuntos de dados depositados no *e-cienciaDatos*.

Quadro 15 - Princípios FAIR no *e-cienciaDatos*

Princípios FAIR	Descrição dos Princípios	Metadados	Repositório <i>e-cienciaDatos</i>
Findable	F1- (meta) dados são atribuídos a um identificador globalmente exclusivo e persistente	Title	DOI
	F2- os dados são descritos com metadados ricos (definidos por R1 no princípio Reusable)	Title Author	nomeados autores dos datasets indicação das instituições aos quais estão vinculados
Reusable	R1.1 - (meta) dados são liberados com uma licença de uso de dados clara e acessível	Rights Terms of use	Creative commons CC0, CC-BY, CC-BY-SA Open Data Commons (ODC)
			Public Domain Dedication and Licence (PDDL) Open Source Initiative (OSI) GNU General Public License Versions MIT license

Fonte: Elaborado pela autora.

Mediante a análise sobre os princípios FAIR foi possível identificar os princípios *Findable* e *Reusable* conforme apresentado no quadro 15. Evidencia-se nos princípios FAIR, o seu princípio *Reusable* – “R1.1 - (meta) dados são liberados com uma licença de uso de dados clara e acessível”. Em todos os conjuntos de dados do repositório há indicação das licenças aplicadas.

Relacionado ao princípio *Findable*, - “F1 - (meta) dados são atribuídos a um identificador globalmente exclusivo e persistente” todos os conjuntos de dados tem

³⁵ <https://www.force11.org/group/fairgroup/fairprinciples>

um identificador persistente. No princípio “F2 - os dados são descritos com metadados ricos (definidos por R1 no princípio *Reusable*)” destaca-se que há o metadado com indicação dos nomes dos autores e para o ORCID. Um identificador único e persistente deve ser atribuído de forma que os dados possam ser inequivocamente referenciados e citados em comunicações de pesquisa. Nesse aspecto, a aplicação desses princípios pelo repositório corrobora com as indicações do Relatório Final e Plano de Ação Europeu no que tange a “[...] um identificador único e persistente deve ser atribuído de forma que os dados possam ser inequivocamente referenciados e citados em comunicações de pesquisa.” (EUROPEAN COMMISSION, 2018, p. 19, tradução nossa).

Ademais, em relação ao princípio *Reusable* - R1.1 são aplicadas licenças de uso, também corroborando com o mesmo relatório no que diz respeito a reutilização dos conjuntos de dados, pois indica que os dados sejam liberados com uma licença de uso clara e acessível (EUROPEAN COMMISSION, 2018).

No documento *Declaración del Consorcio Madroño en apoyo a la Ciencia Abierta a la información académica y científica* é recomendado o uso das boas práticas da gestão dos dados para torna-los FAIR (CONSÓRCIO MADROÑO, 2017).

Os entrevistados foram questionados sobre o entendimento dos Princípios FAIR indicando o quanto estavam de acordo com cada afirmação conforme quadro 16.

Quadro 16 – Respostas dos pesquisadores sobre Princípios FAIR

	Pergunta	P1	P2	P3	P4
(a)	Eu conheço seus princípios	Não conheço o tema	Totalmente de acordo	Totalmente de acordo	Não conheço o tema
(b)	Evidência como compartilhar dados	Totalmente de acordo	Totalmente de acordo	Totalmente de acordo	Não conheço o tema

Fonte: Elaborado pela autora.

Os pesquisadores P2 e P3 conhecem o tema e estão totalmente de acordo com a afirmação de que os princípios FAIR evidenciam quem são os autores dos conjuntos dos dados afirmando a aderência com o princípio *Findable*, e igualmente de acordo que os princípios deixam claro qual é a indicação para compartilhar os dados, corroborando com o princípio *Reusable*.

Salienta-se que P2 considera bons princípios e bem claros sobre como compartilhar dados. O pesquisador faz uma relevante observação que pode ser

associada com o princípio *Acessible*:

Mas eu acredito que, certamente sim, são bons princípios sobre como compartilhar os dados, bastante claros. Os dados devem estar acessíveis pois todo conteúdo que serve para a pesquisa deve estar acessível. (P2).

O pesquisador faz uma analogia com a biblioteca na qual podemos salientar a importância da organização da informação e a utilidade que o catálogo tem para localização do material, além do tempo que poderia ser despendido se o usuário tivesse que percorrer por todas as estantes a procura do material que procura. Essa analogia pode ser refletida nos dados: para serem localizáveis precisam, inclusive, de uma rica descrição, conforme apresentado nos princípios FAIR:

Se a biblioteca não tiver catálogo é inútil, não posso ir percorrendo as prateleiras aleatoriamente porque não é tudo que está no acervo que serve para a pesquisa. (P2).

Outro aspecto levantado pelo pesquisador foi sobre a interoperabilidade que, na sua visão, é o maior desafio, havendo muita incompatibilidade entre os formatos.

Obviamente têm de ser acessíveis a partir da interoperabilidade porque [...] também, talvez seja o maior desafio para o futuro; é verdade que há muita incompatibilidade. Pode haver o uso de alguns formatos proprietários, se deveria avançar mais nos formatos, em formatos abertos, e tudo isso. Em nosso caso em questão, nós vamos com formatos padrão em todos os casos e claro, porque é reutilizável. (P2).

Os pesquisadores P1 e P4 declararam que não conheciam o assunto, inclusive P4 perguntou o que significava cada uma das questões relacionadas ao tema. Como ao longo da entrevista a autora da Tese foi dialogando com os pesquisadores, P4 fez a seguinte pergunta:

Atuam como diretrizes para melhorar reutilização de dados?.

Com as declarações, a autora dessa tese fez uma breve explanação sobre o tema para informa-los sobre o assunto. Após a explanação, P1 destaca que são bons princípios e que os dados usados para uma publicação deveriam ser encontráveis e reutilizáveis. O pesquisador P4 também destaca que “parecem ser bons princípios” e

mais uma enfatiza que não os conheciam.

Embora o repositório dispõe de material de instrução e faça a divulgação dos princípios FAIR, ainda existem pesquisadores que não conhecem o tema. Após os esclarecimentos aos pesquisadores sobre o tema, todos concordaram ser bons princípios, principalmente para a indicação de autoria e licença de uso.

Considerando os fatores levantados, a hipótese **H3**: Princípios FAIR influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados é confirmada.

8.3.4 Plano de gestão de dados

Condizente com o *Pilar Normativo*, o PGD é um documento solicitado pelas agências de fomento, com descrição de como os dados serão coletados ou gerados no decorrer de um projeto e como serão tratados e disponibilizados publicamente. No documento é possível articular alguns dos aspectos da gestão dos dados pelos pesquisadores e que contribuem tanto com os bibliotecários que fazem a representação e organização dos conjuntos de dados no repositório quanto com os usuários que vão acessar e reutilizar os mesmos.

O *Consórcio Madroño* disponibiliza manuais e documentos com orientações a respeito e a página *Investigam: PAGODA* com diversas informações sobre o PGD e material de apoio, além de indicar serviço de assessoramento em cada uma das bibliotecas cooperantes. O *Consórcio Madroño* e suas bibliotecas dispõem de orientações sobre os PGDs, com definição, benefícios, requisitos de financiadores, recomendações, indicação de estrutura e indicação da ferramenta DMPOnline, desenvolvida pelo *Digital Curation Center* do Reino Unido, para sua elaboração. A ferramenta foi traduzida para o espanhol, o PaGoDa³⁶, acrônimo de *Plan de Gestión de Datos*.

Entre os resultados obtidos destaca-se que a projeção da elaboração do PGD já está apresentada no documento *Directrices para la gestión de datos en Horizonte 2020* em sua versão 1 de 11 de dezembro de 2013 no programa de trabalho do Horizonte 2020 para o período de 2014/2015.

Nas orientações sobre os benefícios obtidos com a elaboração do PGD

³⁶ Disponível em: <http://www.consorciomadrone.es/investigam/crear-su-pgd/>

destaca-se o direito moral em seu teor “Outros pesquisadores que usam os dados podem citá-los e a pesquisa ganhará mais prestígio.” (CONSÓRCIO MADROÑO, c2018, tradução nossa). Nesse aspecto, o repositório ratifica que os conjuntos de dados devam ser citados, inclusive o prestígio de ser reconhecido pelo trabalho com os dados pelos pares. Ressalta-se que o aspecto também se alinha ao índice Benefícios.

Nas orientações sobre o teor do Plano de Gestão de Dados destaca-se a orientação “Como os dados serão explorados e / ou compartilhados / acessíveis para verificação e reutilização. Caso os dados não estejam disponíveis, é necessário explicar o motivo.” (CONSÓRCIO MADROÑO, c2018, tradução nossa). A orientação indica o possível uso de licenças, que está relacionado com o direito patrimonial.

O documento *Directrices para la Gestión de Datos en Horizonte 2020 Versión 1.0 11 de diciembre de 2013* deixa claro que o PGD

[...] é da maior importância em todos os aspectos para os projetos que participam no Piloto de Datos de Investigación en Abierto de Horizonte 2020, bem como para qualquer outra pesquisa. Se o projeto participar do piloto, deve ter uma descrição clara e fundamentada sobre o regime de acesso a ser fornecido para os conjuntos de dados coletados. (CONSÓRCIO MADROÑO, 2014, p. 4, tradução nossa).

No mesmo documento há referência de que os dados de pesquisa devam estar abertos e destaca-se que esses dados devam estar acessíveis. Para o ponto, o documento indica a questão abaixo para ser respondida, mencionando sobre a utilização de licenças e informando que as mesmas podem ser aplicadas quando necessário:

a. Pergunta PGD: os dados e o *software* associado produzidos e utilizados no projeto são acessíveis?; através de quais modalidades, escopo, licenças (por exemplo, estrutura de licenças para pesquisa e educação, períodos de embargo, exploração comercial, etc.)?” (CONSÓRCIO MADROÑO, 2014, p. 7, tradução nossa).

Referente ao índice Plano de Gestão de Dados, os pesquisadores foram solicitados a responder à questão: “O quanto você concorda com cada afirmação sobre Plano de Gestão de Dados:

(a) É um elemento útil para indicar os autores e usar as licenças

(b) Foi desenvolvido para descrever a gestão de dados

Quadro 17 – Respostas dos pesquisadores sobre PGD

	Pergunta	P1	P2	P3	P4
(a)	É um elemento útil para indicar os autores e usar as licenças		Totalmente de acordo Não sei do tema		
(b)	Foi desenvolvido para descrever o gerenciamento de dados		Totalmente de acordo Moderadamente de acordo Não sei do tema		

Fonte: Elaborado pela autora.

Optou-se em apresentar o quadro 17 sem as respostas individuais para as identidades não serem reveladas conforme as transcrições das respostas.

Referente a afirmativa a) É um elemento útil para indicar quais são os autores dos conjuntos de dados, dois pesquisadores estão totalmente de acordo e dois não conheciam o tema.

Destaca-se a fala de um dos pesquisadores:

É que eu não o conheço, sem dúvida que estamos por fora do assunto [...] ao declarar este plano, eles o pedem para que?”.

Após a explicação pela autora dessa tese sobre o que é um PGD e os aspectos que estão relacionados a sua elaboração, um dos pesquisadores responde que está totalmente de acordo com a pergunta acima e o outro fez o seguinte comentário deixando claro a utilização de licenças:

[...] em nosso caso, termina muito rapidamente, não temos nenhum plano de gestão de dados explícito, digamos nenhum documento nem nada porque não nos exigem, então a impressão é que estamos mais ou menos, concordamos em não elaboramos nada, e a única coisa que temos em uma página web é a referência para a licença que nós colocamos e **nós sempre colocamos**. (Um dos pesquisadores, grifo nosso).

Os quatro pesquisadores relataram que agências de fomento não solicitaram a elaboração do PGD. Apesar de não ter sido elaborado o PGD, um dos pesquisadores menciona que faz falta sua elaboração e destaca que se deve colocar a indicação da

licença. Interessante destacar que para o projeto coordenado por outro pesquisador não foi solicitado a elaboração de PGD entretanto, foi decidido pela sua elaboração.

Bem, eu estaria totalmente de acordo e elaborado para descrever a gestão de dados, o fazem, mas o fazem sem ser completamente sistemáticos. Nos últimos anos, incentivam a introdução da publicação dos dados e introduzir no projeto do financiador, mas se você não o apresenta, não há problema, digamos que não é obrigatório, digamos que se em seu projeto você tiver feito o plano de gestão de dados, ok. Eu fiz isso, mas o fiz por livre [e espontânea vontade], eu sou muito veterano, então eu tenho um pouco de excesso de liberdade. (Um dos pesquisadores).

Na fala do pesquisador entende-se que foi percebido a importância do PGD, por isso foi elaborado e, também, a constatação do incentivo em sua elaboração, fato não percebido pelos outros pesquisadores.

Outro ponto relevante diz respeito ao projeto coordenado por um dos pesquisadores. Para tal projeto não foi solicitada a elaboração de PGD entretanto, foi decidido pela sua elaboração devido ao prestígio agregado ao projeto e aos conjuntos de dados, além de mencionar o incentivo recebido pela universidade. Esse ponto se alinha a dois índices, Benefícios e Suporte dos bibliotecários.

Eu fiz isso porque eu decidi, não porque o disseram, porque foi um projeto de publicação [a área do conhecimento] e **uma das funções também de prestígio profissional** era publicar um website, porque temos um website, e publicado que nos incentivaram muito a fazê-lo desde a própria universidade e que temos na universidade, esta é nossa principal fonte de dados para onde vamos mudando. A cada [período de tempo] temos um plano de mudança nas datas para estudá-lo e mudá-lo, introduzir as novidades, porque nosso caso há novas publicações, pessoas que têm trabalhado, podemos incorporar novos dados que podemos ver o outro é que seja uma base de dados que carregamos e que é deixado lá e fazemos um novo trabalho, mas neste caso, é algo renovável onde ao mesmo tempo, vamos incorporar, por exemplo, agora vamos incorporar [determinados dados] porque há um período em que estamos levando um ano e meio trabalhando em estende-los, digamos, nossos casos, cada caso eu acho que é diferente da maioria dos casos, são feitos uma pesquisa e os dados permanecem anexados à pesquisa e você os coloca em aberto e lá eles ficam no próximo projeto de pesquisa com novos dados. (Um dos pesquisadores, grifo nosso).

Bom para o projeto isso não foi necessário escreve-lo, para os projetos europeus e que eles te pedem mas no fundo levam coisas semelhantes da rastreabilidade, da documentação e da permanência. (Um dos pesquisadores).

[nós temos documentos que] fala sobre o que são os dados em si, se descreve a informação para os que querem usar. No repositório estão os metadados, mas o plano de gestão de dados não está lá. (Um dos pesquisadores).

Não sei se nos recorda agora se isso foi incluído na memória do projeto e claro, embora não esteja incluído, então estamos respeitando os princípios de arquivamento dos dados e, ademais, porque é isso que, digamos, que a agência leva em consideração. (Um dos pesquisadores).

O *Consórcio Madroño* e as bibliotecas fornecem diversas informações sobre o PGD e suporte aos pesquisadores. Mesmo que três pesquisadores não tenham elaborado o PGD, e dois não conheciam o tema, os pesquisadores são favoráveis a sua elaboração mesmo que não foram solicitados pela agência de fomento.

Dessa forma, a hipótese levantada: **H4.**: Plano de Gestão de Dados influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório, pode ser refutada.

8.3.5 Agências de fomento

O índice está vinculado ao Pilar Normativo e fortemente relacionado ao índice PGD. De forma gradual, os financiadores solicitam o PGD junto ao projeto no momento da submissão ou após a aprovação do projeto. Os documentos analisados apontam como principais programas de financiamento o Programa de investigação e inovação da União Europeia (UE) *Horizon 2020*³⁷ e *Horizon Europe*³⁸, e, também, apontam a influência das agências na mudança de cultura dos atores envolvidos.

Suas orientações estão voltadas para disponibilização dos conjuntos de dados de forma aberta maximizando a reutilização. Ressalta-se que na política do *Horizon 2020* os direitos de acesso e a proteção da propriedade intelectual são apresentados.

Os pesquisadores que vão submeter projetos ou já os que já foram beneficiados pelo *Horizonte 2020*, com destaque às propostas apresentadas como “Pesquisa e Inovação” e “Ações de inovação”, devem articular seus projetos e seguir suas

³⁷ <https://ec.europa.eu/programmes/horizon2020/en>

³⁸ https://ec.europa.eu/info/designing-next-research-and-innovation-framework-programme/what-shapes-next-framework-programme_en

diretrizes, incluindo uma seção sobre gestão dos dados de pesquisa. Diversas agências indicam que os projetos que são financiados devam apresentar um PGD refletindo a gestão dos conjuntos de dados.

Ressalta-se que na política do *Horizon 2020* os direitos de acesso e a proteção da propriedade intelectual são apresentados e que as orientações das agências indicam a necessidade de disponibilização dos conjuntos de dados para reuso.

Nas orientações sobre as vantagens em participar do *Horizon 2020* há menção sobre a melhoria da imagem da entidade: visibilidade e prestígio, investigação e inovação, demonstrando o estímulo à participação dos pesquisadores. Nas orientações sobre o *Piloto de Datos de Investigación en Abierto* é indicado o uso da licença Creative Commons (CC-BY o CC0) (CONSÓRCIO MADROÑO, c2018).

Cada pesquisador tem um nível de conhecimento sobre as diretrizes das agências de fomento relacionadas aos dados.

Referente ao indicador Agências de fomento, os pesquisadores foram solicitados a responder a questão: “O quanto você concorda com cada afirmação sobre as AGÊNCIAS DE FOMENTO:

- a) Estou de acordo com as políticas obrigatórias do financiador para compartilhar dados
- b) As agências devem usar métricas de citação de dados

Quadro 17 – Respostas dos pesquisadores sobre Agências de Fomento

Pergunta		P1	P2	P3	P4
(a)	Estou de acordo com as políticas obrigatórias do financiador para compartilhar dados	Totalmente de acordo	Moderadamente em desacordo	Totalmente de acordo	Totalmente de acordo
(b)	As agências devem usar métricas de citação de dados	Totalmente de acordo	Moderadamente de acordo	Totalmente de acordo	Totalmente de acordo

Fonte: Elaborado pela autora.

Os pesquisadores foram indagados o quão estavam de acordo com as políticas obrigatórias sobre partilha de dados do financiador, considerando a liberação com

menos restrição possível. Destaca-se que os projetos por eles liderados receberam financiamento. Os pesquisadores P1, P3, e P4 estão totalmente de acordo e P2 moderadamente em desacordo.

Uma das reflexões apontadas pelo pesquisador P2 envolve diretamente os preceitos da Ciência Aberta, o acesso aberto às publicações científicas e apresenta relevante preocupação em relação as práticas dos periódicos tradicionais de prestígio e de alto fator de impacto, com o sistema de negócio vigente e altamente rentável de pagamento para acesso aos artigos, e o interesse no modelo *Open Science* pois continuam recebendo o pagamento para liberar os artigos. Esse é um tópico que tem gerado muitas discussões na comunidade científica:

Aqui em nosso caso particular, no caso particular da Espanha, o problema que existe é que, todavia, hoje se seguem predominando as publicações e revistas, editores específicos, de editoras concretas, revistas com seus índices, de suas citações de impacto que, especialmente em nossos campos, é verdade que existem nos campos de ciência mais pura, bem está avançando um pouco mais no modelo do *Open Science* porque interessa aos editores, ou seja, [os editores falam] “me sigam pagando o mesmo, para pagar 4.000 e 5.000 euros, que eu libero seus artigos” que eu ganhei com isso?

Na continuação do relato de P2, é revelado um aspecto contundente que não raro, fica em evidência, nos modelos de negócio vigentes dos periódicos científicos inclusive, uma preocupação que recai sobre os pesquisadores pois, nesse modelo, eles pagam para publicar seus artigos e, dependendo da área, não tem público suficiente que justifique pagar um valor alto para tal. O relato evidencia, ainda, a pressão das agências de fomento sobre o pesquisador para que se tenha publicações de alto impacto:

Em nosso caso em que **pagar para publicar pois é praticamente uma utopia porque temos muitos eventos de curso e muito menor público leitor que justifique**. Nós nos encontramos com que nossas agências de financiamento nos perguntam: quantas revistas [de alto impacto] você publicou? esta publicação, normalmente, me acaba excluindo, limitando, o que logo eu então, talvez, eu tenha que compartilhar os dados estou sujeito a embargos, estou sujeito a limites para publicar e para publicação. É verdade que eu posso fazer, mas é em certo modo ambíguo os pré-prints e publica-los, os pré-prints, antes que eles tenham sido avaliados, ou há alguns casos, após serem avaliados, mas antes de ser, de haver sido publicado [...] mas seu exemplo para mim é uma complicação enorme porque eu publico um print, mas antes de fazer, por exemplo, a revisão de provas, foi corrigido alguns pontos, resulta que estou disponibilizando ou o

que eu disponibilizo os dados que posso obter são inferiores, digamos, ao que eu posso ter publicado na revista porque as revistas estão revisados, não? Isso para os dados secundários, para os dados elaborados, para os dados primários, o material primário e os outros, e procuro os campos onde vemos, certamente, nenhum tipo de tradição da publicação dos materiais primários. É verdade que os campos científicos, pois é muito comum que comprometa a publicar o artigo, e ele tem que deixar ao alcance de qualquer pessoa que queira verificá-lo, nos conjuntos de dados e inclusive as amostras de laboratório. Em nosso caso não é assim, o compartilhamento de materiais primários não é algo que se exija do financiador, em absoluto a mim me exigem e aí e esse foi o problema que eu estava comentando antes de compartilhar os dados secundários os artigos sim que eu tenho que publicá-los após 12 meses abertos os primários. (P2, grifo nosso).

Nos apontamentos do pesquisador P2 é relevante destacar o papel das agências de fomento no sistema da comunicação científica. No cenário atual, e nas discussões das temáticas apresentadas nessa tese, advoga-se pela publicação em aberto de todo o fluxo da pesquisa. Entretanto, nota-se a exigência de publicações em revistas de alto impacto e, as publicações de pré-prints e de dados, ainda não tem suas métricas consideradas nas avaliações em muitas comunidades. Destaca-se que no relatório final e plano de ação da Comissão Europeia a fala de Hodson, *Chair da European Commission's FAIR Data Expert Group*, em que são apresentadas as métricas e indicadores para os dados e para contribuições de pesquisa e sua importância “[...] precisam ser reconsiderados e enriquecidos para garantir que funcionem como incentivos convincentes para Open Science e FAIR. Reconhecimento e recompensas eficazes são vitais para a mudança de cultura.” (EUROPEAN COMMISSION, 2018, p. 8, tradução nossa).

Referente a disponibilização dos seus conjuntos de dados resultantes dos projetos financiados e publicados nos repositórios, foi perguntado qual foi a indicação do patrocinador referente ao compartilhamento dos dados do projeto:

P1 responde que esse requisito não era obrigatório na chamada do financiamento, porém, foi incentivado a publicar os dados em aberto.

P2 destaca que foi sugerido pelo financiador a disponibilização dos dados e não obrigatório.

P3 destaca que a área dele segue as diretrizes e regulamentos da União Europeia e, no caso do projeto dele, foi seguido, e enfatiza que os dados são gratuitos para acesso a todos mesmos quando não registrados (publicados formalmente):

É a política [do nosso financiador] que segue as diretrizes da ciência europeia. Neste sentido, essa é uma instituição pública e é europeia e segue, portanto, os regulamentos da União Europeia em esta questão. Assim que nós seguimos também essas indicações e como eu digo, os dados são livres, são gratuitos para baixar e acessíveis para todo o mundo, incluindo sem necessidade de registro.

P4 responde que não foi obrigatório disponibilizar os dados, mas o fez porque pensa ser a coisa certa a fazer, e esse entendimento foi desde o princípio do projeto. O pesquisador complementa:

[...] já que eles devem ser compartilhados, se forem dados que foram financiados, principalmente com fundos públicos, pois devem ser compartilhados. (P4).

Relevante destacar que os pesquisadores foram incentivados ou tiveram o entendimento de compartilhar os dados, mas o fizeram depois das publicações dos resultados da pesquisa. Outro aspecto relevante é a menção de que os dados coletados com financiamento público das pesquisas devem ser publicados.

No que diz respeito às métricas de citação, foi perguntado aos pesquisadores se concordavam com que as agências de fomento deveriam usar as métricas de citação de dados. Os pesquisadores P1, P3 e P4 estão totalmente de acordo e P2 está moderadamente de acordo. Apesar de P3 estar totalmente de acordo com a utilização das métricas, o mesmo ainda não percebe que as métricas são colocadas em prática e menciona que, no caso de *downloads* dos conjuntos de dados, esses não são considerados para avaliação do pesquisador. Com essa afirmação do pesquisador, surgiu a dúvida se o mesmo tivesse o entendimento da abrangência das métricas de dados. O pesquisador P4, inclusive, pergunta o que queria dizer sobre as métricas de dados, demonstrando que não conhecia o tema.

Assim, para sanar as dúvidas, a autora da Tese fez um esclarecimento sobre a temática.

O relato do pesquisador P2 apresenta uma visão parecida com do outro pesquisador referente às métricas dos *downloads* dos conjuntos de dados. O pesquisador chama a atenção pelo fato de os *downloads* não serem valorizados ou não servirem de medição de impacto da pesquisa ou do pesquisador, uma vez que pode ampliar a visibilidade e estão publicados em aberto:

De dados primários, em nosso campo, porque não seria correto de alguma forma eu coloquei recompensa, ou se dissesse oficialmente que o publicou em aberto, e se resulta que sua página web [ou em] tal repositório foi feito download não sei quantas vezes, **é um índice de impacto que não é valorizado de forma alguma** porque o que se vê sobre os artigos do projeto, os artigos que se publicam nas revistas, nas boas revistas, bem, então são isso e pronto. (P2, grifo nosso).

