

**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

ADRIANA CARLA SILVA DE OLIVEIRA

**DESVENDANDO A AUTORALIDADE COLABORATIVA NA *E-SCIENCE* SOB
A ÓTICA DOS DIREITOS DE PROPRIEDADE INTELECTUAL**

**JOÃO PESSOA
2016**

ADRIANA CARLA SILVA DE OLIVEIRA

**DESVENDANDO A AUTORALIDADE COLABORATIVA NA *E-SCIENCE* SOB
A ÓTICA DOS DIREITOS DE PROPRIEDADE INTELECTUAL**

Tese de Doutorado apresentada ao Programa de Pós-graduação em Ciência da Informação do Centro de Ciências Sociais Aplicadas da Universidade Federal da Paraíba, como requisito final para obtenção do título de Doutora em Ciência da Informação.

Linha de pesquisa: Memória, Organização, Acesso e Uso da Informação.

Orientador: Professor Dr. Guilherme Ataíde Dias

**JOÃO PESSOA
2016**

O48d Oliveira, Adriana Carla Silva de.
Desvendando a autoralidade colaborativa na *e-Science* sob
a ótica dos direitos de propriedade intelectual / Adriana Carla
Silva de Oliveira.- João Pessoa, 2016.
297f. : il.
Orientador: Guilherme Ataíde Dias
Tese (Doutorado) - UFPB/CCSA
1. Ciência da informação. 2. Ciência aberta. 3. *e-Science*.
4. Quarto paradigma científico. 5. Propriedade intelectual -
direitos. 5. Dados científicos. 6. Autoralidade colaborativa.

UFPB/BC

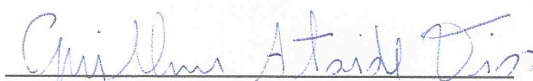
CDU: 02(043)

ADRIANA CARLA SILVA DE OLIVEIRA

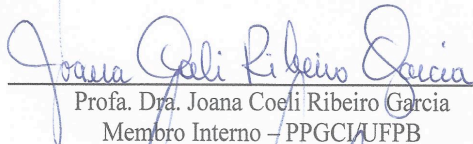
**DESVENDANDO A AUTORALIDADE COLABORATIVA NA *E-SCIENCE*
SOB A ÓTICA DOS DIREITOS DE PROPRIEDADE INTELECTUAL**

Tese de Doutorado apresentada ao Programa de Pós-graduação em Ciência da Informação do Centro de Ciências Sociais Aplicadas da Universidade Federal da Paraíba, como requisito final para obtenção do título de Doutora em Ciência da Informação.

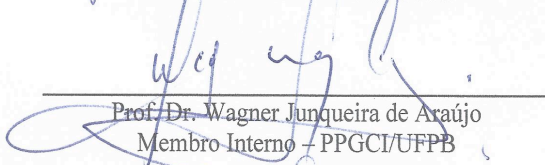
Linha de pesquisa: Memória, Organização, Acesso e Uso da Informação.



Prof. Dr. Guilherme Ataíde Dias
Orientador – PPGCI/UFPB



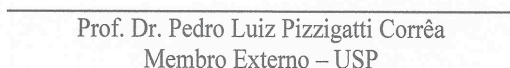
Profa. Dra. Joana Coeli Ribeiro Garcia
Membro Interno – PPGCI/UFPB



Prof. Dr. Wagner Junqueira de Araújo
Membro Interno – PPGCI/UFPB



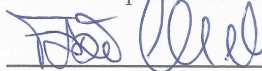
Profa. Dra. Patrícia Borba Vilar Guimarães
Membro Externo – UFRN



Prof. Dr. Pedro Luiz Pizzigatti Corrêa
Membro Externo – USP



Profa. Dra. Dulce Amélia Neves
Suplente Interno – PPGCI/UFPB



Prof. Dr. Fabio Mascarenhas e Silva
Suplente Externo – UFPE

Dedico a minha gratidão a Deus pela misericórdia e graça concedidas.

Dedico a minha mãe, meu pai (*in memoriam*) e aos meus filhos.

AGRADECIMENTOS

Eis que um ciclo se concretiza em minha vida. Após quase quatro anos de dedicação exclusiva à vida acadêmica e ao curso de doutorado muitos aprendizados e frutos foram conquistados. O processo de aprendizagem de um doutorado nos traz realizações e conquistas, apesar de ser árduo e exaustivo, exigindo muito foco, esforço, disciplina e dedicação. E isso nunca me faltou. Contudo, ter contado com a ajuda de muitos coadjuvantes me levou à concretização desse caminho. Coadjuvantes apenas no sentido figurativo, pois cada um teve sua importância imensurável e foi protagonista nesse processo.

Primeiramente minha gratidão a Deus pela vida. Aos meus pais, Antonio Carlos de Oliveira (*in memoriam*) e Maria da Penha Silva de Oliveira o meu eterno agradecimento. Meu pai foi e minha mãe é o maior exemplo de hombridade que já tive na minha trajetória de vida até o momento. Eles foram batalhadores incansáveis para que meus estudos sempre fossem prioridade, mesmo diante de muitas adversidades que a vida nos trouxe. Eles me oportunizaram o legado perpetuo, o conhecimento.

Aos meus valiosos filhos, Geniffer S. de Oliveira Martins e Gyandson S. de Oliveira Martins pela compreensão, cumplicidade e apoio, pois mesmo diante das minhas ausências e abdições sempre foram minha mola propulsora para as batalhas e desafios em nossas vidas. Meu amor incondicional por vocês é eterno.

A minha única e preciosa irmã, Alessandra Oliveira pelo incentivo, motivação e colaboração em todo o processo de estudo e de trajetória de vida. Muito obrigada.

A minha irmã de coração, Alba Lígia pelo carinho, companheirismo e cumplicidade durante a minha trajetória de vida nos últimos 20 anos. Obrigada por tudo.

A Marcelo Aragão pelo incentivo, ajuda e companheirismo nessa longa trajetória. A você eu expresso a minha mais eterna gratidão.

À minha família, tios, tias, primos e primas, as quais nomeio em representação, minhas tias Lourdes Oliveira e Rosângela Silva. Obrigada pela acolhida e carinho de vocês.

À minha tia do coração, Gracinda Soares pela cumplicidade, acolhida e ajuda. A minha admiração e gratidão eternas.

Às minhas inestimáveis amigas-irmãs conquistadas ao longo da minha vida. Algumas são de longas datas e outras tão recentes, mas todas são igualmente importantes em minha vida. Cada uma com sua porção de cumplicidade e colaboração Edilene Silva, Iris Martins, Lourdes Teixeira, Maria Elza, Maria Rodrigues e Maria Isabel.

Aos amigos, André Filipe, Caaci Lima, Carol Mayer, Débora Koshiyama, Dilma

Moraes, Edcleide Silva, Janeide Dantas, Moniki Janegitz, Mônica Carvalho, Myrta Simões e Shirley Mesquita que sempre me apoiaram e acreditaram no meu potencial. Especial agradecimento aos que contribuíram para as revisões e colaborações na feitura da tese. Obrigada a todos vocês.

Aos meus companheiros de jornada doutoral, Ana Córdula, Deise Santos, Derek Tavares, Fabiana França, Fellipe Brasileiro, Hamilton Tabosa, Isaac Newton, Lucas Almeida, Julianne Teixeira, Raissa Carneiro e Rosilene Agapito e aos tantos outros não nominados, mas que foram importantes nessa fase acadêmica. Agradeço a vocês pelo companheirismo em sala de aula, compartilhamento de conhecimentos, alegrias, angústias e trocas de experiências.

Aos professores da banca de qualificação que contribuíram grandemente para que a tese fosse conduzida com qualidade e consistência. A esses professores, muito obrigada. Em especial, dedico meu agradecimento à minha eterna professora e orientadora em momentos passados, Joana Coeli. Agora o destino me presenteou com seu apoio, conselhos e orientações. Obrigada de coração.

Aos professores da banca de defesa agradeço imensamente pela disponibilidade, contribuições atribuídas ao meu trabalho e ao término dessa trajetória.

Ao meu orientador, Guilherme Ataíde pela confiança e orientações que colaboraram para a feitura desta tese. Muito obrigada.

À Capes pela bolsa de financiamento para o doutorado-sanduíche no exterior. Um sonho juvenil de estudar no exterior era impossível quando realizei a graduação em Biblioteconomia na UFPB. Após 23 anos, o sonho foi concretizado na University of Tennessee, Knoxville, Estados Unidos. Não tenho palavras para expressar o aprendizado intelectual e pessoal. Faria tudo novamente. Again!

Meu agradecimento especial, ao professor Luiz Renato Lima pela confiança e oportunidade em intervir junto à University of Tennessee para que me fosse concedida a carta de aceitação. Meu eterno respeito e agradecimento. Amplio o agradecimento a sua esposa Ingrid Lima e suas filhas pela acolhida e apoio durante minha permanência em Knoxville.

À Coordenação do Programa de Pós-graduação em Ciência da Informação da UFPB, professores e servidores que foram imprescindíveis em toda a trajetória acadêmica, contribuindo com novos conhecimentos e aprendizados, além do apoio administrativo.

Ao meu órgão Tribunal de Justiça do Estado do Rio Grande do Norte e a Escola de Magistratura do RN pela concessão do afastamento para estudos durante a realização deste doutorado.

I would like to thank Professor Carol Tenopir from University of Tennessee in

Knoxville (UTK) for her immeasurable intellectual support and for showing me the new possibilities regarding the scenario of open science and scientific data. I had the opportunity to enlarge the scope of my research with the Professor's guidance and lessons. Thank you so much!

I would like to thank everyone I met at the University of Tennessee, Knoxville (UTK) that supported me through this journey, specially the DataONE team, Professor Suzie Allard and Mike Frame for their receptiveness. Thank you!

Special thanks to Misty Jones and Bobbie Suttles for helping me and introducing me to academic support in the UTK.

I also thank my friends at Bridges International in the UTK, specially Eva Mutunga and Michael Henderson for social support.

Os desígnios de Deus nem sempre são compreendidos de imediato, mas com certeza o caminho que Ele nos oferece é sempre o melhor, mesmo diante de uma trajetória longa e árdua. Mas ao final, brotam os sentimentos de dever cumprido e realização profissional e pessoal, mesmo diante de tropeços e dificuldades. As amizades conquistadas e o aprendizado são imensuráveis. A felicidade dessa conquista não está no título, mas simplesmente na minha satisfação interior.

“Não haverá borboletas se a vida não passar por longas datas e silenciosas metamorfoses”.
Rubem Alves (1933-2014)

“Librarians do research too!”
Carol Tenopir (2015)

RESUMO

O estudo traz uma temática inovadora no cenário da ciência contemporânea. A perspectiva apresenta um novo espectro de uma ciência aberta com mudanças nas práticas científicas vigentes. Essas práticas estão sendo aprimoradas, ressignificadas e reconduzidas para as novas dinâmicas no fazer e publicar a pesquisa científica. O quarto paradigma científico conduz essa ciência que é baseada no uso intensivo de dados científicos através das práticas do modelo emergente da *e-Science*. A *e-Science* reflete um ambiente científico de colaboração, compartilhamento, convergência, conectividade, interatividade, uso e reuso de dados científicos. Esse ambiente constitui os pressupostos da ciência aberta e do modelo emergente. O ciclo de vida dos dados é adotado para conduzir e apoiar o gerenciamento de dados científicos. Dessa forma, a tese traz um estudo multidimensional e multidisciplinar através da confluência entre a Ciência da Informação e as Ciências Jurídicas e suas interseções com a Economia e Tecnologia. Teoricamente, o estudo apoia-se na vertente do *commons* preconizado pela Teoria do *Commons* e Economia Criativa, nas regulações e legislações da Propriedade Intelectual vigentes e em diretrizes internacionais para as novas dinâmicas da *e-Science*. O argumento de tese propõe que na *e-Science* a autorialidade é colaborativa e promovida pela recompensa autoral. O objeto de estudo está centrado na autorialidade dos dados científicos como bens intelectuais e o objetivo geral busca elaborar padrões que promovam a recompensa autoral na *e-Science*. A pesquisa é predominantemente qualitativa e adotou a análise de conteúdo de Bardin para a categorização, codificação e inferências do *corpus* de análise. Foi substantiada pelas seis dimensões (epistemológica, teórica, técnica, morfológica, política e ética) do estudo de Bufrem. Tal estudo foi norteador para o desenvolvimento dos capítulos, análise de conteúdo e constituição do modelo conceitual. Utilizou-se para a categorização e codificação do *corpus* de análise o *software NVivo*. A visão multidimensional e conexões temáticas resultaram em cinco categorias e treze subcategorias que ajudaram a alcançar o objetivo e constituir os padrões do modelo proposto para a representação da autorialidade na *e-Science*. Concluiu-se que a autorialidade no contexto da *e-Science* é colaborativa. A autorialidade colaborativa é garantida mediante a recompensa autoral através da atribuição, citação e responsabilização. A atribuição e citação são procedimentos usuais, contudo nas práticas contemporâneas a responsabilidade é atribuída a cada colaborador na proporcionalidade de sua participação. Assim, a tese se confirmou e está representada pelo modelo conceitual de autorialidade colaborativa na *e-Science*. O modelo é composto por padrões multidimensionais que representam o cenário da ciência aberta colaborativa, compartilhada e acessível orientada ao uso e reuso dos dados científicos. Por fim, cada padrão do modelo constitui-se em axiomas norteadores que auxiliarão autores, pesquisadores, curadores, bibliotecários e demais colaboradores, bem como instituições acadêmicas, científicas e agências de fomento a conduzirem projetos de compartilhamento e gerenciamento de dados científicos no contexto da *e-Science* com garantia mínima à autorialidade de todos os envolvidos.

Palavras-chave: Ciência Aberta. *E-Science*. Quarto Paradigma Científico. Direitos de Propriedade Intelectual. Dados Científicos. Autorialidade Colaborativa.

ABSTRACT

This study deals with an innovative theme regarding the scenario of contemporary science. This perspective presents a new spectrum related to open science and changes that occur in current scientific practices. These practices are being improved and present new meanings, towards the new dynamics related to scientific outcomes and publishing. The Fourth Scientific Paradigm leads to a science that is based on intensive use of scientific information through the practices of the emerging model called e-Science. This type of science reflects a collaborative scientific environment that considers sharing, convergence, connectivity, interactivity, use and reuse of scientific data. This environment is based on the assumptions of a more open science and emerging models. In this context, the data life cycle model is adopted in order to drive and support scientific data management. Thus, this thesis constitutes a multidimensional and multidisciplinary study that relies on the confluence between Information Science and Law Sciences and its intersections with Economy and Technology. Theoretically, the study is supported by Commons Theory and Creative Economy; considering current intellectual property regulations and legislations as well as international guidelines for the new dynamics of e-Science. The core argument of this thesis is that in e-Science authority is collaborative practice; promoted by authorship rewards. The study object of the research is centered on authorship of scientific data considered as an intellectual asset. The work aims to elaborate standards that promote reward towards collaborative authority in e-Science. The research is predominantly qualitative. Bardin's content analysis was used for categorizing, coding and performing inferences. The research also relied on the six dimensions (epistemological, theoretical, technical, morphological, political and ethical) according to study of Bufrem that guided the development of the chapters, content analysis and the conceptual model. The *NVivo* software was used for categorization, codification and corpus analysis. The multidimensional view and thematic connections resulted in five categories and thirteen subcategories that helped achieve the objective and indicate the standards of the proposed model for the representation of authorship in e-Science. It concludes that in the context of e-Science, authority is collaborative and ensured by copyright rewards through attribution, citation and accountability. Authorship attribution and citation are usual procedures, but in contemporary practice the responsibility is assigned to each collaborator proportionally. Thus, the thesis is confirmed and is represented by the conceptual model of collaborative authority in e-Science. The model is composed of multi-dimensional patterns that represent the scenario of collaborative open science that focuses on sharing, accessibility; oriented towards to the use and reuse of scientific data. Finally, each standard model represents guiding axioms that will help authors, researchers, curators, librarians, stakeholders, academic institutions, scientific and development agencies to conduct and share scientific data management projects in the context of e-Science to minimally guarantee authorship behalf of all the parties involved.

Keywords: Open Science. E-Science. Fourth Scientific Paradigm. Intellectual Property Rights. Scientific data. Collaborative Authority.