O pesquisador reforça mais uma vez a questão da publicação dos dados depois das publicações dos resultados da pesquisa e traz um exemplo de outra área do conhecimento que publica os dados antes da finalização da pesquisa, fato que o pesquisador considera “estranho”

O estudo de sociologia, pois é quase obrigatório que eles tenham que depositar os dados no instituto de estatística do Instituto de Ciências Sociológicas e no serviço de apoio, não? tem que ser os formatos, em nosso caso é muito estranho que essa, especialmente aquela, sejam publicados durante o período de vigência do projeto. É mais comum que seja publicado após, na medida que vão editando e trabalhando, então não é habitual com os dados primários. (P2).

O pesquisador P2 faz uma sugestão para continuação da pesquisa dessa tese:

Entendo que também em neste caso, este questionário tem que cobrir campos científicos muito diversos com políticas editoriais muito diferentes e com exigências de financiadores muito diferentes. (P2).

Apesar de que todos os pesquisadores terem disponibilizados seus dados no repositório, o fez com ressalvas, disponibilizando somente após a publicação dos resultados das análises.

Considerando o exposto a hipótese **H5**.: As agências de fomento influenciam positivamente a perspectiva dos pesquisadores fomentando o compartilhamento dos dados via repositório é refutada.

8.3.6 Repositório

O índice Repositório está relacionado ao Pilar Normativo. Os Repositórios de dados são ambientes digitais com infraestrutura para gestão, armazenamento e curadoria dos dados, almejando a disponibilização com potencial de reutilização por

outros pesquisadores.

A análise dos documentos revela que as ações das bibliotecas e dos repositórios de dados encontram-se direcionadas para atender as indicações do programa *Horizonte 2020*, o maior programa de Pesquisa e Inovação da União Européia com aproximadamente 80 bilhões de euros para financiamento disponível entre os anos de 2014 a 2020 e, por conseguinte, serem aplicados investimentos privados que são suscitados visando garantir a competitividade global da Europa (EUROPEAN COMMISSION, 2017).

O repositório permite o compartilhamento dos conjuntos de dados, PGDs e documentação de dados em uma ampla variedade de formatos, incluindo texto, planilhas, áudio, vídeo e imagens, em diversos campos da ciência, bem como, exibi modelos de referência dos conjuntos dos dados incentivando a citação e contribuindo para que os pesquisadores obtenham crédito, facilitando para que os resultados da pesquisa sejam citáveis.

Adicionalmente as características apresentadas do repositório na seção 5, destaca-se que os dados estão em livre acesso por *default*, ao menos que existam razões legais, éticas, econômicas, geográficas ou de transferência de tecnologia que o impeça, ou que implique com a conformidade com os critérios de utilização, reutilização e redistribuição sem restrições.

Referente ao indicador Repositório, os pesquisadores foram solicitados a responder a questão: “O quanto você concorda com cada afirmação sobre o REPOSITÓRIO e-cienciaDatos?”

- (a) Eu conheço suas diretrizes para o uso de licenças e de citações;
- (b) Os pesquisadores são incentivados a depositar;
- (c) Me preocupa a violação de direitos autorais;
- (d) Eu não deposito outros conjuntos de dados porque tem pouca visibilidade;
- (e) Participei em capacitações ou orientações que abordam os direitos autorais de dados

Quadro 18 – Respostas dos pesquisadores sobre Repositório e-cienciaDatos

	Pergunta	P1	P2	P3	P4
(a)	Eu conheço suas diretrizes para o uso de licenças e de citações	Moderadamente de acordo	Moderadamente de acordo	Totalmente em desacordo	Totalmente de acordo
(b)	Os pesquisadores são incentivados a depositar	Totalmente de acordo	Moderadamente de desacordo	Totalmente em desacordo	Totalmente de acordo
(c)	Me preocupa a violação de direitos autorais	Totalmente em desacordo	Totalmente em desacordo	Totalmente em desacordo	Totalmente de acordo
(d)	Eu não deposito outros conjuntos de dados porque tem pouca visibilidade	Moderadamente de acordo	Totalmente em desacordo	Totalmente em desacordo	Neutro
(e)	Participei em capacitações ou orientações que abordam os direitos autorais de dados	Moderadamente de acordo	Totalmente em desacordo	Totalmente em desacordo	Moderadamente de desacordo

Fonte: Elaborado pela autora.

A primeira resposta dos pesquisadores chama a atenção por alguns fatores: primeiro pelo fato do pouco esclarecimento sobre as orientações do repositório; segundo por um dos pesquisadores não ter conhecimento que os conjuntos de dados de seu projeto estavam disponibilizados via repositório; terceiro, outro pesquisador não sabia o nome do repositório *e-cienciaDatos* e o confundiu com outro repositório. Quando o pesquisador se deparou com a pergunta sobre a temática, perguntou sobre qual repositório estava se referindo, visto que na pergunta tinha o nome do repositório.

Estar ciente da existência do repositório de dados de sua instituição deveria ser unânime entre os pesquisadores, principalmente pelo fato de seus dados estarem armazenados nele, mesmo que os pesquisadores prefiram depositar seus dados em repositórios temáticos ou repositórios externos à universidade. Essa insciência é, ainda, mais preocupante devido ao fato do trabalho que é feito pelo *Consórcio Madroño* e as bibliotecas na gestão, organização e divulgação dos dados, do tempo de sua existência e da relevância que tem dentro das universidades. Diante das respostas, novamente houve a intervenção da autora dessa tese para esclarecimentos.

Em relação a afirmação “a) Eu conheço suas diretrizes para o uso de licenças

e de citações” as respostas dos pesquisadores foram:

P2 que acha que conhece mais ou menos

P3 não conhece

P4 declara que não sabe se as conhece

As repostas dos pesquisadores deixam evidente que os mesmos não se informaram a respeito das políticas do repositório, fato constatado com as respostas das perguntas seguintes.

Referente a afirmativa “b) Os pesquisadores são incentivados a depositar” foi identificada uma lacuna na divulgação e até mesmo na conscientização da relevância do compartilhamento dos dados e da visibilidade que é possível alcançar. Outro aspecto identificado foi a observação feita pelo pesquisador sobre o papel do bibliotecário nas instituições.

E enquanto ao incentivo dos pesquisadores. Em nosso caso, **não houve muito, e o caso da universidade [...] não tem muito incentivo não**, especialmente lá. Também é **verdade que isso deve ser sempre avaliado/valorizado dependendo, por exemplo, do pessoal bibliotecário que cada instituição tem**. Então, bem [a universidade] tampouco é que tenha muita capacidade para que, de repente, todos os pesquisadores se coloquem a depositar e possa processar esse material que, em nosso caso, pois sim é verdade que eles nos ofereceram porque antes havia surgido uma conversa com a biblioteca a propósito da obtenção de ISSN para nosso corpus e, então, pois se havia, de certa forma, como acordado conosco e quando se tratava de buscar para e-cienciaDatos pois todos nos contactaram, então transferimos alguns materiais. (Um dos pesquisadores, grifo nosso).

Um dos pesquisadores declara que o repositório facilitou para o depósito dos dados do projeto. Sua resposta chama atenção ao fato de terem depositado os dados por ser um local para armazenamento e, devido ao custo que teriam se não tivesse depositado no repositório. Esse fato abre outra discussão sobre o compartilhamento de dados: pesquisadores podem depositar seus dados em ambientes que não disponibilizam os dados em aberto. Em sua fala, o pesquisador menciona que ao depositar os dados nesses ambientes o pesquisador tem que aceitar suas políticas.

[...] e a questão que talvez seja isso mais problemático é como compartilhar os dados porque, por exemplo, em nosso caso, é verdade que a forma que **o repositório e-cienciaDatos foi aberto, uma perspectiva de poder armazenar os dados gratuitamente**, mas no

nosso caso, armazenar [nossos dados] custa dinheiro e poder colocá-los ao alcance de qualquer pessoa, pois, tem um custo, são arquivos enormes que eu não posso hospedar em uma página web pessoal, então eu creio que um primeiro problema que existe em relação a comparar para compartilhar os dados é que de fato nos projetos de pesquisa, pelo menos os espanhóis, começa a dizer honestamente que não se podem financiar as despesas de hospedagem de página da web e outras contas para compartilhar materiais. Eles justificam como sempre, os burocratas são muito criativos em suas justificativas, em que claro, não podem saber se eu uso dropbox para armazenar os dados ou se é para compartilhar fotos de gatinhos, e eu não sei que encontro na internet, ele os armazenou lá, mas é claro que nos restringe ou podemos encontrar alguns repositórios, mas temos que **adaptar-nos às regras do repositório** para sua disponibilidade para os meios de convite ou, no nosso caso particular, custa dinheiro e não nos ajudam as instituições [...] (Um dos pesquisadores, grifo nosso).

Ao que se refere a afirmação “c) Me preocupa a violação de direitos autorais” o pesquisador P2 não se preocupa, diferentemente de P4 que é enfático:

Não me preocupa nada. (P2).

Sim, estou porque eles podem usá-lo sem mencionar-me. (P4).

No que concerne a depositar outros dados no repositório, foi perguntado se não depositariam porque tem pouca visibilidade - “d) Eu não deposito outros conjuntos de dados porque tem pouca visibilidade”

No caso, estes são os únicos conjuntos de dados que tenho, se eu não deposito outros é, basicamente, pelo que eu dizia, que **nossa disciplina por um lado, não há esse costume de depositar e, por outro lado, tampouco, como diria em muitos casos, são pesquisas tão específicas que a possibilidade de que alguém vá acessar estes conjuntos de dados e utilizá-lo pois é muito, muito remoto**, porque são trabalhos sobre obras específicas [...] que cinco, dez, vinte pessoas que, hoje em dia, mesmo em todo o mundo pode trabalhar e, sobre isso sabemos quem somos e onde vemos problemas no intercâmbio de dados, mas onde chegaríamos e a instituição em todo o trabalho de processamento e nós, comunidade científica, temos muito desempenho nesse nome é um caso muito distinto do que creio que essa é a famosa sequência de o famoso coronavírus, este que anda solto, é evidentemente para publica-lo e mantê-lo aberto [os dados a respeito do coronavírus] porque me permite acesso a milhões de pessoas interessadas em trabalhar nisso, então eu **não tenho visibilidade no repositório**. (Um dos pesquisadores, grifo nosso).

A fala do pesquisador enfatiza que seus dados não tem visibilidade no

repositório porque seus dados são muito específicos e com poucos pesquisadores com interesse no tema, e faz uma analogia com a visibilidade que dados do Coronavírus atingiria. Esse fator está atrelado, também, com o que o pesquisador menciona de “dados muito específicos”. Nesse contexto de incentivos e trabalhos entorno do compartilhamento de dados e das práticas da Ciência aberta, é coerente que os pesquisadores tenham uma visão positiva e a percepção dos benefícios em disponibilizar seus dados publicamente, mesmo que os dados sejam específicos e que o possível público interessado seja pequeno.

O depoimento de outro pesquisador sobre a visibilidade do repositório está correlacionado com a visibilidade de pesquisadores, assim como prestígio profissional e incentivos na carreira. Sua fala está fortemente relacionada com a Categoria *Carreira* e seus índices apresentados na subseção seguintes. O pesquisador menciona o fato da obrigação da publicação em aberto.

Esta é uma questão importante, incentivos na carreira acadêmica tem a ver com seu prestígio profissional, tem a ver com sua visibilidade, logo, se te obrigam a publica-los [os dados] em aberto em um site onde não há visibilidade, talvez o incomode e te prejudique. Depois de fazer um esforço o que querem é visibilidade e, neste caso então, se a obrigação de o fazer de uma certa maneira te podem prejudicar. (Um dos pesquisadores, grifo nosso).

A continuidade das respostas dos pesquisadores, P4 indica a resposta “neutro”. A autora da Tese pergunta se tem outros conjuntos de dados em outros repositórios e P4 responde que tem em uma página *Web*.

No tocante a visibilidade, a resposta de P4 corrobora com a resposta de P2. Durante a entrevista P4 fez uma busca, no que pensou ser o repositório *cienciaDados*, com seu nome e o título dos conjuntos de dados, e faz uma declaração demonstrando pouca clareza referente a atribuição da autoria dos dados:

Sim sempre se deve reconhecer uma autoria intelectual e é algo que não me parece claro que tenha no repositório. Aparecem os dados mas nem sempre aparece claramente a quem está associado, isto é, se alguém vai para o arquivo pessoal, quando eles me veem em espiral, eu não sei se em todo momento está claro. (P4, grifo nosso).

Após sua declaração, a autora dessa tese comentou que um dos objetivos do repositório é dar visibilidade à pesquisa, e o repositório, aparentemente, tem a

organização e catalogação dos dados com as informações corretas, inclusive com a indicação correta da autoria para que os autores sejam reconhecidos e citados. Após, P4 destaca que recebeu o link do repositório do pessoal da biblioteca que dá acesso aos conjuntos de dados, enfatizando a preocupação do não reconhecimento de sua autoria nos dados e complementa:

[...] por exemplo, aqui se eu clico em qualquer uma dessas coisas de projeto estaria [menciona o nome de outro documento] e parece que o copyright é [menciona que o copyright está vinculado a outrem] **mas, onde diz que eu cuidei de dirigir e incentivei esse projeto?** porque não está em nenhum lugar, você sabe que não seguimos. E se você tiver a publicação associada a esses dados? me foi mostrado, mas eu não consigo encontrar. Eu não saberia como encontrá-los [os dados] porque me forneceram o link. (P4, grifo nosso).

Na declaração é evidenciada a pouca clareza de P4 sobre o repositório. No entendimento de P4, ao pesquisar no repositório de dados, os conjuntos de dados seriam recuperados juntamente a outros documentos de sua autoria.

Nesse momento ocorre a segunda intervenção da autora dessa tese perguntando onde o pesquisador estava fazendo sua busca, tendo como resposta, no repositório de dados. A autora da tese faz umas buscas e responde que os dados foram recuperados quando buscados pelo nome do projeto e do autor.

A continuação do depoimento do pesquisador, após algumas tentativas e buscas no repositório, o qual P4 diz ser o repositório de dados, o pesquisador menciona não ter certeza se todos os conjuntos de dados armazenados são recuperados, nem se aparecem todos nos resultados de busca, devido ao fato de não terem sido recuperados quando o pesquisador fez sua busca. Ademais, o pesquisador evidencia certo desconforto ao verificar os resultados de sua busca, repetindo várias vezes que não via os dados, demonstrando certa incredulidade, indicando perda das referências que havia tido:

E estou me procurando, **eu não sei se saem todos esses dados como eu os compartilhei**, e não aparece meu, meu quer dizer, **não vejo que aqui apareçam os dados** [...] **não, eu não vejo isso** [os dados], eu vejo que aqui estão meus artigos que eu digo autorizá-los a ser. Aqui vejo que meus artigos [...] a tese que orientei, **mas não vejo, não vejo os dados, não vejo, não vejo os dados. Não vejo os dados. Não vejo os dados, os dados, olha eu não os vejo**, eu sei que há de alguma forma. Mas não, **eu não me encontro aqui**, ou seja, estão as teses que orientei, e, se talvez, um texto no qual me cita, **mas eu não vejo que aqui estejam**

os dados desse projeto associado ao meu nome. (P4, grifo nosso).

Diante da situação, a pesquisadora dessa tese perguntou mais uma vez em qual ambiente ele estava pesquisando e informou que em sua busca os dados foram recuperados e que as publicações subjacentes não aparecem no repositório, apenas as informações relacionadas, inclusive tentou novamente averiguar se o pesquisador estava pesquisando no repositório de dados.

A autora dessa tese constatou que P4 estava pesquisando no repositório institucional e não no repositório de dados. Com essa constatação, são levantados dois pontos relevantes: Primeiro, tanto o repositório institucional quanto o catálogo da biblioteca não recuperam os dados no repositório de dados, aspecto negativo já que os dois são fontes de informação onde, normalmente, os usuários fazem a busca; se não recupera os dados do repositório, os dados e o repositório têm menos visibilidade. O aspecto foi confirmado pela autora dessa tese que, juntamente com o pesquisador, fez os testes de busca com a estratégia de busca que P4 estava usando. Segundo, P4 não tinha clareza sobre a informação de que o link que havia recebido não era do e-cienciaDatos, ou por ter se confundido quando foi informada ou por não ter sido mostrado o repositório de dados, pois recebeu o link e não fez nenhum treinamento oferecido pela biblioteca e não teve nenhuma instrução e, também, tem pouca habilidade e instrução para fazer pesquisas no repositório.

Os dois pontos levantados pelo pesquisador demonstram a importância de dispor de informações claras e corretas sobre o funcionamento dos catálogos e dos repositórios de dados, da relevância das instruções e treinamentos oferecidos pela biblioteca, da confiança adquirida pelo pesquisador de que seus dados estão disponibilizados e atingindo ambos propósitos, a perspectiva do pesquisador ao compartilhar e a função do repositório na gestão e disseminação dos conjuntos de dados. No caso de o pesquisador perceber que o seu propósito não foi atingido ou sentir insegurança no compartilhamento de seus dados há possibilidade de retirar os dados do repositório.

Com os resultados de sua busca, P4 faz uma menção a indexação do sistema pois são recuperados documentos que diz não ter relação com seu nome:

Não sei se ele, *e-ciencia-Datos*, tem, ou seja, isso é ao autor, mas não se estende aos dados científicos. No *e-ciencia-Datos* sim estarei recuperável, mas não estaremos aqui, se você ver aqui saem todos de

minha autoria, mas não vejo que aqui estão os dados [...] estão as teses que orientei, logo aparecem artigos que eu não estou ciente de ter participado, mas, talvez, algum em meu nome. **Algo não está bem indexado porque aparecem coisas que não têm nada a ver comigo.** (P4, grifo nosso).

A Continuação, sem mencionar em qual repositório estava, P4 demonstra, mais uma vez, a preocupação de não estar claro que os dados são de sua autoria

Claro que isso não está bem, e esses dados, se alguém chegar a esses dados para este arquivo [...] para expirar enquanto se pensa que esses quatro são completamente em aberto, o que é verdade, mas eles não têm um mandante, **mas a pessoa que liderou o projeto para que esses dados sejam abertos sou eu, e ainda não parece.** Procuro por [coloca o nome do projeto] vamos ver se assim funciona, mas não vejo que estão, ou seja, **o que eu não vejo é que esses dados estejam associados ao meu nome**, é verdade que na hora em que aparecem claramente os artigos que estão associados ao projeto sim, mas, **mas eu vejo novamente, mas não estão associados em meu nome.** (P4).

Com a fala do pesquisador, a autora dessa tese fez uma busca no repositório institucional. Nos resultados de busca aparecem os metadados e o local onde há o *link* para o repositório de dados fica no final da página e só localizado após identificação de uma possível informação que conteria um *hiperlink*. Destaca-se que não é um link fácil de localizar e que para localiza-lo foi preciso verificar em todas as informações disponíveis na página. Com efeito, um usuário que não está acostumado com o ambiente teria dificuldade em localizar.

Toda a ausência de clareza do funcionamento dos repositórios, confusão entre um repositório e outro e a dificuldade em localizar os dados no repositório institucional revela o despreparo do pesquisador e intensifica a preocupação dos motivos que podem levar os pesquisadores à retenção de seus dados.

Após a busca no repositório institucional, a autora dessa tese explicou para o pesquisador o que havia acontecido, a diferença entre os repositórios e onde estava a informação sobre os direitos autorais associado ao seu nome. Mediante os esclarecimentos, P4 demonstrou menor desconforto sobre a situação e enfatizou a importância da participação em treinamentos e de informações corretas.

Sobre participação em orientações ou treinamentos que abordam os direitos autorais de dados, foi apresentada a afirmação “e) Participei em capacitações ou orientações que abordam os direitos autorais de dados” obtendo as respostas:

P2: Não participei em nada disso.

P3: Não participei, porque eu nessas coisas eu não me dedico. Eu dedico a parte científica.

P4: Mas eu não participei de nenhum curso.

Os pesquisadores foram indagados sobre o motivo pelo qual disponibilizaram seus conjuntos de dados do projeto no repositório *e-cienciaDatos*.

A razão está, em geral, que havíamos dito sobre os repositórios, que **foram intitulados a vantagem, de um ponto de vista materialista, que é gratuito**, que para mim outros repositórios especializados, porque significaria armazenar os dados uns 8.000, 10.000 15.000 euros para poder armazenar os dados, coisas que, então aqui, não me pedem. Eu creio que também porque no início não estavam conscientes ou não somos muito conscientes da quantidade de dados que vai ser um ditado por entrevista, o vídeo, o áudio, as diferentes versões pesquisadas, entrevista a ser feita pode ser discretamente 10, 15 gigabytes de dados se apenas multiplicamos as 300-400 quantidades muito, muito grandes. (Um dos pesquisadores, grifo nosso).

Vale a pena porque, vamos dizer a verdade, para maior visibilidade, ou seja, para a **parte de incentivo da carreira**, [...] em nome de uma **maior visibilidade de um projeto de investigação e para os dados**, em nosso caso, e a crítica aos dados melhora o banco de dados, e agora nosso objetivo geral é a visibilidade de nossa pesquisa. (Um dos pesquisadores, grifo nosso).

Porque era uma boa plataforma para compartilhar todos aqueles dados de forma aberta e, também, **eles cuidavam da manutenção dos dados**. (Um dos pesquisadores, grifo nosso).

A partir das falas dos pesquisadores percebe-se diferentes motivos que os levaram a depositar seus dados no repositório: financeiro, visibilidade e preservação. Conforme uma das falas, verifica-se a quantidade de dados que é possível gerar e indaga-se a limitação de espaço que os repositórios de universidades podem disponibilizar a cada um de seus pesquisadores. Nesse aspecto, cabe uma reflexão: se o pesquisador demandar de mais espaço do que o repositório disponibiliza e o pesquisador não tem financiamento que cobre os gastos de armazenamento o que serão feitos com os dados? É possível ocorrer de esses dados ficarem armazenados nos computadores pessoais e, com o tempo, serem perdidos por obsolescência dos suportes ou serem descartados. É viável a orientação aos pesquisadores para prever

gastos de armazenamento dos dados quando submetem solicitações às agências de fomento. Cabe ressaltar que existem repositórios de dados gratuitos, porém, muitos também têm limites de armazenamento. Quando excede o limite, há opção de pagamento.

Outra observação apresentada é a opinião de um dos entrevistados ao mencionar que depositou os dados no repositório para ter maior visibilidade dos dados. De certo que quando os dados estão em repositório é possível ter maior visibilidade. A fala do pesquisador se contrapõe ao já apresentado sobre visibilidade. De certo que esse pesquisador não fez a mesma dinâmica, de procurar seus dados durante a entrevista. Pode-se inferir que o mesmo teve uma experiência ou percepção diferente do pesquisador anterior.

Os pesquisadores foram indagados se depositariam outros dados no Repositório *e-cienciaDatos* e por quê. A fala de um pesquisador expõe uma perspectiva que impacta negativamente o compartilhamento de dados: definir o pensa ser de interesse para a comunidade. Na perspectiva do pesquisador, corroborando com respostas anteriores, compartilharia seus dados se acreditasse o que é e o que pode não ser de interesse da comunidade. Outro aspecto que gera indagações: Pensando que os dados disponibilizados publicamente em repositórios podem ser acessados por qualquer pessoa em nível global com possibilidades e fins de uso não previstos, como um pesquisador mediria o grau de interesse da comunidade sobre seus dados?

Pois sim, eu os depositaria no caso de ser os dados que **eu acreditasse que poderiam ter um interesse**, uma pesquisa um pouco mais ampla sobre algumas coisas, algum registro de materiais de determinado autor, bom, pois sim o coloco a lista porque alguém pode trabalhar com ela. Como no momento não tenho, mas sim estaria aberto, vai ser publicado em qualquer repositório. (Um dos pesquisadores, grifo nosso).

Sim, sim eu os depositaria, porque não? (Um dos pesquisadores).

À continuação, outro pesquisador tem uma perspectiva diferente e declara que disponibilizaria para que outros pesquisadores avancem nas pesquisas e não precisem fazer trabalhos repetidos e, no caso, repetir a coleta de dados. O pesquisador chama a atenção para o fato de pesquisadores brasileiros da mesma área estarem publicarem em português e muitas dessas publicações serem parecidas

com outras publicações internacionais. Acompanhando raciocínio do pesquisador, dados disponibilizados nos repositórios, ou até mesmo, dados publicados em *data journals*, no Brasil podem estar no idioma português e, conseqüentemente, pode acontecer de se fazer coletas parecidas e, inclusive, conforme a fala do pesquisador, se perder a parte inovadora da pesquisa.

Sim, porque eu acho que o que é um bom sistema para tornar sua pesquisa visível e para melhorar, em termos gerais, ao tornar sua pesquisa mais visível, **melhorar as possibilidades de outros pesquisadores avançarem com um dos problemas que temos em muitas pesquisas: a repetição de trabalhos parecidos. Há realmente um choque, porque cada lugar, um chega antes a um lugar que aquele que vem depois não publica, sim porque ele perdeu a parte inovadora da pesquisa**, mas pode ser que tenham sido publicados tudo em línguas [diferentes], se publica em inglês, são revistas de muito alto nível porque normalmente são, mas eles publicam em inglês, são revistas de alto nível porque as vezes se unem. **No Brasil estão publicando coisas similares ao que eu estou fazendo**, isto é cada vez menor porque o mundo está mais globalizado, a comunicação é maior, mas ainda está existindo, isso acontece muito comigo com a América Latina, em países muito, muito fechados como o Brasil, e as pessoas não tem em conta o que se publica fora porque o próprio país, porque eles não estão seguros e porque, bem, também por razões que tudo depende do quanto, do porquê as pessoas publicam o brasileiro muito, **[publicando no idioma português]** estou falando de um setor [...] nos vê falando um setor muito especializado como na biologia odontológica porque estão, como o resto do mundo. (Um dos pesquisadores, grifo nosso).

Outro ponto relacionado ao depósito no repositório foi sobre quais dados o pesquisador não depositaria no repositório *e-cienciaDados* e seus motivos

A priori não sei, não consigo pensar em uma maneira de dizer, talvez em o único caso, e fazer alguma pesquisa que estivesse em desenvolvimento e que se tenha que ficar parada a um tempo [...] mas, tampouco, esteja muito elaborado para publicação, pode ser melhorado se for insuficiente ou poderia trabalhar mas **decidiu deixar em aberto porque talvez já tornaram-se obsoleto**, porque alguém vem trabalhar neles e depois não posso voltar a explorar a fundo os materiais primários. Seriam apenas qualquer questão de esse tipo de coisa. (P2, grifo nosso).

Bom, se forem publicações elaboradas, por exemplo, que serão objeto de uma publicação comercial [não publicaria], mas eu sou, em princípio, favorável a depositar tudo sim. (P4).

A fala do pesquisador chama atenção para o fato de depositar dados que estão obsoletos ou pelo fato de não estar bem preparados para publicação.

Foi proposto pela autora dessa tese relatar a dificuldade encontrada na busca pela autoria e recuperação dos resultados para os responsáveis pelo repositório. Assim, o pesquisador agradeceu e indicou que é muito importante o relato e correção do problema.

Mediante o exposto, a hipótese H6.: Repositório influencia positivamente o compartilhamento dos dados via repositório e confirmada.

A hipótese H7.: Repositório influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório é refutada.

8.3.7 Leis e normas

Relacionado à categoria *Regulador*, este componente abrange a legislação e normativas da União Europeia, da Espanha, das universidades e do repositório, incluindo às políticas do repositório, à gestão de dados, às instruções e exigências dos financiadores, aos aspectos da Ciência aberta e os componentes que estão em seu entorno, aos princípios FAIR, especificando cada um desses princípios, aos aspectos éticos e legais envolvendo propriedade intelectual e privacidade dos dados. Destaca-se que esses documentos elaborados pelo *Consórcio Madroño* e as bibliotecas estão seguindo, principalmente, orientações da União Europeia. Cabe ressaltar que a motivação para implementação do repositório foi a Política Nacional de Acesso Aberto.

O *Consórcio Madroño* e suas bibliotecas dispõem de vários documentos alinhados a essas leis e normativas que regulamentam o ciclo de vida dos dados e, conseqüentemente, sua gestão pelos pesquisadores e pelos repositórios (QUADRO 11). Embasado nessas leis e normativas foram elaborados diversos informes e manuais.

As entrevistas revelaram que Bibliotecários auxiliam os pesquisadores a entender as exigências das Políticas das agências de fomento, os aspectos da propriedade intelectual e privacidade dos dados. Conhecimento sobre as leis e normativas pelos Bibliotecários são fundamentais para a gestão dos dados tanto pelos pesquisadores quanto pelo repositório.

A figura 25 apresenta medidas na União Europeia com proposta de reforma, melhoria ou adequação na Lei e nas normativas sobre direitos de autor e *copyright* que incluem dados.

Figura 25 – Propostas de reformas na Lei de Propriedade Intelectual



Fonte: Elaborado pela autora.

As atualizações nas normativas demonstram que as instituições que direta e indiretamente transferem dados requerem modificações do dispositivo jurídico que regula suas transferências. Os detentores de direitos autorais recorrerem à justiça na ocasião em que o resultado de sua atividade intelectual ou científica é utilizado sem a devida autorização, mostrando-se desejosos de obter formas de proteção legal (UNESCO, 1981).

Seguindo a análise, é possível detalhar os aspectos indicados na *Open Science Taxonomy - Open Science Policies - Organisational Mandates* - subdividido em: *Funders Policies, Governmental Policies; Institutional Policies* conforme apresentados na figura 19. Os aspectos indicados na *Open Science Taxonomy - Open data* subdividido em: *Intellectual Property Rights, Law e Licenses* (FIGURA 20) também são regulamentados pelas políticas do repositório.

O *Consórcio Madroño* e suas bibliotecas dispõem de vários documentos que regulamentam o CVD que indica as políticas governamentais e das agências de fomento (FIGURA 21). Os documentos representados na figura 22 são: Políticas - instruções gerais sobre os dados depositados; *Declaración del Consorcio Madroño en apoyo a la Ciencia Abierta a la información académica y científica; Política para la Gestión de datos de Investigación; Normativa Vigente y Principales*

Recomendaciones; Licencia de uso; Licencia de depósito.

Referente ao indicador Leis e Normas, os pesquisadores foram solicitados a responder à questão: “O quanto você concorda com cada afirmação sobre as leis e normativas da União Europeia e a PROPRIEDADE INTELECTUAL da Espanha?”

Quadro 19 – Respostas dos pesquisadores sobre Propriedade Intelectual

Pergunta	P1	P2	P3	P4
(a) Eu conheço toda a legislação e diretrizes	Neutro	Neutro	Neutro	Totalmente em desacordo
(b) Leis e normas facilitam o desenvolvimento científico e tecnológico	Neutro	Moderadamente em desacordo	Moderadamente de acordo	Totalmente em desacordo

Fonte: Elaborado pela autora.

Referente a afirmativa “a) Eu conheço toda a legislação e diretrizes”, três dos pesquisadores responderam neutro e um está totalmente em desacordo.

Eu a vi muito por acima [...] no sentido de que nem todas as diretrizes da União Europeia, ou seja, estamos falando sobre diretrizes daqueles emitidas pelo governo europeu, por entendermos pela Comissão e, depois, eles transpõem e se estivesse transposta na Espanha, não sei que vou me ver desde logo que **eu não conheço toda a legislação, principalmente porque é uma coisa terrivelmente imensa a da propriedade intelectual.** Porque, além disso, o modelo legislativo espanhol, em esse aspecto, **é que ao invés de eliminar tudo e fazer algo novo bem, eles vão adicionando, modificando uma coisa à outra, adicionando coisas que vêm da Europa,** mas eles as colocam em vista, eles as completam também, então não, eu não a conheço em particular [...] (P2, grifo nosso).

Seguramente não [conheço toda a legislação]. (P3).

É que **eu não sei quais são as regras de propriedade intelectual da Europa e da Espanha.** Eu sei que há uma tendência que os dados devem ser abertos, é a única coisa que eu sei, mas toda a legislação e diretrizes eu não conheço, não conheço. (P4, grifo nosso).

As repostas dos pesquisadores revelam a ausência ou perda das referências normativas. Destaca-se a menção de que a legislação é extensa e existem várias modificações que, ao ver do pesquisador, dever-se-ia eliminar a lei existente e fazer uma nova, e, também, que um dos pesquisadores sabe da tendência do acesso aberto

aos dados, sem fazer referência a Lei.

A continuação, o pesquisador P2 menciona aspectos que podem levar a confusão das normas ou descumprimento das mesmas

Eu sei que toda vez, ao meu modo de ver, **vai se tornando mais complicada questões que nos afetam como o direito de citação e o direito de reprodução que, alguns anos atrás, era manifestado de uma forma bastante mais aberta**, se podia remover e reproduzir um trabalho científico ou acadêmico, um grande fragmento sempre foi para o comentário e nos últimos anos vai se complicando, portanto vai adicionando condições, sempre que dentro de um curso universitário, sempre que é colocado em uma página da web de acesso, sempre que é dizer que se foi complicando bastante nesse sentido. (P2, grifo nosso).

Interessante destacar que o pesquisador P4 enfatiza que se deve respeitar o direito intelectual e que os dados devem ser compartilhados abertamente

Se são sobre os direitos, bem sem dúvida se deve respeitar o direito intelectual, mas os dados de pesquisa eu acho que eles deveriam ser compartilhados abertamente. (P4).

Referente a opção “b) Leis e normas facilitam o desenvolvimento científico e tecnológico” há disparidade entre as respostas. Em uma das respostas, o pesquisador menciona mais uma vez que não as conhece.

Sim, eu entendo que eles facilitam, mas tampouco conheço o tema com muito detalhe, essa é a verdade. (P3).

A análise dos documentos propiciou, também, verificar diversas orientações e informações que correlacionam a categoria *Carreira* com seus índices: *Benefícios*, *Riscos* e *Esforços* relacionados a Dimensão Pesquisador, apresentado nas próximas subseções.

Mediante o exposto, a hipótese H8.: Normas e Leis influenciam positivamente o compartilhamento dos dados via repositório foi refutada.

8.3.8 Benefícios

Relacionado à Categoria Carreira corresponde aos benefícios percebidos ou esperados pelo pesquisador com o compartilhamento de seus conjuntos de dados. O índice *Beneficio* está relacionado ao direito moral visto que engloba o direito do autor sobre seus conjuntos de dados. O compartilhamento dos conjuntos de dados possibilita maior número de citações de modo a propiciar aumento das métricas e reconhecimento entre seus pares.

O aspecto vai ao encontro das recomendações do Relatório Final e Plano de Ação Europeu para os princípios FAIR. Em seu conteúdo são discutidas as necessidades de desenvolvimento e adoção de conjuntos de métricas FAIR direcionadas à dados de pesquisa e em apoio à Ciência Aberta. Embora esse aspecto esteja em desenvolvimento, a prática de citação de dados é fomentada pelos repositórios e por instituições como FORCE 11 e *DataCite* e aplicada pela comunidade acadêmica.

Assim como demonstrado no índice *Princípios FAIR*, para todos os conjuntos de dados do repositório *e-cienciaDatos* são descritos os metadados nome e ORCID³⁹ dos autores evidenciando a autoria e induzindo a citação dos mesmos quando reutilizarem os dados. De forma similar, oferece uma citação e referência normalizada para cada conjunto de dados. Simultaneamente, o compartilhamento de conjuntos de dados pode propiciar parcerias em novas pesquisas e publicações.

O documento *Directrices para la Gestión de Datos en Horizonte 2020: Versión 1.0 11 de diciembre de 2013* traz em seu teor as diretrizes para que o pesquisador descreva sua política de gestão de dados para solicitar financiamento do *Horizonte 2020* (CONSÓRCIO MADROÑO, 2014).

Para compreender a percepção dos pesquisadores sobre os benefícios relacionados ao compartilhamento dos dados foi apresentada a seguinte questão: “O quanto você concorda com cada afirmação sobre os benefícios ao compartilhar seus dados?”

Quadro 20 – Respostas dos pesquisadores sobre Benefícios

	Pergunta	P1	P2	P3	P4
(a)	Aumenta as minhas citações as métricas	Totalmente de acordo	Moderadamente de acordo	Totalmente de acordo	Totalmente de acordo
(b)	Melhora meu reconhecimento entre pares	Totalmente de acordo	Moderadamente de acordo	Totalmente de acordo	Moderadamente de acordo
(c)	Permite colaboração e publicação coletiva	Totalmente de acordo	Totalmente de acordo	Totalmente de acordo	Moderadamente de acordo

Fonte: Elaborado pela autora.

Referente a afirmação “a) Aumenta as minhas citações as métricas” três pesquisadores estão totalmente de acordo e um moderadamente de acordo. Com a resposta do pesquisador P1, é possível relacionar ao índice Esforços. O pesquisador comenta que a avaliação dos dados pelos pares permite ao pesquisador melhorar as análises tornando-os mais confiáveis e mais rigorosos cientificamente.

Ao mesmo tempo, ao publica-los, **implica um esforço rigoroso, e seus pares [analisaram os dados minunciosamente]** e você [pesquisador] pode até mesmo receber críticas e, a partir daí, você refinar [melhorar a análise, deixar a análise mais exata]. Isso faz sentido, ou seja, **os dados são mais confiáveis e mais rigorosos em termos científicos, se já tiverem sido avaliados** [pelos pares] (P1, grifo nosso).

O problema sempre neste tipo de consideração é que, geralmente é muito difícil ser capaz de demonstrá-lo, isso significa que você teria que ter um universo paralelo no qual eu não tivesse publicado estes materiais e saber se eu era mais conhecido ou não por publicá-lo, por publicá-lo aí. **Eu suponho que sim, que algo aumentou. Particularmente eu não tenha visto.** Eu também não faço muito acompanhamento de ver se foi citado e se em alguma citação é mencionado no repositório, então possivelmente aumenta, mas isso é, digamos, que **eu diga por intuição minha, não porque eu possa manifestá-lo com informações objetivas.** (P2, grifo nosso).

Eu imagino que sim [que citam os dados]. Claro, não fazemos um monitoramento totalmente detalhado porque a política [da instituição onde trabalha] é permitir o acesso sem maiores requisitos, simplesmente, inclusive não é imprescindível que as pessoas se inscrevam para baixar os dados e, em outros casos, sim é solicitado um registro, uma senha etc., então, também, controle sobre quem baixa os dados, para o caso específico do projeto, não é essencial. Nós recomendamos que se registrem para ter uma estatística de quem os usa, com que fins etc, mas não, não temos total garantia de que as pessoas fazem isso. Até certo ponto isso facilita para os usuários baixarem, mas aqueles que geram

dados cria problemas porque tem menos controle sobre como se utilizam, com que finalidade é, por tanto, pois **aí se perde um pouco a possibilidade de enriquecer e adaptar melhor a informação às necessidades dos usuários porque não sabemos com todos detalhes o perfil desses usuários.** (P3, grifo nosso).

Os comentários dos pesquisadores P1 e P3 apresentam relevantes observações sobre o trabalho colaborativo, aspecto que se alinha a afirmação “c) Permite colaboração e publicação coletiva”. Na perspectiva do pesquisador P1, a avaliação dos pares sobre os dados, colabora com o pesquisador pois, com as críticas e observações, os dados podem ser melhorados, enriquecendo a qualidade dos dados. Já na perspectiva do pesquisador P3, a colaboração está sobre o ponto de vista dos usuários, que se beneficiam com os dados disponibilizados. Relevante destacar a preocupação do pesquisador ao mencionar que conhecendo quem são os usuários poderia adaptar e enriquecer as informações para melhor reuso dos dados. Esse ponto se alinha a fala do pesquisador P2 que também considera colaboração sob o ponto de vista de quem vai reusar os dados.

Colaboração coletiva porque aí estão os materiais, é poder pegar sem a necessidade de permissões e autorizações nem de nada, é uma melhoria. (P2).

Outro quesito relacionado aos benefícios é o reconhecimento entre pares. Sendo assim, foi apresentada a afirmação: “b) Melhora meu reconhecimento entre pares”. Entre as respostas, dois pesquisadores estão de totalmente de acordo e dois moderadamente de acordo. Destaca-se que P2 está moderadamente de acordo pois, conforme sua fala anterior, considera que aumenta por própria intuição e não por acompanhar as métricas.

Na mesma linha de pensamento relacionada ao Índice Esforços, para o pesquisador P3, os dados devem ter autoria associada para dar referência ao trabalho, assim como ter referência para as atualizações ou correções dos dados.

Minha opinião sobre direitos autorais e de dados de pesquisa bem, parece-me que é razoável que **dados que requerem muito trabalho gerar e requer muito esforço, tenham uma referência da pessoa que gerou, a equipe de pesquisa que gerou** essa informação, por um lado **para dar reconhecimento ao trabalho** que foi feito e, por outro lado, ter uma referência para possíveis atualizações ou correções daqueles dados. Muitas vezes os dados têm várias versões e vão sendo atualizados porque

se descobrem erros ou problemas no processo e são reprocessados, então é importante que haja uma referência para que se tenha sempre a versão mais recente disponível para pessoas que queiram usá-los. (P3, grifo nosso).

Sim, eu creio que sim, que isso melhora no sentido de que: bom alguém vê que você tenha feito isso e vendo diga, **“olhe eu não conhecia este repositório”, e se interessam [...]** Como venho dizendo, **você faz algo que não é usual em nosso campo, que é publicá-los [os dados] neste tipo de repositório** e publicar os materiais primários e, então, olhe é uma pessoa que, no que eu sei, compilação é feita na medida que faz algo um pouco diferente sim, mas certamente favorece. (P2, grifo nosso).

Bem, eu diria que isso pode aumentar as citações, melhora o reconhecimento entre os meus pares. (P4).

Interessante observar na fala de P2 dois assuntos: primeiro, a visibilidade que os dados podem proporcionar ao repositório. Segundo, a declaração que no campo de conhecimento dele não é habitual publicar em repositórios. No primeiro assunto, diferente do apresentado no índice Repositório onde os pesquisadores opinaram sobre a visibilidade que o repositório proporciona aos dados, a visibilidade vem em sentido contrário, os dados proporcionam a visibilidade ao repositório. Com efeito, em caso de pesquisadores fazerem buscas em metabuscadores de dados ou, até mesmo, de itens documentários, ter entre os resultados os dados do repositório ou os itens que linkam esses dados, os repositórios serão conhecidos por outros e terão maior visibilidade.

Diante do exposto, a hipótese H9.: Benefícios na carreira influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório foi confirmada.

8.3.9 Riscos

Relacionado à Categoria *Carreira* corresponde aos riscos percebidos pelo pesquisador com o compartilhamento de seus conjuntos de dados. Com a análise dos dados, identifica-se que o *Consórcio Madroño* e as bibliotecas em seus diversos documentos mostram os aspectos que podem amenizar os riscos no compartilhamento dos dados.

Referente ao indicador Riscos, os pesquisadores foram solicitados a responder a questão: “O quanto você concorda com cada afirmação sobre RISCOS obtidos ao compartilhar seus dados?”

Quadro 21 – Respostas dos pesquisadores sobre Riscos

	Pergunta	P1	P2	P3	P4
(a)	Não ser citado	Totalmente em desacordo	Moderadamente de acordo	Moderadamente de acordo	Totalmente de acordo
(b)	Perder oportunidades de novas publicações ou produção de novos produtos	Totalmente em desacordo	Totalmente em desacordo	Moderadamente de acordo	Neutro
(c)	Fazer uso dos dados para fins não previstos na licença aplicada	Totalmente em desacordo	Moderadamente em desacordo	Totalmente de acordo	Moderadamente de acordo
(d)	Perder oportunidade de comercialização (patente)	Totalmente em desacordo	Totalmente em desacordo	Totalmente em desacordo	Neutro

Fonte: Elaborado pela autora.

Referente a afirmação “a) Não ser citado”, P1 está totalmente em desacordo, P2 e P3 estão moderadamente de acordo e P4 está totalmente de acordo. Os pesquisadores chamam atenção em seus comentários sob prismas diferentes:

No nosso caso é o problema que seria o mesmo se o estudo seguisse o repositório específico, não o e-cienciaDados, no sentido de que **nós também temos os materiais publicados por outras vias, então quem vai nos citar pois não nos citariam igual por um ou por outro. É, infelizmente, uma prática bastante usual**, eu não vejo isso muito ruim, mas eu não cito, eu não acho que neste caso específico, em nosso caso específico nos afetam, é melhor que eu já diga que o temos publicado em outras fontes e que poderia afetar uma pessoa que não o tivesse publicado, não que fosse inédito, pois ai eles te poderiam te embargar, não mais posso colocar moderadamente de acordo, mas possivelmente outras pessoas que este seja sua única forma de publicar esses dados, que seriam mais totalmente de concordo. (P2, grifo nosso).

Sim, desde o início acho que é **um direito à justiça [ser citado], digamos que se faça menção às pessoas que geraram dados quando utilizados e, por outro lado, é muito conveniente para garantir a rastreabilidade de informações** e que as mais recentes, as mais atualizadas, as mais precisas estejam acessíveis. (P3, grifo nosso).

O prisma apontado por P1 menciona práticas de plágio, o que é observado em pesquisas científicas, onde é possível que o mesmo se estenda aos dados disponibilizados no repositório ou em qualquer outra fonte. Já P3 destaca que a prática da citação dos autores é um direito a justiça, considerando ser um trabalho e esforço dos pesquisadores, além de colaborar com a rastreabilidade dos dados. Com a fala do pesquisador P4 percebe-se sua preocupação de não ter seu nome vinculado com a propriedade intelectual dos dados, aspecto que poderia deixar de ser citado. A continuação dos comentários, destaca-se a fala do pesquisador P4 com a não observância da autoria dos dados, aspecto apresentado no índice Repositório.

Eu acho que sim porque parece que esses dados foram deixados lá por alguém depositado, mas, é claro, os dados em realidade, fazem parte de um projeto de pesquisa que financiaram a mim Sim, estou dando a conhecer [os dados] com todo o prazer, **mas minha autoria não está clara** [...] que não aparece, ou seja, **se o procurar no repositório não sabe que esses dados correspondem a mim**, ao facto de compartilhar os dados, **então isso é incômodo** porque sim, realmente deveria citar por essa referência e não se é citado assim. Eu digo é que se você procurar meu nome em nenhum lugar parece que eu deposei esses dados [...] (P4, grifo nosso).

Referente a afirmação “b) Perder oportunidades de novas publicações ou produção de novos produtos”, destaca-se a fala de um dos pesquisadores o qual comenta que não perderia oportunidades de publicação pois não tem a previsão de novas análises dos dados:

Sim, se pode perder, mas, também não. (Um dos pesquisadores).

Nosso caso não, não nos afeta porque as publicações que nós faríamos seria análise desses materiais, dos materiais primários e, em todo caso, se fôssemos em algum momento editar algum material seriam seleções, **e não publicaríamos um livro que tivesse as trezentas e ímpares entrevistas feitas** em toda a fronteira senão então alguns fragmentos, então seria complementar em qualquer caso. (Um dos pesquisadores, grifo nosso).

Referente a afirmação c) Fazer uso dos dados para fins não previstos na licença aplicada, os pesquisadores comentam que:

Também não sei, é que podem fazer muitos, muitos. (P4).

Mas, como em nosso caso a licença é aberta e a única, é aquela que nós consideramos, **não sei agora mesmo com qual está publicada no repositório e-cienciaDados**, mas digamos que como estamos em conformidade com qualquer tipo de uso, pois não nos afeta que detecta, porque o único cumprimento de nossa licença **que poderia fazer seria a de não citar**, o que falamos, o resto estaria autorizado, incluindo transformação e uso comercial. (P2, grifo nosso).

Assim que, em princípio, estou de acordo que [os dados] **sejam compartilhados com condições determinadas**, não que elas garantam, por um lado, porque a documentação que está bem documentada e que seja, digamos, acessível facilmente, que não requeira custo se levou a outro tipo, não me parece se os dados são adquiridos com projetos públicos deveriam poder ser acessados publicamente também. (P3, grifo nosso).

Referente a afirmação “d) Perder oportunidade de comercialização (patente)”, é interessante ressaltar que para acesso aos dados, e caso seu controle seja estabelecido, é preciso esclarecer se haverá o embargo, e depois de quanto tempo os dados ficarão disponibilizados.

Então para tratar de ter a oportunidade de comercialização ou de transferência de conhecimento, em nosso caso, quando os conteúdos patenteáveis nos aplicariam, e, por outro lado, não são nosso tipo de conteúdo, nada que possa ser comercializado. No caso de outros setores, imagino que o normal, eleger a prática normal seria totalmente compatível porque se eu vou solicitar uma patente esses acessos desses domínios científicos, sabem ou deveriam saber os técnicos, os outros e **a primeira coisa que eles têm que fazer é solicitar a patente e, em seguida, publicar os materiais para que não haja o problema da divulgação**, é dizer que os vejo também compatíveis com esse tipo, de mesmo **para materiais sujeitos a patentes teriam que realizar a patente com a licença e depois publicar em aberto**. (P2, grifo nosso).

Mediante o exposto, a hipótese **H10**.: Riscos na carreira influencia negativamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório é confirmada.

8.3.10 Esforços

Relacionado à categoria *Carreira*, corresponde aos trabalhos decorrentes, o tempo e a energia despendidos com a gestão dos dados e com sua preparação para o compartilhamento.

O *Consórcio Madroño* e o repositório enfatizam a relevância em documentar os dados e fornecem instruções a respeito com indicação, inclusive, de autoria e direitos de uso: “Todos os conjuntos de dados devem ser documentados para serem compreensíveis e reutilizáveis: **autoria**, título, descrição, metodologia, projetos de financiamento, cobertura temporal e espacial, **direitos de uso** e privacidade, etc.” (E-CIENCIADATOS, c2020, grifo nosso, tradução nossa).

No tocante a documentação, é disponibilizado um template, *Readme.txt template*³⁹, para que os pesquisadores preencham e enviem para suas respectivas bibliotecas juntamente com os conjuntos de dados e PGDs. Relevante destacar que essa documentação auxilia os bibliotecários na descrição dos conjuntos dos dados no repositório. Em entrevista com os bibliotecários, foi enfatizado a relevância dessa documentação para a descrição dos dados. Como não há um vocabulário controlado, a documentação favorece a indexação de assuntos com a utilização dos metadados fornecidos pelos autores.

Para atender aos requisitos das agências de fomento, às indicações dos princípios FAIR, as recomendações da Ciência aberta, as instruções do Repositório, a elaboração de PGD e considerar os aspectos éticos e legais é demandada um rol de atividades para o pesquisador na gestão e disponibilização dos conjuntos de dados. Caso o pesquisador solicite financiamento para sua pesquisa, terá em mente o dever em atender os requisitos da agência de fomento relacionados à gestão e compartilhamento dos dados e, por conseguinte, as orientações do repositório em que armazenará seus conjuntos de dados.

Com efeito, ao longo de pesquisas os pesquisadores terão que despender tempo, energia e trabalho com o ciclo de vida de seus dados mesmo que não os disponibilizem ao público. Publicar os dados ou torná-los públicos antes das publicações dos resultados das pesquisas e das análises dos dados pode depender do quanto de esforço o pesquisador teve e como ele avalia o cenário para a publicação. Pode-se verificar esse aspecto na fala do pesquisador:

[...] **os publica [os dados] dependendo do tipo de esforço.** Em nosso caso que tem sido uma longa pesquisa de anos e publicamos [os dados] abertamente em uma página *Web* **depois de ter feito uma série de publicações, não foi uma publicação foi um grupo de publicações.** (P1, grifo nosso).

³⁹ <https://edatos.consorciomadrone.es/readme.xhtml>

Destaca-se que ao ser indagado sobre sua opinião sobre o compartilhamento dos dados, o pesquisador P1 respondeu:

Eu tenho posição muito favorável, uma vez publicado os resultados da pesquisa, ao menos que, possa haver investigações como no nosso caso [...] que publicamos [tipos de dados específicos da área], pode acontecer que a investigação em si seja puramente empírica e o objetivo da pesquisa seja a própria publicação dos dados. (P1, grifo nosso).

Referente ao indicador Esforço, os pesquisadores foram solicitados a responder à questão: “O quanto você concorda com cada afirmação sobre **ESFORÇOS** dedicados ao compartilhamento de seus dados?”