LISTA DE FIGURAS

Figura 1 – Categorias temáticas da tese	24
Figura 2 – Dimensões da pesquisa	25
Figura 3 – Matriz dos padrões multidimensionais	33
Figura 4 – Padrões multidimensionais do modelo conceitual	35
Figura 5 – Paradigmas da ciência	43
Figura 6 – Cenário das publicações científicas	58
Figura 7 – Meios de comunicação	60
Figura 8 – Ciclo de vida de informação científica através do periódico acadêmico	61
Figura 9 – Modelo de ciclo da informação de periódico acadêmico	62
Figura 10 – Gerenciamento e compartilhamento de dados geo-espaciais	64
Figura 11 – Cenário dos dados primários	67
Figura 12 – Imagem da HR 8799 no sistema planetário	68
Figura 13 – Modelo de ciclo de vida dos dados de pesquisa do ICPSR	71
Figura 14 – Modelo de ciclo de vida dos dados de pesquisa da USGS	73
Figura 15 – Papeis e responsabilidades do Modelo USGS	73
Figura 16 – Modelo de ciclo de vida e curadoria	77
Figura 17 – Modelo de ciclo de vida dos dados	79
Figura 18 – Modelo de ciclo de vida dos dados	80
Figura 19 – Modelo para infraestrutura tecnológica	85
Figura 20 – Framework do livre acesso à publicação científica e dados de pesquisa	140
Figura 21 – Dimensões de Infraestrutura	153
Figura 22 – Subcategoria domínio público	163
Figura 23 – Subcategoria acesso aberto	165
Figura 24 – Nuvem de palavras dos princípios norteadores da OECD	182
Figura 25 – Nuvem de palavras da subcategoria autoralidade	209
Figura 26 – Word tree da subcategoria atribuição e citação	216
Figura 27 – Padrões multidimensionais do modelo de autoralidade	230
Figura 28 – Modelo conceitual de autoralidade colaborativa da <i>e-Science</i>	247

LISTA DE QUADROS

Quadro 1 – Procedimentos da análise de conteúdo	31
Quadro 2 – Categorias temáticas da análise de conteúdo	32
Quadro 3 – Tipologias dos dados	51
Quadro 4 – Classificação dos objetos digitais	55
Quadro 5 – Benefícios do compartilhamento de dados	65
Quadro 6 – Níveis dos padrões de metadados	91
Quadro 7 – Orientações para a descrição de metadados	91
Quadro 8 – As cinco gerações do estudo da criatividade	106
Quadro 9 – Políticas de acesso aberto em repositórios institucionais	120
Quadro 10 – Banco de dados da WIPO	128
Quadro 11 – Políticas de <i>copyright</i> e de autoarquivo de editores	136
Quadro 12 – Recomendações para a proteção de dados científicos	191

LISTA DE ABREVIATURAS E SIGLAS

ACTA	<i>Anti-Counterfeiting Trade Agreement</i>
ALA	<i>Atlas of Living Australia</i>
ANDS	<i>Australian National Data Service</i>
ASCII	<i>American Standard Code for Information Interchange</i>
CA	Ciência Aberta
CCS	<i>Cascade Style Sheets</i>
CDI	<i>Community for Data Integration</i>
CILIP	<i>Chartered Institute of Library and Information Professionals</i>
COPE	<i>Committee on Publication Ethics</i>
CUB	Convenção Unionística de Berna
DANS	<i>Data Archiving and Networked Services</i>
DataONE	<i>Data Observation Network for Earth</i>
DCC	<i>Digital Curation Centre</i>
DMPTool	<i>Data Management Plan Tool</i>
DMWG	<i>Data Management Working Group</i>
DOI	<i>Digital Object Identifier</i>
DPA	<i>Data Protection Act</i>
DPIs	Direitos de Propriedade Intelectual
DVN	<i>Dataverse Network®</i>
EACH/USP	Escola de Artes, Ciências e Humanidades da Universidade de São Paulo
EBLIDA	<i>European Bureau of Library, Information and Documentation Associations on Copyright Issues</i>
EC	Economia Criativa
EPUSP	Escola Politécnica da Universidade de São Paulo
FCCN	Fundação para a Computação Científica Nacional
FCT	Fundação para a Ciência e Tecnologia
FDTE	Fundação de Desenvolvimento Tecnológico da Engenharia
FINEP	Financiadora de Estudos e Projetos
FNDE	Fundo Nacional de Desenvolvimento da Educação
FTP	<i>File Transfer Protocol</i>
GII	<i>Global Information Infrastructure</i>
GIZ	<i>Deutsche Gesellschaft für Internationale Zusammenarbeit</i>

Gold OA	<i>Gold Open Access</i>
Green OA	<i>Green Open Access</i>
I2	<i>Institutional Identification</i>
IBICT	Instituto Brasileiro de Informação, Ciência e Tecnologia
ICMBio	Instituto Chico Mendes de Conservação da Biodiversidade
ICPSR	<i>Inter-university Consortium for Political and Social Research</i>
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
IFLA	<i>International Federation of Library Associations and Institutions</i>
IGCE/UNESP	Instituto de Geociência e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”
INDA	Infraestrutura Nacional de Dados Abertos
INPI	Instituto Nacional de Propriedade Intelectual
ISNI	<i>International Standard Name Identifier</i>
ISO	<i>International Organisation for Standardization</i>
JISC	<i>Joint Information Systems Committee</i>
KNAW	<i>Royal Netherlands Academy of Arts and Sciences</i>
LACA	<i>Libraries and Archives Copyright Alliance</i>
LAI	Lei de Acesso à Informação
LCSH	<i>Library of Congress Subject Headings</i>
LDA	Lei de Direitos Autorais
MCTi	Ministério de Ciência, Tecnologia e Inovação
MEC	Ministério da Educação
MeSH	<i>Medical Subject Headings</i>
MMA	Ministério do Meio Ambiente
NAE	<i>National Academic of Engineering</i>
NAM	<i>National Academy of Medicine</i>
NARCIS	<i>National Academic Research and Collaborations Information System</i>
NAS	<i>National Academic of Science</i>
NISO	<i>National Information Standards Organization</i>
NLeSC	<i>Netherlands e Science Center</i>
NRC	<i>National Research Council</i>
NSF	<i>National Science Foundation</i>
NWO	<i>Netherlands Organisation for Scientific Research</i>
OA	<i>Open Access</i>

OAI-ORE	<i>Open Archives Initiative Object Reuse and Exchange</i>
OAI-PMH	<i>Open Archives Initiative Protocol for Metadata Harvesting</i>
OAIS	<i>Open Archival Information System</i>
OECD	<i>Organization for Economic Co-operation and Development</i>
OGP	<i>Open Government Partnership</i>
OJS	<i>Open Journal System</i>
OMC	Organização Mundial do Comércio
OMPI	Organização Mundial de Propriedade Intelectual
ORCID	<i>Open researcher and Contributor Identification</i>
PCT	<i>Patent Cooperation Treaty</i>
PI	Propriedade Intelectual
PIDs	<i>Persistent Identifiers</i>
PKP	<i>Public Knowledge Project</i>
PREMIS	<i>Preservation Metadata Maintenance Activity</i>
RCAAP	Repositório Científico de Acesso Aberto de Portugal
SDLM	<i>Scientific Data LifeCycle Model</i>
SDMX	<i>Statistical Data and Metadata Exchange</i>
TI	Tecnologia da Informação
TICs	Tecnologias de Informação e Comunicação
TPP	<i>Trans-Pacific Partnership</i>
TRIPS	<i>Trade Related Aspects of Intellectual Property Rights</i>
UE	União Europeia
UMinho	Universidade do Minho
UNF	<i>Universal Numerical Fingerprint</i>
URI	<i>Uniform Resource Identifier</i>
USGS	<i>United States Geological Survey</i>
USP	Universidade de São Paulo
WCT	<i>WIPO Copyright Treaty</i>
WIPO	<i>World Intellectual Property Organization</i>
WPPT	<i>WIPO Performances and Phonograms Treaty</i>
WTO	<i>World Trade Organization</i>
XML	<i>Extensible Markup Language</i>

SUMÁRIO

1	INTRODUÇÃO	18
2	TRAJETÓRIA METODOLÓGICA NUMA VISÃO MULTIDIMENSIONAL	23
2.1	MÉTODO CIENTÍFICO	25
2.2	DESCREVENDO AS DIMENSÕES DA PESQUISA	26
2.2.1	Dimensão Epistemológica	26
2.2.2	Dimensão Política	27
2.2.3	Dimensão Teórica	27
2.2.4	Dimensão Morfológica	28
2.2.5	Dimensão Ética	29
2.2.6	Dimensão Técnica	29
2.3	PROCEDIMENTOS DE ANÁLISE	30
2.3.1	Etapas de Codificação e Categorização da Análise de Conteúdo	31
2.4	CONSTITUIÇÃO DO MODELO CONCEITUAL	34
3	CIÊNCIA ABERTA E O QUARTO PARADIGMA CIENTÍFICO	36
3.1	CIÊNCIA ABERTA E <i>E-SCIENCE</i>	36
3.2	(R)EVOLUÇÃO DOS PARADIGMAS DA CIÊNCIA	41
3.3	PESQUISA ABERTA BASEADA NO QUARTO PARADIGMA CIENTÍFICO	47
3.3.1	<i>Constructo</i> Epistemológico da Pesquisa Aberta	47
3.3.2	Proveniência e Tipologia dos Dados de Pesquisa	49
4	MODELOS DE PUBLICAÇÕES CIENTÍFICAS	57
4.1	MODELOS DE COMUNICAÇÃO CIENTÍFICA: ABORDAGEM TRADICIONAL	57
4.2	MODELOS CONTEMPORÂNEOS DE PUBLICAÇÃO CIENTÍFICA	63
4.2.1	Gerenciamento e Compartilhamento dos Dados Científicos	63
4.3	MODELOS DE CICLO DE VIDA DOS DADOS	70
4.3.1	Modelo de ciclo de vida do ICPSR	70
4.3.2	Modelo de ciclo de vida da USGS	71
4.3.3	Modelo <i>Curation Lifecycle</i> do <i>Digital Curation Centre</i> (DCC)	76
4.3.4	Modelo <i>Data Lifecycle</i> do DataONE	78

5	MODELAGEM TECNOLÓGICA PARA O GERENCIAMENTO DOS DADOS CIENTÍFICOS	83
5.1	INFRAESTRUTURA TECNOLÓGICA BASEADA NO MODELO DE CICLO DE VIDA DOS DADOS	83
5.1.1	Primeiro estágio do ciclo: Planejar (<i>Plan</i>)	86
5.1.2	Segundo estágio do ciclo: Coletar (<i>Collect</i>)	87
5.1.3	Terceiro estágio do ciclo: Assegurar (<i>Assure</i>)	89
5.1.4	Quarto estágio do ciclo: Descrever (<i>Describe</i>)	90
5.1.5	Quinto estágio do ciclo: Preservar (<i>Preserve</i>)	92
5.1.6	Sexto estágio do ciclo: Descobrir (<i>Discover</i>)	94
5.1.7	Sétimo estágio do ciclo: Integrar (<i>Integrate</i>)	95
5.1.8	Oitavo estágio do ciclo: Analisar (<i>Analyze</i>)	96
6	DIREITOS DE PROPRIEDADE INTELECTUAL E AS TENDÊNCIAS CONTEMPORÂNEAS DA <i>E-SCIENCE</i>	98
6.1	A CIÊNCIA DO COMUM E COLABORATIVA	98
6.2	ARCABOUÇO POLÍTICO-LEGAL BRASILEIRO DOS DIREITOS DE PROPRIEDADE INTELECTUAL	108
6.1.1	Normatização Legal dos Direitos de Propriedade Intelectual	108
6.2.2	Cenário Brasileiro do Acesso Aberto à Informação e aos Dados Governamentais	117
6.2.3	Acesso Aberto aos Dados Científicos	119
7	DIRETRIZES INTERNACIONAIS POLÍTICO-NORMATIVAS PARA UMA CIÊNCIA ORIENTADA AOS DADOS CIENTÍFICOS	123
7.1	ESCOPO INTERNACIONAL POLÍTICO-NORMATIVO	123
7.1.1	Convenção Unionística de Berna (CUB)	124
7.1.2	<i>Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS)</i>	125
7.1.3	<i>World Intellectual Property Organization (WIPO)</i>	127
7.2	DIRETRIZES POLÍTICO-NORMATIVAS PARA AS PRÁTICAS CIENTÍFICAS DA <i>E-SCIENCE</i>	130
8	<i>STATUS DA AUTORALIDADE NA E-SCIENCE</i>	143
8.1	PONTO DE VISTA LEGAL	145
8.2	AUTORIA ORIGINÁRIA E DERIVADA: UMA REFLEXÃO SOB O PONTO DE VISTA DA PROPRIEDADE	147

8.3	RESSIGNIFICANDO A AUTORIA CIENTÍFICA	149
8.4	AUTORIA COLETIVA OU COLABORATIVA: ONDE RESIDE A AUTORALIDADE NA <i>E-SCIENCE</i> ?	150
8.5	AUTORIA COLABORATIVA E ATRIBUIÇÃO COMPARTILHADA	151
9	NOVAS ACEPÇÕES DO MODELO EMERGENTE DA <i>E-SCIENCE</i>: AUTORALIDADE COLABORATIVA E RECOMPENSA AUTORAL	157
9.1	CONCEPÇÕES CONTEMPORÂNEAS DA CIÊNCIA ABERTA E <i>E-SCIENCE</i>	157
9.2	ENTENDIMENTO DOS DADOS CIENTÍFICOS SOB A PERSPECTIVA DO <i>COMMONS</i>	160
9.2.1	Instrumentos Internacionais Regulatórios	160
9.2.2	Domínio Público	167
9.2.3	Acesso Aberto	164
9.2.4	Bem Intelectual Público e Comum	167
9.3	DIRETRIZES POLÍTICO-NORMATIVAS ORIENTADAS AOS DADOS CIENTÍFICOS	176
9.3.1	Princípios Normativo-legais para Dados Científicos	181
9.3.1.1	Princípio Norteador 1: Proteção da Propriedade Intelectual	183
9.3.1.2	Princípio Norteador 2: Conformidade Legal	189
9.3.1.3	Princípio Norteador 3: Responsabilidade Formal	196
9.3.2	Diretrizes Político-legais Brasileiras	196
9.4	SISTEMATIZANDO E GERENCIANDO OS DADOS CIENTÍFICOS	200
9.5	AUTORALIDADE COLABORATIVA E RECOMPENSA NA <i>E-SCIENCE</i>	207
9.5.1	Autoralidade na <i>e-Science</i>	208
9.5.2	Autoria Colaborativa	213
9.5.3	Atribuição e Citação como Recompensa Autoral	216
10	ENIGMA DESVENDADO: AUTORALIDADE COLABORATIVA NA <i>E-SCIENCE</i>	229
10.1	PADRÃO EPISTEMOLÓGICO-TEÓRICO	230
10.2	PADRÃO POLÍTICO-NORMATIVO	231
10.2.1	Axiomas para a formalidade da política de dados científicos	232
10.2.2	Axiomas para a <i>práxis</i> da política de dados científicos	233
10.3	PADRÃO MORFOLÓGICO-TÉCNICO	233
10.3.1	Axiomas morfológicos para a sistematização técnica e tecnológica	234

10.4	PADRÃO ÉTICO-LEGAL-CULTURAL	237
10.4.1	Primeiro Axioma abrange a permuta dos requisitos autorais vigentes nos sistemas de Propriedade Intelectual para o modelo emergente da <i>e-Science</i>	237
10.4.1.1	Proteção da Propriedade Intelectual	238
10.4.1.2	Conformidade Legal	240
10.4.1.3	Responsabilidade Formal	241
10.4.2	Segundo Axioma direciona a nova acepção de autoria originária e colaborativa	242
10.4.3	Terceiro Axioma baseado no paradigma subjacente do commons e nos dados científicos como bens intelectuais	243
10.4.4	Quarto axioma com ênfase na recompensa autoral	243
10.4.5	Quinto axioma com ênfase na responsabilidade da autoria colaborativa	245
11	CONCLUSÃO	248
	REFERÊNCIAS	256
	APÊNDICES	278

1 INTRODUÇÃO

A tese discute sobre ciência aberta, práticas científicas da *e-Science* e direitos de propriedade intelectual sob uma visão multidimensional e abordagem multidisciplinar. A primeira norteia a construção epistemológica e teórica da tese com base nos estudos de Bufrem (2013) sobre as dimensões da pesquisa científica. A condução do estudo multidimensional traz um panorama de que a ciência aberta se constitui em dimensões políticas, normativas, legais, científicas, técnicas, tecnológicas, éticas e socioculturais. Enquanto a multidisciplinaridade permite compreender as interseções entre a Ciência da Informação, as Ciências Jurídicas, a Economia e a Tecnologia.

O fio condutor da pesquisa tem como ponto de partida o entendimento das diferentes nuances do modelo vigente da comunicação científica – ênfase na publicação de resultados científicos – e do emergente nomeado de *e-Science* – com foco no uso e reuso de dados de pesquisa. Consecutivamente, também estuda os Direitos de Propriedade Intelectual (DPIs) permeados pela ideia da ciência colaborativa e do *commons* que preconizam o bem intelectual como comum, público e aberto.

Na comunicação científica um novo limiar desponta com mudanças no modo de fazer e publicar os resultados científicos. Os processos e práticas da comunicação científica, que por séculos, foram apoiados pela cultura impressa que tornava públicos os resultados dos processos de investigação, atualmente, ampliaram-se por meio dos movimentos *Open Access* e *Open Science*.

Tais movimentos provocaram alterações na geração, acesso e uso de publicações científicas impressas e eletrônicas e avançam para o uso intensivo de dados científicos como produto primário de pesquisa. Novos horizontes estão despontando para uma ciência aberta que pressupõem transformações nas práticas científicas no cenário mundial, a qual concentra em seus esforços diferentes colaboradores e ambientes com ativistas, pesquisadores e instituições de fomento, acadêmicas, científicas e governamentais direcionadas à promoção da produção científica coletiva, colaborativa, produtiva e inovadora.

Ainda na seara científica as transformações são notadamente mais inovadoras nas práticas baseadas no novo paradigma científico nomeado de *Fourth Paradigm*, o qual conduz as práticas da *e-Science*. Esse paradigma pressupõe que as investigações científicas aproveitem seus dados oriundos de pesquisas como fontes primárias do processo investigativo. Ainda, direciona para que tais práticas sejam conduzidas sob a perspectiva da ciência compartilhada, colaborativa, acessível, digital e aberta. Essas práticas estão sendo

aprimoradas, ressignificadas e reconduzidas para as novas dinâmicas do fazer científico baseado na gestão, uso, reuso e reprodutibilidade dos dados científicos.

A apropriação e adoção de uma visão multidisciplinar que envolve as transformações da comunicação científica sob a ótica da Ciência da Informação e dos Direitos de Propriedade Intelectual foi um desafio que resultou em um alargamento da compreensão sobre o objeto de estudo: dados científicos na ciência aberta. A Economia trouxe uma nova concepção dos bens intelectuais como bens comuns e públicos sob a ótica da Teoria do *Commons* e da Economia Criativa. A tecnologia é um suporte imprescindível às práticas da *e-Science*.