Quadro 22 – Respostas dos pesquisadores sobre Esforços

	Pergunta	P1	P2	P3	P4
(a)	Não tenho as habilidades ou conhecimentos necessários para gerenciar e disponibilizar meus dados	Moderadamente de acordo	Moderadamente em desacordo	Moderadamente de acordo	Moderadamente de acordo
(b)	Terei muito trabalho e outros pesquisadores podem ter mais benefícios como publicação em periódicos com maior fator de impacto	Totalmente em desacordo	Moderadamente de acordo	Moderadamente em desacordo	Totalmente de acordo

Fonte: Elaborado pela autora.

Referente a afirmação “a) Não tenho as habilidades ou os conhecimentos necessários para gerenciar e disponibilizar meus dados” a maioria dos pesquisadores estão moderadamente de acordo. Os pesquisadores comentam que:

Eu, neste momento eu tenho [certa habilidade]. Em um tempo eu tive algum problema de preguiça por falta de habilidades porque significa um esforço e um esforço de tempo significa, tal que como dissemos antes, **recompensa em termos de prestígio**. Estou moderadamente de acordo porque não sou especialmente habilidoso. (P1, grifo nosso).

Bem, é claro, concordo moderadamente. Concorde porque as possibilidades do repositório são maiores das que eu posso gerar e ter. (P4).

Destaca-se que P1 faz uma declaração que se alinha ao índice Benefício, pois seus esforços trazem recompensas como seu prestígio entre seus pares. Outro aspecto levantado pelo P4 é a capacidade do repositório na gestão de dados que, para o pesquisador, sua equipe está melhor preparada.

Referente a afirmação “b) Terei muito trabalho e outros pesquisadores podem ter mais benefícios como publicação em periódicos com maior fator de impacto” há divergência entre as respostas, sendo:

P1 totalmente em desacordo

P2 Moderadamente de acordo

P3 Moderadamente em desacordo

P4 Totalmente de acordo

A respeito do assunto, faz jus serem ressaltados alguns comentários dos pesquisadores:

Digamos que não, não é um assunto que me preocupe, pelo contrário, digamos, eu acho que para as pessoas que faço um esforço e que alguém possa lucrar com meu esforço, acima do meu no caso e, neste caso, digamos que sempre serei feliz, mais do que feliz que as pessoas usam os dados e **para mim é um êxito**. (Um dos pesquisadores, grifo nosso).

Muito bem eu atualmente não suponho problemas com essa suposição. Podem supor um problema, mas, geralmente não é. A pessoa ou nível profissional o que está aqui, quem pode decidir como gerenciar dados de um projeto e uma pessoa em estágios anteriores que publica, libera os materiais e **acontece que outra pessoa publica, e essa pessoa é uma concorrente direta em uma bolsa de pós-doutorado ou em contrato de professores acima dessa pessoa, tem utilizado dados e eu a ajudei para que tenha mais publicações, não vai prejudicar tanto, é para que eu pudesse me arrepender**, aqui agora se alguém com isso ele obteve um fator maior de impacto, estou feliz por ele, ótimo, mas me dá mais satisfação saber que meus materiais foram usados e que eles servem como algo que ele não inveja que tem uma revista de um fator de impacto um ou dois e eu tenho que me contentar com uns quatro ou mais, sim bem, então damos isso. (Um dos pesquisadores, grifo nosso).

[...] Você não entende, eu estou dando esses dados gratuitamente porque eu estou cuidando de [fazer todo o trabalho de organização] e o autor intelectual está relacionado [a outra pessoa], mas minha participação eu não sei reflete aqui. (Um dos pesquisadores).

A fala de um dos pesquisadores revela preocupações no compartilhamento dos dados quando outros se beneficiaram em casos como concorrente em bolsa de estudo ou contrato profissional. Percebe-se que são fatores que também influenciam a carreira e que os pesquisadores podem reter os dados pois também percebem os riscos associados com seus esforços. Em outra fala, o pesquisador novamente demonstra sua preocupação em não ter seu nome associado a autoria dos dados, aspecto que pode causar entendimento ao autor que outros utilizaram seus dados e assinarem como autores. Outro pesquisador declara que se seu esforço favorecer outro pesquisador, ele fica feliz. Todas as falas se conectam com os outros índices da carreira do pesquisador.

Sendo assim a hipótese **H11**.: Esforços na carreira influenciam negativamente no compartilhamento dos dados via repositório foi refutada.

8.3.11 Suporte dos Bibliotecários

Relacionado à categoria *Recursos*, corresponde ao suporte dos bibliotecários na gestão dos dados, no trabalho do repositório, nos treinamentos dos pesquisadores e nas orientações dos pesquisadores sobre suas necessidades de dados. O suporte dos bibliotecários está presente no cenário apresentado pelo *framework*. Relevante destacar que há manuais, infografias e vídeos elaborados pelas bibliotecas com informações e instruções aos usuários que incluem a propriedade intelectual.

A análise das entrevistas com os bibliotecários revelou o potencial de seu trabalho no contexto dos repositórios de dados. Os bibliotecários do *e-cienciaDatos* participaram das reuniões, do diagnóstico, da projeção do repositório e elaboração do texto do projeto, da implementação do *software*, dos testes e calibração para implementação do repositório, da elaboração das políticas e manuais. Destaca-se que suas habilidades e competências inerentes ao trabalho com os repositórios institucionais e suas atividades foram primordiais nesse processo.

Os bibliotecários oferecem treinamentos e orientações a sua comunidade sobre a gestão de dados assim como depósito de dados e, a questão dos direitos autorais é apresentada, expondo as possibilidades da cessão, da disponibilização e da reutilização dos dados bem como a relevância da citação, e podem exercer influência

positiva sobre os pesquisadores no compartilhamento dos dados. Para tal, as orientações e treinamentos dos bibliotecários devem ser claros e abordarem os diversos aspectos do direito autoral, assim como a visibilidade que o repositório propicia, as formas de recuperação dos dados, os metabuscadores, os repositórios institucionais e os catálogos das bibliotecas.

O papel dos bibliotecários foi apresentado juntamente com o índice *Repositório* devido ao contexto das discussões, inclusive as que surgiram com um dos pesquisadores. O índice *Suporte dos bibliotecários* está na *Dimensão Pesquisador* pois é possível que a percepção dos pesquisadores sobre papel e trabalho dos bibliotecários sejam diferentes da que tem sobre os repositórios e, também, porque o Repositório está alinhado à categoria *Normativo*.

Referente ao indicador Suporte dos Bibliotecários, os pesquisadores foram solicitados a responder a questão: “O quanto você concorda com a afirmação?”

Quadro 23 – Respostas dos pesquisadores sobre Suporte dos Bibliotecários

Pergunta		P1	P2	P3	P4
(a)	Participei em capacitações ou orientações que abordam os direitos autorais de dados	Moderadamente de acordo	Totalmente em desacordo	Totalmente em desacordo	Moderadamente de desacordo

Fonte: Elaborado pela autora.

Destaca-se que a pergunta acima foi a apresentada na seção 8.3.6.

Sendo assim, a hipótese **H12**: Suporte dos Bibliotecários influencia positivamente no compartilhamento dos dados via repositório foi refutada.

8.3.12 Benefícios para a sociedade

Índice relacionado à categoria *Social*, tanto sociedade civil quanto científica. Está alinhado às necessidades de disponibilização dos dados, por exemplo, para a agilidade em respaldar determinada demanda da sociedade, rapidez na análise situacional, tomada de decisão, elaboração de políticas, avanço em outras pesquisas,

início de novas pesquisas, entre outros. A disponibilização de dados contribui com a sociedade científica para maximizar o reuso dos dados. Esse aspecto se relaciona com o direito patrimonial pois na disponibilização dos dados podem ser aplicadas licenças de uso.

As políticas de informação trabalham com a transferência da informação científica no qual se tem o Estado como agente na sua elaboração e estabelecimento e, a ciência e a tecnologia, como domínio de seu exercício, além de definição e implementação de diretrizes para as necessidades que emergem na sociedade da informação. Essas políticas respaldam a elaboração e atualização dos documentos apresentados do *Consórcio Madroño* e das bibliotecas.

Na sociedade que necessita de dados e informações, políticas de informação, normas e leis voltadas para a proteção da propriedade intelectual se apresentam como instrumental relevante para o desenvolvimento científico e tecnológico.

A coleta de dados revelou que os bibliotecários devem dialogar constantemente com os usuários para identificar suas necessidades de dados, além de auxiliar para identificação das palavras chaves dos conjuntos de dados para que os mesmos sejam melhor indexados no repositório. Conhecer a sua comunidade também auxilia o bibliotecário na preparação dos materiais de orientação, instrução e *marketing*, pois refletem os aspectos de maior interesse ou dificuldade dos pesquisadores.

Destaca-se que na entrevista, os bibliotecários enfatizaram a importância do diálogo constante com os pesquisadores, assim como a necessidade de se fazer um questionário para conhecer a opinião dos mesmos sobre o repositório.

Ao longo das entrevistas os pesquisadores mencionaram diversos fatores que beneficiam a sociedade e que foram apresentados nas discussões dos outros índices. Inclui-se às discussões outras opiniões dos pesquisadores sobre o compartilhamento dos dados:

[...] para que mais pessoas possam manejá-los e tenham acesso mais fácil e rápido [aos dados] e **possam olhar para pesquisas que, em princípio, não estavam previstas** quando eu os disponibilizei para o Comunidade. (Um dos pesquisadores, grifo nosso).

Bom, em princípio minha ideia é que esses dados sejam úteis para outras pessoas, **algo de que não precisem voltar a fazer** o que já fizemos e, também, na medida em que alguns dados podem não ser completamente conhecidos por essas pessoas, pois que isso **também pode dar lugar a projetos em conjuntos ou colaborações e que**

gerou alguns dados, pois conhece bem seus pontos fortes e suas limitações e uma pessoa que irá usar externamente, mesmo que se leia a documentação, pode não ser tão ciente disso. Assim, geralmente os dados públicos também são fruto de colaboração porque as pessoas que vão usá-los são especialistas em outros temas e pedem ajuda, pedem colaboração. Então bom, em muitos casos, pois isso permite fazer publicações conjuntas com outros especialistas e bem isso, também eu creio que tenha seu interesse, mas muito além disso no nosso caso, pelo menos tem mais, existem mais de 2.000 downloads dos produtos e obviamente não temos participado de 2.000 artigos com essas pessoas. Assim, que bom pois tem uso muito variado, muitas vezes que nós mesmos não conhecemos. (Um dos pesquisadores, grifo nosso).

Interessante notar a visão dos pesquisadores sobre as colaborações de pesquisas. Na fala de um pesquisador, podem surgir projetos em conjuntos abordando a internacionalização da ciência, o trabalho cooperativo, troca de informações e de conhecimentos, aspecto presentes na Ciência aberta.

Sendo assim, a hipótese **H13**.: Benefícios para a sociedade influencia negativamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório é refutada.

A hipótese **H14**: Benefícios para a sociedade influencia positivamente no compartilhamento dos dados via repositório é confirmada.

8.3.13 Ética

Relacionado à categoria Social indica, o conjunto de valores de uma área do conhecimento e seu comportamento considerado apropriado ou impróprio pela comunidade, juntamente aos deveres, os princípios e as normas que são compatíveis aos grupos dos pesquisadores em seu contexto. Oportuno enfatizar que o índice *Ética* perpassa os outros índices do *framework*, assim como o índice *Cultural*.

As condições éticas e legais permeiam uma séria de fatores que devem ser levados em consideração ao longo do ciclo de vida dos dados. Nesse aspecto, as orientações do repositório respondem a seguinte questão em seu FAQ: “Que questões éticas e legais existem sobre os dados?”

A produção, recolha, gestão e divulgação dos dados obedecem a um quadro legal que deve ser respeitado. Os conjuntos de dados podem ser afetados por disposições legais referentes, entre outras, à **propriedade intelectual**, propriedade industrial, privacidade ou sigilo e proteção de dados pessoais.

(CONSÓRCIO MADROÑO, c2018, tradução nossa, grifo nosso).

No documento *Directrices para la gestión de datos en Horizonte 2020* em sua versão 1 de 11 de dezembro de 2013 em seu *Anexo 1: Plantilla del Plan de Gestión de Datos (PGD)* no item *Datos compartidos* há orientações sobre o índice Ética: “[...] Caso os dados não possam ser compartilhados, as razões devem ser mencionados (por exemplo, **ética**, jurídica, proteção de dados pessoais, **propriedade intelectual**, comercial, privacidade e segurança ...).” (CONSÓRCIO MADROÑO, 2014, p. 6, tradução nossa, grifo nosso).

O compartilhamento pode seguir os aspectos éticos relacionados à propriedade intelectual, respeitando as licenças aplicadas aos dados, assegurando os interesses da ciência, da comunidade científica, da sociedade e do pesquisador, além da citação dos autores seguindo as boas práticas científicas.

Com a análise dos dados foram identificadas informações que orientam tanto no uso de licenças quanto na citação dos dados. A citação dos dados está ilustrada em uma das infografias disponibilizadas conforme figura 26, elemento relacionado ao índice Benefícios. A infografia fornece modelo de citação dos conjuntos de dados, informações sobre a importância da citação e das boas práticas da citação. Interessante destacar que no repositório para cada conjunto de dados é fornecida a referência normalizada, o DOI e a publicação em que os dados estão relacionados.

Figura 26 – Orientações sobre citação de dados

Cita tus datos de investigación

Por qué es importante citar los datos:

- Los conjuntos de datos también son resultados de investigación como los artículos, monografías, etc.
- Facilita la identificación y el acceso a los datos y de esta forma su localización, validación y reutilización.
- Permite reconocer la autoría de sus creadores.
- Facilita la métrica e impacto de los datos.
- Favorece la transparencia de la investigación científica.

Buenas prácticas para citar datos:

- Se debe facilitar la identificación, localización y el acceso a los datos mediante un identificador único y persistente (DOI, Handle, etc.).
- Cada conjunto y subconjunto de datos (dataset) debe citarse de forma independiente.
- Las citas de los datos utilizados han de aparecer en la sección de referencias bibliográficas de la publicación resultante.
- Se recomienda incluir un identificador único de autor (ORCID, etc.).

Elabora los datos con los documentos resultado de investigación y viceversa, y crea las referencias bibliográficas de los mismos.

Elaboración de la cita

Existen elementos mínimos obligatorios (O) y otros recomendados (R) que se combinan para elaborar la cita en cualquier estilo estándar (APA, MLA, Chicago, etc.) o los propuestos por los principales repositorios de datos (Dataverse, Dryad, etc.).

Elementos de la cita:

- Autor(es) (O)
- Identificador autor (R)
- Fecha (O)
- Título (O)
- Identificador único persistente (O)
- Tipo de recurso (O)
- Versión y/o Edición (O)
- Repositorio de datos (R)
- Publicación (R)
- Productor (R)
- Ámbito geográfico (R)
- Ámbito temporal (R)

Ejemplo de cita estilo APA

Autoría: Remesar Betllloch, X., Antelo, A., Llivina, C., Alba, E., Berdié, L., Agnelli, S.,... Alemany, M.
 Título: (2015). *Influence of a hyperlipidic diet on the composition of the non-membrane lipid 6 pool of red blood cells of male and female rats*. [Dataset]. Versión de 22 de junio de 2015.
 Diposit digital de documents de la UAB: <http://hdl.handle.net/2445/66010>

Componentes de la cita:

- Fecha
- Repositorio
- Tipo de recurso
- Identificador único y persistente
- Versión

El personal de tu Biblioteca te puede asesorar

crue Universidades Españolas Red de Bibliotecas REBUN

CC BY

Fonte: REBUIN (2016b).

No documento *Cómo depositar sus datos de investigación en e-cienciaDatos* em seu ítem *Cómo citar los datos*, é ressaltada a importância da citação dos dados: “É importante citar corretamente os dados para favorecer sua reutilização em novas pesquisas, em atividades acadêmicas e para outros fins da sociedade em geral.” (CONSÓRCIO MADROÑO, 2019, p. 3, tradução nossa). Ademais, junto ao documento é indicado um endereço eletrônico com instruções para modelo de citação dos dados: “Como citar dados de pesquisa corretamente: <http://www.rebiun.org/acceso-abierto/cita-tus-datos-de-investigacion>” (CONSÓRCIO MADROÑO, 2019, p. 3, tradução nossa). Pertinente enfatizar que esse aspecto está diretamente relacionado ao índice *Benefício*.

Fatores como pressão da comunidade científica ou das exigências da agência de fomento podem levar os autores a renunciarem ou não a todos os direitos de autor colocando seus dados em domínio público. Em relação ao aspecto apresentado, na fala de um dos pesquisadores é possível notar ambos fatores: pesquisadores que

concordam em colocar seus dados em domínio público e a pressão institucional:

Nós, como muitos, podemos aspirar [aos direitos autorais] e o fazemos mais por estas obrigações acadêmicas de usar os direitos morais de autoria. Temos uma página na web, porque [os usuários] têm que nos citar dessa forma. Digamos que seria a postura que adotamos e a postura institucional como projeto, **mas, a verdade é que muitas vezes eu estaria incluso mas perto da postura de liberação para o domínio público, inclusive abrir mão do direito de citar** [...] digamos que nós mantemos, acima de tudo, para esse assunto acadêmico, e que depois veem e nos perguntam: como se cita? como colocamos? que impacto tem? e como é que nós mantemos? (P2, grifo nosso).

No argumento do pesquisador P2, há dois fatores: primeiro, não é unânime que pesquisadores estão preocupados em serem citados quando terceiros utilizarem seus dados, aspecto apresentado no índice *Ciência Aberta* quando indagados sobre aplicar a licença CC0; segundo, as instituições demonstram a relevância do reconhecimento dos pesquisadores, com a citação, pois as citações aumentam as métricas e o impacto das pesquisas e, inclusive, propiciam aumentar a visibilidade tanto de pesquisadores quanto das universidades nas quais estão vinculadas.

Os pesquisadores foram indagados sobre o compartilhamento de dados em acesso aberto. As seguintes afirmações foram apresentadas:

Quadro 24 – Respostas dos pesquisadores sobre Ética

	Pergunta	P1	P2	P3	P4
(a)	É dever ético dos pesquisadores compartilhar dados em acesso aberto	Totalmente de acordo	Totalmente de acordo	Totalmente de acordo	Totalmente de acordo
(b)	Os dados compartilhados podem ser usados para fins comerciais	Totalmente em desacordo	Totalmente de acordo	Moderadamente de acordo	Moderadamente em desacordo

Fonte: Elaborado pela autora.

Referente a afirmação “a) É dever ético dos investigadores compartilhar os dados em acesso aberto”. Nesse aspecto, todos os pesquisadores estão totalmente de acordo. Destaca-se que nos depoimentos anteriores, os pesquisadores enfatizaram que só compartilharam seus dados depois das publicações e, em algumas falas, foram enfatizadas que só depois.

Então deixando de lado essa dificuldade prática [da dificuldade de financiamento para armazenamento dos dados] pois, desde logo, nossa postura é que devem ser, se **devem compartilhar, mas não mobilizar moralmente, eu acho que deveria ser obrigatório, mas não posso exigir a partir do momento em que vejo, que descobri que não há infraestrutura suficiente** para compartilhar esses dados e imediatamente torná-los disponíveis para todo mundo e sem custo, sim, sim pode parecer que a prática continua. (Um dos pesquisadores, grifo nosso).

No depoimento de um dos pesquisadores destaca-se o fato de ser obrigatório o acesso aberto, mas salienta que para isso carece de infraestrutura gratuita. Novamente foi levantada a discussão da gratuidade tanto da publicação quanto do acesso aos dados. Em depoimento de outro pesquisador foi apresentada a preocupação de o sistema da publicação científica dos dados se alinhar ao mesmo da publicação tradicional em que periódicos comerciais fiquem como detentores dos dados e a comunidade científica ter como opção o mesmo: pagar para publicar os dados ou para acessá-los.

Na continuação de seu depoimento, o pesquisador entrelaça o compartilhamento e a obsolescência tecnológica. Na visão do autor, depositar os dados no repositório contribui para diminuir possibilidades de perder os dados por não os migrar de suporte.

É óbvio que hoje em dia tem guardado um disquete de computador de 20 anos atrás não sabe o que fazer com ele, ter a possibilidade de abrir o arquivo e ser capaz de lê-lo é muito, muito escassa. Em nosso caso que podemos tornar conhecido como se conservarão os dados dentro de 20 anos, se eu salvar meus dados em CD-Rom, por exemplo, e **o armazeno no meu “armário” poderá ser aberto em 20 anos? difícil não?** porque hoje em dia computadores de fábrica ou os próprios CDs são gravados por toda parte de tempo então, nossa perspectiva, embora busquemos em um repositório por uma iniciativa desse tipo de compartilhamento, que esse tipo de repositório, em princípio, são garantidos por 70 a 80 anos de permanência. **Fazendo nós mesmos nos custariam muitíssimo ter em uma página web** e, também, possivelmente teríamos morrido antes desses 80 anos porque certamente iria pular e também muitos casos de não, não é o caso eu entendo da ciência de dados, mas sim repositórios mais especializados de linguística sim que estão garantindo. Existem técnicos que vão periodicamente vendo quais avanços tecnológicos no formato dos arquivos vão realizando e adaptando que transformam esses meios, isso significa dizer que, si em cinco anos aparecer um novo padrão de arquivos de áudio que ultrapassem a web, pois, possivelmente esses técnicos se encarregaram de transferi-lo com o qual sempre estaria acessível a qualquer pessoa com uma tecnologia habitual desse período. Digamos que seria o terceiro nível, a terceira coisa que nós buscamos nos dados, a durabilidade ao longo do tempo, e que o acesso siga sendo possível dentro de 15, 20, 40, 60 anos, [essa visão também é da minha equipe] Em geral nós concordamos bastante com essa abordagem, mais especificamente, sobre a última, sobre a parte tecnológica, um membro de nossa

equipe trabalha precisamente em um arquivo linguístico em um repositório, mas não generalista, como é o e-cienciaDatos, mas específico para documentação linguística o qual é uma pessoa que não apenas no olhar de citações ideológicas sobre o acesso aos dados em si coincide, mas sabe muito bem os problemas de obsolescência tecnológica, e o que podemos deparar dentro de 5 ou 10 anos, de não poder acessar os dados. (Um dos pesquisadores, grifo nosso).

Um ponto de discussão entre a comunidade científica é o uso de dados que foram coletados com financiamento público ou por pesquisadores de instituições públicas serem usados para fins comerciais, aspecto constatado em muitos casos de publicações de resultados das pesquisas ou mesmo publicação dos dados em periódicos comerciais. Para conhecer a opinião dos pesquisadores sobre a temática, foi apresentada a seguinte questão: b) Os dados compartilhados podem ser usados para fins comerciais

P1 está totalmente em desacordo;

P2 está totalmente de acordo;

P3 está moderadamente de acordo;

P4 está moderadamente em desacordo.

Nota-se uma divergência entre as respostas dos pesquisadores. Relevante destacar as considerações apresentadas. Um dos pesquisadores fez uma reflexão sobre uso comercial, porém com enfoque diferente do sistema de negócio vigente dos periódicos científicos. Na reflexão do autor, o uso comercial não seria apenas de editoras comerciais e, também, não incluiria, necessariamente, o lucro. Se forem aplicadas licenças que restringem fins comerciais, como a licença CC BY-NC-AS, não é possível trabalhos como o destacado pelo pesquisador.

Este formato para fins comerciais, especialmente considerando que, ademais, é isso que muitas vezes as coisas das quais pensamos não ter uma finalidade comercial ou legal, o são. Nós, por exemplo, estávamos preocupados, tínhamos muito presente o caso de nossas comunidades de fronteira, pensamos, por exemplo, que **se alguém decidiu fazer um livro de compilação, contos e histórias da aldeia, e utilizou nossas entrevistas [nossos dados] e foi o livro típico que se vendia nas festas da aldeia, assim a preço de custo, mas não deixava de ser um fim comercial**, levando em consideração a legalidade da alínea, então foi mais um estímulo, mas digamos para eliminar essa cláusula. (Um dos pesquisadores, grifo nosso).

Outro pesquisador reflete sobre o papel das instituições públicas em nível nacional e internacional para o financiamento de pesquisas e a utilização de materiais de pesquisa com fins comerciais, além da forte cultura enraizada que gira em torno dessa dinâmica. O pesquisador faz um destaque de o sistema ser “imoral”:

Então esse é um sistema [apresentado acima], que as universidades lhes dizem: tem que ser feito a nível global; e depois as faculdades pagam com os impostos dos espanhóis, pagam as assinaturas de revistas e a universidade tem que pagar a assinatura de uma revista que nós fazemos. **É completamente ridículo, ridículo não, é imoral, mas é um sistema muito difícil de desarticular.** É um grande debate, está de fato sendo tentado pelo União Europeia, mas com muito pouco, muito pouco eficaz porque os próprios pesquisadores não querem parar de fazer isso até que nos garantam que teremos financiamento, e como esse financiamento é um tema muito complicado, é um tema de competitividade, há revistas que têm 40 anos 50, 60, 100 anos, sim e esse capital humano acumulado dentro da revista e o prestígio que adquire é um capital público que são explorados pelos distribuidores privado, e ao público é feito com pesquisadores públicos ao longo de um século, até talvez 50 anos, em nosso caso, um fundador que já morreu da revista trabalhou para a universidade pública da qual está obtendo benefícios [...] sim e o esforço que fazemos gratuitamente, porque nós fazemos a revisão por pares, fazemos todo o processo. (Um dos pesquisadores, grifo nosso).

Além do exposto, o pesquisador fez uma reflexão sobre o modelo vigente dos periódicos comerciais, a necessidade de pagamento pelas universidades para ter acesso aos artigos publicados, em muitos casos, por pesquisadores da própria instituição. Ademais, destaca a forte cultura desse sistema com possível extensão para os dados.

É um tema para o qual vamos... **é um tema muito, muito forte** porque eu tenho sido editor de revista que está [em um dos grandes editores comerciais] e nos financia e nos permite financiar a publicação, e ele o faz vendendo nossa revista em distribuição internacional e temos um pagamento por isto, que é, um não sei se é, suponho que seja cerca de 20%, 30% de tudo o que eles ganham, mas isso nos permite financiar uma publicação científica. **Isso também pode ocorrer com os dados.** É dizer que há pessoas que financiam websites, etc. **Ocorre sim mantendo os dados no sagrado e eu acredito que é dizer que isso é muito difícil de mudar porque, para isso mudar, e o público, é dizer a União Europeia o estado espanhol, em nosso caso, deveria financiar os periódicos,** se não querem que nós vamos para o privado tem que financiá-las. (Um dos pesquisadores, grifo nosso).