As novas dinâmicas do fazer científico apresentam uma confluência entre as práticas científicas vigentes e emergentes em direção ao aprimoramento de teorias, métodos, modelos, processos, funcionalidades e tecnologias. A *e-Science* utiliza um novo modelo científico chamado de ciclo de vida dos dados, conhecido na literatura internacional como *Data Lifecycle* para a sistematização das práticas científicas. Esse modelo é constituído em estágios que conduzem à gestão dos dados científicos e estimulam a colaboração entre pesquisadores e instituições através do compartilhamento e de parcerias institucionais.

Na contemporaneidade, as dinâmicas científicas caminham para uma reflexão acerca dos bens intelectuais como abertos, comuns e públicos e, conseqüentemente, para a necessidade de mudanças no escopo dos sistemas vigentes dos DPIs. Os bens intelectuais podem ser constituídos da informação, do conhecimento e dos dados de pesquisa como produtos primários do processo criativo, científico e cultural.

Os DPIs vigentes são postos à prova mediante o questionamento do seu paradigma proprietário econômico – que enfatiza a autoria individual e exclusiva regida pelos atuais sistemas - frente a um novo paradigma que apregoa o compartilhamento e a colaboração – incitado pela cultura do *commons*. Krokosz (2015) e Albagli (2015) explicitam que o paradigma do *commons* é alicerçado por práticas relacionadas com a autoria colaborativa e coletiva através das cocriações e colaborações que estão representadas na *e-Science*, ciência aberta, inovação aberta, produções *peer-to-peer*, *wiki* e *crowdsourcing*.

A cultura colaborativa e compartilhada provoca reflexões no entendimento da autoria tradicional e, por conseguinte, no *status* do autor dentro do espectro dos sistemas autorais vigentes. O *status* do autor na cultura colaborativa se volta para a ampliação da autoria para a autorialidade. A autorialidade em um sentido *strictu sensu* abrange a coletivização da autoria, construída por um processo de criação intelectual, criativo e cultural com diversos sujeitos, que vai além da figura do autor ou pesquisador e estende-se aos diferentes colaboradores e

instituições partícipes, onde cada um exerce sua função individual ou condição de autor, em um espaço amplo e coletivo.

A interseção com a tecnologia proporciona infraestrutura sustentável com ferramentas, protocolos, licenças, repositórios e *software* que permeiam todo o processo de gerenciamento de dados científicos durante o seu ciclo de vida. Essa infraestrutura é a sustentabilidade técnica e tecnológica para o compartilhamento e gestão que envolve os processos de coleta, curadoria, preservação, visualização e análise ampliando as possibilidades para o uso, reuso e reprodutibilidade de dados científicos em longo prazo.

Assim, a tese se pauta no *constructo* de uma ciência aberta direcionada aos dados científicos como um bem intelectual comum, aberto e público. Apoia-se na vertente da cultura do *commons* e parte da primeira premissa que as práticas científicas da *e-Science* são baseadas na colaboração, no compartilhamento, na coletividade e na autoria colaborativa.

Os pressupostos que conduzem a tese versam sobre os novos contornos dos direitos do autor que assumem a acepção baseada na autoralidade colaborativa quando se tratar de práticas que envolvem dados científicos, em contraponto ao DPIs arraigados e apoiados nos requisitos de originalidade e exclusividade centrados na individualidade autoral.

O objeto de estudo dessa tese é a **autoralidade colaborativa dos dados científicos na *e-Science***. Estão intrínsecas na acepção da autoralidade colaborativa, as características de pluralidade e múltiplas colaborações que permeiam o processo de comunicação científica, tornando-se necessário repensar a proteção autoral neste novo ambiente. A ideia precípua é que a proteção autoral revestida pelos requisitos tradicionais dá espaço a uma recompensa que emerge do processo intelectual coletivo e é baseado nos requisitos de originalidade relativa e novidade temática, quando se tratar do uso e reuso de dados científicos no contexto da *e-Science*.

A autoralidade colaborativa parte das premissas que a criação primária está relacionada com a origem (proveniência dos dados) e a criação derivada com o reuso (dados derivados ou secundários). Assim, a proteção é reconfigurada para a recompensa autoral por meio da atribuição e citação. É através da recompensa que o reconhecimento e a responsabilidade autorais são promovidas.

Embora, a atribuição e a citação sejam práticas vigentes na comunicação científica tradicional, o entendimento na *e-Science* está sendo reconduzido para que recompensa autoral seja proporcional à colaboração e à responsabilidade de cada envolvido. A responsabilização se volta para a contribuição individual e a conduta ética com que cada colaborador se envolve e participa no processo. No Direito a nova configuração se volta para a proteção do direito de

autor, garantida pela recompensa e licenciamento, a partir de um ambiente coletivo, colaborativo e interativo, sem a predominância dos requisitos de exclusividade e individualidade que protegiam a propriedade intelectual do autor.

Isto posto, o recorte da pesquisa é referente à autoralidade colaborativa dos dados científicos. O argumento de tese parte da afirmação que na *e-Science* a autoralidade é **colaborativa e promovida pela recompensa autoral**. Com base nesse argumento de tese levanta-se a seguinte questão de pesquisa: **como a recompensa autoral é promovida na e-Science considerando que a autoralidade é colaborativa?**

Em observância a essa questão, a hipótese arguida é que **a recompensa autoral seja promovida, por meio da atribuição, citação e responsabilização dos envolvidos proporcionalmente à sua contribuição no processo de pesquisa**. O contexto contemporâneo da comunicação científica, traz mudanças em sua forma de atribuição e citação, e por conseguinte, a responsabilidade de cada envolvido é proporcionalmente vinculada a sua participação. Essa proporcionalidade é entendida como uma contribuição individual em um processo que envolve múltiplas colaborações e uma pluralidade de autorias. Portanto, a hipótese se fundamenta em um espaço de múltiplas autorias e novas formas de recompensa autorais.

Assim, a busca pela resposta ao problema proposto e a hipótese arguida, considerou a contextualização e as premissas apresentadas para o seguinte objetivo geral: **elaborar padrões que promovam a recompensa autoral na e-Science**.

Os objetivos específicos são os seguintes:

- a) Contextualizar as dinâmicas contemporâneas científicas da ciência aberta, *e-Science* à luz do quarto paradigma científico e do ciclo de vida dos dados;
- b) Compreender o escopo político-normativo-legal à luz dos direitos de propriedade intelectual e do movimento do *open access* e *open science*;
- c) Entender os dados científicos como bens intelectuais sob a perspectiva da Teoria do *Commons* e da Economia Criativa;
- d) Apresentar as diretrizes internacionais político-normativas orientadas à ciência dos dados;
- e) Representar por meio de um modelo conceitual os padrões que promovem a recompensa da autoralidade colaborativa na *e-Science*.

Com base no exposto, a tese se apoia em três aspectos motivantes. A **inovação** por trazer à tona assuntos relevantes que estão emergindo no cenário mundial da comunicação

científica como uma temática de pesquisa inovadora, multidisciplinar e multidimensional pautada na ciência colaborativa e aberta.

A **novidade** é refletida pela contextualização e compreensão de questões epistemológicas, teóricas, legais e morfológicas relacionadas com um ambiente de práticas científicas, colaborativas e compartilhadas. À comunidade científica e em especial, autores, pesquisadores e colaboradores, a tese aponta para diferentes perspectivas no modo de fazer, usar e reusar os dados primários e resultados (preliminares e finais) de investigação científica.

A **originalidade** da tese traz uma essência única, singular e original relacionada ao objeto de estudo com a reflexão de um tema emergente e a interseção entre disciplinas distintas do conhecimento. O estudo é singular quando aborda a autoralidade colaborativa do reuso de dados científicos na perspectiva da ciência aberta. Essa originalidade apresenta a abordagem multidimensional como norteadora para uma política para dados científicos orientada à ciência aberta e direciona o entendimento da *práxis* da *e-Science*. Tal abordagem oportunizará uma visão abrangente e dimensional da ciência contemporânea à comunidade científica, instituições de ensino e pesquisa, agências de fomento e órgãos governamentais no cenário nacional, quiçá ao cenário internacional. Para ilustrar a abordagem contemporânea a tese finaliza com um modelo conceitual constituído por padrões dimensionais. O modelo é aplicável mediante a adoção de diretrizes mínimas e de conformidade internacional que garantam a recompensa autoral na *práxis* da *e-Science*. A conformidade internacional aplicada ao modelo foi pensada em uma abrangência local e internacional.

A tese considera duas categorias básicas, a ciência aberta e a propriedade intelectual acrescidas da Economia e da Tecnologia como temas agregadores e transversais que constituiu os capítulos. O primeiro capítulo é a introdução. O segundo capítulo descreve o percurso metodológico empregado na tese. Os capítulos terceiro e quarto contextualizam as temáticas ciência aberta, *e-Science*, quarto paradigma científico e modelos de publicações científicas. O quinto capítulo trata da infraestrutura tecnológica que apoia o ciclo de vida do gerenciamento de dados científicos. O sexto capítulo discorre sobre os dados científicos como bens intelectuais na perspectiva da Teoria do *Commons* e da Economia Criativa e o escopo político-normativo-legal dos direitos de propriedade intelectual no cenário brasileiro. O sétimo capítulo trata das diretrizes internacionais político-normativas. O oitavo capítulo traz reflexões sobre a autoralidade na *e-Science*. A análise constitui o nono capítulo com as novas acepções da autoria e da recompensa autoral na *e-Science*. Por fim, o capítulo final apresenta o modelo conceitual que traz a representação da autoralidade colaborativa na *e-Science*. Em seguida, a tese é finalizada com a conclusão, referências, apêndices e anexos.

2 TRAJETÓRIA METODOLÓGICA NUMA VISÃO MULTIDIMENSIONAL

A trajetória metodológica que conduz a tese tem como ponto de partida o objeto de estudo centrado na **autoralidade colaborativa dos dados científicos**. Essa trajetória é constituída pelos procedimentos metodológicos que abrangem o método científico, a análise de conteúdo e a proposta do modelo conceitual de autoralidade colaborativa.

A multidisciplinaridade encontra-se na ideia de que o conhecimento pode ser dividido em partes (disciplinas), [...] a disciplina é um tipo de saber específico e possui um objeto determinado e reconhecido, bem como conhecimentos e saberes relativos a este objeto e métodos próprios (MENEZES; SANTOS, 2015, *on-line*).

constitui-se, então, a partir de domínios específicos e relações com objetos distintos de conhecimento. Assim, a contribuição de cada domínio específico e sua relação com outros conhecimentos se direciona para o objeto de estudo como recorte multidisciplinar dessa tese. Esse recorte foi constituído em duas categorias temáticas. Filosoficamente, entende-se que categorias são unidades de classificação utilizadas para ordenar e relacionar fatos e experiências, portanto ajudam a classificar e ordenar as coisas. Constitui o sistema de orientação com o qual pensamentos, procedimentos e *insights* são organizados e sistematizados dentro do processo investigativo.

A primeira categoria é denominada de **Ciência Aberta (CA)** direcionada as questões emergentes relacionadas com a *e-Science*, quarto paradigma científico, modelos e dados científicos. A segunda categoria é definida como **Propriedade Intelectual (PI)** que versa sobre os direitos de propriedade intelectual (DPIs), instrumentos regulatórios internacionais, legislações específicas e diretrizes internacionais político-normativas. A confluência entre essas duas categorias constitui o *constructo* da pesquisa e sustenta o argumento de tese.

Além das categorias, a tese apresenta dois pontos transversais que conduzem para a interseção entre a Tecnologia e a Economia. A transversalidade de domínios específicos com outras áreas é “a abordagem epistemológica dos objetos de conhecimento [...] e diz respeito à compreensão dos diferentes objetos de conhecimento possibilitando a referência a sistemas distintos [...]” (MENEZES; SANTOS, 2015, *on-line*). A complexidade entre objetos do conhecimento em disciplinas distintas aponta para as interseções transversais da tese.

A interseção com a **Tecnologia** visa discorrer sobre a infraestrutura tecnológica sustentável em uma cultura colaborativa. A transversalidade com a **Economia** está centrada na relação entre os DPIs e a *e-Science* com o direcionamento dos dados científicos como bens intelectuais em uma cultura colaborativa. Nessa direção, as categorias temáticas direcionou

a construção dos capítulos com sustentabilidade epistemológica, teórica e morfológica conforme demonstradas na Figura 1.

Figura 1 – Categorias temáticas da tese



Fonte: Elaborado pela autora (2016).

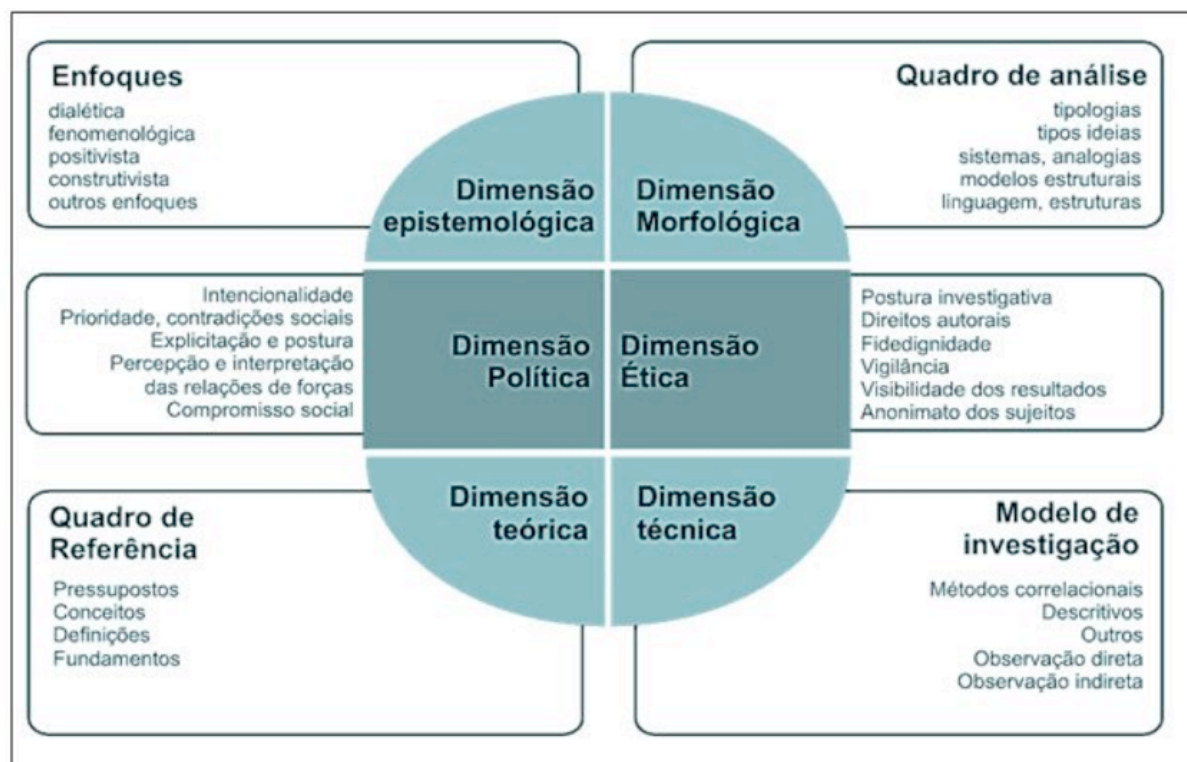
Acerca da visão multidimensional a pesquisa aponta para duas abordagens. A primeira abordagem norteia a construção da tese através das dimensões de pesquisa apresentadas no estudo de Bufrem (2013). A segunda abordagem foi adotada para a compreensão das múltiplas dimensões ou aspectos das práticas científicas da *e-Science* e os DPIs no contexto da ciência aberta. A ideia de trazer uma visão multidimensional como parte do *constructo* do estudo ilustra as dimensões (políticas, normativos, epistemológicos, teóricos, morfológicos, legais, éticos e culturais) necessárias para um ambiente colaborativo voltado para uma ciência aberta e práticas da *e-Science*. As seções seguintes discorrem acerca do método científico, análise de conteúdo e modelo conceitual.

2.1 MÉTODO CIENTÍFICO

O método científico está apoiado nas seis dimensões da pesquisa de Bufrem (2013), que tem como base epistemológica o método quadripolar idealizado por Paul de Bruyne, J. Herman e M. de Schoutheete (1974). No contexto da Ciência da Informação a abordagem quadripolar tem a finalidade de compreender o objeto de estudo como uma construção social, “a partir das transformações da realidade, das condições para o fazer científico, dos objetos reais, ideais ou mentais, representados por dados, informações, conhecimentos e modelos [...]” (BUFREM, 2013, p. 2). A dinâmica de investigação resulta de uma interação dialética entre dimensões que possibilita versar entre a teoria e a prática em diferentes contextos.

Bufrem (2013) elaborou um modelo alternativo e multidimensional para análise e discussão de pesquisas científicas, ressaltando natureza, gêneros e dimensões (SILVA, 2016). O modelo apresentado na Figura 2 demonstra uma visão dimensional de como os estudos podem ter diferentes abordagens e perspectivas em um mesmo *constructo* de pesquisa.

Figura 2 – Dimensões da pesquisa



Fonte: Bufrem (2013, p. 36).

O modelo apresentado por Bufrem foi constituído a partir de quatro polos (epistemológico, teórico, morfológico e técnico) e acrescido com os polos (político e ético). A

atualização do método quadripolar trouxe novos direcionamentos e abordagens que constituíram esse modelo a partir de dimensões.

Dessa forma, o modelo multidimensional traz uma ampliação dos aspectos e características da pesquisa moderna e foi adotado como norteador para a construção dos capítulos teóricos, das inferências emergidas da análise de conteúdo e na construção do modelo conceitual que reflete as novas dinâmicas científicas.

2.2 DESCRREVENDO AS DIMENSÕES DA PESQUISA

Como apresentado na seção anterior o método científico é constituído pelas dimensões da pesquisa: epistemológica, política, teórica, morfológica, ética e técnica.

2.2.1 Dimensão Epistemológica

Essa dimensão delimita o *constructo*, o objeto de estudo e a problemática de investigação. Os pesquisadores exprimem e exercem uma vigilância crítica e a objetivação (produção e explicitação do objeto e problemática) da pesquisa (BRUYNE; HERMAN; SCHOUTHEETE, 1977). A dimensão discursiva não remete diretamente ao conteúdo propriamente dito, mas a apreensão de ideias, linguagens e reflexões sobre o objeto de estudo. A dimensão epistemológica exerce o domínio sobre o objeto científico em permanente construção e reformulação. A convicção e atitude científica do pesquisador sobre o objeto de estudo conduz a sua aceitação ou crítica com a produção de novos sentidos e significados sobre a ciência contemporânea.

A dimensão epistemológica nessa tese é expressa a partir das reflexões: dos paradigmas da ciência observacional, experimental e computacional e do quarto paradigma da *e-Science*; os dados científicos como bens intelectuais; e, legislações e iniciativas políticas e normativas frente ao objeto de estudo. Assim, a partir da compreensão epistemológica, a nova acepção da autorialidade colaborativa aflora da *e-Science* e provoca reflexões no cerne dos DPIs no bojo de uma ciência aberta.

A noção de autorialidade nasce em função do processo criativo e coletivo, “avança além dos territórios individuais que produzem fragmentos criativos a serem inseridos na obra: ela se projeta no espaço que se estabelece como território existencial coletivo. A concepção sai da privacidade para o espaço público da sala” (TROTTA, 2006, p. 163). Na perspectiva jurídica a concepção de autoria está em se ao ‘eu-autor’ plural vem se reunir um ‘nós-

autores’, pode-se falar então em produção de uma subjetividade, fonte para a coletivização da autoria” (TROTТА, 2006, p. 163). Assim na perspectiva contemporânea da *e-Science* o termo autoralidade abrange pluralidade, coletividade e colaboração de sujeitos.

2.2.2 Dimensão Política

Na tese a opção de dar ênfase a essa dimensão está associada ao contexto específico que abrange a ciência aberta e os DPIs. A noção do termo política “pertence ao domínio do conhecimento prático e é de natureza normativa, estabelecendo os critérios da justiça [...]” (JAPIASSÚ; MARCONDES, 2006, p. 220).

O arcabouço legal é insuficiente no provimento de regulamentações e instrumentos normativos que abarquem as necessidades advindas da sociedade contemporânea, digital e interativa. Assim, na ciência moderna os instrumentos jurídicos e as legislações específicas utilizam-se do viés político para dar sustentabilidade as estratégias e diretrizes que devem ser desenvolvidas e adotadas para a promoção da pesquisa científica acessível, aberta, gratuita e sustentável.

A estrutura política utiliza-se de instrumentos normativos e regulatórios para dar sustentabilidade às diretrizes político-normativas. Tais instrumentos são princípios, prescrições e regras que conduzem às decisões em um determinado contexto, os quais exercem um poder simbólico e normativo. A ‘norma’ serve como um “conjunto de prescrições relativas, as quais podemos decidir se algo é verdadeiro ou falso [...]” (JAPIASSÚ; MARCONDES, 2006, p. 220), ajudando a nortear os direcionamentos em um determinado contexto específico.

Assim, a dimensão política visa direcionar as diretrizes da ciência aberta e complementa o escopo político-normativo-legal dos DPIs. As diretrizes político-normativas brasileiras são direcionadas ao acesso aberto dos dados governamentais e ao governo eletrônico e no cenário internacional as iniciativas são abrangentes e contemplam políticas com diretrizes orientadas aos dados de pesquisa.

2.2.3 Dimensão Teórica

A dimensão teórica orienta o pesquisador na elaboração de conceitos, pressupostos ou hipóteses, interpreta fatos ou fenômenos e define provisoriamente soluções de pesquisa (BUFREM, 2013). A pesquisa progride concomitante ao processo de elaboração do

referencial teórico. As noções preliminares que partem do senso comum e se estabelecem sistematicamente no escopo teórico são baseadas em conceitos, definições, enunciados e regras metodológicas direcionadas pela dimensão teórica.

Nesse estudo, a dimensão é direcionada para explicar e compreender o *constructo* teórico do objeto de estudo, pressupostos e conceitos norteadores em todos os capítulos da tese. A dimensão busca aproximar o referencial teórico, as inferências da análise de conteúdo e o modelo conceitual voltados para o objeto de estudo. As indagações postas no problema e objetivos são investigadas, refletidas e argumentadas na construção teórica.

A teoria é constituída por um ‘feixe de conceitos’ ou sistema conceitual que reflete um conjunto coerente de conceitos (MACHADO, 2007). É uma visão sistemática e encapsulada de conceitos que forma um todo coerente. O conceito, dessa feita, é a:

existência de uma questão, a formulação de um problema, a teoria apresenta determinada resposta, sugere uma solução. Privilegiar o conceito significa valorizar a ciência como processo. Esse aspecto dinâmico que caracteriza o conceito - e faz da ciência o domínio operatório [...] (MACHADO, 2007, p. 21).

O conceito é, pois, a abstração mental que se propõe em diferentes abordagens à reconstrução e à ressignificação de um contexto específico. Os conceitos devem ser formulados por nós – não os recebemos prontos – possibilitam o trabalho científico – a mediação do saber. Assim, a dimensão teórica contribui para a construção dos conceitos chaves na tese, a saber: ciência aberta, *e-Science*, paradigmas, dados científicos e autoridade colaborativa. Além disso, o modelo conceitual que representa o objeto de estudo é composto por elementos conceituais que refletem essa dimensão.

2.2.4 Dimensão Morfológica

A dimensão morfológica é a estruturação e objetivação do objeto científico, o qual, adquire um espaço de representação. A representação formal do objeto científico, compreende: exposição do objeto do conhecimento, estilo pelo qual o investigador expressa os resultados e construção dos modelos de interpretação (BUFREM, 2013; SILVA, 2016).

Os resultados da investigação são formalizados e representam o objeto científico estudado e explica todo o processo de pesquisa e análise científica ao redor do objeto. É o pensar do objeto num espaço configurativo, um espaço qualquer de representação, por articular os conceitos, os elementos, as variáveis numa arquitetônica mais ou menos

rigorosamente construída (BRUYNE et al., 1977). O objeto passa a ser configurado em uma articulação teórica, operacional, representacional, estrutural e prática.

Nesse estudo, a dimensão é composta pelos modelos científicos que estão presentes nas práticas da *e-Science* e reflete também a representação do modelo conceitual. No referencial teórico é possível analisar a dimensão morfológica sob o prisma do modelo de ciclo de vida dos dados e infraestrutura tecnológica. Tal modelo apresenta uma sistemática de estágios que estrutura as práticas no gerenciamento de dados científicos no ambiente da *e-Science*. No modelo conceitual a dimensão articula teorias, modelos e inferências para explicitar o objeto de estudo e atender ao objetivo geral.

2.2.5 Dimensão Ética

A dimensão ética é contemplada pela preocupação com a fidedignidade, veracidade e honestidade que o pesquisador deve ter no processo de pesquisa e nos resultados finais. Nas práticas de pesquisa aberta, a preocupação é intensificada pela conduta ética, comportamento, honestidade, fidedignidade e veracidade científicas de pesquisadores, colaboradores e instituições. A postura e conduta ética investigativa são aspectos importantes que devem ser vigiados no processo de compartilhamento e colaboração científica.

Na tese a reflexão foi construída com base em dois espectros ético-legais. O primeiro reflete as questões legais relacionadas com os dados científicos de natureza pessoal, sigilosa e restrita, que na maioria das vezes obedecem instrumentos normativos ou legislações específicas. Para nessa reflexão os critérios de veracidade, fidedignidade, autenticidade, sigilo e anonimização que estão presentes no escopo ético-legal. O segundo espectro conduz a tese para a recompensa autoral. A mudança no *status* da autoridade dentro da *e-Science* intensifica preocupações legais quanto as condições e restrições de uso e reuso associadas à recompensa autoral de dados científicos.

Isto posto, a dimensão ética contemplada durante o desenvolvimento da tese, norteou a pesquisadora a ter uma conduta baseada na fidedignidade, veracidade e honestidade para a feitura do conteúdo. Além de direcionar para as reflexões de questões ético-legais nas práticas científicas emergentes.

2.2.6 Dimensão Técnica

A dimensão técnica estabelece a relação entre a construção do objeto científico e a

realidade através das abordagens, sistemáticas e operações técnicas. Nessa dimensão, os dados sobre o mundo real são obtidos mediante estratégias de pesquisa e técnicas de coleta de dados e evidências pertinentes à problemática da investigação, aos objetivos da pesquisa, ao universo e à amostra (BUFREM, 2013; SILVA, 2016). Possui a finalidade de explicar procedimentos, técnicas, experimentações que são adotadas no decorrer da pesquisa. É quando o objeto de estudo entra em contato com a realidade objetivada e se materializa por meio de instrumentos, dispositivos metodológicos, teorias formuladas, conceitos operacionais e modelos representacionais.

2.3 PROCEDIMENTOS DE ANÁLISE

A pesquisa é predominantemente qualitativa e adotou a análise de conteúdo de Bardin (2011) para a categorização, codificação e inferências do *corpus* de análise. Foi substantiada por categorias e norteadas pelas seis dimensões do estudo de Bufrem (2013). Com base nas dimensões de pesquisa, a análise partiu dos pressupostos e foi guiada pelos conceitos-chaves: ciência aberta, *e-Science*, quarto paradigma, dados científicos, propriedade intelectual, ciência do *commons*, autoralidade colaborativa e recompensa autoral.

A análise de conteúdo é definida como um conjunto de técnicas que norteiam as interpretações e apresenta duas funções: heurística e prova. “A função heurística conduz a análise de conteúdo e enriquece a tentativa exploratória, aumentando a propensão para a descoberta. A prova se apresenta na forma de afirmações provisórias e diretrizes” (BARDIN, 2011, p. 35). Assim, adotou-se a análise de conteúdo temática em virtude das funcionalidades que possibilitam: a gestão de ideias, dados e informações para sistematização do raciocínio epistemológico e conceitual a partir de categorias representadas por nós. Os nós e sub-nós são as categorias e sub-categorias, respectivamente, emergidas dos conteúdos (textuais, audiovisuais e interativos) e em diferentes formatos (Word, PDF., Excel, MP3., JPG.); e a modelagem visual a partir do *software* NVivo oportunizou a elaboração de tabelas, diagramas, nuvens de palavras, mapas e construção de modelos.

A análise temática foi construída por uma grade de categorias assentadas no *corpus* de análise, direcionada ao objeto de estudo e guiada pelos pressupostos e objetivos. O processo de codificação também denominado por unidade de registro, orientou na codificação e categorização dos conteúdos. Bardin (2011) afirma que a categorização é uma operação de classificação de elementos constitutivos de um conjunto por diferenciação e, em seguida, por reagrupamento com os critérios previamente definitivos. Podem ser reunidos por tópicos,

temas ou casos, configurando no critério de categorização semântica ou temática. Com base na categorização temática, as unidades de registro foram codificadas por temas, frases, parágrafos e sites.

Para a execução dos procedimentos de codificação e categorização foi utilizado o *Software NVivo* (versão 11 Mac). Os resultados obtidos a partir dos procedimentos gerados no *Software Nvivo* ajudaram a constituir o capítulo da análise de conteúdo e gerar representações a partir das categorias (*nodes*) e subcategorias.

2.3.1 Etapas de Codificação e Categorização da Análise de Conteúdo

A pré-análise, descrição analítica e interpretação inferencial compuseram a sistemática do *corpus* de análise realizada no período de seis meses (fevereiro a julho de 2016). Esse *corpus* de análise foi guiado pelo referencial teórico e pelas categorias temáticas: ciência aberta e propriedade intelectual. As fontes de pesquisa foram constituídas por artigos científicos (nacionais e internacionais), base de dados (*Elsevier*, *Proquest*, *Wiley Online Library*), legislações nacionais e documentos resultantes de diretrizes, normativas, iniciativas e instrumentos regulatórios (acordos, tratados e convenções). Também fizeram parte do escopo portais governamentais, organizações não governamentais, repositórios de dados (científicos e governamentais), sites de instituições de ensino, de pesquisa e agências de fomento. Todo o levantamento bibliográfico foi por meio eletrônico, com predominância do conteúdo no idioma inglês, em virtude das fontes de pesquisa serem oriundas de diferentes países, em especial, os que já possuem práticas consolidadas de ciência aberta. O escopo brasileiro foi analisado por meio de legislações específicas constituídas pela Lei de Direito Autoral (LDA), Lei de Acesso à Informação (LAI), decretos, instruções normativas e algumas iniciativas no âmbito dos órgãos governamentais e instituições científicas voltadas para o acesso aberto, dados abertos (governamentais e científicos). Os procedimentos adotados para a sistemática da análise de conteúdo estão demonstrados no Quadro 1.

Quadro 1 – Procedimentos da análise de conteúdo

ETAPAS	PROCEDIMENTOS	DESCRIÇÃO
1ª. Etapa PRÉ-ANÁLISE	COLETA E ORGANIZAÇÃO	Análise preliminar: - Coletou-se o material de análise (artigos, legislação, livros, diretrizes, vídeos e sites); - Realizou-se a separação do conteúdo e transferência dos arquivos relacionados com o referencial teórico da tese para <i>Sources - Internals</i>; - Criou-se pastas seguindo a sistemática do referencial teórico e suas correlações temáticas.

2ª. Etapa DESCRIÇÃO ANALÍTICA	CATEGORIZAÇÃO TEMÁTICA	Criação das categorias: - Criou-se as categorias temáticas a partir das Categorias: Propriedade Intelectual (PI) e Ciência Aberta (CA) definidas previamente na estruturação temática da tese; - Relacionou-se uma leitura exaustiva de todo o arcabouço teórico de PI, com o intuito de identificar as temáticas de estudo; - Adotou-se para nomenclatura das categorias, as unidades de análise ou registro baseada em palavras-chaves representativas que embasam o estudo. - Gerou-se os <i>nodes</i> (categorias temáticas) foram criados a partir: referencial teórico, observação empírica, problema, hipóteses (pressupostos) e objetivos. - Realizou-se para cada <i>node</i> criado uma descrição (<i>Get Info</i>) com o intuito de orientar a análise de conteúdo. - Criou-se a partir das temáticas pré-selecionadas as categorias de análise (<i>nodes</i>).
3ª. Etapa: DESCRIÇÃO ANALÍTICA	CODIFICAÇÃO (Interpretação Inferencial em <i>Codes</i>)	Recorte e codificação: - Escolheu-se as unidades de registro e contexto; - Criou-se os <i>codes</i> (unidades de registro) a partir dos conteúdos previamente coletados; - Utilizou-se os critérios: pertinência, relevância e internacionalidade. - Realizou-se a classificação e agregação pela escolha das categorias temáticas.
4ª. Etapa: DESCRIÇÃO ANALÍTICA	INFERÊNCIAS	Enumeração dos <i>codes</i>: <i>Query</i>: Foi utilizado o recurso <i>word frequency</i> para descobrir as palavras que mais se repetem no <i>corpus</i> de análise e nos <i>codes</i> separadamente. Realizou-se as inferências nas 20 principais palavras que emergiram do recurso <i>word frequency</i>. O recurso <i>text search</i> foi utilizado para identificar as palavras-chave.

Fonte: Elaborado pela autora (2016).

As conexões temáticas resultaram em categorias e subcategorias dispostas no Quadro 2 e que direcionaram para a constituição dos padrões do modelo conceitual.

Quadro 2 – Categorias temáticas da análise de conteúdo

CATEGORIAS DA TESE	CATEGORIAS DE ANÁLISE	SUBCATEGORIAS	DIMENSÕES	PADRÕES
Ciência Aberta Economia Propriedade Intelectual Tecnologia	Estruturação e concepções da Ciência Aberta e <i>e-Science</i>	Ciência Aberta e <i>e-Science</i>	Epistemológica e Teórica	Epistemológico-Teórico
	Entendimento dos dados científicos sob a perspectiva do <i>commons</i>	Instrumentos Internacionais Regulatórios	Política e Legal	Político-normativo
		Bem Intelectual Público e Comum		
		Domínio Público		
		Acesso Aberto		
	Diretrizes político-normativas orientadas aos dados científicos	Princípios Normativo-legais para Dados Científicos:	Política e Legal	Político-normativo Ético-Legal
		Princípio Norteador 1: Proteção da		

		Propriedade Intelectual		
		Princípio Norteador 2: Conformidade Legal		
		Princípio Norteador 3: Responsabilidade Formal		
		Diretrizes Político-legais Brasileiras		
	Sistematização e gerenciamento dos dados científicos	Modelo Ciclo de Vida dos dados	Morfológica e Técnica	Morfológico-Técnico
	Atribuição e Citação como Recompensa Autoral	Autoralidade na <i>e-Science</i>	Ética e Legal	Ético-Legal-Cultural
		Autoria Colaborativa		
		Recompensa Autoral: atribuição, citação e responsabilização.		

Fonte: Elaborado pela autora (2016).

A ideia precípua foi estabelecer padrões multidimensionais a partir da análise com sustentabilidade para o modelo conceitual. A *e-Science*, por si só remete a uma essência fluida, assim inicialmente pensar em padrões traz uma noção de rigidez, o que requer uma ampliação sobre o termo. O termo **padrão**, quando usado em um modelo, tem a finalidade de estabelecer diretrizes e regras para aplicação prática. Ele visa além das regras, diretrizes e procedimentos mínimos, direcionar a execução de elementos necessários para garantir qualidade e conformidade. Assim, os padrões foram constituídos em uma matriz a partir das dimensões da pesquisa conforme a Figura 3.