O pesquisador termina com a apreensão de acontecer o mesmo com os dados e menciona a necessidade de conscientização dos pesquisadores e a relevância

das discussões da Tese para o atual contexto, e menciona mais uma vez que a apropriação dos dados com fins comerciais é uma prática “imoral”:

Sim, parte disso pode acontecer com os dados, não é um pressuposto nacional, mais porque eu digo que isto não é exatamente sobre os dados, e no caso de dados que eu não posso aprender esta lição, que é **não deixar o privado explorar comercialmente do esforço público de pesquisa**. Esta é a escolha, por isso tem uma ligação: as publicações científicas com a publicação científica dos dados tem uma conexão de aprendizagem porque os dados estão em voga, ora cada vez mais coisas estão sendo feitas com dados, os dados se tornaram o big data, é seu negócio e sim portanto, **esta questão é uma questão de grande relevância, é por isso que eu quis participar nesse tema** [se referindo a participar da pesquisa da tese] **é um tema de grande relevância para o futuro** porque tem a ver com tudo, tem a ver não somente com publicações científicas e de dados de todos os tipos, o **que fazer com os dados se a quem pertence as informações é uma questão muito importante** há um tema, **há um tema que é cultural, é verdade e mais importante porque é o que nos leva ao senso comum** e há um tema em alguns países mais do que em outros e de não estar consciente de que o financiamento da pesquisa, portanto a **maioria dos dados, é pública, o que torna imoral tornar os dados privados, é apropriar-se do esforço do público para obter o benefício ou rentabilidade disso**, e isso me parece ser um assunto de amplitude, se você pega as porcentagens de pesquisa que o público financia, na maior parte dos países superam 80%, pessoas onde há países onde elas estão nos 90%, outros estão em 79% em 81% e há pessoas que estão em 90%, a maioria do grosso do financiamento sanitário, [por exemplo]. (Um dos pesquisadores, grifo nosso).

O depoimento de um dos pesquisadores revela a possibilidade de colaboração na pesquisa, vantagem para a equipe e para outros pesquisadores.

Em nosso caso sim, nós temos a vontade disso [de publicar outros dados que serão coletados ao longo do projeto]. Nós conversamos, fomos confrontados com o problema, bem, até que ponto vamos publicando os dados antes que o projeto termine? bem, alguém pode vir e utiliza-los não sei para quê, para as coisas deles mesmo, bem se veem não acreditamos porque progredimos e antes somos poucos nesse campo, mas bom se vier alguém bom estupendo que vantagens nos ofereceríamos para poder publica-los durante o projeto? Por um lado do ponto de vista ético, o que já tínhamos conversado sobre mostrar às pessoas, eles têm que perder o medo de publicar os dados e como é ótimo Espanhol que diz o melhor pregador é exemplo de batalha, nós fazendo, pois mostramos a nossa aparência, importante, e nós o fazemos e não temos nenhum problema em fazê-lo enquanto se está fazendo o projeto. Em segundo lugar, não tem sido o caso, **mas bem poderia que alguma pessoa teria, e que bom nos sugerissem formas de tratar os materiais, se juntar à equipe a partir de outras perspectivas**. Outras perspectivas sim, já nos aconteceu, mas bem, não diretamente, mas **por causa dele e não apenas por causa da publicação dos dados**, mas já houveram algumas vezes comunicados de imprensa da universidade que falavam sobre o projeto e, também, foi mencionado onde os dados estão publicados, publicado na página web da universidade e pessoas da própria universidade nos contactaram e disseram

bem, eu sou dessa cidade que o senhor entrevistou, e que tem isso, podemos oferecer-nos para colaborar, se chegar a um dia em que apresentado às pessoas, então pode acontecer que o nosso caso tenha sido por causa de comunicados de imprensa, mas poderia acontecer que alguma pessoa que isso não está lá e, pelo olhar, está a falar aqui sobre o meu povo, vou escrevê-lo e vou busca-lo e se isso for feito durante o desenvolvimento do projeto, pois é melhor que essa pessoa possa ajudar-nos a completar as coisas, podem ajudar-nos a interpretá-lo de outra forma, uma vez encerrado o projeto. Então, não há forma de voltar atrás. (Um dos pesquisadores).

O compartilhamento dos dados, a valorização da produção coletiva do conhecimento e a colaboração em pesquisas são pontos presentes na ciência contemporânea, aspectos essenciais para o avanço do conhecimento.

Diante do exposto, a hipótese H15.: Ética influencia positivamente o compartilhamento dos dados via repositório é confirmada.

Quadro 25 – Resultado das hipóteses

ÍNDICE		HIPÓTESE	CONFIRMADA	REFUTADA
H1	Cultura	A cultura influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório.		X
H2	Ciência Aberta	Ciência Aberta influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório.		X
H3	Princípios FAIR	Princípios FAIR influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados.	X	
H4	Plano de Gestão de Dados	Plano de Gestão de Dados influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório.		X
H5	Agências de fomento	As agências de fomento influenciam positivamente a percepção dos pesquisadores fomentando o compartilhamento dos dados via repositório.		X
H6	Repositório	Repositório influencia positivamente o compartilhamento dos dados.	X	
H7	Repositório	Repositório influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados.		X
H8	Leis e Normas	Normas e Leis influenciam positivamente o compartilhamento dos dados via repositório.		X
H9	Benefícios	Benefícios na carreira influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório.	X	
H10	Riscos	Riscos na carreira influencia negativamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório.	X	
H11	Esforços	Esforços na carreira influenciam negativamente no compartilhamento dos dados via repositório.		X
H12	Suporte dos Bibliotecários	Suporte dos Bibliotecários influencia positivamente no compartilhamento dos dados via repositório.		X
H13	Benefícios para a sociedade	Benefícios para a sociedade influencia negativamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório.		X
H14	Benefícios para a sociedade	Benefícios para a sociedade influencia positivamente no compartilhamento dos dados via repositório.	X	
H15	Ética	Ética influencia positivamente o compartilhamento dos dados via repositório.	X	

Fonte: Elaborada pela autora.

Os resultados foram apresentados considerando os aspectos indicados ao longo das discussões da tese.

A tese investigou como fatores institucionais, recursos, crenças, atitudes e normas afetam as percepções e atitudes de pesquisadores no compartilhamento de dados de pesquisa. Os resultados mostraram que a cultura perpassa os diversos índices apresentados na Tese e, na comunidade científica, influencia significativamente os comportamentos de pesquisadores. As influências cognitiva, normativa e reguladora demonstraram ter uma relação expressiva com as atitudes dos pesquisadores.

As instituições consistem de estruturas cognitivas, normativas e reguladoras, os três pilares apresentados na Tese e que compuseram o *framework* (FIGURA 13) e provêm estabilidade nos comportamentos sociais onde cada pilar fornece uma base para a legitimidade (SCOTT, 1995). Segundo Scott (1995) a legitimidade não é apenas uma “*commodity*” a ser possuída ou trocada, mas uma condição que reflete o alinhamento cultural, suporte normativo ou combinação com regras e leis. Estabelecendo um paralelo entre as discussões de Scott e as discussões sobre a Teoria do Consenso de Merton, o último considera que a sociedade alcança a ordem através de um consenso e, a partir desse ponto, todos os participantes comungam de valores e normas comuns. Para o autor, as leis são criadas e aplicadas de forma imparcial e todos são tratados de forma igualitária, e é nisso que a teoria se baseia (MERTON, c1968).

Scott (1995) defende que é possível instituições serem suportadas por estruturas sociais, que estão estabelecidas em expectativas padronizadas em que indivíduos tem papéis e posições sociais determinados. As estruturas forçam e fortalecem o comportamento dos atores ao mesmo tempo em que eles são reproduzidos e transformados por esse comportamento (SCOTT, 1995).

Em uma sociedade há valores dominantes, valores culturalmente aceitos, os quais a maioria elege como ideais para uma convivência harmônica. Nesse aspecto, a comunidade acadêmica, em muitas áreas do conhecimento, tem como valores dominantes ou como cultura a retenção dos dados da pesquisa e publicação dos resultados na comunicação científica tradicional, como livros, capítulos de livros e artigos científicos.

A influência dos valores culturais demonstrou ter relação significativa com a percepção e atitudes em relação ao compartilhamento dos dados, fato observado pelo

número de conjuntos de dados e número de projetos que depositaram no repositório em relação a quantidade de universidades participantes no *Consórcio Madroño*. Isso sugere que a influência cultural influencia direta e indiretamente os comportamentos reais dos pesquisadores.

Com relação à categoria Normativo (Ciência aberta, Princípios FAIR, PGD, Agências de fomento e Repositórios) percebe-se que a introdução e ampla divulgação dos princípios e concepções da Ciência aberta e princípios FAIR, além das orientações e exigências das agências de fomento em seus vários tópicos estão primando pelo modelamento da cultura das comunidades científicas para o compartilhamento dos dados, sendo sua disponibilização fortemente incentivada via repositórios ou publicação de acesso aberto. Cabe ressaltar que a Ciência envolve sinergia de governos, agências de fomento, comunidade científica e pesquisadores a fim de tornar o processo da comunicação científica e seus resultados, incluindo os dados, acessíveis e com o mínimo de restrições ao público (ALBAGLI, 2015; ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT, 2015). A Ciência aberta e os princípios FAIR demandam mudanças na cultura de todos os envolvidos e no *modus operandi* da pesquisa e sua comunicação. Isso significa que a comunidade científica presencia transformações com a nova configuração. Os resultados mostram que, apesar de todos os índices do *framework* estarem presentes há anos no contexto espanhol, os pesquisadores têm pouca clareza, conhecimento ou incompreensão a respeito de suas regras, condutas e funcionamentos, se instaurando a anomia.

Os achados da análise dos dados se fundamentam na ideia da teoria da anomia de Robert Merton e Emile Durkheim. A teoria de Durkheim teve sua gênese no período pós-revolução Industrial, final do século 19 e início do século 20, momento em que ocorriam mudanças sociais com o êxodo rural, ocasionando transformação social e, também, o enfraquecimento das normas, as quais a maior parte da população não conseguiam compreender as mesmas normas e os mesmos valores. Para o autor, o conceito de anomia enfatiza a incapacidade sofrida pelo indivíduo de reconhecer as regras sociais que, depois de afetadas por mudanças repentinas na estrutura normativa, dificultam indivíduos a se submeterem a elas. Correlativamente as considerações de Durkheim, o contexto dos pesquisadores revela transformações com as novas práticas científicas, incluindo a gestão dos dados para o compartilhamento. As respostas dos pesquisadores mostram a pouca aptidão do trabalho com a gestão dos dados de

pesquisa em seu ciclo de vida a fim de prepara-los para o compartilhamento, assim como a limitação (incapacidade) em reconhecer as regras e cumpri-las como foram propostas.

De forma semelhante a Durkheim, Merton apresenta a anomia enfatizando, também, a incapacidade da estrutura social em proporcionar a certas pessoas o que precisam para superarem seus desafios sociais. Esse aspecto ampara os resultados da Tese principalmente em uma das falas dos pesquisadores onde menciona que o compartilhamento de dados deveria ser obrigatório, mas na medida que se tenha infraestrutura suficiente para todos os dados. Da mesma forma, outra fala enfatiza que os repositórios não teriam profissionais suficientes ou capacitados para todo o trabalho e suporte aos pesquisadores. Seguindo os preceitos dos autores a anomia é, também, a incapacidade de os pesquisadores atingirem as metas propostas devido a insuficiência de meios institucionalizados.

Como teoria estrutural funcionalista, a teoria de Durkheim estuda as estruturas pela sua função na sociedade a qual é vista como um todo orgânico articulado que, para funcionar perfeitamente, os indivíduos precisam interagir no ambiente com valores e regras comuns. Émile Durkheim analisava a sociedade através do fato social, conceito marcante dos seus estudos, dos instrumentos sociais ou culturais que determinam o modo de agir, de pensar e de sentir de um indivíduo que se relaciona com outras pessoas e, dessa interação, surge uma consciência coletiva. Scott (1995) argumenta que as estruturas cognitivas e normativas são constituídas de compreensões internalizadas de cada indivíduo a partir de sua interpretação da realidade social em que atua com seus componentes. As argumentações de Durkheim e Scott nos faz entender que os papéis sociais propiciam diferentes interpretações subjetivas do indivíduo, cujas características sociais variam de acordo com o tempo e o espaço.

Sob este ponto de vista, Merton relaciona a estrutura social e os valores da estrutura cultural. Para o autor, a estrutura cultural, com valores normativos, serve para estabelecer as metas da sociedade, e a estrutura social estabelece quais são as normas institucionais mostrando quais são os meios encontrados na sociedade para chegar até as metas.

Para o autor,

[...] a estrutura social pressiona os valores culturais, tornando possível e fácil aos indivíduos que ocupam determinadas situações sociais dentro da sociedade, agirem de acordo com os ditos valores e impossível para outros

indivíduos. A estrutura social age como barreira ou como porta aberta para o desempenho dos mandatos culturais. Quando a estrutura social e cultural estão mal integradas, a primeira exigindo um comportamento que a outra dificulta, há uma atenção rumo ao rompimento das normas ou ao seu completo desprezo. Evidentemente, não se conclui que este seja o único processo que favoreça a condição social da anomia; as teorias e pesquisas posteriores estão orientadas à procura de outras fontes padronizadas, geradoras de um alto grau de anomia. (MERTON, c1968, p. 237).

Os resultados demonstram que os pesquisadores tem interpretações diferentes dos índices da categoria Normativo demonstrando confusão de suas regras, das normas (sociais) apresentadas, descumprimento das regras, perda das referências normativas ou incapacitação de atingi-las de acordo com o proposto pelas instituições representadas na Tese pelo índice da categoria Normativo. A relação entre a influência normativa e os comportamentos de compartilhamento de dados dos pesquisadores não teve uma relação significativa. Isso sugere que em vez de pesquisadores terem uma percepção direta sobre ela, o índice normativo, os pesquisadores que compartilharam seus dados tiveram influências pessoais que inferiram diretamente em seus comportamentos reais. Entretanto, pela baixa adesão ao repositório, infere-se que o entendimento e consenso da comunidade científica em relação ao compartilhamento de dados é fundamental para facilitar os comportamentos de compartilhamento de dados dos pesquisadores. A Tese não encontrou relação significativa entre a pressão normativa das instituições e atores da categoria Normativo e os comportamentos de compartilhamento de dados dos pesquisadores. A comunidade científica pode, ainda, estar no estado de anomia relacionada aos índices, o que influenciaria negativamente seus comportamentos reais de compartilhamento de dados.

Merton vincula a teoria da anomia com a filosofia do sonho americano. Segundo o autor aqueles que não conseguem ascender aos níveis de bem-estar desejado seriam mais pressionados e cometeriam condutas irregulares com a finalidade de alcançar aquelas metas que são cobiçadas por toda a sociedade.

A análise de Merton corrobora com o direcionamento das discussões da Tese na medida em que o ideal alcançado do mérito e reconhecimento acadêmico (sonho da comunidade) é concretizado com as publicações, principalmente, em periódicos de auto impacto, muito deles comerciais. Com as novas vias de publicações demandadas com as filosofias abertas, os canais de publicação, ainda morosos para atende-las, levam pesquisadores a continuar no sistema tradicional. O ecossistema que está sendo

operacionalizado vem direcionando políticas e normas para contornar o modelo vigente. Os pesquisadores que não se sentem seguros, com pouca clareza ou negação do proposto tendem a continuar seguindo o tradicional e o cultural na sua comunidade gerando o estado de anomia.

Robert Merton chama atenção sobre as metas a serem alcançadas e os meios institucionalizados para tal. Tal situação de metas a serem conquistadas (advogado pela ciência aberta) e meios para se alcançar as metas (infraestrutura, repositórios, periódicos, *data journals*, princípios FAIR, políticas e normas) vai ocasionar uma tensão entre a estrutura Cultural, com seus valores que estabelecem as metas até então que prevalecem, e a estrutura social que são as normas que institucionalizam os meios para se atingir as metas, e que forçam o indivíduo a agir de acordo com seus valores, com suas crenças, sua cultura. Para Merton a anomia significava uma descontinuidade das metas culturais e dos meios institucionalizados para alcançá-las. Percebe-se um desequilíbrio e descumprimento as metas entorno das categorias, Cognitivo, Normativo e Regulador, e os caminhos legitimados para atingi-las.

Da mesma forma, a percepção e crenças dos pesquisadores sobre carreira (Benefícios, Riscos e Esforços), aspectos sociais (Benefícios para a sociedade e Ética) e recursos disponíveis (Suporte dos bibliotecários) influenciam os comportamentos dos pesquisadores em relação ao compartilhamento de dados e tem potencial de alinhar com os ideais apresentados da comunicação científica. Merton (1973, p. 274, tradução nossa) preconizava que “A concepção institucional da ciência como parte do domínio público está ligada ao imperativo de comunicação das descobertas. O sigilo é a antítese dessa norma”. Os resultados demonstram que as atitudes de compartilhamento dos pesquisadores corroboram com os preceitos de Merton quando relacionadas à comunicação de seus dados e o faz, após a comunicação de suas descobertas com a publicação científica tradicional.

Berto (2003) fez uma afirmação sobre publicações científicas que é presenciado no contexto de dados. O autor afirma que essas publicações “São conteúdos especializados resultantes de informações e procedimentos técnicos, talentos e restrições, experiência acumulada, atitudes, normas e valores.” (BERTO, 2003, p. 3). Da mesma forma os dados, desde o planejamento da coleta, perpassando pela análise e organização, são resultados de procedimentos ou metodologias, experiências e conhecimentos.

Conforme apresentado no índice *Carrera* (Benefícios, Riscos e Esforços), a gestão dos dados em seu ciclo de vida está associada às publicações científicas. A comunicação científica é imprescindível para o desenvolvimento da ciência, produção de novos conhecimentos e reconhecimento das pesquisas desenvolvidas por outros pesquisadores. A comunicação dos dados fornece *insights* para novas análises e descobertas contribuindo para novas comunicações. Os dados são, portanto, parte de um ciclo de credibilidade, que lhes confere um valor de uso (para a produção de novas publicações) e um valor de troca (como moeda de troca no contexto de colaborações com parceiros).

A comunicação científica consiste em assegurar o intercâmbio de informações e de dados sobre os trabalhos em andamento ou finalizados e colocar os cientistas em contato entre si e com as outras pesquisas concomitantes (LE COADIC, 2004). Ademais, tem o intuito de disseminar o conhecimento produzido ou em andamento e, assim, proporcionar visibilidade, reconhecimento da pesquisa e do autor e credibilidade ao pesquisador, a pesquisa e a sua Instituição. Merton (1973, p. 273, grifo nosso) destaca que “A pressão pela difusão dos resultados é reforçada pelo objetivo institucional de avançar as fronteiras do conhecimento e pelo **incentivo ao reconhecimento** que, obviamente, depende da publicação. Adicionalmente, a comunicação científica é uma forma de colaboração em larga escala, com a possibilidade de reprodução de resultados das pesquisas e experimentos, além de oferecer evidências para a qualidade do trabalho científico (LYNCH, 2011).

A comunicação de dados permite o reconhecimento dos autores e a menção de seu nome todas as vezes que os dados forem utilizados. Van Rossum (2017) destaca que um dos desafios na comunicação científica é a atribuição da autoria nos fluxos das pesquisas, isso observado em contextos digitais e perante o dinamismo e a produtividade presente em tal ambiente. Assevera o autor que um agravante é o fato de que “A ciência é essencialmente uma atividade não comercial, mas ironicamente o negócio da comunicação acadêmica é uma das indústrias mais lucrativas do mundo.” (VAN ROSSUM, 2017, p. 6).

Os resultados apresentados enfatizam a indústria lucrativa com o depoimento de um dos pesquisadores em seu trabalho de editor de periódico. Interessante destacar que entre suas falas é mencionado que esse é um negócio difícil de desarticular. Ao longo da entrevista é possível observar que os pesquisadores têm os resultados das

pesquisas embasados pelos dados disponibilizados no repositório e em periódicos, sendo alguns comerciais. Em outro depoimento, um dos pesquisadores menciona a preocupação de os dados se tornarem “privados”, aspecto já observado com a publicação de *Data papers* de editores comerciais.

Na teoria apresentada por Merton (1973) com o valor da ciência, a motivação para que os cientistas alcancem reputação e prestígio entre seus pares é um valor preponderante em muitas comunidades científicas. Nos últimos anos, a progressão na carreira de pesquisadores acadêmicos é fortemente dependente de métricas vinculadas à publicações científicas. No relatório final e plano de ação do *European Commission Expert Group on FAIR Data* é apontado que os pesquisadores que dedicam tempo e experiência às atividades da gestão de dados e curadoria atualmente não são recompensados por métricas de progressão de carreira tradicionais e, para sanar tal aspecto, estão formando grupos para desenvolver métricas que serão utilizadas em apoio à Ciência Aberta (EUROPEAN COMMISSION, 2018).

Pesquisadores que percebem menos benefícios, mais riscos na carreira e esforços no compartilhamento de dados são mais propensos a ter atitudes negativas em relação ao compartilhamento de dados gerando o estado de anomia. Uma atitude mais positiva em relação ao compartilhamento levará a mais comportamentos de compartilhamento de dados.

Com relação aos recursos, a disponibilidade de serviços oferecidos pelos bibliotecários, não foram identificados relacionamentos significativos com as atitudes dos pesquisadores em relação ao compartilhamento. Além disso, também se descobriu que é possível ter uma influência menos positiva na medida em que as instruções não forem claras aos pesquisadores, como no ocorrido com um dos pesquisadores e sua falta de clareza e confusão com o uso dos repositórios levando ao estado de anomia. Isso sugere que o trabalho do bibliotecário deve ser proativo e dar ampla divulgação em seus serviços de dados, outro fator identificado com o relato de um dos pesquisadores que só depositou os dados depois de um contato com a biblioteca para solicitação de ISSN. Os bibliotecários e os repositórios de dados podem apoiar o compartilhamento de dados dos pesquisadores de várias maneiras, aumentando a percepção e a atitude positiva de compartilhamento de dados.

Os repositórios de dados representam um elemento com forte potencial de contribuição na gestão de dados. Os diversos requisitos funcionais de tais repositórios

podem ajudar na gestão do ciclo de vida dos dados de pesquisa e no apoio técnico aos pesquisadores. Os resultados demonstram a possibilidade de colaboração científica quando os dados são compartilhados, inclusive, em repositórios e sugerem que os repositórios precisam considerar como conscientizar pesquisadores para o compartilhamento, assim como oferecer serviços apropriados e direcionados para as necessidades de dados dos pesquisadores e treinamentos que possibilitem reduzir esforços na gestão dos dados.

A colaboração científica, um dos aspectos incentivados no contexto da Ciência aberta, é essencial para o avanço do conhecimento, pois gera oportunidades de interação e de contribuição em projetos, no entanto, ainda há dificuldades em identificar e se conectar à colaboradores (MCKIERNAN et al., 2016).

Referente aos índices Benefícios para a sociedade e Ética, os mesmos influenciam os comportamentos dos pesquisadores em relação ao compartilhamento de dados. Isso sugere que tais índices podem apoiar a conscientização dos pesquisadores em comportamentos que levam ao compartilhamento de dados, mesmo antes da finalização dos projetos e publicação de todos os resultados.

Considerando os fatores apresentados sobre anomia e os apontamentos finais de Durkheim (1977) de que a anomia pode se referir a problemas não apenas sobre falta de normas em contextos específicos e em determinados períodos, mas com a impossibilidade da sociedade em estabelecer claramente os limites das ações e os sujeitos terem entendimento para tal; e, no caso das normas convencionalmente aceitas sejam descumpridas, com a ineficiência de punição, na Tese, o argumento de que há um estado de anomia nos pesquisadores condiz com o contexto evidenciado, uma vez que os pesquisadores têm pouca clareza sobre os índices apresentados afetando o cumprimento das orientações e regras propostas e, assim, desequilíbrio entre as metas e os caminhos legítimos para alcançá-las.

9 CONCLUSÕES

A Tese apresentou a operacionalização de repositório de dados em relação aos aspectos do direito autoral, citação e licenças de uso aplicados aos dados de pesquisa e os fatores institucionais e pessoais que podem influenciar os pesquisadores no compartilhamento de seus conjuntos de dados.

O *Consórcio Madroño* e seus membros subscrevem seu compromisso com a Ciência Aberta e o repositório *e-cienciaDatos* armazena, gerencia, preserva e dissemina os dados resultantes das pesquisas de sua comunidade de forma aberta; suas políticas estão em consonância com as diretrizes da Ciência Aberta e princípios FAIR, um dos pilares da Agenda Europeia de Ciência Aberta. A Ciência Aberta é um movimento com a premissa de que a documentação científica e os conjuntos de dados estejam em acesso aberto respeitando os aspectos legais e éticos, elementos seguidos pelo *e-cienciaDatos*.

A operacionalização do repositório *e-cienciaDatos* tem potencial para influenciar positivamente o compartilhamento dos dados. A análise dos dados revela que o *e-cienciaDatos* tem infraestrutura e profissionais para atuação com a gestão de dados, entretanto, a quantidade de conjuntos de dados armazenados é pequena relacionada a quantidade de universidades e pesquisadores que fazem parte do *Consórcio Madroño*. Tal aspecto refuta a hipótese de que o repositório influencia positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados, mas confirma a hipótese de que influencia positivamente o compartilhamento dos dados pois o repositório está povoado com conjuntos de dados. Os pesquisadores tiveram motivos variados para a disponibilização via repositório abrindo possibilidades para visibilidade de suas pesquisas, suas instituições e do repositório e, por outro lado, fortalecendo as premissas do compartilhamento e reuso de dados.