Figura 3 – Matriz dos padrões multidimensionais



Fonte: Elaborado pela autora (2016).

Os padrões foram pensados para constituir elementos diretivos para a promoção da autoralidade em ambientes científicos que utilizam dados como produtos de pesquisa.

2.4 CONSTITUIÇÃO DO MODELO CONCEITUAL

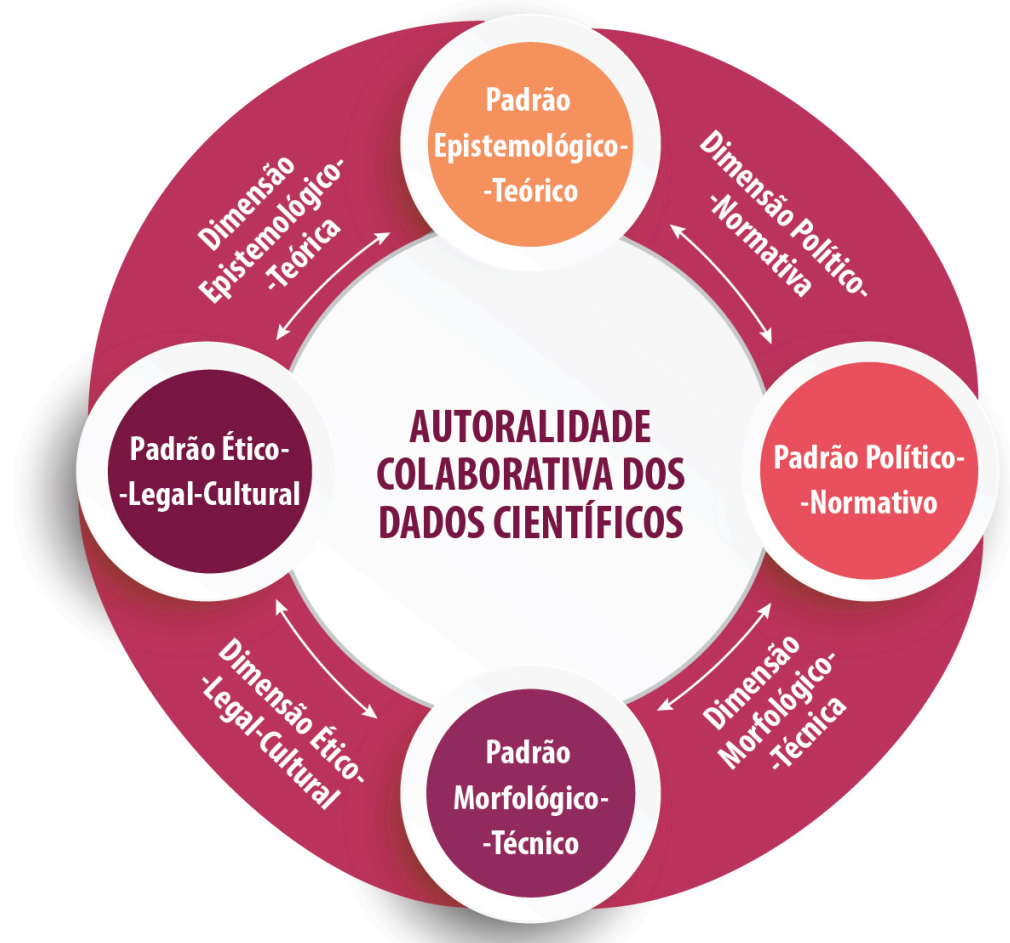
O modelo conceitual foi pensado para dar uma forma e representação morfológica do objeto de estudo com acepções contemporâneas do novo fazer científico. Além de permitir uma interface entre o referencial teórico, as inferências emergidas a partir dos conteúdos categorizados, percepções e reflexões da autora para uma aplicação real. As reflexões epistemológicas e teóricas da autora foram sendo constituídas desde o início da pesquisa de doutorado, que passou por um período de experiência junto ao Projeto DataONE na *University of Tennessee*, Estados Unidos e se consolidou por meio da elaboração do referencial teórico, análise e modelo conceitual durante a elaboração da tese.

A partir das apreensões teóricas e práticas vigentes e emergentes, a concepção para o modelo conceitual partiu de duas premissas teóricas: a representação do que está latente em conteúdos intrínsecos contidos nas publicações científicas, legislações, instrumentos normativos, diretrizes, padrões, indicadores, entre outros; e os aspectos extrínsecos relacionados com modelos, tecnologias e práticas institucionais que dão sustentabilidade ao processo de gerenciamento de dados científicos. As inferências emergidas ao se conectarem com os capítulos, fizeram brotar um corpo de ideias que foram materializadas e integradas conceitualmente em padrões.

Para a construção de tal modelo foram considerados aspectos conceituais, lógicos e práticos. Sob o aspecto conceitual a palavra conceito significa a abstração do pensamento, ideia; é uma noção abstrata ou ideia geral designada a um objeto único ou uma classe de objetos. Na Filosofia, o termo conceito é compreendido como elementos em construção (JAPIASSÚ; MARCONDES, 2006). O cerne do modelo será constituído pelos conceitos apreendidos durante o processo de construção do referencial e da análise.

Adicionou-se as premissas, o *constructo* elaborado no referencial teórico, análise de conteúdo e a convergência multidimensional que resultaram em quatro padrões demonstrados na Figura 4. Do ponto de vista lógico, ele é caracterizado por sua extensão e compreensão. Desvendar o enigma das questões de autorialidade na perspectiva da ciência aberta e das práticas científicas da *e-Science* e sua transversalidade como os direitos de propriedade intelectual vigentes foi o caminho norteador para corroborar com o argumento da tese. Sob o aspecto prático, o modelo direciona a noção geral de que cada elemento constitutivo seja compreendido e aplicado pelo pesquisador-autor, grupos de pesquisa e instituições nas práticas da *e-Science*, com vistas a atender os padrões mínimos para que a autorialidade seja garantida a todos os envolvidos.

Figura 4 – Padrões multidimensionais do modelo conceitual



Fonte: Elaborado pela autora (2016).

O intuito primordial do modelo é auxiliar pesquisadores, autores, colaboradores, instituições acadêmicas, científicas e governamentais e agências de fomento que ao adotar ou financiar projetos de compartilhamento e gerenciamento de dados científicos as garantias mínimas de autorialidade de todos os envolvidos sigam diretrizes norteadoras. Essas diretrizes foram constituídas e agrupadas em padrões que compõem o modelo conceitual.

3 CIÊNCIA ABERTA E O QUARTO PARADIGMA CIENTÍFICO DA *E-SCIENCE*

A comunicação científica está evoluindo para a promoção de uma ciência aberta e inovadora apoiada no quarto paradigma científico e nas práticas científicas emergentes. O quarto paradigma científico preconiza o uso intensivo de dados gerados nas investigações científicas como insumos primários. O percurso metodológico deste capítulo foi construído pelas dimensões epistemológica e teórica. A distinção conceitual entre a ciência aberta e a *e-Science* se fez necessária neste estudo com vistas a evidenciar a visão multidimensional que envolve esse contexto emergente. Assim, a abordagem conceitual segue na direção que a ciência aberta é o ambiente inovador, colaborativo, compartilhado, aberto e digital da ciência moderna e a *e-Science* é a representação de uma nova configuração para as práticas científicas.

3.1 CIÊNCIA ABERTA E *E-SCIENCE*

A ciência aberta é uma abordagem para investigações no contexto da comunicação científica e da ciência-cidadã. É um termo derivado da expressão *Open Science* que surgiu no século XVI após a revolução pós-renascentista, com a proprietização do bens intelectuais, a partir do surgimento da imprensa e ressurgiu nas últimas décadas como um grande esquema emergido do movimento de acesso aberto (FROSIO, 2014). Apresenta como premissa um modelo que busca o equilíbrio entre a propriedade privada e o acesso aberto do capital intelectual.

Assim, a ciência aberta passa a constituir um termo guarda-chuva, amplo e multidimensional, que inclui acesso livre a “publicações científicas, dados científicos abertos, ferramentas científicas abertas, *hardware* científico aberto, cadernos científicos abertos, *wikipesquisa*, ciência cidadã e educação aberta” (ALBAGLI; 2015, p. 15). É uma ciência voltada para grandes conjuntos de dados que são gerados em sofisticadas bases de dados e software que agregam uma ampla gama de simulações, metodologias e modelos efetivados em diversas áreas do conhecimento. Por conseguinte, esta ciência conduz a novas pesquisas, investigações e hipóteses a partir do uso, reuso e reprodutibilidade dos dados científicos, sob a égide do quarto paradigma, o qual fomenta o surgimento de um fazer científico baseado no compartilhamento e na colaboração.

O movimento da ciência aberta avança com o surgimento de “mecanismos regulatórios que visam reverter assimetrias resultantes da apropriação e distribuição privada dos

conhecimentos científicos, sobretudo em áreas sensíveis e com forte apelo social, como saúde, agricultura, alimentação e meio ambiente” (ALBAGLI, 2015, p. 18).

Esse novo fazer científico pressupõe que regras e práticas sejam ensejadas pela comunidade científica, agências de fomento, governo e instituições de pesquisa e ensino. Esse fenômeno demanda que se pense amplamente nas dimensões epistemológica, política, ético-legal-cultural, morfológica e tecnológica que sustentem a efetivação da ciência aberta.

Pressupõe uma prática científica de tal forma que outros possam colaborar e contribuir sempre que os dados de pesquisa estejam disponíveis. Engloba, ainda, ferramentas e recursos tecnológicos digitais e colaborativos, mecanismos alternativos para a propriedade intelectual, como o uso de licenças *copyleft*, que promovem a reutilização dos resultados de pesquisas (DELFANTI; PITRELLI, 2015). As práticas científicas e a infraestrutura para apoiar o fazer científico está no escopo da *e-Science*.

O termo conhecido como *E-Science* ou *eScience* é definido como a ciência de computação intensiva, realizada em ambientes de rede altamente distribuídos que utilizam volumosos conjuntos de dados, os quais permitem a coleta, processamento, preservação, análise e armazenamento de grande quantidade de dados em diferentes disciplinas. Representa, pois, as práticas do quarto paradigma científico.

Sua origem está relacionada à Segunda Guerra Mundial quando houve uma explosão das publicações científicas. O crescimento exponencial do conhecimento científico e das disciplinas especializadas, criou o “fenômeno que deu origem à chamada *Big Science*. No Século XXI, a pesquisa colaborativa que produz grande volume de dados deu origem aos termos: *Big Data*, *cyberscience*, *cyber infrastructure* e *e-Science*” (COSTA; CUNHA, 2015, p. 128). As inovações, descobertas e teorias sempre fizeram parte do escopo da ciência.

A inovação científica foi ampliada pela colaboração e compartilhamento. Na *e-Science* “as expectativas em torno de um mundo rico em dados são imensas e incluem desde descobertas de novas drogas, passando por um melhor entendimento sobre as mudanças climáticas, até metodologias mais apuradas para examinar a história e a cultura” (BORGMAN, 2010, p. 2, tradução nossa). Já é possível ver mudanças significativas em ambientes que envolvem as disciplinas das ciências naturais e da terra, biotecnologia, medicina genética, biotecnologia e o avanço cada vez mais especializado das engenharias, tecnologias, a exemplo da nanotecnologia, engenharia genética e inteligência artificial.

Além de que, “a relevância dos dados no contexto das “big Sciences”, como Astronomia, Física e Biologia, conduziu não somente ao surgimento de novos modelos de ciência - coletivamente chamados de “quarto paradigma científico” ou “eScience” [...]”

(BORGMAN, 2010, p. 2, tradução nossa), mas a novos modelos e práticas a partir da utilização e do valor que os dados de pesquisa passaram a ter em áreas consideradas de alta complexidade e especializadas, dessas chamadas grandes ciências.

Em contraponto aos estudiosos que conduzem as investigações na perspectiva da *e-Science*, Frické (2014) afirma que as pesquisas baseadas em dados científicos são superficiais e não oferecem explicações suficientes para resolução de problemas na ciência, “o ponto central é que os dados não falam. O que é necessário é uma enorme quantidade de conhecimento de fundo, ou suposições, ou pesquisa anterior de um tipo ou outro” (FRICKÉ, 2014, p. 654). Com essa afirmação o autor descreve em seu estudo que algumas áreas, como a ciência da Biodiversidade ou Ciência Genômica possam realmente fazer descobertas científicas e evitar a replicação de pesquisas.

Entretanto, os argumentos do autor contrapõem-se com outros estudiosos dessa temática. Para Hey e Trefethen (2003), Gray (2007), Marcum e George (2010) e Costa e Cunha (2015) a *e-Science* refere-se à coleção de instrumentos, metodologias, modelos e tecnologias necessárias para apoiar a pesquisa científica colaborativa do século XXI. É um movimento que ascende e tem aprimorado a pesquisa tradicional com outras formas de fazer científico.

Delfanti e Pitrelli (2015, p. 59, tradução nossa) afirmam que Michael Nielsen¹: “estamos no meio de uma transição para uma nova era científica, uma era comparável à da revolução científica do século XVII e à da transição para a Idade Moderna”. Pois, as práticas científicas da *e-Science* trazem na sua essência a mudança no *modus operandi* de fazer, compartilhar, divulgar, acessar, usar, reusar e reproduzir dados de pesquisa com possibilidades de compartilhamento, interatividade, convergência e colaboração entre pesquisadores. A natureza colaborativa e multidisciplinar corrobora como o novo fazer científico através do gerenciamento de dados de pesquisa. A prática científica pressupõe uma infraestrutura que envolva aspectos políticos, institucionais, gerenciais, metodológicos e tecnológicos. No cenário internacional, os esforços para uma ciência orientada aos dados de pesquisa são apoiados por políticas, diretrizes, modelos, tecnologias e boas práticas.

Sob o ponto de vista tecnológico a *e-Science* é apoiada por uma infraestrutura tecnológica denominada de ciberinfraestrutura. A Tecnologia da Informação (TI) é a interseção entre a *e-Science* e os cientistas. A ciberinfraestrutura que suporta o ambiente da pesquisa científica aberta é um dos componentes de sustentabilidade da *e-Science*. Utiliza alta tecnologia, computação intensiva, ferramentas e recursos para lidar com dados (estruturados e

¹ Especialista em computação quântica e defensor da ciência aberta.

desestruturados). Deve apoiar a captura, armazenagem, gerenciamento, preservação, visualização e análise de dados de pesquisa, “tornando os sistemas de computação vitais para o moderno ambiente de pesquisa” (COSTA; CUNHA, 2015, p. 130).

O termo *cyberinfrastructure* está enraizado na iniciativa das agências de financiamento dos Estados Unidos, como orientação aos cientistas que buscavam obtenção de financiamento de pesquisa na *National Science Foundation* (NSF). Nesse sentido o relatório intitulado *Revolutionizing Science and Engineering Through Cyberinfrastructure: report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure* publicado em 2003, define que: “*cyberinfrastructure* refere-se à infraestrutura baseada em computador distribuído, tecnologia de informação e comunicação. Se a infraestrutura é necessária para uma economia industrial, então *Cyberinfrastructure* é necessária para a economia do conhecimento” (ATKINS, 2003, p. 5, tradução nossa).

Assim o termo Ciberinfraestrutura foi adotado pelas agências de financiamento para descrever ambientes de pesquisa que suportam aquisição avançada de dados, armazenamento de dados, gerenciamento de dados, integração de dados, mineração de dados, visualização de dados e outros serviços de computação e processamento de informação através da *internet* para além do âmbito institucional (ATKINS, 2003). É a plataforma que sustenta as soluções tecnológicas nos estágios do ciclo de vida de gerenciamento dos dados.

Sob o aspecto científico, a *e-Science* é uma síntese da tecnologia da informação e da ciência com métodos científicos cuja evolução se dar por uma ciberinfraestrutura que sustenta modelos, ferramentas, padrões, sistemas e repositórios abertos. Pauta-se em uma ciência codificada e objetivada e enfrenta um dilúvio de dados (GRAY, 2007). A mudança nas práticas da comunicação científica estão provocando um processo híbrido das publicações científicas. Assim a *e-Science* conduz as práticas científicas para novos modelos científicos.

Borgman, em 2003, já tinha um despertar para essa nova era digital, a qual chamou de *Global Information Infrastructure (GII)*, comparando o momento ao feito de *Gutenberg*. Afirmar que “Gutenberg aperfeiçoou a impressão com tipo móvel. Foi a abordagem holística que o possibilitou aperfeiçoar cada fase do processo [...], contribuindo, portanto, significativamente para a tecnologia ou comunicação humana” (BORGMAN, 2007, p. 30, tradução nossa). A infraestrutura global de informação que se refere a autora aponta para a reflexão na seara da pesquisa tradicional para uma nova concepção de pesquisa aberta.

A transição de práticas científicas do modelo tradicional para a *e-Science* está apoiada em três aspectos: o primeiro é a adoção das novas práticas a partir de um modelo de ciclo de vida dos dados que garanta o efetivo gerenciamento dos dados primários; o segundo é

proporcionar uma ciberinfraestrutura sustentável para uso, reuso e reprodutibilidade dos dados de pesquisa; e o terceiro são aspectos relacionados ao compartilhamento e colaboração científica entre diferentes pesquisadores, profissionais, instituições e até países.

Neste sentido Tenopir e King (2004, p. 11, tradução nossa) afirmam que a “inovação nunca acontece em um vácuo; inovação requer comunicação. Os autores sugerem realizar todos os estágios de pesquisa, desenvolvimento, desenho, e produção tão eficientes e efetivos quanto possíveis, para que a informação possa ser clara e entendida”. Este processo de comunicação requer inovação, não só no entendimento das descobertas científicas, mas no seu processo operacional de mudança do *modus operandi* científico.

Compartilham desse entendimento Borgman (2010) e Medeiros e Caregnato (2012), quando afirmam que a *e-Science* visa a colaboração entre cientistas a partir do compartilhamento e gerenciamento de dados científicos primários. O acesso aberto à informação e aos dados é um processo de novas formas de se fazer e divulgar a ciência. A utilização das tecnologias sustenta as mudanças e apoia o alcance em larga escala. Esta infraestrutura global proposta precisa do estabelecimento de alianças institucionais e do desenvolvimento de políticas governamentais para que transformações possam ser realizadas.

Sales, Sayão e Souza (2013) afirmam que a comunidade científica anseia por novos modelos de publicações que possam expressar a complexidade e a dinâmica da pesquisa científica contemporânea. Os modelos científicos da *e-Science* apresentam possibilidades que explicitam mais claramente a pesquisa desenvolvida, seus métodos, seu material, os dados e conhecimentos gerados, além de revelar aspectos intrínsecos da pesquisa científica, a exemplo da intenção do pesquisador e a avaliação dos resultados de forma mais eficaz e interativa.

Sob o ponto de vista dos dados científicos, a *e-Science* apoia o compartilhamento e colaboração para uso e reuso. O reuso possibilita dar continuidade a uma determinada pesquisa, gerar novas hipóteses, inferências, direcionamentos e alcançar novas descobertas a partir dos dados primários de pesquisa de outros pesquisadores. Esse processo de reuso é possível quando ocorre a efetiva colaboração científica.

De acordo com Borgman (2010, p. 10, tradução nossa), “o novo paradigma reflete a capacidade de inferir significado aos dados de pesquisa e tomar medidas para um novo sentido [...]. Essa divisão é um aspecto do Quarto Paradigma”. Dar sentido e gerar novos significados com o uso e reuso de dados científicos é uma das facetas do quarto paradigma. É possível entender que a ideia de um novo paradigma passa a ser revolucionário no cenário da comunicação científica, quando ocorre mudanças significativas e altera a produção, gestão e acesso das publicações geradas nesse contexto (OLIVEIRA et al., 2015).

A seção trouxe à tona a compreensão sobre a ciência aberta e a *e-Science*. A ciência aberta pensada como um guarda-chuva com um escopo mais abrangente sobre as políticas, diretrizes e normatização em um contexto abrangente. Assim, a *e-Science* é o arcabouço das práticas metodológicas, técnicas e tecnológicas que dão sustentabilidade as novas dinâmicas científicas baseada no quarto paradigma científica.

3.2 (R)EVOLUÇÃO DOS PARADIGMAS DA CIÊNCIA

A evolução da ciência é pautada por revoluções científicas que representam suas práticas e provocam mudanças no *status quo* e no *modus operandi* do pensar e fazer científico. Tradicionalmente, “se a ciência é uma constelação de fatos, teorias e métodos coletados em textos atuais; então os cientistas são as pessoas que, de forma bem-sucedida ou não, têm lutado para contribuir com um ou outro elemento para aquela constelação específica” (KUHN, 1970, p.1, tradução nossa). Assim, a ciência é conduzida por um conjunto de fatos, fenômenos, métodos científicos e por processos de comunicação científica que constituem o arcabouço que Kuhn denominou de constelação.

Na modernidade, a ciência e a tecnologia interagem com a comunicação científica e “têm sido responsáveis pelas grandes transformações pelas quais a humanidade passa, alterando ainda de forma desigual, o cotidiano das pessoas” (GASQUE, 2008, p. 30). Tal interação é primordial para o avanço científico e social. As transformações estão cada vez mais dinâmicas e velozes devido aos meios tecnológicos que dão suporte à ciência.

Numa concepção epistemológica

a ciência tem o potencial, mais do que qualquer outro empreendimento humano, para revelar verdades epistemológicas sobre as quais o universo e tudo que nele dependem; mas é somente através do processo de ceticismo, crítica e criatividade que nos aproximamos de um entendimento qualificado da natureza (CARVALHO; MORRIS, 2014, p. xvi, tradução nossa).

Tal potencial é impulsionado por esforços de cientistas que trabalham incessantemente para que práticas científicas evoluam e novas teorias, dinâmicas e modos de fazer e comunicar a ciência possam surgir, despontando para a mudança de paradigmas. Muito embora o significado do termo paradigma remonte à época de Platão, sua definição foi inicialmente adotada nos estudos da Gramática, por Merriam-Webster (1900), na aplicação da Linguística, por Ferdinand de Saussure (1857-1913) e na ciência como um padrão científico, por Thomas Kuhn (1922-1996).

Mas foi o filósofo Thomas Kuhn, o primeiro a adotar o termo paradigma para o contexto da ciência, sugerindo que a investigação científica não progride em direção a verdades, mas está sujeita a dogmas e agarrada a velhas teorias. Kuhn defende que a noção de paradigma é reforçado pelos dogmas, práticas arraigadas e teorias antigas que se mudam ou alteram de acordo com a evolução histórica, científica, econômica, social e cultural na sociedade. Nas práticas científicas alterações e mudanças de paradigmas são representadas por melhorias, ressignificados de teorias, modelos, exemplos e no *modus operandi* existentes.

O pressuposto de Kuhn se deu na tentativa de compreender as divergências e manifestos ocorridos nas práticas científicas. A sua inquietação pairava nas soluções e modelagens conduzidas nas diversas ciências. Afirmou que, “particularmente, fiquei impressionado com o número e extensão das divergências manifestas entre os cientistas sociais sobre a natureza dos problemas científicos legítimos e métodos” (KUHN, 1970, p. *viii*, tradução nossa).

Nesse sentido, o surgimento de um paradigma está na perda, mudança e transição de práticas e modelos tradicionais para uma nova forma de pensar e fazer algo. Com base no estudo de Kuhn, “o fenômeno da perda, no seu ponto de vista, elimina a imagem cumulativa tradicional do progresso. A pesquisa é revolucionada quando o paradigma existente não soluciona as anomalias importantes” (BIRD, 2013, *on-line*, tradução nossa). A transição está na mudança das práticas antigas e na busca por soluções existentes em novos modelos. Nessa revolução científica o modo como a ciência interage com o mundo e a opinião científica se torna um fator preponderante na mudança de paradigma.

Shuttleworth (2008, *on-line*, tradução nossa) afirma que a mudança de paradigma sugerida por Thomas Kuhn, “não é apenas uma pequena mudança na ciência ou a modificação de uma teoria, mas uma revolução científica e muda completamente o modo como a ciência olha para o mundo. E muitas vezes dita como o público olha para o mundo”.

Um paradigma pode absorver alguns erros, mas eles eventualmente tornam-se intransponíveis e resultam em uma transição para novas formas de pensar e fazer a ciência. Nesse sentido, os paradigmas se distinguem entre si devido aos contextos culturais, sociais que ocorrem na sociedade influenciando também na natureza científica.

Atualmente a ciência está experimentando um novo paradigma. O cientista computacional Jim Gray² foi o precursor do quarto paradigma científico (*Fourth Paradigm*). A afirmação de Gray (2007) conduz para o entendimento de que o quarto paradigma reflete o

² Vencedor do prêmio *Turing*.

atual *status* da ciência aberta e inovadora baseada no uso intensivo de dados de pesquisa para a exploração científica. No entendimento de Jim Gray (2007) o paradigma atual reforça a exploração intensiva de dados unificando os paradigmas anteriores (experiência, teoria e simulação computacional) em uma nova prática científica.

Com base no exposto, os paradigmas da ciência representam a evolução histórica e científica dos paradigmas científicos conforme demonstrados na Figura 5.

Figura 5 – Paradigmas da ciência



Fonte: Baseado nos estudos de Gray (2007) e Hey (2010, *on-line*).

O primeiro paradigma da ciência está baseado na experimentação. Ocorrido há mais de 1000 (mil) anos com a ciência experimental, teve origem na Grécia antiga e China, onde os filósofos e estudiosos tentavam explicar as suas observações através de leis naturais, ao invés de causas sobrenaturais.

O segundo paradigma baseou-se na ciência teórica com o surgimento de modelos e generalizações. Ocorrido nos últimos cem anos, por volta do século XVII, cientistas como Isaac Newton tentaram fazer previsões para os novos fenômenos e verificação de novas hipóteses através da realização de experimentos, originando novas teorias.

O terceiro paradigma, refletido nas últimas décadas, é representado na ciência computacional com base na simulação de fenômenos complexos. Com o advento dos computadores de alto desempenho na segunda metade do século XX, o ganhador do Prêmio Nobel, Ken Wilson, identificou a computação e a simulação como um terceiro paradigma para a exploração científica. Simulações computacionais detalhadas, capazes de solucionar equações em grande escala, permitiram aos cientistas explorar campos de investigação que eram inacessíveis para experimentação e teoria, tais como: modelagens climáticas ou a formação de galáxias, bem como resolver numericamente os sistemas de equações em grande escala e de forma detalhada (HEY, 2010).

A informação digital vem fluindo de todos os tipos de sensores, instrumentos e simulações, sobrecarregando a nossa capacidade de organizar, analisar e armazená-la [...], a um ponto, no entanto, onde até mesmo os especialistas estão se afogando em dados” (HEY, 2010, *on-line*, tradução nossa). A revolução multicentral chegou apenas à medida que se enfrenta um aumento exponencial de dados. Esse aumento não é um desafio a ser enfrentado apenas com ajustes e atualizações, é um repensar de toda abordagem para a ciência de dados intensivos.

Quanto ao quarto paradigma afloram a geração intensiva de uma grande quantidade de dados que emergem do processo de investigação científica. A mudança está na utilização dos dados primários adjacentes de diferentes disciplinas e utilizados como insumos na investigação científica. Afirmam Tenopir e outros (2011, p. 1, tradução nossa) que “seguindo os paradigmas de investigação anteriores (experimentais, teóricos e computacionais), esta nova era tem sido chamada de ‘o quarto paradigma: descoberta científica do uso intensivo de dados’”. É a junção dos paradigmas anteriores para apoiar o processamento dos dados intensivos na era digital. Complementam ainda que

toda a literatura científica está *on-line*, todos os dados da ciência estão *on-line*, e eles interoperam uns com os outros. Os dados digitais não são apenas as saídas de investigação, mas fornecem subsídios para novas hipóteses, permitindo novos conhecimentos científicos e promovendo a inovação (TENOPIR et al., 2011, p. 1, tradução nossa).

Então, além do uso dos dados como insumos primários na pesquisa e a incorporação das tecnologias e de práticas consolidadas nos paradigmas anteriores estão sendo aprimoradas e adicionadas a nova era digital. Portanto, “o quarto paradigma não está tentando substituir cientistas ou as outras três metodologias, mas exige um conjunto diferente de habilidades” (HEY, 2010, *on-line*, tradução nossa). Sem a capacidade de aproveitar as ferramentas

sofisticadas de computadores que manipulam dados, mesmo o perito mais altamente treinado nunca conseguiria descobrir os *insights* que agora estão começando a emergirem. As ferramentas utilizadas para analisar, visualizar, minerar e manipular os dados científicos sistematicamente buscam soluções para desafios globais (HEY, 2010).

A interseção entre os paradigmas reforça que a ciência está sempre em um processo evolutivo e dinâmico. Aprimora e modifica o *status quo* e o *modus operandi* do fazer científico na cadeia científica. De certa forma, o quarto paradigma preconizado por Gray proporciona uma estrutura integrativa que permite os três primeiros a interagirem e reforçarem um ao outro, “a exemplo do ciclo científico tradicional no qual a teoria ofereceu predições que poderiam ser experimentalmente testadas, e esses experimentos identificaram fenômenos que exigiam explanação teórica” (DLYNCH, 2009, p. 177, tradução nossa). Envolve o crescimento intensivo dos dados, a integração entre disciplinas complexas, a exemplo da Astrologia, Genética e Biologia com a Informática e a Computação, a derivação de outras concepções interdisciplinares e a necessidade de uma infraestrutura técnica, metodológica e tecnológica global para abarcar este contexto.

Para Bo-Christer Björk (2007, p. 20, tradução nossa), a mudança está na “performance da pesquisa, financiamento, comunicação dos resultados e aplicação do conhecimento”. O conhecimento é comunicado e os resultados são divulgados. Além de uma nova dinâmica e infraestrutura, este contexto abrange diferentes escopos e políticas.

A ciência aberta mobiliza múltiplos níveis e escopos de abertura, remetendo tanto a um **sentido pragmático**, de permitir maior dinamismo às atividades de ciência, tecnologia e inovação, quanto a um **sentido democrático**, de possibilitar maior diversidade de perspectivas no âmbito da própria ciência e, ainda, maior participação e intervenção da sociedade (ALBAGLI, 2015, p. 17, grifo nosso).

Envolve instituições governamentais, educacionais, científicas e a sociedade em geral. A tecnologia cada vez mais presente nos processos científicos,

A atividade de pesquisa científica do século XXI produz uma quantidade extraordinária de dados, principalmente em formatos digitais [...]. Isto acontece essencialmente porque a tecnologia digital se torna cada vez mais um elemento onipresente nos processos da construção do conhecimento científico, seja por aumentar a capacidade dos instrumentos científicos, seja por reconstruir realidades por meio de simulação, ou ainda inaugurando formas inéditas de colaboração e compartilhamento de dados e informações (SAYÃO; SALES, 2013, p. 2).

Este cenário, denominado de *e-Science*, traz à tona discussões sobre a mudança na geração e disponibilidade de publicações científicas e eletrônicas, conectadas pelas

tecnologias digitais de informação e comunicação através da *internet*, abrindo possibilidades de compartilhamento e colaboração coletiva a partir de dados de pesquisa.

O movimento pela ciência aberta, em seu formato atual, reflete, na verdade, novos modos de pensar e de exercer a cientificidade do quarto paradigma, com repercussões diretas sobre os compromissos, normas e arcabouços institucionais que interferem diretamente na prática científica e nas suas relações com a sociedade (ALBAGLI, 2015). Acompanha o entendimento de Kuhn com a geração de novos modelos, técnicas, tecnologias que alterem o padrão e práticas dos paradigmas anteriores e proporciona o surgimento de novas hipóteses e inovação científica.

A literatura científica internacional aponta para as vantagens da pesquisa colaborativa aberta e muitas iniciativas e experiências já estão em andamento. Entretanto, estudos também apontam para ineficiências que surgem do movimento *open science*. O estudo de Flické (2014) traz reflexões que contrapõem ao movimento e afirma que “a ciência baseada em dados, o ‘quarto paradigma’, é uma quimera. A ciência precisa de problemas, pensamentos, teorias e experiências projetadas. Se alguma coisa, a ciência precisa de mais teorias e menos dados” (FLICKÉ, 2014, p. 660). Afirma que as pesquisas precisam de aprofundamento metodológico e epistemológico como existe na ciência tradicional. O autor aponta para a fragilidade de replicação de pesquisas observacionais e experimentais que, por vezes, alteram os resultados fidedignos, contudo, acredita que a coleta da grande quantidade de dados, a exemplo do fenômeno *big data*, traz vantagens como a redução de custos e a avaliação contínua.

Entretanto, pensar na ciência aberta na sua essência ainda é um argumento positivo que se contrapõem com o argumento de Flické. O quarto paradigma científico apresenta um aprimoramento de métodos, procedimentos e tecnologias com possibilidades de avanços e acessibilidade à pesquisa.

Pondera-se, que não há um rompimento de paradigma, mas uma confluência e aprimoramento de teorias, métodos, modelos, práticas, habilidades e funcionalidades diferentes e revalidadas. A ciência é dinâmica e resultante de um acúmulo histórico, social, cultural, teórico e científico.

A abordagem deste capítulo se deu para a compreensão da ciência aberta, *e-Science* e paradigmas da ciência no contexto científico contemporâneo. A dimensão epistemológica trouxe o entendimento da relação dos paradigmas da ciência com o quarto paradigma científico e a contextualização teórica norteou a distinção entre a ciência aberta e a *e-Science*.

3.3 PESQUISA ABERTA BASEADA NO QUARTO PARADIGMA CIENTÍFICO

O capítulo busca compreender as novas dinâmicas científicas a partir dos dados como produtos primários do processo de investigação. Nessas dinâmicas diferentes conceitos, nomenclaturas e tipologias fazem parte do novo *constructo* da ciência aberta e das práticas científicas conduzidas pelo quarto paradigma científico da *e-Science*.

3.3.1 *Constructo* Epistemológico da Pesquisa Aberta

Tradicionalmente, a pesquisa científica é o meio pelo qual descobertas, teorias e conhecimentos são gerados no bojo da ciência. O registro e a propagação da pesquisa perpassa pela publicação em diversos suportes. Por séculos esses suportes foram impressos, e mais recentemente, passaram a ser diversificados em meios eletrônicos e digitais. Tal mudança se deu no suporte e nos meios de gerenciamento, armazenamento e acesso à publicação científica.

A publicação científica tradicional ainda é considerada o principal meio de publicização dos feitos da ciência, contudo, as mudanças contemporâneas nos processos da comunicação científica apresentam novas abordagens e dinâmicas que trazem como *constructo* científico os dados primários. Pois, na pesquisa tradicional vigente, os dados primários não são processados como produtos finais. Historicamente, os dados gerados durante a pesquisa científica na maioria das vezes são desprezados e inacessíveis pelos canais formais de comunicação. Geralmente no processo tradicional os resultados finais são divulgados por meio de publicações científicas, sem o aproveitamento do valor e potencial que os dados possuem dentro do processo de investigação científica. Comumente, os resultados dessa investigação são divulgados nos canais formais, antecipando parcialmente ou encerrando o ciclo tradicional da pesquisa científica (BORGMAN, 2015).