O fluxo de trabalho dos bibliotecários demonstra que tais profissionais devem ter estratégias e habilidades para implementar as práticas de Ciência aberta e princípios FAIR, para fomentar a cultura e prática dos pesquisadores na gestão e compartilhamento de dados e nas orientações desse contexto. A constatação nos resultados da participação dos bibliotecários desde a elaboração do projeto, perpassando pela implementação do repositório e gestão dos dados chamou atenção

por causa do papel que profissional tem nas universidades. Sua proatividade é crucial não só na divulgação dos serviços e produtos de dados, como também nos treinamentos e na conscientização dos pesquisadores sobre a relevância do compartilhamento dos dados, que pode influenciar positivamente o compartilhamento dos dados via repositório, aspecto recomendável para minimizar efeitos contrários aos objetivos e às metas dos mesmos.

A hipótese da influência positiva dos bibliotecários foi refutada na Tese, pois há a possibilidade de que falte compreensão das reais necessidades de dados dos pesquisadores, dos caminhos para alcança-los e sua sensibilização para o compartilhamento, sendo necessário fazer um estudo de usuários para identificar as demandas reais e potências de dados, qual tipo de conteúdo deve ser apresentado, os horários de treinamentos ou orientações e estratégias de divulgação dos produtos e serviços do repositório. Relevante destacar que mesmo que os pesquisadores escolham outros repositórios para arquivarem seus dados é oportuno que pelo menos os metadados estejam no *e-cienciaDados*.

As vertentes dos direitos autorais apresentam-se explicitadas nos documentos analisados, os quais seguem as normas, as diretivas e a lei de propriedade intelectual, corroborando que os conjuntos de dados têm direitos morais e patrimoniais. As orientações contribuem para suscitar os pesquisadores a citarem os dados e evitar o plágio. Por outra parte, apontam como os dados estão liberados para uso e replicação. A proteção da propriedade intelectual se apresenta como instrumental relevante para o desenvolvimento científico e tecnológico.

Ao contextualizar os aspectos envolvidos com o Direito Autoral foi considerado o Direito moral e Econômico de conjuntos de dados perpassando pela citação de dados e licenças de uso como também a propriedade intelectual na União Europeia e na Espanha. Verifica-se que a legislação espanhola teve atualizações e as instituições elaboraram documentos que resguardam as necessidades do contexto digital, da sociedade da informação e da inclusão de dados. Com o levantamento dos documentos para a análise, foi constada quantidade significativa com manuais e infografias disponibilizados ao público, os quais com ilustrações que facilitam a visualização e compreensão, entretanto os pesquisadores conhecem pouco o teor das leis, normativas e dos documentos, aspecto que refutou a hipótese de que Normas e Leis influenciam positivamente o compartilhamento dos dados via repositório. Os pesquisadores muitas vezes têm o foco no desenvolvimento de suas pesquisas de

modo que legislações, normas e manuais não são incluídos em sua rotina de leitura, estudo e consultas.

Foi desenvolvido e apresentado um *framework* com 13 índices referentes às categorias Cognitivo, Normativo, Regulador, Carreira, Recursos e Sociais que têm potencial para influenciar a atitude dos pesquisadores em compartilhar ou reter seus conjuntos de dados. A constituição dos índices partiu das temáticas levantadas nos documentos analisados e das entrevistas e questionários aplicados aos bibliotecários. Cada índice foi relacionado com as vertentes do direito autoral de dados. O *framework* apresentado fornece um cenário de fatores basilares de prováveis influências que podem instigar ou inibir os pesquisadores. Portanto, a Tese oferece o *framework*, o qual é possível direcionar estudos futuros sobre diversos aspectos da gestão de dados e seu compartilhamento e em diferentes áreas do conhecimento. Ademais, é possível relacionar temáticas transversais além do direito autoral.

Os resultados demonstram que o *framework* é adequado para explicar o comportamento de compartilhamento de dados de pesquisadores nas diversas áreas do conhecimento. O *Framework* será validado com aplicação entre pesquisadores de universidades. O mesmo pode servir como uma ferramenta instrumental para medir quantitativamente e qualitativamente efeitos em populações ou ambientes específicos de pesquisadores. Além disso, é um instrumento para formular políticas de repositórios, manuais, produtos e serviços nas bibliotecas.

Os pesquisadores envolvidos no contexto apresentado na Tese precisam de orientações sobre direitos autorais tanto moral quanto patrimonial e, assim, evitar a anomia que impacta negativamente a aderência entre suas percepções, perspectivas, a operacionalização de repositórios de dados e o compartilhamento de dados. Os pesquisadores demonstraram não ter habilidades para preparar seus dados para reutilização e, ainda, têm pouca clareza das normas e condutas necessárias relacionados aos agentes envolvidos nos índices do framework: Ciência aberta, Princípios FAIR, Agências de fomento e Plano de gestão de dados, o que refutou as hipóteses de que Ciência Aberta, Agências de fomento e Plano de gestão de dados influenciam positivamente a relação entre a percepção dos pesquisadores e o compartilhamento dos dados via repositório, sendo possível notar a imprescindibilidade de *upgrade* nos serviços de dados. Tais refutações constataam que embora as temáticas discutidas na tese são há alguns anos apresentadas no contexto europeu com forte presença e o compartilhamento dos dados, em muitos casos, são

estimulados ou até mesmo estabelecidos pelas agências de fomento, existem pesquisadores que estão no estado de anomia pois desconhecem ou não tem clareza de suas normas e orientações.

Nesse sentido, surge a reflexão sobre quais estratégias são imprescindíveis para ampla divulgação, conhecimento e internalização das políticas, normas e instruções no âmbito apresentado na Tese? O que está evidente é que o foco atual é na construção de infraestrutura de dados, visto o ecossistema do *Consórcio Madroño*, que igualmente vislumbrado em outras instituições. A partir disso, o foco deve ser ampliado para a sensibilização e compromisso de disponibilização pública dos dados seguindo a máxima “tão aberto quanto possível, tão fechado quanto necessário”. Ressalta-se que o compartilhamento de dados faz parte do fluxo de trabalho da pesquisa e evidencia a centralidade do objetivo do ecossistema disponibilizado aos pesquisadores.

A partir de 2015 o projeto do repositório foi sendo articulado até sua implementação. Apesar de anos de funcionamento os pesquisadores não o conhecem ou dispõe de pouca clareza a respeito. Tais aspectos foram vislumbrados ao longo das entrevistas com inquietação pois a previsão era que os pesquisadores estivessem familiarizados com as temáticas apresentadas. Ademais, instiga a reflexão sobre onde está a possível falha do limitado conhecimento dos pesquisadores.

Em âmbito geral observa-se que a cultura tem forte influência no comportamento dos pesquisadores, na medida em que é seguido ou praticado nas comunidades científicas com a publicação dos resultados, muitos em editores comerciais, e pouca ou nenhuma disponibilização pública dos dados, fato constatado com as declarações dos entrevistados. Por outro lado, os pesquisadores participantes da pesquisa comprovaram seguir o lado oposto de seus pares ao disponibilizarem seus dados via repositório. Tal atitude pode configurar consequência das transformações da ciência, da comunicação científica e das filosofias abertas incluindo ampliação da visibilidade dos resultados de pesquisa.

O trabalho de campo revelou inestimável contribuição para a pesquisa, pois com a entrevista e o questionário sendo respondido na presença da autora da Tese propiciou-se identificar lacunas, ouvir as indagações, dialogar, esclarecer as dúvidas e apresentar os temas da Tese, além de mostrar a relevância de tal contexto. Ao longo das entrevistas foram feitas intervenções explicando e esclarecendo dúvidas sobre os assuntos apresentados, fato que colaborou para o entendimento dos pesquisadores,

sua conscientização sobre a relevância das temáticas e do compartilhamento dos dados e, principalmente sobre direitos de autor e para a divulgação e esclarecimentos sobre o repositório. A quantidade de intervenções não estava prevista no roteiro da coleta de dados devido ao fato de ser considerado que os pesquisadores conheciam as temáticas já que há anos são apresentadas e discutidas em âmbito europeu e espanhol, diferentemente da realidade brasileira. Ademais, propiciou a disseminação das temáticas que os pesquisadores não estavam familiarizados.

A análise dos dados mostrou a anomia no contexto dos pesquisadores com a pouca clareza ou confusão das normas institucionalizadas, sociais e morais, descumprimento das orientações das instituições, desequilíbrio entre as metas e os caminhos para atingi-las e pouca habilidade para a gestão dos dados.

Ao apresentar a natureza do compartilhamento de dados dos pesquisadores, a Tese enfatizou a relevância do repositório de dados, seus serviços de dados e seus bibliotecários para ajudar e orientar os pesquisadores na gestão dos dados e seu compartilhamento e reutilização. A integração entre os repositórios institucionais das universidades e o repositório de dados é imprescindível, assim, quando o usuário fizer a pesquisa no primeiro, os conjuntos de dados aparecem nos resultados. Além disso, os usuários precisam conhecer os repositórios e saber fazer as buscas. Ensinar os caminhos da localização dos dados evita desconfortos e inseguranças aos pesquisadores que depositaram seus dados, fatos relatados na seção dos resultados.

Um ponto fundamental observado durante a pesquisa foi que os pesquisadores tem disponível um ecossistema, com infraestrutura tecnológica, repositório, recursos humanos com profissionais preparados para atender as necessidades dos pesquisadores e, apesar de ter tais recursos, nota-se que a disponibilização dos dados via repositório é pequena comparada ao número de universidades e pesquisadores participantes do *Consórcio Madroño*. Constata-se que a pouca adesão ao repositório com o compartilhamento de dados não resulta da falta de um ecossistema. Infere-se, também, com tal aspecto a pouca disposição ou falta de habilidades dos pesquisadores na gestão dos dados, a lacuna na percepção da relevância do compartilhamento, os benefícios pessoais, sociais e éticos. As hipóteses de tais percepções foram confirmadas devido aos pesquisadores entrevistados estarem propícios ao compartilhamento. Nota-se, também, que é primordial a conscientização dos pesquisadores sobre a relevância do compartilhamento de seus dados, não compartilhar só porque são obrigados, e dos benefícios que são alcançados para a

sociedade e para o pesquisador.

Os resultados da Tese contribuem para o entendimento a respeito da percepção de pesquisadores e seus comportamentos de compartilhamento de dados por meio de repositórios, bem como suscita algumas questões relevantes que podem ser entendidas para promover melhores práticas de compartilhamento de dados.

Uma vez que os benefícios do compartilhamento de dados são tão bem conhecidos, um pesquisador pode desejar compartilhar seu banco de dados/dados com outros pesquisadores.

A percepção e atitude de compartilhamento de pesquisadores podem sofrer alterações dependendo da sua região de atuação, políticas, normas institucionais e características da sua comunidade. É provável que estudos dessa mesma natureza encontre indicadores diferentes, destacando particularidades intrínsecas de cada contexto social.

REFERÊNCIAS

ABUBAHIA, A. M.; COCEA, M. Exploiting vector map properties for GIS data copyright protection. *INTERNATIONAL CONFERENCE ON TOOLS WITH ARTIFICIAL INTELLIGENCE (IEEE)*, 27., 2015, Vietre Sul Maré.

Proceedings [...]. Mumbai, 2015. Disponível em:

<https://ieeexplore.ieee.org/document/7372185/>. Acesso em: 10 jan. 2020.

ACKOFF, L. R. "From Data to Wisdom." Georgia: University of North Georgia, 1999. Disponível em: <http://faculty.ung.edu/kmeltton/Documents/DataWisdom.pdf>. Acesso em: 20 nov. 2020.

AGIER, M. Distúrbios identitários em tempos de globalização. **Mana**, Rio de Janeiro, v. 7, n. 2, p. 7–33, out. 2001. DOI 10.1590/S0104-93132001000200001. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-93132001000200001&lng=pt&tlng=pt. Acesso em: 4 set. 2018.

AJZEN, I. The theory of planned behavior. **Organizational Behavior and Human Decision Process**, Maryland Heights, 52, n. 2, p. 179–211, 1991.

ALBAGLI, S. Ciência aberta em questão. *In*: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (org.). **Ciência aberta, questões abertas**. Brasília, DF: IBICT; Rio de Janeiro: UNIRIO, 2015. Cap. 1, p. 9–25, 2015. Disponível em: <https://livroaberto.ibict.br/handle/1/1060>. Acesso em: 10 nov. 2020.

ALEIXO, D. V. B. **O estado de anomia dos dados no acesso aos dados governamentais abertos no Brasil**. 2020. 260 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2020. Disponível em: https://repositorio.unesp.br/bitstream/handle/11449/191686/aleixo_dvbs_dr_mar.pdf?sequence=3&isAllowed=y. Acesso em: 25 fev. 2021.

ALVES, R. A. V. **Metadados como elementos do processo de catalogação**. 2010. 132 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010. Disponível em: http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/alves_rachel.pdf. Acesso em: 20 ago. 2015.

ALTMAN, M. A Proposed standard for the scholarly citation of quantitative data. **D-Lib Magazine**, Reston, v. 13, n. 3-4, 2007. Disponível em: <http://www.dlib.org/dlib/march07/altman/03altman.html>. Acesso em: 20 ago. 2020.

ALTMAN, M.; BORGMAN, C.; CROSAS, M.; MATONE, M. An introduction to the joint principles for data citation: an introduction to the Joint Principles for Data Citation. **Bulletin of the American Society for Information Science and Technology**, Silver Spring, v. 41, n. 3, p. 43–45, fev. 2015. DOI 10.1002/bult.2015.1720410313. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/bult.2015.1720410313>. Acesso em: 20 ago. 2020.

APPEL, A. L. **A e-Science e as atuais práticas de pesquisa científica**. 2014. 88 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Comunicação, Universidade Federal do Rio de Janeiro; Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2014. Disponível em: https://ridi.ibict.br/bitstream/123456789/872/1/Pesquisa_Andre_Appel_2014-06-26_final.pdf. Acesso em: 10 ago. 2020.

APPOLINÁRIO, F. **Dicionário de metodologia científica**. 2. ed. rev. e ampl. São Paulo: Atlas, 2011.

APPOLINÁRIO, F. **Metodologia da ciência: filosofia e prática da pesquisa**. 2. ed. rev. e ampl. São Paulo: Cengage Learning, 2012.

ASSOCIAÇÃO PAULISTA DA PROPRIEDADE INTELECTUAL. **Propriedade intelectual**. São Paulo, c2016. Disponível em: <http://www.aspi.org.br/pt-br/propriedadeintelectual.aspx>. Acesso em: 20 set. 2016.

AUSTRALIAN NATIONAL DATA SERVICE. **FAIR data training**. Austrália, [201-]. Disponível em: <https://www.andis.org.au/working-with-data/fairdata/training>. Acesso em: 3 dez. 2020.

BARDIN, L. **Análise de conteúdo**. Rev. e atual. Lisboa: Edições 70, 2010.

BAUMAN, Z. **Modernidade Líquida**. Rio de Janeiro: Jorge Zahar, 2001.

BERTO, R. M. V. S. **Novas práticas de comunicação e produção de publicações científicas**. 2003. Disponível em: http://repositorio.portcom.intercom.org.br/bitstream/1904/5281/1/EDOCOM_BERTO.pdf. Acesso em: 12 abr. 2006.

BEZJAK, S. *et al.* **Manual de formação em Ciência Aberta**. FOSTER (2018). [S. l.]: Zenodo, 2018. Disponível em: book.fosteropenscience.eu/pt. Acesso em: 31 set. 2020.

BORGMAN, C. L. The Conundrum of sharing research data. **Journal of the American Society for Information Science and Technology**, New York, p. 1-40, 2011. DOI 10.2139/ssrn.1869155. Disponível em: <http://www.ssrn.com/abstract=1869155>. Acesso em: 12 jun. 2020.

BORGMAN, C. L. *et al.* Knowledge infrastructures in science: data, diversity, and digital libraries. **International Journal on Digital Libraries**, New York, v. 16, n. 3, p. 207-227, 2015. Disponível em: <http://link.springer.com/article/10.1007%2Fs00799-015-0157-z#page-1>. Acesso em: 6 nov. 2015.

BRASIL. Ministério da Cultura. **Direitos autorais**. Brasília, DF, 2007. Disponível em: <http://cultura.gov.br/documents/18021/130362/direito-autoral-27-11-2007.pdf/d3967b82-3f51-4fe5-9021-4825af66f668>. Acesso em: 20 set. 2016.

CAFÉ, L. *et al.* Repositórios institucionais digitais: nova estratégia para publicação

científica na Rede. *In*: ENCONTRO NACIONAL DE INFORMAÇÃO EM CIÊNCIAS DA COMUNICAÇÃO, 12, 2003, Belo Horizonte. **Anais [...]**. Belo Horizonte: [PUC MG], 2003. p. 1-12.

CAMARGO, L. S. A.; VIDOTTI, S. A. B. G. Arquitetura da informação para repositórios científicos digitais. *In*: SAYÃO, L. F. *et al.* (org.). **Implantação e gestão de repositórios institucionais**: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p. 55-82.

CARROLL, M. W. Sharing research data and intellectual property law: a primer. **PLOS Biology**, San Francisco, v. 13, n. 8, p. e1002235, 27 ago. 2015. DOI 10.1371/journal.pbio.1002235. Disponível em: <https://dx.plos.org/10.1371/journal.pbio.1002235>. Acesso em: 24 nov. 2020.

CASTELLS, M. **O poder da identidade**. São Paulo: Paz e Terra, 1999.

CHIZZOTTI, A. A pesquisa qualitativa em ciências humanas e sociais: evolução e desafios. **Revista Portuguesa de Educação**, Braga, v. 16, n. 2, p. 221-236, 2003. Disponível em: <https://www.redalyc.org/articulo.oa?id=374/37416210>. Acesso em: 25 maio 2020.

COMISSION EUROPEA. **Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones**. Bruselas: COMISSION EUROPEA, 2016. (COM(2016) 178).

CONSÓRCIO MADROÑO. **Directrices para la Gestión de Datos en Horizonte 2020**: Versión 1.0 11 de diciembre de 2013. Madrid, 2014. Disponível em: http://www.consorciomadrono.es/wp-content/uploads/2017/05/directrices_gestion_datos_horizon_2020_es.pdf. Acesso em: 17 ago. 2019.

CONSÓRCIO MADROÑO. **Publicación de datos científicos**. Madrid, 2016. Disponível em: <https://www.consorciomadrono.es/investigam/curso-publicacion-de-datos-cientificos/>. Acesso em: 10 jan. 2019.

CONSÓRCIO MADROÑO. **Declaración del Consorcio Madroño en apoyo a la Ciencia Abierta a la información académica y científica**. Madrid, 2017. Disponível em: https://www.consorciomadrono.es/docs/declaracion_ciencia_abierta.pdf. Acesso em: 15 dez. 2018.

CONSÓRCIO MADROÑO. **El Plan de Gestión de Datos**. Madrid, c2018. Disponível em: <http://www.consorciomadrono.es/investigam/faq/>. Acesso em: 20 jan. 2020.

CONSÓRCIO MADROÑO. FAQ: **¿Qué cuestiones éticas y legales hay sobre los datos?** Madrid, c2018. Disponível em: <http://www.consorciomadrono.es/investigam/faq/>. Acesso em: 13 abr. 2019.

CONSÓRCIO MADROÑO. **Cómo depositar sus datos de investigación en e-cienciaDatos**. Madrid, 2019. Disponível em: https://www.urjc.es/images/Biblioteca/Servicios/AccesoAbierto/guia_gestion_datos_investigacion.pdf. Acesso em: 20 abr. 2020.

COUTO CORRÊA, F. **Gestión de datos de investigación**. Barcelona: Editorial UOC, 2016. Disponível em: <http://bit.ly/2uwefAX>. Acesso em: 2 jul. 2017.

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: <http://lis.sagepub.com/content/46/4/299.full.pdf+html>. Acesso em: 27 set. 2016.

CREATIVE COMMONS (CC). **Frequently asked questions**. California, [2016]. Disponível em: <https://creativecommons.org/faq/#what-is-creative-commons-and-what-do-you-do>. Acesso em: 20 set. 2016.

CREATIVE COMMONS (CC). **Sobre as licenças**. California, [20--]. Disponível em: <https://creativecommons.org/licenses/>. Acesso em: 20 set. 2016.

CROSAS, M. The Dataverse Network®: an open-source application for sharing, discovering and preserving data. **D-Lib Magazine**, Reston, v. 17, n. 1/2, Jan./Fev. 2011. Disponível em: <http://www.dlib.org/dlib/january11/crosas/01crosas.html>. Acesso em: 10 nov. 2015.

DATA CITE METADATA WORKING GROUP. **DataCite Metadata Schema Documentation for the Publication and Citation of Research Data**. 2017. Disponível em: https://schema.datacite.org/meta/kernel-4.1/doc/DataCite-MetadataKernel_v4.1.pdf. Acesso em: 20 nov. 2020.

DAHRENDORF, Ralf. **A lei e a ordem**. Brasília, DF: Instituto Tancredo Neves, 1987.

DATA Management Plan 4: data access and security. Utrecht: Utrecht University, 2016. Disponível em: <http://libguides.library.uu.nl/c.php?g=202272&p=1330220&preview=1776d74a25e3418c4d63aafdbf99598f#s-lg-box-wrapper-4739017>. Acesso em: 18 ago. 2016.

DIGITAL CURATION CENTRE. **What are metadata standards**. Edinburgh, c2004-2015a. Disponível em: <http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards>. Acesso em: 31 jul. 2015.

DUDZIAK, E. **Dados de Pesquisa agora devem ser armazenados e citados**. 2016. Disponível em: <http://www.sibi.usp.br/?p=6189>. Acesso em: 10 ago. 2016.

DE JONG, F. M. G. *et al.* CLARIN: towards FAIR and responsible data science using language resources. *In*: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION (LREC 2018), 11., 2018, Miyazaki. **Proceedings** [...]. Utrecht: Utrecht University Repository, 2018. p. 3259 - 3264.

Disponível em: <https://dspace.library.uu.nl/handle/1874/364776>. Acesso em: 7 mar. 2019.

DE SMEDT, K. *et al.* Towards an open science infrastructure for the digital humanities: the case of CLARIN. *In: CONFERENCE ON DIGITAL HUMANITIES IN THE NORDIC COUNTRIES*, 3., 2018, Helsinki. **Proceedings** [...] Aachen: CEUR Workshop Proceedings, 2018. p. 139-151. Disponível em: <http://ceur-ws.org/Vol-2084/paper11.pdf>. Acesso em: 7 mar. 2019.

DURKHEIM, E. **Da divisão do trabalho social**. São Paulo: Martins Fontes, 1999.

DURKHEIM, E. **O suicídio**: estudo de Sociologia. Lisboa: Editorial Presença/Martins Fontes, 1977.

E-CIENCIADATOS. **Modelo Readme.txt**. Madrid, c2020. Disponível em: <https://edatos.consorcioadrono.es/readme.xhtml>. Acesso em: 26 fev. 2021.

E-CIENCIADATOS. **Repositorio de Datos del Consorcio Madroño**. Madrid, c2021. Disponível em: <https://edatos.consorcioadrono.es/>. Acesso em: 26 fev. 2021.

ELLIOT, R. Who owns scientific data? The impact of intellectual property rights on the scientific publication chain. **Learned Publishing**, West Sussex, v. 18, n. 2, p. 91-94, Apr. 2005. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1087/0953151053584984>. Acesso em: 5 jan. 2017.

ENIS, M. **Figshare debuts repository platform for Institutions**. New York: The Digital Shift, 2013. Disponível em: <http://www.thedigitalshift.com/2013/09/publishing/figshare-debuts-repository-platform-for-institutions/>. Acesso em: 22 nov. 2017.

EUROPEAN COMMISSION. **Turning FAIR into reality**. Luxemburgo: Final Report and Action Plan from the European Commission Expert Group on FAIR Data. Publications Office of the European Union, 2018. Disponível em: https://ec.europa.eu/info/sites/.../turning_fair_into_reality_1.pdf. Acesso em: 10 jan. 2019.

FITZGERALD, A.; FITZGERALD, B.; PAPPALARDO, K. O futuro da política de dados. *In: HEY, A. J. G. et al. O quarto paradigma*: descobertas científicas na era da eScience. Tradução Leda Beck. São Paulo: Oficina de textos, 2011. p. 209-216.

FONSECA JÚNIOR, W. C. Análise de conteúdo. *In: DUARTE, J.; BARROS, A. Métodos e técnicas de pesquisa em comunicação*. 2. ed. São Paulo: Atlas, 2014. p. 280-304.

FOSTERING THE PRACTICAL IMPLEMENTATION OF OPEN SCIENCE IN HORIZON 2020 AND BEYOND. **Open Science taxonomy**, 2018. Disponível em: https://www.fosteropenscience.eu/themes/fosterstrap/images/taxonomies/os_taxonomy.png. Acesso em: 19 mar. 2019.

FRANCO, M. L. P. B. **Análise de conteúdo**. 2. ed. Brasília, DF: Liber Livros, 2008.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO (FAPESP). **Programa FAPESP de Pesquisa sobre Mudanças Climáticas Globais**: chamada de propostas de pesquisa 2017. São Paulo, 2017. Disponível em: <http://www.fapesp.br/11068>. Acesso em: 15 dez. 2017.

GANDELMAN, H. **De Gutenberg à internet**: direitos autorais das origens à era digital. 5. ed. Rev. e atual. Rio de Janeiro: Record, 2007.

GODOY, A. S. Pesquisa qualitativa: tipos fundamentais. **Revista de Administração de Empresas**, São Paulo, v. 35, n. 3, p. 20-29, maio/jun. 1995a. Disponível em: <http://www.scielo.br/pdf/rae/v35n3/a04v35n3.pdf>. Acesso em: 30 jun. 2018.

GODOY, A. S. Introdução à pesquisa qualitativa e suas possibilidades. **Revista de Administração de Empresas**, São Paulo, v. 35, n. 2, p. 57-63, mar./abr. 1995b. Disponível em: <http://www.scielo.br/pdf/rae/v35n2/a08v35n2.pdf>. Acesso em: 30 jun. 2018.

GRAY, J. Jim Gray on ecience: a transformed scientific method. *In*: HEY, T.; TANSLEY, S.; TOLLE, K. (ed.). **The fourth paradigm**: data-intensive scientific discovery. Washington: Microsoft Research, 2009. Disponível em: <http://language.log.idc.upenn.edu/myl/JimGrayOnE-Science.pdf>. Acesso em: 05 maio 2015.