Diante disto, entende-se que atualmente ocorre um processo híbrido entre as práticas tradicionais vigentes da comunicação científica e as emergentes na configuração da *e-Science*, notadamente no cenário internacional. Mesmo utilizando tecnologias da informação e comunicação para a gestão, o acesso e uso das publicações científicas em meios eletrônicos, ainda prevalece o modo de fazer científico tradicional. Enquanto que, a dinâmica emergente é conduzida por dados primários coletados no início do processo de investigação e gerenciados preliminarmente como produtos de pesquisa.

A acepção geral do termo ‘dados’ pode se constituir de fatos representados por uma unidade de análise, agrupada e constituída de significado. Tais significados são construídos a partir de fatos, fenômenos, símbolos, unidades, caracteres ou *bits*, desde que possuam uma representação legível por linguagem computacional.

As práticas emergentes da *e-Science* os dados são insumos primários de investigações e podem apresentar resultados preliminares de pesquisa (BORGMAN, 2015). Parte assim do pressuposto que a estrutura é dinâmica e abrange diferentes dimensões (epistemológica, política, teórica, morfológica, técnica e teórica). Os dados são a infraestrutura da ciência moderna, a qual está se tornando a ciência de dados intensivos e colaborativos (TENOPIR et al., 2011). Desta forma, a reflexão preliminar conduz ao *constructo* epistemológico que dados científicos são insumos do processo de investigação, e por conseguinte, são considerados produtos primários de pesquisa.

Na perspectiva da *e-Science*, a noção de ‘dado’ também foi conceituada pela *Open Archival Information System* (OAIS) como “uma representação reinterpretável de informações de uma maneira formalizada adequada para comunicação, interpretação ou processamento” (OAI, 2015, *on-line*, tradução nossa). Essa definição está bem associada as condições em que os dados digitais devem ser acessíveis, interoperáveis e reinterpretáveis semanticamente. Para Frické (2014, p. 652) “dados é qualquer coisa gravável em um banco de dados relacional de forma semântica e pragmaticamente sólida. A semântica requer que as gravações sejam entendidas como declarações verdadeiras ou falsas”. A representação semântica garante o processamento, manipulação, reinterpretação e reutilização dos dados digitalmente.

Essa compreensão se alarga na pesquisa científica contemporânea. Dados científicos são produzidos, coletados, manipulados, reusados e reinterpretados a partir de qualquer formato, suporte e contexto. O contexto dos dados exige pragmatismo “naturalmente, nosso conhecimento de dados, ou o que constitui dados, é falível. Os dados são, portanto, conjectural. Os dados também estão carregados de teoria no seguinte sentido” (FRICKÉ, 2014, p. 652). Essa afirmação aponta para que os instrumentos de medição e métodos utilizados para a interpretação dos dados são passíveis de influências e equívocos advindos do contexto e do ambiente em que se encontram. Essa é uma afirmação que deve ser considerada quando práticas científicas envolvem dados científicos.

A *National Science Foundation* (NSF) adota a definição de ‘dados científicos’ para se referir a qualquer informação que possa ser armazenada em forma digital, “incluindo texto, números, imagens, vídeos, áudio, *software*, algoritmos, equações, animações, modelos,

símbolos e simulações geradas por vários meios observacionais, computacionais e experimentais” (NSF, 2005, p. 20, tradução nossa). Corroboram com essa afirmativa Borgman (2010) e Sayão e Sales (2013) quando afirmam que essa variedade e amplitude dos dados científicos expressos por inúmeras representações descrevem objetos, ideias, condições, situações, fatos, fenômenos e contextos.

Os dados científicos possuem diversos suportes e formatos. Essa diversidade depende da proveniência dos dados científicos de acordo com a área de conhecimento, domínio específico da disciplina e contextos em que foram criados e processados. Desta forma, os dados científicos podem variar consideravelmente entre pesquisadores e áreas do conhecimento. A constatação é que os dados são gerados para diferentes propósitos, por distintas comunidades acadêmicas e científicas e por meio de diversos processos (SAYÃO; SALES, 2013).

Corroboram nesse entendimento Kaase (2001), Borgman (2010), Sayão e Sales (2013) e Nielsen e Hjørland (2014) quando afirmam que o valor dos dados científicos é relativo, pois atendem diferentes perspectivas, propósitos e contextos. Destinam-se aos mais variados pesquisadores, colaboradores e instituições que se utilizam e reusam os dados científicos como o intuito de produzir novos sentidos e significados. Na ciência a relatividade também se destina ao valor, relevância e domínio específico da disciplina científica.

E ainda, é pensar nos dados como uma publicação científica primária. A “noção de ‘publicação’ nessa direção, inclui os dados que sustentam a publicação e os resultados, também conhecidos como dados subjacentes” (EUROPEAN COMMISSION, 2015, *on-line*, tradução nossa). Assim, pensar em ‘dados’ é assumir a condição que a forma bruta de uma ideia materializada já é um produto primário de investigação.

3.3.2 Proveniência e Tipologia dos Dados de Pesquisa

A classificação de dados de pesquisa está relacionada à proveniência (origem) de diferentes contextos de pesquisas. Os estudos da NSF (2005), Borgman (2010) e Sayão e Sales (2015) afirmam que a proveniência dos dados é oriunda de três contextos de pesquisas: observacionais, computacionais ou experimentais.

Os dados observacionais são provenientes de observações diretas de pesquisas, associadas a lugares e tempo específicos. São dados coletados de pesquisas com recursos naturais, como exemplo a erupção de determinado vulcão (numa data específica), temperatura do oceano, amostras de solo, medições meteorológicas, que estão associadas a coleta em

vários lugares e tempos, como é o caso de estudos longitudinais e transversais. Envolve ainda pesquisas de comportamento eleitoral que também podem estar associadas a locais, dias e horários específicos.

A coleta observacional utiliza instrumentos diretos ou indiretos, mecânicos, manuais, computacionais e de campo. Possui natureza instantânea, pois o fenômeno observacional é efêmero, e por vezes, pode não ser repetido. Portanto, “guardam uma importância crítica que os qualificam como registros históricos” (SAYÃO, SALES, 2015, p. 2), que não permitem uma coleta secundária posterior e, portanto, devem ser submetidos a curadoria preliminar que inicialmente será a curto prazo, mas que durante os estágios do ciclo de vida são preservados a longo prazo.

O ciclo de vida deve contemplar “o processo de coleta de dados combinados com atividades pré-planejadas e decisões no campo” (WALLIS et al., 2008, p. 117) quando os dados de pesquisas são oriundos de fenômenos observacionais. O planejamento prévio para a coleta de dados observacionais colabora com a tomada de decisão e estratégias que podem ser demandadas de alterações inesperadas durante o fenômeno.

Os dados experimentais oriundos de experimentos primários, tal como nos fenômenos observacionais, nem sempre podem ser reproduzidos com a mesma precisão, em virtude das condições experimentais únicas. Os experimentos são provenientes de situações controladas em bancadas de laboratórios, como por exemplo, “medidas de uma reação química [...] e, nem sempre é possível reproduzir precisamente todas as condições experimentais, particularmente onde algumas variáveis experimentais não podem ser conhecidas e quando os custos de reprodução do experimento são proibitivos” (SAYÃO; SALES, 2015, p. 8).

As investigações experimentais são realizadas em laboratórios e no campo. Independentemente da área do conhecimento, os dados de pesquisa possuem uma essência muito específica. O ambiente e as condições em que foram gerados podem variar de acordo com o clima, a região, o país e as variáveis ambientais do contexto específico. Dependendo do campo específico “várias constatações são de particular importância para a coleta de dados. [...] A calibração é notoriamente difícil, à medida que as medições do sensor derivam, e os sensores experimentais são resistentes em condições de campo de maneiras imprevisíveis” (WALLIS et. al., 2008, p. 124).

Notadamente, as variações e condições ocorridas nos campos experimentais interferem diretamente na coleta e interpretação dos resultados. Dai a importância de um projeto experimental bem elaborado e pré-testes realizados, a exemplo dos testes de calibração, que auxiliam nas decisões e interpretação de dados coletados *a priori*. Ademais, os

dados e documentação para a reprodução do experimento devem ser mantidos, pois poderão ocorrer variações futuras de custos, reprodutibilidade experimental e condições ambientais que estão condicionadas a natureza desses dados.

Um conjunto diferente de considerações aplica-se aos dados computacionais. Estes são gerados pela interação homem-máquina, através de simulações, realidade virtual ou resultados provenientes de manipulações computacionais. Sayão e Sales (2015, p. 3) afirmam que os dados computacionais “são resultados da execução de modelos computacionais ou de simulações, seja, por exemplo, no domínio da física ou para a criação de ambientes virtuais culturais ou educacionais”. A utilização das ferramentas computacionais auxiliam os pesquisadores e a coleta dos recursos científicos são possíveis de ser realizados. A grande quantidade de dados de pesquisa são gerados e processados por meio de diversos recursos computacionais. Dessa forma, a replicação de modelos, simulações e ferramentas computacionais, agregadas a uma extensa documentação relacionada a entrada dos metadados em *software* e repositórios específicos fazem parte do processo de investigação.

Independente da tipologia dos dados, a origem está relacionada com a procedência e proveniência, caracterizando-os pela natureza, valor semântico e significado na prática de pesquisa. As interpretações, análises, validações e novas hipóteses quando se tratar de reuso e reprodutibilidade serão conduzidas por novas práticas científicas. A acessibilidade é outra característica associada ao universo dos dados que emergem de processos de pesquisa ou investigações científicas. A tipologia está relacionada à natureza, essência, espécie ou condição. A acessibilidade relaciona-se as condições de uso e restrições condicionadas ao acesso e controladas por licenças, direitos de propriedade intelectual e normas de conduta ética. O Quadro 3 apresenta uma sinopse das diferentes tipologias adotadas no contexto das práticas da *e-Science*.

Quadro 3 - Tipologias dos dados

TIPO DE DADOS	DESCRIÇÃO DOS DADOS
Dado de Pesquisa	São unidades de informação criadas ou coletadas durante o processo de investigação científica. Geralmente são dados primários, que a partir de uma interpretação preliminar são agrupados como unidades informacionais, e frequentemente formatadas de maneira a torná-las adequadas à comunicação, re-interpretação e processamento legível por computador.
Dado Aberto	São dados de pesquisa disponíveis livremente para reuso e republicação sem restrições de <i>copyright</i> , patentes ou controle de propriedade intelectual. Devem ser disponíveis gratuitamente na <i>internet</i> , permitindo a qualquer usuário baixar, copiar, analisar, reprocessar, capturá-los via <i>software</i> , utilizá-los, reusar e reproduzir para qualquer finalidade, sem barreiras financeiras, legais ou técnicas adicionais.
Dado Primário ou Originário	São coletados em pesquisas científicas, opinião pública ou governamental. Geralmente, são oriundos de coletas observacional, experimental e computacional.
Dado Derivado ou	São originados a partir dos dados originais e podem ser relacionados com outros

Secundário	objetos e documentos externos. Compartilhados por meio de <i>links</i> e <i>hyperlinks</i> .
Dado Científico	São produzidos no contexto da pesquisa científica, coletados, interpretados, acessíveis em processos investigativos.
Dado Bruto	É chamado de dado cru ou primário, coletado diretamente de instrumentos científicos ou da fonte sem nenhuma manipulação ou processamento prévio.
Dado Derivado	É o dado resultante do processamento ou combinação de dados brutos ou de outros dados. Também chamado de dado secundário.
Dado Restrito	É a restrição de acesso ao conteúdo do dado, como restrições quanto ao sigilo do conteúdo e período de embargo. Pode estar associado aos direitos de propriedade intelectual vigentes ao <i>copyright</i> ou patente.
Dado Embargado	São dados que precisam obedecer um período de tempo razoável e específico para o compartilhamento e acesso. É o chamado período de tempo de embargo entre dois a cinco anos, após a coleta primária ou resultados submetidos para publicação sejam aceitos. Agências de fomento permitem os períodos de embargo por razões políticas, comerciais, direitos de propriedade intelectual vigentes, a exemplo de processos de patenteamento em andamento.

Fonte: Adaptado pela autora de Sayão e Sales (2015) e Dicionário de Preservação Digital (2015).

Os dados de pesquisa geralmente são originados em ambientes de pesquisa, e mais especificamente, em processos de investigação científica. Tais dados possuem intrínseca natureza generalista, pois são aplicados para diferentes propósitos e contextos. Assim, o termo ‘dados de pesquisa’ será balizado no sentido *lato sensu* no bojo da ciência aberta e das práticas científicas da *e-Science*.

No contexto científico os dados de pesquisa são conceituados como unidades de informação ou conjuntos de agrupados de dados, os quais são, condensados em fatos ou números específicos, coletados, examinados e considerados como base para o raciocínio, discussão ou cálculo. A *European Commission* (EC) define que “os dados incluem estatísticas, resultados de experimentos, medidas e observações resultantes do trabalho de campo, gravações de entrevistas e imagens” (EC, 2016, p. 3, tradução nossa).

A acessibilidade aos dados refere-se ao direito de acesso, uso e reuso. Geralmente estão condicionados aos direitos, termos e condições de acesso, contratos ou licenças públicas ou de concessão. Essa é uma questão que está diretamente ligada aos sistemas de propriedade intelectual vigentes e as práticas da *e-Science*. A importância de estabelecer políticas e diretrizes que norteiem as práticas científicas e normas acadêmico-científicas e institucionais é uma recomendação internacional (EC, 2016). Prever quais os tipos de dados serão acessados, as condições de acesso e suportes, além das questões legais para uso e restrições são fundamentais para que a acessibilidade e o compartilhamento se realizem.

Assim, o entendimento que os dados abertos são àqueles aptos ao acesso ilimitado (baixar, copiar, analisar, reprocessar, usar, reusar e reproduzir). Ademais o compartilhamento e reuso para os dados abertos requerem que estejam acessíveis, interoperáveis, legalizados, abertos, compartilhados e disponíveis para uso, reuso e reprodutibilidade. Entende-se que

dados abertos são aqueles “disponíveis livremente para reuso e republicação sem restrições de *copyright*, patentes ou outros mecanismos de controle de propriedade intelectual” (SALES; SAYÃO, 2015, p. 80).

Requer que sejam conduzidos por um rol de princípios políticos e éticos, geralmente regidos por instrumentos regulatórios, normativos e códigos de conduta que assegurem os aspectos de fidedignidade, sigilosidade, veracidade e anonimização. Portanto, para que os dados assumam o *status* de aberto, mesmo que produzidos em diversos contextos (científicos, governamentais, institucionais, privados e públicos) é condicionante que estejam acessíveis gratuitamente, com permissão legal ou licenciados para uso, cópia, *download*, visualização, análise, uso, reuso e na forma digital.

Independente dos aspectos relacionados à proveniência e ao acesso, os dados científicos também são classificados como primários (originários) e derivados (secundários). Dados primários são dados brutos, crus ou originários de pesquisas de campo, observacionais, experimentais e computacionais. Portanto, os “dados primários devem ser associados aos seus metadados e armazenados em um ou mais repositórios de dados externos. Esses dados compõem o arcabouço da infraestrutura científica da *e-Science* e alimentam os estágios do ciclo de vida” (TENOPIR et al., 2011, p. 1, tradução nossa). São os primeiros elementos coletados do campo de pesquisa. Essa é uma premissa importante nas práticas que são conduzidas no entendimento da *e-Science*. Pois, as práticas preconizam que dados primários sejam coletados e armazenados preliminarmente para que no futuro estejam disponíveis e acessíveis para uso e reuso.

Frequentemente, os dados primários são reusados em novos processos investigativos com novos objetivos e hipóteses distintas do processo de pesquisa original, mas que passaram por refinamentos, verificações e novas análises gerando resultados, versões e direcionamentos distintos. Comumente, os dados derivados são provenientes da forma bruta (original) que anteriormente foram constituídos por dados primários, documentos e métodos científicos e destinados às novas hipóteses, investigações ou continuidade da investigação original.

A partir de um novo processo de compartilhamento, preservação, reuso e outros procedimentos específicos adotados na investigação primária, novas interpretações, ressignificados e análises são geradas. Assim, a partir do processo de reuso dos dados primários emerge os dados derivados. Dessa forma, os dados derivados são caracterizados por transformações, ressignificações e reconduções científicas (NSF, 2005). Embora acredite-se que os dados primários sejam mais completos, os dados derivados podem ser utilizados por outros e trazer novas acepções para o domínio da pesquisa.

Para tornar os dados acessíveis e preservados “é necessário preservar a documentação adequada relativa ao conteúdo, estrutura, contexto e fonte (parâmetros experimentais e condições ambientais) da coleta de dados, chamados coletivamente de metadados” (NSF, 2005, p. 20, tradução nossa). A preservação deve ser garantida independente das formas, circunstâncias e contextos. A iniciativa de preservar e reutilizar os dados não é um empreendimento isolado. Requer que seja fomentada por instituições científicas, governamentais e de ensino. Nesse quadro, os canais formais de comunicação também merecem atenção já que podem divulgar dados de pesquisa utilizados como matéria-prima no processo científico (BORGMAN, 2015).

De acordo com os princípios e diretrizes estabelecidos no relatório da *Organisation for Economic Co-Operation and Development* (OECD),

os dados científicos são registros factuais utilizados como fontes primárias da investigação científica, comumente aceitos na comunidade científica como necessários para validar resultados investigativos. Um conjunto de dados de pesquisa constitui uma representação sistemática e parcial do assunto que está sendo investigado (OECD, 2007, p. 13, tradução nossa).

Reside nessas afirmações que dados científicos resultam do processo investigativo que obedeceram os rigores científicos e metodológicos e são definidos como “dados digitais e a moeda do universo da coleta de dados, que, como moeda na esfera financeira, vem em formas diferentes. Estas diferenças incluem a natureza dos dados, a sua reprodutibilidade e o nível de processamento a que foram submetidos” (NSF, 2005, p. 19, tradução nossa).