GREENBERG, J. *et al.* A metadata best practice for a scientific data repository. **Journal of Library Metadata**, Philadelphia, v. 9, n. 3-4, p. 194-212, dec. 2009. Disponível em: <http://www.tandfonline.com/doi/full/10.1080/19386380903405090>. Acesso em: 22 nov. 2017.

GRUPO de Trabajo de “Depósito y Gestión de datos en Acceso Abierto” del proyecto RECOLECTA. La conservación y reutilización de los datos científicos en España. **Informe del grupo de trabajo de buenas prácticas**. Madrid: Fundación Española para la Ciencia y la Tecnología, FECYT, 2012. Disponível em <http://www.fecyt.es>. Acesso em: 14 jan. 2019.

GUIA prático de inovação: propriedade intelectual e transferência de tecnologia. São Paulo: Agência Unesp de inovação, [20--].

GUIMARÃES, J. A. C. **Propetip 22 - Gestão de dados de pesquisa**. São Paulo:

Unesp, 2018. Disponível em: <https://www2.unesp.br/portal#!/prope/propetips>. Acesso em: 9 maio 2018.

HAGUETTE, T. M. F. **Metodologias qualitativas na Sociologia**. 10. ed. Petrópolis: Vozes, 2003.

HEDOUX, J. Dubar, (Claude). La socialisation: construccìon de identités sociales et professionnelles. **Revue française de pédagogie**, París, v. 100, p. 117-121, 1992. Disponível em: https://www.persee.fr/docAsPDF/rfp_0556-7807_1992_num_100_1_2509_t1_0117_0000_2.pdf. Acesso em: 19 maio 2020.

HERNÁNDEZ-PÉREZ, T.; MÉNDEZ RODRÍGUEZ, E. **Improving health research in EU through FAIR Data**. European Union: FAIR4Health Consortium, 2018. Disponível em: <https://osf.io/download>. Acesso em: 30 jan. 2020.

JORENTE, M. J. V.; SANTOS, P. L. V. A. C. Mídias de pós-vanguarda, direito de autor, cultura livre e produtos de criação contemporânea. *In*: GUIMARÃES, J. A. C.; MOLINA, J. C. F. (org.). **Aspectos jurídicos e éticos da informação digital**. São Paulo: Cultura acadêmica, 2008, p. 113-132.

JORGE, V. D. A. **Abertura e compartilhamento de dados para pesquisa nas situações de emergência em saúde pública: o caso do vírus Zika**. 2018. 263 f. Tese. (Doutorado em Ciência da Informação) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018. Disponível em: https://www.arca.fiocruz.br/bitstream/iciict/32604/2/Tese_VanessaJorge.pdf. Acesso em: 29 abr. 2020.

KALINAUSKAITĖ, D. **To be findable, accessible, interoperable and reusable: language data and technology infrastructure for supporting the FAIR data approach**. 2017. Disponível em: https://pdfs.semanticscholar.org/fe23/3a6acf062719d9834d4ba71d8aa0acee82b4.pdf?_ga=2.154795704.1673960049.1568136138-885200410.1566584314. Acesso em: 29 jun. 2019.

KARAFILOSKI, E.; MISHEV, A. Blockchain solutions for big data challenges: a literature review. *In*: IEEE EUROCON International Conference on Smart Technologies, 17., 2017. **Proceedings** [...]. Mumbai, 2017. p. 763-768. Disponível em: <https://ieeexplore.ieee.org/document/8011213>. Acesso em: 20 set. 2018.

KIM, H. M.; LASKOWSKI, M. Toward an ontology-driven blockchain design for supply-chain provenance. **Intelligent System in Accounting Finance e Management**, Hoboken, v. 25, n. 2, p. 18-27, Jan./Mar. 2018. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1002/isaf.1424>. Acesso em: 20 set. 2018.

KIM, Y. Fostering scientists' data sharing behaviors via data repositories, journal supplements, and personal communication methods. **Information Processing & Management**, Elmsford, v. 53, n. 4, p. 871-885, 2017. <https://linkinghub.elsevier.com/retrieve/pii/S0306457316305908>. Acesso em: 20 set. 2018.

KIM, Y.; ADLER, M. Social scientists' data sharing behaviors: Investigating the roles

of individual motivations, institutional pressures, and data repositories. **International Journal of Information Management**, Guildford, v. 35, n. 4, p. 408–418, ago. 2015. DOI 10.1016/j.ijinfomgt.2015.04.007. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0268401215000432>. Acesso em: 16 dez. 2020.

KIM, Y.; STANTON, J. M. Institutional and individual influences on scientists' data sharing practices. **Journal of Computational Science Education**, Hoboken, v. 3, N. 1, p. 47-56, 2012. Disponível em: <http://www.shodor.org/media/content/jocse/volume3/issue1/Kim>. Acesso em: 29 set. 2018.

KIM, Y.; STANTON, J. M. Institutional and individual factors affecting scientists data-sharing behaviors: a multilevel analysis. **Journal Of the Association for Information Science and Technology**, Hoboken, v. 67, n. 4, p. 776–799, 2016. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1002/asi.23424>>. Acesso em: 29 set. 2018.

KIM, Y.; ZHANG, P. Understanding data sharing behaviors of STEM researchers: the roles of attitudes, norms, and data repositories. **Library & Information Science Research**, Noorwoold, v. 37, n. 3, 189–200, 2015. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0740818815000584>. Acesso em: 21 mar. 2020.

LE COADIC, Y. F. **A ciência da informação**. 2. ed. Brasília, DF: Brinquet de Lemos livros, 2004.

LEMOS, R. Prefácio à edição brasileira. In: LESSIG, L. **Cultura livre**: como a grande mídia usa a tecnologia e a lei para bloquear a cultura e controlar a criatividade. São Paulo, Trama, 2005. p. 15-24

LIBEREUROPE. **Research data services in europe's academic research libraries**. The Hague, 2016. Disponível em: <https://libereurope.eu/blog/2016/10/13/research-data-services-europes-academic-research-libraries/>. Acesso em: 8 maio 2018.

LUBIDA, A.; PILESJÖ, P.; ESPLING, M.; RUNNSTRÖM, M. Applying the theory of planned behavior to explain geospatial data sharing for urban planning and management: cases from urban centers in Tanzania. **African Geographical Review**, Abingdon, v. 34, n. 2, p. 165–181, 4 maio 2014. DOI10.1080/19376812.2014.892433. Disponível em: <http://www.tandfonline.com/doi/full/10.1080/19376812.2014.892433>. Acesso em: 16 dez. 2020.

LYNCH, C. A. Institucional Repositories: essential infrastructure for scholarship in the digital age. **ARL Bimonthly Report** 26. 2003. Disponível em: <http://www.arl.org/news/226/ir.html>. Acesso em: 25 out. 2004.

LYNCH, C. O quarto paradigma de Jim Gray e a construção do registro científico.

In: HEY, T.; STEWARD, T.; TOLLE, K. (org.). **O quarto paradigma**: descobertas científicas na era da eScience. Tradução Leda Beck. São Paulo: Oficina de textos, 2011. p. 187-193.

MALO DE MOLINA T. e-cienciaDatos: el repositorio de datos de investigación del consorcio madroño. **RUIDERAe**: Revista de Unidades de Información, Ciudad Real, v. 15, p. 1–16, 2019. Disponível em: <https://revista.uclm.es/index.php/ruiderae/article/view/2170>. Acesso em: 3 abr. 2021.

MCKIERNAN, E. C. *et al.* How open science helps researchers succeed. **eLife**, Cambridge, v. 5, p. e16800, 7 jul. 2016. DOI 10.7554/eLife.16800. Disponível em: <https://elifesciences.org/articles/16800>. Acesso em: 19 abr. 2021.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos da metodologia científica**. 7. ed. São Paulo: Atlas, 2016.

MANTOVANI, O.; DIAS, M. H. P.; LIESENBERG, H. Conteúdos abertos e compartilhados: novas perspectivas para a educação. **Educação & Sociedade**, Campinas, v. 27, n. 94, p. 257-276, jan./abr. 2006.

MÁRDERO ARELLANO, M. A. **Critérios para a preservação digital da informação científica**. 2008. 358 f. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2008

MARTONE M. (ed.) **Data Citation Synthesis Group**: Joint Declaration of Data Citation Principles. San Diego CA: FORCE11, 2014. Disponível em: <https://www.force11.org/datacitationprinciples>. Acesso em: 10 jan. 2019.

MAYER- SCHONBERGER, V.; CUKIER, K. **Big data**: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana. Rio de Janeiro: Elsevier, 2013.

MEADOWS, A. J. **A comunicação científica**. Brasília, DF: Brinquet de Lemos livros, 1999.

MERTON, R. K. **Sociologia**: teoria e estrutura. São Paulo: Mestre Jou, c1968.

MERTON, R. K. The normative structure of science. In: MERTON, R. K. **The Sociology of Science**: theoretical and empirical Investigations, University of Chicago Press, Chicago 1973. p. 223-280.

MOLFINO, M. R.; GONZÁLEZ, C. M. Open Access to scientific literature and research data: a window of opportunity for Latin America. In: II ETHICOMP LATINOAMÉRICA, 2., 2012, Bahía Blanca. **Anais** [...]. Bahía Blanca: Memoria Académica, 2012. p. 1–11. Disponível em: http://www.memoria.fahce.unlp.edu.ar/trab_eventos/ev.2514/ev.2514.pdf. Acesso em: 2 abr. 2021.

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: 30 abr. 2017.

MONTEIRO, E. C. D. 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**, Florianópolis, v. 23, n. 53, p. 160–173, 6 set. 2018. DOI 10.5007/1518-2924.2018v23n53p160. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p160>. Acesso em: 18 out. 2020.

MONTEIRO, E. C. S. A.; SANT'ANA, R. C. G. Direitos autorais e Ciência Aberta: estudo do Repositório de dados do Consórcio Madroño. *In*: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 28., 2019, Vitória. **Anais [...]**. São Paulo: FEBAB, 2019. Disponível em: portal.febab.org.br/anais/article/download/2074/2075. Acesso em: 10 jan. 2020

MONTEIRO, E. C. S. A.; SANT'ANA, R. C. G. HERNÁNDEZ-PÉREZ, A. Direitos autorais de dados científicos no contexto da Ciência Aberta: estudo do repositório de dados do Consórcio Madroño. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 20., 2019, Florianópolis. **Anais [...]**. Florianópolis: UFSC, 2019. v. 20, Disponível em: <https://conferencias.ufsc.br/index.php/enancib/2019/paper/view/1128/756>. Acesso em: 10 jan. 2020.

MONTEIRO, E. C. S. A. *et al.* Proteção autoral de dados: uso de Blockchain. *In*: SANT'ANA, R. C. G.; DUTRA, M. L.; DIAS, G. A. (org.). **Workshop de informação, dados e tecnologia**. João Pessoa: Editora UFPB, 2018. p. 304-311. Disponível em: <https://www.ufpb.br/widat/contents/documentos/anaiswidat2018.pdf>. Acesso em: 12 jun. 2019.

MONTEIRO, E. C. S. A. *et al.* Proteção autoral de dados e ciência aberta na blockchain. **Ciência da Informação**, Rio de Janeiro v. 48, n. 3, 2020. Disponível em: <http://revista.ibict.br/ciinf/article/view/4872>. Acesso em: 10 nov. 2020.

MORA, F. (coord.). **Compromisos de las universidades ante la Open Science**. Madri: CRUE Universidades Españolas, 2019. Disponível em: http://www.crue.org/Documentos%20compartidos/Informes%20y%20Posicionamientos/2019.02.20-Compromisos%20CRUE_OPENSCIENCE%20VF.pdf. Acesso em: 24 jun. 2019.

MORIGI, V. J.; PAVAN, C. Entre o “tradicional” e o “virtual”: o uso das tecnologias de informação e comunicação e as mudanças nas bibliotecas universitárias. **Revista ACB: Biblioteconomia em Santa Catarina**, Florianópolis, v. 8/9, p. 54, 2003/2004.

MORIN, E. **Introdução ao pensamento complexo**. 3. ed. Porto Alegre: Sulina,

2007.

MUELLER, S. P. M. A ciência, o sistema de comunicação científica e a literatura científica. *In*: CAMPELLO, B. C.; CENDÓN, B. V.; KREMER, J. M. (org.) **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte: UFMG, 2000. p. 21–34.

NAÇÕES UNIDAS. **Declaração Universal dos Direitos Humanos**. Rio de Janeiro: UNIC, 2009. Disponível em: <http://www.dudh.org.br/wp-content/uploads/2014/12/dudh.pdf>. Acesso em: 22 set. 2016.

OPENDOAR. **OpenDOAR Statistics**. 2020. Disponível em: https://v2.sherpa.ac.uk/view/repository_visualisations/1.html. Acesso em: 02 dez. 2020.

OPEN KNOWLEDGE FOUNDATION. **Open Data Commons**. [201-]. Disponível em: <https://opendatacommons.org/>. Acesso em: 12 dez. 2019.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). **OECD Principles and guidelines for access to research data from public funding**. Paris: OECD, 2007. Disponível em: <http://www.oecd.org/sti/scitech/38500813.pdf>. Acesso em: 20 set. 2016.

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD). **Making Open Science a Reality**. OECD Science, Technology and Industry Policy Papers, n. 25. [S. l.]: OECD, 2015. DOI 10.1787/5jrs2f963zs1-en. Disponível em: https://www.oecd-ilibrary.org/science-and-technology/making-open-science-a-reality_5jrs2f963zs1-en. Acesso em: 20 jun. 2020.

PELLS, R. Open access: Plan S launch delayed until 2021. **Times Higher Education**, London, May 2019. Disponível em <https://www.timeshighereducation.com/news/open-access-plan-s-launch-delayed-until-2021>. Acesso em: 31 maio 2019.

PETERS, I. *et al.* Research data explored: an extended analysis of citations and altmetrics. **Scientometrics**, Amsterdam, v. 107, p. 723–744, 2016. Disponível em: <https://link.springer.com/article/10.1007/s11192-016-1887-4>. Acesso em: 28 out. 2017.

PINFIELD, S; COX, A. M.; SMITH, J. Research data management and libraries: relationships, activities, drivers and Influences. **PLoS ONE**, San Francisco, v. 9, n. 12, p. 1-28, 2014. Disponível em: <http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=1&sid=fbf9aa61-6c97->. Acesso em: 28 out. 2017.

PIWOWAR, H. A.; CHAPMAN, W. W. Public sharing of research datasets: a pilot study of associations. **Journal of Informetrics**, Amsterdam, v. 4, n. 2, p. 148–156, abr. 2010. DOI 10.1016/j.joi.2009.11.010. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1751157709000881>. Acesso em: 2 jun.

2019.

PLOMAN, E. W.; HAMILTON, L. C. **Copyright:** intellectual property in the information age. London: Routledge & Kegan Paul, 1980. Disponível em: 445e-a489-9d76b4383641%40sessionmgr107&hid=128. Acesso em: 27 set. 2016.

PONTIKA, N. *et al.* Fostering Open Science to Research using a Taxonomy and an eLearning Portal. *In*: INTERNATIONAL CONFERENCE ON KNOWLEDGE TECHNOLOGIES AND DATA DRIVEN BUSINESS, 15., 2015, Graz. **Anais** [...]. Milton Keynes: The Open University, 2015. Disponível em: http://oro.open.ac.uk/44719/2/kmi_foster_iknow.pdf. Acesso em: 10 jan. 2019.

PRYOR, G. (ed.) **Managing research data**. United Kingdom: Facet Publishing, 2012. Disponível em: <http://www.dcc.ac.uk/news/book-managing-research-data>. Acesso em: 27 set. 2016.

QUADERI, N. *et al.* **The Plan S footprint:** Implications for the scholarly publishing landscape. Jersey: Web of Science Group, 2018. Disponível em: <https://clarivate.com/g/plan-s-footprint/>. Acesso em: 22 mar. 2019.

QUEIROZ, V. **Anomia social e alienação:** Émile Durkheim. 2015. Disponível em: <https://colunastortas.com.br/anomia-social/>. Acesso em: 6 abr. 2021.

REBOUILLAT, V. **Ouverture des données de la recherche:** de la vision politique aux pratiques des chercheurs. 2019. Disponível em: <https://tel.archives-ouvertes.fr/tel-02447653/document>. Acesso em: 29 nov. 2020.

RED DE BIBLIOTECAS UNIVERSITARIAS ESPAÑOLAS (REBUIN). **Infografía 3:** Conserva tus derechos de autor frente a las editoriales. España, REBUIN, 2015. Disponível em: <https://www.rebiun.org/node/168>. Acesso em: 18 dez. 2018.

RED DE BIBLIOTECAS UNIVERSITARIAS ESPAÑOLAS (REBUIN). **Infografía 8:** Ciencia Abierta: la investigación y los datos científicos accesibles y abiertos a todos los ciudadanos. España, REBUIN, a2016. Disponível em: <https://www.rebiun.org/node/179>. Acesso em: 18 dez. 2018.

RED DE BIBLIOTECAS UNIVERSITARIAS ESPAÑOLAS (REBUIN). **Infografía 10:** Cita tus datos de investigación. b2016. Disponível em: <https://www.rebiun.org/node/184>. Acesso em: 18 dez. 2018.

RESEARCH DATA MANAGEMENT SERVICE GROUP. **Introduction to intellectual property rights in data management**. [201-]. Disponível em: <https://data.research.cornell.edu/content/intellectual-property>. Acesso em: 18 ago. 2020.

RIBEIRO, F. M. V. Nuances da sociologia do desvio em Émile Durkheim. **Revista Cadernos de Ciências Sociais da UFRPE**, Recife. v. 1, n. 1, jul./dez, p. 1–19, 2012. Disponível em: <http://www.journals.ufrpe.br/index.php/cadernosdecienciasocias/article/view/233/204>. Acesso em: 6 abr. 2021.

RODRIGUES, E. *et al.* **RepositóriUM**: criação e desenvolvimento do Repositório da Universidade do Minho. 2004. Disponível em: <http://www.apbad.pt/pdf8cong/com/com14.pdf>. Acesso em: 13 set. 2004.

ROSETTO, M. **Metadados e formatos de metadados em sistemas de informação**: caracterização e definição. 2003. 95 f. Dissertação (Mestrado em Ciências da Comunicação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2003.

ROZADOS, H. B. F. Uso de indicadores na gestão de recursos de informação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 3, n. 1, p. 60-76, jul./dez. 2005. Disponível em: http://basessibi.c3sl.ufpr.br/brapci/_repositorio/2015/12/pdf_4248c73f66_0000015274.pdf. Acesso em: 10 abr. 2017.

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. 264 f. Tese (Doutorado em Ciência da Informação) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014. Disponível em: <http://carpedien.ien.gov.br/bitstream/ien/853/1/LUANA%20SALES%20D.pdf>. Acesso em 19 set. 2016.

SALES, L. F.; SAYÃO, L. F. O impacto da curadoria digital dos dados de pesquisa na comunicação científica. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 17, p.118-135, 2012. Especial 2. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17nesp2p118/23573>. Acesso em: 21 set. 2016.

SANT'ANA, R. C. G. A importância do papel do profissional da ciência da informação nos processos de recuperação de conteúdos digitais estruturados. *In*: GUIMARÃES, J. A. C.; FUJITA, M. S. L. (org.). **Ensino e pesquisa em biblioteconomia no Brasil**: a emergência de um novo olhar. Marília: Fundepe; São Paulo: Cultura Acadêmica, 2008. p. 145-154.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação e informação**, Londrina, v. 21, n. 2, p. 116-142, maio/ago. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940/20124>. Acesso em: 20 out. 2016.

SANTOS, P. X. (coord.). **Livro Verde - Ciência aberta e dados abertos**: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional. Rio de Janeiro: Fiocruz, 2017. Disponível em: <https://www.arca.fiocruz.br/handle/icict/24117>. Acesso em: 10 jan. 2019.

SAYÃO, L. F. Interoperabilidade das bibliotecas digitais: o papel dos sistemas de identificadores persistentes - URN, PURL, DOI, Handle System, CrossRef e OpenURL. **TransInformação**, Campinas, v. 19, n. 1, p. 65-82, jan./abr., 2007. Disponível em: <http://periodicos.puc->

campinas.edu.br/seer/index.php/transinfo/article/view/621/601. Acesso em: 14 jul. 2016.

SAYÃO, L. F.; SALES, L. F. Curadoria digital e dados de pesquisa. **AtoZ: novas práticas em informação e conhecimento**, Curitiba, v. 5, n. 2, p. 67-71, 2016. Disponível em: <http://revistas.ufpr.br/atoz/article/view/49708/30161>. Acesso em: 2 ago. 2017.

SARACEVIV, T. A natureza interdisciplinar da ciência da informação. **Ciência da Informação**, Brasília, DF, v. 29, n. 1, p. 1-9, 1995.

SESARTIC, A.; TOWE, M. Research data services at ETH-Bibliothek. **Journal IFLA**, Munich, v. 42, n. 4, p. 284-291, Nov. 2016. Disponível em: <http://journals.sagepub.com/doi/abs/10.1177/0340035216674971>. Acesso em: 5 out. 2017.

SCOTT, W. R. Contemporary institutional theory. In: SCOTT, W. R. **Institutions and organizations**. London: Thousand Oaks : SAGE, 1995. Cap. 3, p. 32-62.

SHAKERI, S., GRACY, K. F. A model for data curation research in small sciences. **American Society for Information Science and Technology**, Silver Springer, v. 51, n. 1, p. 1-5, 2014. Disponível em: <http://onlinelibrary-wiley.ez87.periodicos.capes.gov.br/doi/10.1002/meet.2014.14505101094/full>. Acesso em: 7 fev. 2018.

SICILIA, M. A.; GARCÍA-BARRIOCANAL, E.; SÁNCHEZ-ALONSO, S. Community curation in open dataset repositories: insights from Zenodo. In: INTERNATIONAL CONFERENCE ON CURRENT RESEARCH INFORMATION SYSTEMS, CRIS, 13., 2016, Scotland. **Anais [...]** Scotland, 2017. v. 106, p. 54-60. Disponível em: https://ac.els-cdn.com/S1877050917302776/1-s2.0-S1877050917302776-main.pdf?_tid=665f851c-d5ed-11e7-99ca-00000aab0f27&acdnat=1512060306_bc2d26bb63f413fafa61ca1a688ddd76. Acesso em: 28 set. 2017.

SIQUEIRA, M. A.; COSTA SANTOS, P. L. V. A. A versão XML do MARC21 e as formas de representação descritiva na Ciência da Informação. In: VIDOTTI, S. A. B. G. (org). **Tecnologia e conteúdos informacionais**: Abordagens teóricas e práticas. São Paulo: Polis, 2004. p. 95-111.

SOUZA, F. C. O nome profissional “bibliotecário” no brasil: o efeito das mudanças sociais e econômicas dos últimos anos do século XX. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 9, n. 18, 90- 106, 2004. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2004v9n18p90>. Acesso em: 21 ago. 2020.

STEPANOV, I. Introducing a property right over data in the EU: the data producer's right – an evaluation. **International Review of Law, Computers & Technology**,

UNIVERSIDAD CARLOS III DE MADRID (UC3M). **Marco legal de los datos de investigación**. Getafe, [201-b]). Disponível em: <https://www.uc3m.es/biblioteca/propiedad-intelectual-industrial/propiedad-intelectual+>. Acesso em: 10 ago. 2019.

UNIVERSITY OF MELBOURNE. **Management of Research Data and Records Policy (MPF1242)**. Melbourne, 2009. Disponível em: <https://policy.unimelb.edu.au/MPF1242>. Acesso em: 10 jan. 2019.

VAN ROSSUM, J. *et al.* **Blockchain for research**. Science, Nov. 2017

WORLD INTELLECTUAL PROPERTY ORGANIZATION (WIPO). **What is intellectual property?** Geneva, [20--]. Disponível em: http://www.wipo.int/edocs/pubdocs/en/intproperty/450/wipo_pub_450.pdf. Acesso em: 20 set. 2016.

ZENODO. **Coronavirus Disease Research Community - COVID-19**. 2020. Disponível em: <https://zenodo.org/communities/covid-19/?page=1&size=20>. Acesso em: 15 set. 2020.

ZOPE-CHAUDHARI, S.; VENKATACHALAM, P.; BUDDHIRAJU, K. M. Copyright protection of vector data using vector watermark. *In*: IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS), 2017. **Proceedings [...]**. Fort Worth, TX: IEEE, jul. 2017. p. 6110–6113. DOI 10.1109/IGARSS.2017.8128403. Disponível em: <http://ieeexplore.ieee.org/document/8128403/>. Acesso em: 15 dez. 2020.

ANEXO A - TÉRMINO DE CONSENTIMIENTO LIBRE Y ESCLARECIDO - BIBLIOTECÁRIO

Usted está siendo convidado(a) a participar de la investigación titulada Repositorios de datos de investigación: un estudio sobre el consorcio Madroño, realizada por la alumna de doctorado Elizabete Cristina de Souza de Aguiar Monteiro del Programa de Postgrado en Ciencia de la Información (PPGCI), de la Universidad Estadual Paulista (UNESP), Campus de Marília, Brasil. Su participación no es obligatoria. En cualquier momento usted puede desistir a participar y retirar su consentimiento. Su rechazo no traerá ningún perjuicio en su relación con el investigador o con la Institución. Este estudio busca identificar cómo es el trabajo del bibliotecario relacionado con las actividades del Consorcio de Universidades de la Comunidad de Madrid y de la UNED para la Cooperación Bibliotecaria (Madroño).

Su participación en esta investigación consistirá en responder a un cuestionario y a una entrevista, a ser aplicada *in loco*, por la propia investigadora. La información recopilada y el análisis de los datos no identificará individualmente a los participantes manteniendo su privacidad.

Usted recibirá una copia de este Término donde consta el teléfono y el e-mail de la investigadora responsable de la investigación, pudiendo resolver sus dudas sobre la investigación y su participación en cualquier momento.

Declaro que entiendo la información relacionada con la investigación y el término y estoy de acuerdo en participar.

Autorizar,

Fecha: ____/____/____

Firma del participante

Elizabete Cristina de Souza de Aguiar Monteiro
MSc. en Ciencia de la Información y Doctoranda en Ciencia de la Información -
UNESP / Marília, São Paulo, Brasil
ecsamonteiro@gmail.com - whatsapp + 5514-96872597

ANEXO B - TÉRMINO DE CONSENTIMIENTO LIBRE Y ESCLARECIDO – PESQUISADOR

UNIVERSIDADE ESTADUAL PAULISTA - UNESP - BRASIL
FACULDADE DE FILOSOFIA E CIÊNCIAS – FFC/ MARÍLIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
UNIVERSIDAD CARLOS III DE MADRID

TÍTULO DE LA INVESTIGACIÓN: OPERACIONALIZAÇÃO DE REPOSITÓRIOS DE DADOS: uma análise sobre as perspectivas e atitudes dos pesquisadores nas questões de autoria e licença

INVESTIGADORA: Elizabete Cristina de Souza de Aguiar Monteiro

Usted está siendo invitado a participar en esta investigación que tiene la orientación del Prof. Dr. Ricardo César Gonçalves Sant'Ana y co-orientación del Prof. Dr. Antonio Hernández-Pérez.

Término de consentimiento libre y esclarecido.

Acepto la invitación para participar en la investigación titulada "Repositórios de Dados de Pesquisa: um estudo sobre o Consórcio Madroño" de Elizabete Cristina de Souza de Aguiar Monteiro.

Me informaron de que esta investigación tendrá como objetivo general identificar las convergencias y divergencias entre las actitudes y perspectivas de los investigadores y la operacionalización de los repositorios de datos e-cienciaDatos del Consorcio Madroño, consorcio de Universidades de la Comunidad de Madrid y UNED en el factor Derechos de autor (DA).

Soy consciente de que mi participación como sujeto de esta investigación es responder un cuestionario y la información recopilada y el análisis de datos no identifican individualmente a los participantes y que se mantendrá mi privacidad.

Estoy de acuerdo y acepto que los datos obtenidos a través de esta investigación serán parte de los resultados de la Tesis del investigador y pueden hacerse públicos, a través de la publicación de informes y documentos científicos.

En caso de duda, contactar Elizabete, teléfono: +55 14 99687-2597, e-mail: ecsamonteiro@gmail.com o con Comitê de Ética em Pesquisa da UNESP teléfono: +55 14 3402-1300.

APÊNDICE A – APRESENTAÇÃO DA PESQUISADORA E DA PESQUISA

REPOSITORIOS DE DATOS DE INVESTIGACIÓN: UN ESTUDIO SOBRE EL CONSORCIO MADROÑO

Estimada(o) Bibliotecaria (o),

Soy Elizabete Cristina de Souza de Aguiar Monteiro, alumna de doctorado en el Programa de Postgrado en Ciencia de la Información en la línea Información y Tecnología en la Universidade Estadual Paulista (UNESP), Campus de Marília, y vengo por médio de este convidarla (o) para participar de este estudio, bajo orientación del Dr. Ricardo César Gonçalves Sant'Ana, docente del Programa de Postgrado en Ciencia de la Información UNESP, y supervisión del Prof. Dr. Tony Hernández-Pérez del departamento de Biblioteconomía y Documentación de la Universidad Carlos III de Madrid.

Este estudio busca identificar cómo es el trabajo del bibliotecario relacionado con las actividades del Consorcio de Universidades de la Comunidad de Madrid y de la UNED para la Cooperación Bibliotecaria (Madroño). Se trata de un recorte para mi tesis doctoral, la cual aborda la gestión de datos y derechos de autor en los repositorios de datos. Su participación es importante para la investigación pues va a contribuir en conocer la práctica profesional del bibliotecario, y divulgar la realidad de un consorcio de repositorios.

Son 28 cuestiones cerradas y abiertas para ser respondidas. Resaltamos que no serán identificados individualmente los participantes manteniendo la privacidad de estos, y los e-mails institucionales no serán divulgados. Los datos recolectados serán utilizados para presentación y discusión de los resultados de la investigación. Es importante recordar que su participación es muy importante para que podamos describir con propiedad el escenario actual y contribuir con las discusiones teóricas y prácticas profesionales en la implementación y gestión de datos.

Cualquier duda pido que entren en contacto y desde ya agradecemos por su participación.

Atentamente,

Elizabete Cristina de Souza de Aguiar Monteiro

MSc. en Ciencia de la Información y Doctoranda en Ciencia de la Información -

UNESP / Marília, São Paulo, Brasil

ecsamonteiro@gmail.com - whatsapp + 5514-96872597

APÊNDICE B – ROTEIRO PARA O QUESTIONÁRIO E ENTREVISTA
APLICADOS AOS BIBLIOTECÁRIOS

Instrucción: Por favor, responda las siguientes preguntas con la mayor precisión posible y de acuerdo con su conocimiento.

1) Género

☐ Masculino

☐ Femenino

2) Edad

☐ 18-28

☐ 29-39

☐ 40-49

☐ 50-59

☐ 60 o más

3) ¿Está formado en un curso de Biblioteconomía o Ciencia de la información? Si la respuesta es no, ¿cuál es su formación?

☐ Sí

☐ No

4) Formación _____

5) ¿Tiene curso de Postgrado? Si la respuesta es sí, ¿cuál es el nivel?

☐ especialización

☐ maestría

☐ doctorado

6) Usted posee Postgrado en cual área del conocimiento?

☐ Biblioteconomía

☐ Ciencia de la información

☐ ¿Otra?

7) ¿Ha hecho algún curso para trabajar en el repositorio? Si la respuesta es sí, ¿qué curso?

☐ Sí

☐ No

8) ¿Hace cuántos años trabaja en la biblioteca?

☐ 0-5

☐ 6-10

☐ 11-15

☐ 16-20

☐ 21-25

☐ Más de 25 años

9) ¿Hace cuántos años trabaja en el Repositorio de Datos del Consorcio Madroño?

- ☐ menos de 1 año
☐ 1-4
☐ 5-9
☐ 10-13
☐ 14-18

10) Ya había trabajado en otro repositorio de datos? Si la respuesta es sí, ¿en qué repositorio de datos?

- ☐ Sí
☐ No

11) ¿Ha recibido entrenamientos para trabajar en el repositorio? Si es afirmativo, ¿cuáles fueron los cursos?

- ☐ sí
☐ no

12) ¿Cuáles son las actividades que el bibliotecario desarrolla para apoyo de los investigadores con sus datos de investigación?

- ☐ orientan sobre las políticas de los financiadores acerca de la gestión de datos?
☐ orientan sobre las herramientas para la elaboración del plan de gestión de datos?
☐ ayudan en la elaboración del plan de gestión de datos?
☐ ayudan en la planificación de la recolección de datos?
☐ ayudan en la elaboración de los protocolos de recolección de datos?
☐ ayudan en la organización de los datos?
☐ ayudan en la documentación?
☐ ayudan en el análisis de los datos?
☐ ayudan en la preservación de los datos?
☐ ayudan en el depósito de los datos en el repositorio?
☐ ayudan en la publicación de los datos en periódicos?
☐ orientan sobre los requisitos de los periódicos para la publicación de los datos?
☐ orientan sobre el uso compartido de datos?
☐ orientan sobre las cuestiones de derecho de autor y citación de los datos?
☐ orientan sobre las cuestiones de privacidad de los sujetos que participaron en las encuestas?

- ☐ ayudan en la elección de las licencias empleadas para los datos?
- ☐ ayudan en la búsqueda y recuperación de datos en repositorios y periódicos?
- ☐ ayudan en la reutilización de los datos?
- ☐ orientan sobre la citación y referencia de los datos reutilizados?
- ☐ otros. ¿Qué?

13) ¿Cómo se hace la divulgación del servicio ofrecido por la biblioteca?

- ☐ sitio web de la biblioteca
- ☐ Facebook
- ☐ Twitter
- ☐ folders impresos
- ☐ otros. ¿Cuáles:

14) ¿Quién realiza el depósito de los datos en el repositorio?

- ☐ el autor (autoarchivo)
- ☐ el investigador envía por e-mail y el bibliotecario hace el depósito?

15) Indique las actividades que usted desarrolla en el repositorio.

- ☐ Revisión de los datos y envío
- ☐ Archivado
- ☐ Indexación
- ☐ Catalogación
- ☐ Otras:

16) ¿Cuáles considera usted son las habilidades necesarias para trabajar en el Repositorio de Datos del Consorcio Madroño?

17) Los usuarios han sido consultados para saber qué cursos les gustaría que la biblioteca ofreciera?

- ☐ Sí

() No

18) ¿Las bibliotecas dan entrenamientos para los usuarios (docentes, alumnos e investigadores) sobre la utilización del Repositorio de Datos del Consorcio Madroño?

() Sí

() No

19) ¿Qué cursos se ofrecen a los usuarios de la Universidad?

() Autoarchivo en el repositorio

() Gestión de datos

() Plano de gestión de datos

() Busca en otros repositorios de datos

() Otros. Cuales?

20) ¿En su opinión cuál es la diferencia entre el bibliotecario que trabaja con documentos y aquel bibliotecario que trabaja con datos?

21) ¿El bibliotecario de datos necesita nuevas y diferentes habilidades?

() Sí

() No

22) ¿Cuáles son las nuevas y diferentes habilidades que el bibliotecario de datos necesita tener?

23) ¿Cuáles son las competencias (conocimientos, habilidades y actitudes) necesarias para el bibliotecario actuar en un repositorio de datos?

24) ¿Existe un sistema de evaluación/seguimiento de las competencias de los bibliotecarios? Si la respuesta es sí, describa por favor.

() Sí

() No

25) ¿Cuáles son las principales dificultades encontradas en su trabajo?

26) En su opinión, ¿cómo el repositorio de datos contribuye a preservar la memoria de la institución?

27) ¿Cómo piensa que sea el ideal para la representación de los datos visualizando su futura recuperación?

28) ¿Quisiera hacer alguna observación? Describa por favor:

APÊNDICE C - ROTEIRO DE QUESTIONÁRIO E ENTREVISTA APLICADOS AOS PESQUISADORES

1. ¿Cuál es su opinión sobre los derechos de autor de los datos de investigación?

2. ¿Cuál es su opinión sobre compartir los datos de investigación?

3. ¿Cuales son tus perspectivas al compartir los datos de investigación?

Leyes,
normativas
y políticas

Las siguientes declaraciones están relacionadas con las Leyes y normativas que regulan los aspectos de propiedad intelectual de Unión Europea y de España, las políticas de Ciencia abierta, agencias de fomento y principios FAIR.

Por razones de validación, hay preguntas similares en todas las secciones.
Las opciones de respuestas están entre Muy en desacuerdo y Totalmente de acuerdo.

Si no tiene conocimiento de algún tema para marcar una opción por favor, señale la opción "No sé del tema"

Leyes y normativas de la Unión Europea y de la España

4. ¿Cuánto está de acuerdo con cada declaración sobre las leyes y directrices de Unión Europea y la PROPIEDAD INTELECTUAL de España?

Marcar apenas una oval por linha.

	Totalmente en desacuerdo	Moderadamente en desacuerdo	Neutral	Moderadamente de acuerdo	Totalmente de acuerdo
Conozco toda la legislación y directrices	☐	☐	☐	☐	☐
Las leyes y normas facilitan el desarrollo científico y tecnológico	☐	☐	☐	☐	☐
No sé del tema	☐	☐	☐	☐	☐

5. ¿Cuánto está de acuerdo con cada declaración sobre CIENCIA ABIERTA?

Marcar apenas una oval por linha.

	Totalmente en desacuerdo	Moderadamente en desacuerdo	Neutral	Moderadamente de acuerdo	Totalmente de acuerdo
Conozco sus pautas para compartir datos sin restricciones de uso	☐	☐	☐	☐	☐
Estoy dispuesto a compartir datos aplicando la licencia CC0 cediendo todos los derechos de autor	☐	☐	☐	☐	☐
No sé del tema	☐	☐	☐	☐	☐

6. ¿Cuánto está de acuerdo con cada declaración sobre principios FAIR - (Findable, Accessible, Interoperable, Reusable) - actúan como pautas para mejorar la reutilización de datos?

Marcar apenas una oval por línea.

	Totalmente en desacuerdo	Moderadamente en desacuerdo	Neutral	Concordo moderadamente	Totalmente de acuerdo
Conozco sus principios	☐	☐	☐	☐	☐
Evidencia cómo compartir datos	☐	☐	☐	☐	☐
No sé del tema	☐	☐	☐	☐	☐

7. ¿Cuánto está de acuerdo con cada declaración sobre AGENCIAS DE FINANCIACIÓN?

Marcar apenas una oval por línea.

	Totalmente en desacuerdo	Moderadamente en desacuerdo	Neutral	Moderadamente de acuerdo	Totalmente de acuerdo
Estoy de acuerdo con las políticas obligatorias del financiador para compartir datos	☐	☐	☐	☐	☐
Las agencias deben usar métricas de citas de datos	☐	☐	☐	☐	☐
No sé del tema	☐	☐	☐	☐	☐

8. ¿Cuál fue la indicación del patrocinador para compartir datos del proyecto?

9. ¿Cuánto está de acuerdo con cada declaración sobre PLAN DE GESTIÓN DE DATOS?

Marcar apenas una oval por línea.

	Totalmente en desacuerdo	Moderadamente en desacuerdo	Neutral	Moderadamente de acuerdo	Totalmente de acuerdo	No se el tema
Es un elemento útil para indicar los autores y utilizar las licencias.	☐	☐	☐	☐	☐	☐
Fue elaborado para describir la gestión de datos	☐	☐	☐	☐	☐	☐
No sé del tema	☐	☐	☐	☐	☐	☐

Repositorios y Bibliotecarios

Repositório de datos e-cienciaDatos disponibilizado para las universidades públicas da Comunidad de Madrid y la UNED

10. ¿Cuánto está de acuerdo con cada declaración sobre el REPOSITORIO e-ciencia Datos?

Marcar apenas una oval por línea.

	Totalmente en desacuerdo	Moderadamente en desacuerdo	Neutral	Moderadamente de acuerdo	Totalmente de acuerdo
Conozco sus directrices para el uso de licencias y citas	☐	☐	☐	☐	☐
Se incentivan a los investigadores a depositar	☐	☐	☐	☐	☐
Me preocupa la infracción de derechos de autor	☐	☐	☐	☐	☐
No deposito otros datasets porque tiene poca visibilidad	☐	☐	☐	☐	☐
Participé en capacitación o orientaciones que abordan los derechos de autor de los datos.	☐	☐	☐	☐	☐

11. ¿Por qué depositó los conjuntos de datos del proyecto en el repositorio e-cienciaDatos?

12. ¿Depositaría otros datos en el REPOSITÓRIO e-cienciaDatos? ¿Por qué?

13. ¿Qué datos no depositaría en el repositorio e-cienciaDatos? ¿Por qué?

Carrera

Sobre los beneficios, riesgos y el trabajo con la preparación y el intercambio de datos.

14. ¿Cuánto está de acuerdo con cada declaración sobre os BENEFICIOS obtenidos al compartir sus datos?

Marcar apenas uma oval por linha.

	Totalmente en desacuerdo	Moderadamente en desacuerdo	Neutral	Moderadamente de acuerdo	Totalmente de acuerdo
Aumenta mis citas y métricas	☐	☐	☐	☐	☐
Mejora mi reconocimiento entre compañeros	☐	☐	☐	☐	☐
Permite la colaboración y la publicación colectiva.	☐	☐	☐	☐	☐
No sé del tema	☐	☐	☐	☐	☐

15. ¿Cuánto está de acuerdo con cada declaración sobre los RIESGOS obtenidos al compartir sus datos?

Marcar apenas una oval por linha.

	Totalmente en desacuerdo	Moderadamente en desacuerdo	Neutral	Moderadamente de acuerdo	Totalmente de acuerdo
No ser citado	☐	☐	☐	☐	☐
Perder oportunidades de nuevas publicaciones o producción de nuevos productos.	☐	☐	☐	☐	☐
Hacer uso de los datos para fines no previstos en la licencia aplicada	☐	☐	☐	☐	☐
Perder oportunidad de comercialización (patente)	☐	☐	☐	☐	☐
No sé del tema	☐	☐	☐	☐	☐

16. ¿Cuánto está de acuerdo con cada declaración sobre ESFUERZOS dedicados a compartir sus datos?

Marcar apenas uma oval por linha.

	Totalmente en desacuerdo	Moderadamente en desacuerdo	Neutral	Moderadamente de acuerdo	Totalmente de acuerdo
No tengo las habilidades ni el conocimiento necesario para administrar y hacer que mis datos estén disponibles	☐	☐	☐	☐	☐
Tendré mucho trabajo y otros investigadores pueden tener más beneficios como publicación en revistas con un mayor factor de impacto	☐	☐	☐	☐	☐
No sé del tema	☐	☐	☐	☐	☐

Ética y beneficios para la sociedad

Sobre la ética de compartir datos y sus beneficios para la sociedad.

17. ¿Cuánto está de acuerdo con cada declaración?

Marcar apenas uma oval por linha.

	Totalmente en desacuerdo	Moderadamente en desacuerdo	Neutral	Moderadamente de acuerdo	Totalmente de acuerdo
Es deber ético de los investigadores compartir datos en acceso abierto	☐	☐	☐	☐	☐
Os dados compartilhados podem ser usados para fins comerciais	☐	☐	☐	☐	☐

18. ¿Los datos del proyecto fueron depositados en el repositorio durante o después de completar el proyecto? ¿Por qué?

19. Si tiene algún aspecto sobre el derecho de autor que no se ha abordado y que considere importante para la investigación por favor, indique.

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários

APÊNDICE D – PLANO DE GESTÃO DE DADOS

Um Plano de Gestão de Dados criado usando DMPTool

Title: OPERACIONALIZAÇÃO DE REPOSITÓRIOS DE DADOS: UMA ANÁLISE SOBRE AS PERSPECTIVAS E ATITUDES DOS PESQUISADORES NAS QUESTÕES DE AUTORIA E LICENÇA

Criador: Elizabete Cristina de Souza de Aguiar Monteiro

DMP ID: <https://doi.org/10.48321/D1HW2B>

Afiliação: São Paulo State University (UNESP)

Financiador: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (capes.gov.br)

Modelo: Digital Curation Centre

Resumo do projeto:

Repositórios de dados são realidades em muitas instituições internacionais e contribuem para a divulgação e preservação dos dados coletados por sua comunidade. A operacionalização dos repositórios envolve funções diversas e possuem infraestrutura e recursos humanos que estão paulatinamente trabalhando na gestão de dados e nas orientações de estudos sobre a gestão de dados em seu ciclo de vida. Aspectos do direito autoral são preponderantes nas orientações para melhor reuso e citação dos autores e perpassam por diversas temáticas incluindo Ciência aberta, princípios FAIR, agências de fomento, plano de gestão de dados, carreira de investigação, benefícios para os requisitos e aspectos éticos. O objetivo da tese foi apresentar convergências e divergências entre as atitudes e percepções dos pesquisadores e a operacionalização de repositórios de dados no fator Direito de Autor consolidado o grau de aderência dos pesquisadores ao processo de depósito de conjuntos de dados nos repositórios considerando, ainda, a questão da perspectiva para identificar uma possível anomia dos pesquisadores em relação aos repositórios. Foi usar uma pesquisa documental e exploratória para identificação, coleta e análise dos documentos assim como a metodologia Análise de conteúdo para apoiar os procedimentos metodológicos e mapear as fornecidas sobre as questões do direito do autor. A pesquisa foi composta por dois universos: o primeiro composto pelo repositório a questão da perspectiva para identificar uma possível anomia dos pesquisadores em relação aos repositórios. Foi utilizar uma pesquisa documental e exploratória para identificação, coleta e análise dos documentos assim como a metodologia Análise de conteúdo para apoiar os procedimentos metodológicos e mapear as fornecidas sobre as questões do direito do autor. A pesquisa foi composta por dois universos: o primeiro composto pelo repositório a questão da perspectiva para identificar uma possível anomia dos pesquisadores em relação aos repositórios. Foi usar uma pesquisa documental e exploratória para identificação, coleta e análise dos documentos assim como a metodologia Análise de conteúdo para apoiar os procedimentos metodológicos e mapear as fornecidas sobre as questões do direito do autor. A pesquisa foi composta por dois universos: o primeiro composto pelo repositório e *cienciaDados* com seus documentos e bibliotecários; o segundo, composto pelos pesquisadores que depositaram seus conjuntos de dados no repositório. A coleta ocorre por meio da triangulação com a aplicação de questionário, entrevista estruturada e análise documental. Com uma análise do primeiro universo foi proposto um framework com treze índices, com base nas estruturas teóricas, Teoria Institucional de W. Richard Scott, na Teoria do Comportamento Planejado de Icek Ajzen e no modelo de pesquisa de compartilhamento de dados em repositório de dados proposto de Youngseek Kim. O framework é composto por categorias e índices que orientaram as perguntas direcionadas aos pesquisadores. Os resultados demonstram que o repositório conta com infraestrutura tecnológica, recursos humanos, documentos e manuais que instruem pesquisadores sobre os direitos de autor favorecendo o compartilhamento de dados, entretanto, há baixa adesão entre os pesquisadores produzidos pela anomia. A anomia está relacionada a fatores como falta de clareza das leis e normativas, desequilíbrio entre metas institucionais e os caminhos para atingir as, descumprimento de regras e negação da moral. Conclui-se que o Consórcio Madroño e as bibliotecas membro subscrevem seu compromisso com Ciência Aberta, o repositório *cienciaDados* armazena, gerencia e disseminação os dados fornecer das pesquisas de sua comunidade de forma aberta, os conjuntos de dados são protegidos pela propriedade intelectual e as políticas estão em consonância com as diretrizes da Ciência Aberta e princípios FAIR demonstrando a direitos dos direitos de propriedade intelectual para os atores envolvidos. A conscientização dos pesquisadores sobre os benefícios alcançados com o compartilhamento para diminuir a redução de seus dados e a anomia.

Palavras-chave: Repositório de dados. Anomia. Compartilhamento de dados. Direitos autorais. Consorcio Madroño.

ID: 66249

Start date: 07-30-2017

End date: 09-12-2021

Última modificação: 09-09-2021

OPERACIONALIZAÇÃO DE REPOSITÓRIOS DE DADOS: UMA ANÁLISE SOBRE AS PERSPECTIVAS E ATITUDES DOS PESQUISADORES NAS QUESTÕES DE AUTORIA E LICENÇA

DATA COLLECTION

What data will you collect or create?

Foram coletados dados **qualitativos** e qualitativos por meio de entrevistas individuais com roteiro estruturado e com questionário.

How will the data be collected or created?

O questionário foi elaborado com o Formulário do Google e contém 15 páginas. As entrevistas foram realizadas e, as feitas com os pesquisadores foram gravadas pelo *Hangouts* e *Skype*. As entrevistas, em um total de 4 pesquisadores, tem em média 50 minutos cada uma e cada arquivo de vídeo aproximadamente 200 MB. As transcrições das entrevistas foram registradas em documento de texto e os dados do questionário em planilhas. Os arquivos das entrevistas foram nomeados pelo sobrenomes dos entrevistados. As transcrições das entrevistas foram registradas em documento de texto e os dados do questionário em planilhas. Todos os dados serão reutilizados para novas análises e divulgação da pesquisa.

DOCUMENTATION AND METADATA

What documentation and metadata will accompany the data?

Será utilizado o padrão de metadados do repositório da Unesp.

ETHICS AND LEGAL COMPLIANCE

How will you manage any ethical issues?

Com o intuito de cumprir as questões éticas e legais referentes à pesquisa, o projeto foi enviado ao **Comitê de Ética** em Pesquisa da Faculdade de Filosofia e Ciências da Unesp. A pesquisa tem o Termo de Consentimento Livre Esclarecido de todos os participantes. Todas as informações que possam identificar os entrevistados ou violar a privacidade dos dados serão retirados das transcrições ou **anonimizados**. Os questionários respondidos e as transcrições das entrevistas foram armazenados em pasta protegida por senha em computador pessoal da pesquisadora, e uma cópia da pasta armazenadas na nuvem por meio do Google Drive. Os dados serão armazenados no repositório da Universidade, após a anonimização dos dados **sensíveis, identificadores ou semi-identificadores**, e serão acompanhados de uma licença Creative Commons ou open Data Commons. Sempre que os dados forem reutilizados por terceiros, deve-se citar a autora.

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

Os dados pertencem à autora do projeto. Se utilizado em futuras pesquisas a autora deve ser citada pois a citação é uma das boas práticas da gestão de dados. Os dados serão disponibilizados após a publicação dos resultados da pesquisa. A licença Creative Commons ou open Data Commons será aplicada sobre os dados.

STORAGE AND BACKUP**How will the data be stored and backed up during the research?**

O backup é feito pela pesquisadora a cada atualização da análise dos dados. Após a finalização da pesquisa os dados serão depositados no repositório seguindo um protocolo para a anonimização utilizando uma das técnicas mais adequada (supressão, generalização, randomização [adição de ruído] ou troca de dados [swapping]).

How will you manage access and security?

O acesso aos dados será feito apenas pela pesquisadora enquanto estiver no computador pessoal e no Google Drive. Todos os acessos são protegidos por senhas.

SELECTION AND PRESERVATION**Which data are of long-term value and should be retained, shared, and/or preserved?**

Todos os dados coletados serão preservados por tempo indeterminado. No repositório os dados ficarão preservados seguindo sua política.

What is the long-term preservation plan for the dataset?

Seguirá o plano de preservação do repositório.

DATA SHARING**How will you share the data?**

Os dados estarão disponíveis para reutilização por terceiros após a publicação dos resultados da pesquisa. Não serão aplicadas restrições a reutilização dos dados. Os dados disponibilizados no repositório receberão um identificador persistente.

Are any restrictions on data sharing required?

Os dados ficarão disponíveis no repositório assim que os resultados forem tratados e os dados que identifiquem os participantes da pesquisa sejam anonimizados ou retirados.

RESPONSIBILITIES AND RESOURCES**Who will be responsible for data management?**

A pesquisadora foi responsável pela implementação deste Plano de Gestão de Dados assim como por toda gestão do ciclo de vida dos dados. Após o armazenamento no repositório a responsabilidade pela gestão será do mesmo seguindo suas políticas.

What resources will you require to deliver your plan?

O acesso ao repositório.