Dessa forma os propósitos que constituem a infraestrutura do universo, estruturas internas, modelos, metodologias e políticas devem ser tratados de forma individualizada na instituição e no contexto que os proveram. Recentemente, Borgman (2015) apresenta uma visão atualizada para a infraestrutura dos dados de acordo com sua relevância. Defende que na infraestrutura da *e-Science* existem os grandes dados (*big data*), os poucos dados (*little data*) e os dados que não possuem valor para a pesquisa (*no data*). O argumento do estudo versa para o valor e relevância dos dados no contexto que estão inseridos. Os dados quando isolados de contextos específicos não possuem valor, significado e relevância.

Assim a relevância está atribuída ao valor e significado dos dados inseridos em uma “infraestrutura de conhecimento - uma ecologia de pessoas, práticas, tecnologias, instituições, objetos materiais, e relacionamentos” (BORGMAN, 2015, *on-line*, tradução nossa). As práticas dessa infraestrutura são baseadas no gerenciamento, compartilhamento, curadoria e preservação a longo prazo. O processo de gerenciamento de dados é realizado por meio de um

modelo metodológico eficiente e aplicável a diferentes tipos de dados e processos. Este modelo é chamado de ciclo de vida dos dados. Atualmente é aplicado no âmbito das instituições de ensino, pesquisa e governamental para a ciência orientada aos dados em alguns países, tais como, Estados Unidos, Reino Unido, Inglaterra e Holanda.

Ainda como parte do contexto da *e-Science* estão os objetos digitais. O termo ‘objeto digital’ é bastante utilizado quando se trata da preservação digital. Por vezes, eles são comparados com outros recursos, tais como: arquivo de computador, documento digital, artefato digital, recurso digital e material digital. O estudo de Yamaoka e Gauthier (2013, p. 82) afirmam que objeto digital “é a mídia de armazenamento, isto é, o suporte mais as cadeias de bits gravadas nele e o objeto digital é constituído de cadeias de bits mais todas as coisas necessárias para dar sentido a essas cadeias”. São representados em diferentes tipos e formatos sob a forma digital, constituindo-se de objeto de informação por possuir uma carga semântica. No estudo os autores propõem que os objetos digitais herdem as propriedades de três classes que Thibodeau (2005) definiu como: físico, lógico e conceitual (Quadro 4).

Quadro 4 – Classificação dos objetos digitais

OBJETO	CLASSIFICAÇÃO
Objeto físico	É o objeto digital que simplesmente prescreve uma inscrição de sinais em uma mídia. O meio físico determina uma convenção para a gravação de dados com densidade e tamanho de blocos diferentes. A inscrição física é independente do significado e, portanto, deve ser lido pelo computador em linguagem natural. A linguagem natural permite que texto ou foto sejam representados e legíveis pelo computador.
Objeto lógico	É um objeto digital reconhecido e processado por <i>software</i> . No nível lógico, um <i>software</i> aplicativo reconhece o formato do objeto. Por exemplo, os tipos de dados padronizados pelos códigos <i>American Standard Code for Information Interchange</i> (ASCII) ou a formatação para o tipo de fonte, recuos e estilos.
Objeto conceitual	Como objeto conceitual, o objeto digital é reconhecido e entendido por uma pessoa e processado por codificações digitais. Por exemplo, um mesmo documento, gravado e reproduzido no processador de texto MS-Word e no <i>Adobe Portable Document Format</i> (pdf), podem manter o mesmo conteúdo, aparência e estrutura. O conteúdo e a estrutura de um objeto conceitual devem ser contidos na forma (objeto lógico) ou nos objetos (na forma digital).

Fonte: Baseado no estudo de Yamaoka e Gauthier (2013, p. 82).

Os tipos de objetos estabelecidos por Thibodeau (2005) e estudados por Yamaoka e Gauthier (2013) apresentam variações de acordo com a natureza da pesquisa e tipologia dos dados. Dependem de variáveis como área de conhecimento e formato que geram o objeto digital. Por vezes, os objetos digitais estão relacionados em uma mesma ambiência computacional com funcionalidades distintas. Com base no entendimento acerca do objeto digital, para um melhor entendimento conceitual buscou-se aprofundar os termos: dados de pesquisa, dados abertos, dados científicos e dados científicos abertos.

O termo ‘dados de pesquisa’ como dito anteriormente se aplica a uma abordagem generalizada referindo-se ao contexto da ciência aberta. Quanto aos ‘dados abertos’ é proposto ao contexto em que o acesso é aberto, livre e gratuito. Não incide restrições para o acesso, uso, reuso e reprodutibilidade dos dados.

Acerca dos ‘dados científicos’ a compreensão está voltada para aqueles atribuídos e provenientes exclusivamente dos processos de pesquisa científica, validados por meio de um método científico e disponíveis para acesso, uso, reuso e reprodutibilidade. Devem ter valor semântico, científico e relevância agregados aos seus conteúdos. Podem ser representados por diferentes suportes e advindos de disciplinas e contextos variados.

Por fim, o capítulo teve o intuito de aprofundar o entendimento acerca dos dados de pesquisa no contexto de uma ciência aberta e de práticas científicas contemporâneas. Ademais, o compartilhamento, proveniência e tipologia no contexto da *e-Science* levou a compreender a distinção e a acessibilidade entre os dados de pesquisa. Assim, os dados científicos compõem o objeto de estudo e o *status* de acessibilidade (aberto, restrito e embargado) é contextualizado de acordo com a natureza da pesquisa. Muito embora almeja-se que dados científicos gerados em um ambiente de ciência aberta e baseado nas práticas da *e-Science* sejam acessíveis, abertos, interoperáveis e gratuitos.

4 MODELOS DE PUBLICAÇÕES CIENTÍFICAS

Este capítulo tem o intuito de mostrar alguns modelos de comunicação científica formais, ao mesmo tempo que, apresenta os modelos contemporâneos que estão abrindo novos horizontes no fazer científico. Até então, os canais formais de publicação científica eram voltados para o processamento da informação e do conhecimento com o intuito de publicar os resultados finais da pesquisa. Na atualidade novos modelos emergiram para dar espaço ao fazer e publicar com dados científicos. Utiliza-se o modelo contemporâneo da *e-Science* denominado de ciclo de vida dos dados, conhecido internacionalmente como *Data Lifecycle Model*. O ciclo de vida é um modelo de gerenciamento dos dados obrigatório para práticas da *e-Science*. Envolve procedimentos mandatórios em diversos estágios do ciclo e que se relacionam diretamente com as questões autorais e a garantia para a autoralidade dos envolvidos no processo de pesquisa. Assim, a dimensão morfológica norteou a sistemática no sentido de mostrar os aspectos práticos (modelos, técnicas e metodologias) para a implantação do ciclo de vida em pesquisas orientadas aos dados científicos.

4.1 MODELOS DE COMUNICAÇÃO CIENTÍFICA: ABORDAGEM TRADICIONAL

A reflexão desta seção se pauta nos modelos de comunicação científica tradicionais. No processo tradicional de Comunicação Científica, os modelos estão relacionados com o fluxo e comunicação da informação ou conhecimento em diferentes níveis. Esses níveis apresentam uma complexidade no gerenciamento e no acesso por parte dos usuários às informações científicas. Com base em modelos, padrões e sistemas é possível ter um processo científico eficiente, pois a “inovação nunca acontece em um vácuo; inovação requer comunicação” (TENOPIR; KING, 2004, p. 11, tradução nossa). A sistematização do fluxo da comunicação científica envolve diferentes níveis e etapas.

Os modelos de comunicação científica remontam ao século XVII com o surgimento dos primeiros jornais científicos. Os pesquisadores utilizavam os meios de comunicação da época, a carta, para se comunicarem com seus pares, independentemente de fronteiras geográficas. As sociedades formadas por literários e cientistas e o surgimento das academias de ciências, criadas a partir deste século, proporcionaram a consolidação da pesquisa acadêmica, por meio dos jornais para divulgar os feitos científicos. “Essas sociedades começaram a publicar formalmente e distribuir materiais dedicados ao seu campo. A revista acadêmica logo se tornou a peça central da sociedade científica” (TENOPIR; KING, 2004, p.

11, tradução nossa). A partir do surgimento dos periódicos científicos, os canais e modelos de comunicação científica foram se modelando. A Figura 6 apresenta uma representação do cenário da publicação científica em vigência.

Figura 6 – Cenário das publicações científicas tradicionais



Fonte: Adaptado do DataONE (2015).

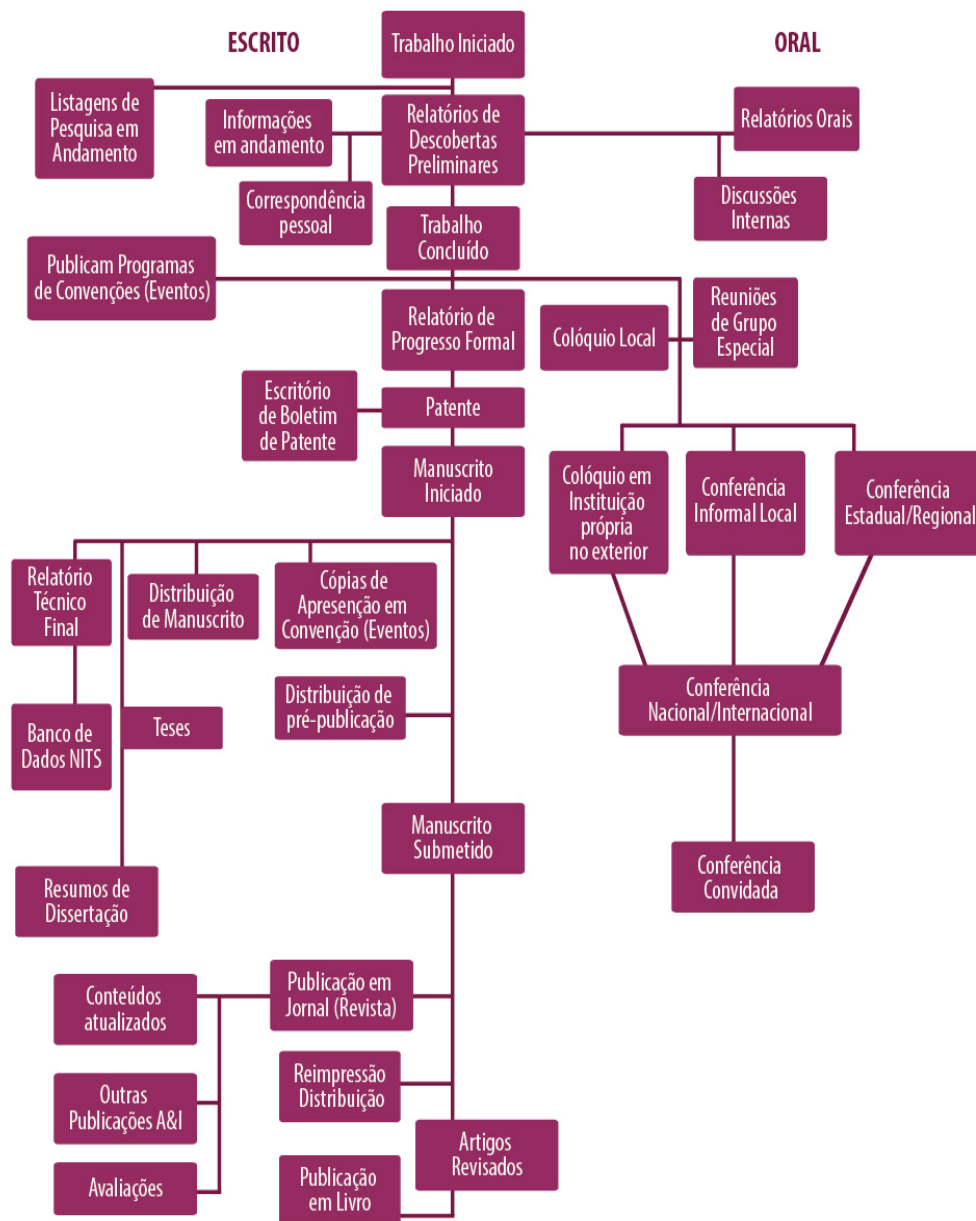
No processo de comunicação científica formal, as publicações científicas são responsáveis pela difusão da pesquisa. No atual cenário da comunicação, devido à complexidade, quantidade e variedade de informações, os modelos diferem muito dos que existiam nos primórdios, muito embora o conceito seja o mesmo, isto é, organizar e comunicar a pesquisa científica.

Os modelos de comunicação científica sofreram alterações e outros novos surgiram a partir de 1990 com a adoção da *internet* na sociedade. A expansão da *internet* foi o meio eletrônico de comunicação utilizado na comunidade acadêmico-científica junto com as Tecnologias de Informação e Comunicação pelo largo espectro nas publicações científicas, especialmente, no formato de publicação científica por meio do periódico científico.

Inicialmente, os periódicos passaram a ser publicados em formato de *CD-ROM* e acompanhavam a versão impressa; em seguida, passaram a ser acessíveis *on-line*; e, atualmente, a maioria das publicações periódicas é acessível eletronicamente, com uma minoria que ainda publica suas versões também de forma impressa. Diferentes sistemas de modelos de comunicação científica são adotados para o gerenciamento do fluxo de informações em diferentes suportes.

O modelo desenvolvido por Garvey, Lin e Nelson (1970) e adaptado por Garvey e Griffith (1979) aborda os canais formais e informais por onde trafega a informação. Este modelo é chamado de ‘meios de comunicação’. O modelo descreve a variedade de canais dos quais os conteúdos de informação técnica e científica são comunicados. Alguns, por sua natureza, são orais (reuniões, palestras, conferências e discussões) e outros são armazenados em documentos (artigos, livros, relatórios, jornais e patentes).

No contexto da comunicação científica, o canal formal registra apenas as informações contidas em suportes tradicionais. Informações que veiculam nos canais informais não são processadas em um fluxo informacional. O conteúdo informacional relacionado com o processo e a pesquisa científica não é processado e armazenado na sua totalidade. A Figura 7 apresenta o modelo denominado meios de comunicação desenvolvido por Garvey e Griffith e apresentado no estudo de Tenopir e King (2004) que traz um fluxo que trata das informações no ciclo formal (escrito) e informal (oral).

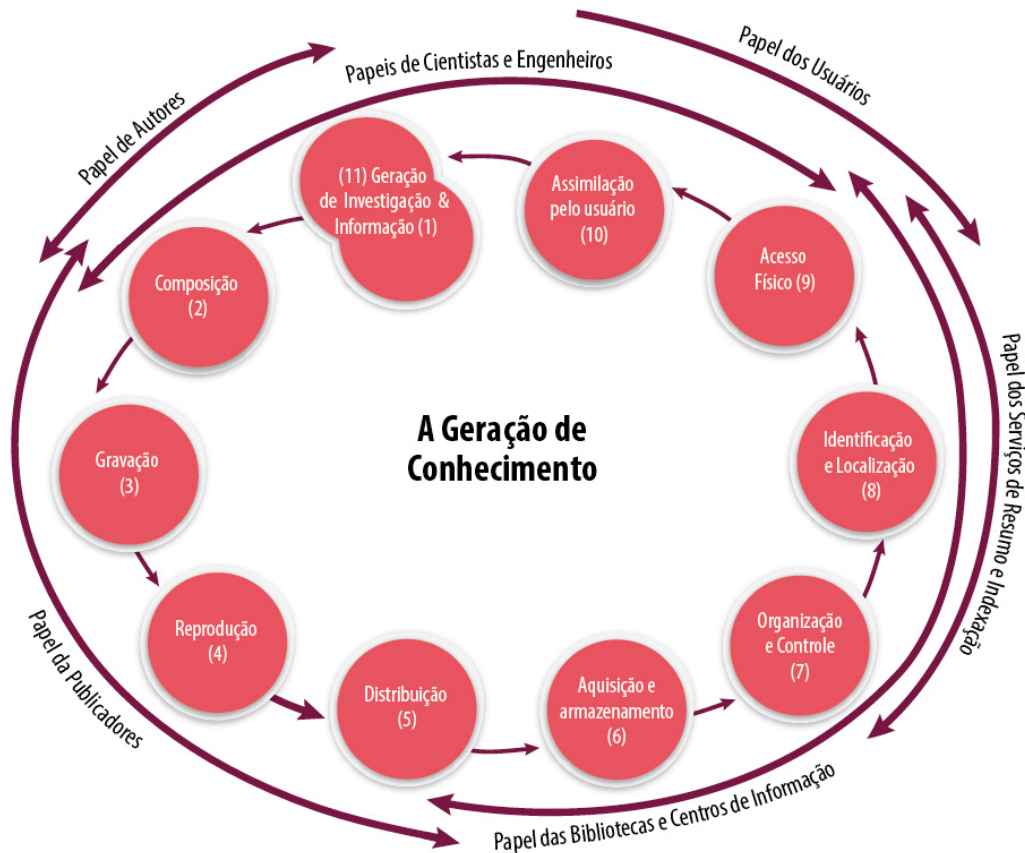
Figura 7 – Meios de comunicação

Fonte: Modelo apresentado por Garvey e Griffith (1979) e traduzido do estudo de Tenopir e King (2004, p. 11).

Estudos também apontam para os aspectos de tempo, custos, duplicação e colaboração científica no processo de comunicação científica. Muito embora ele tenha sido elaborado antes da *internet*, sua abordagem ainda é atualizada e aplicável para um sistema atual.

O modelo desenvolvido por King e Roderer (1981) chamado de *Life Cycle of Scientific Information Through the Scholarly Journal*, tem uma abordagem centrada no ciclo de vida da informação científica através do sistema de periódico acadêmico conforme está representado na Figura 8. É um modelo que se aproxima dos atuais modelos contemporâneos orientados aos dados de pesquisa.

Figura 8 – Ciclo de vida de informação científica através do periódico acadêmico

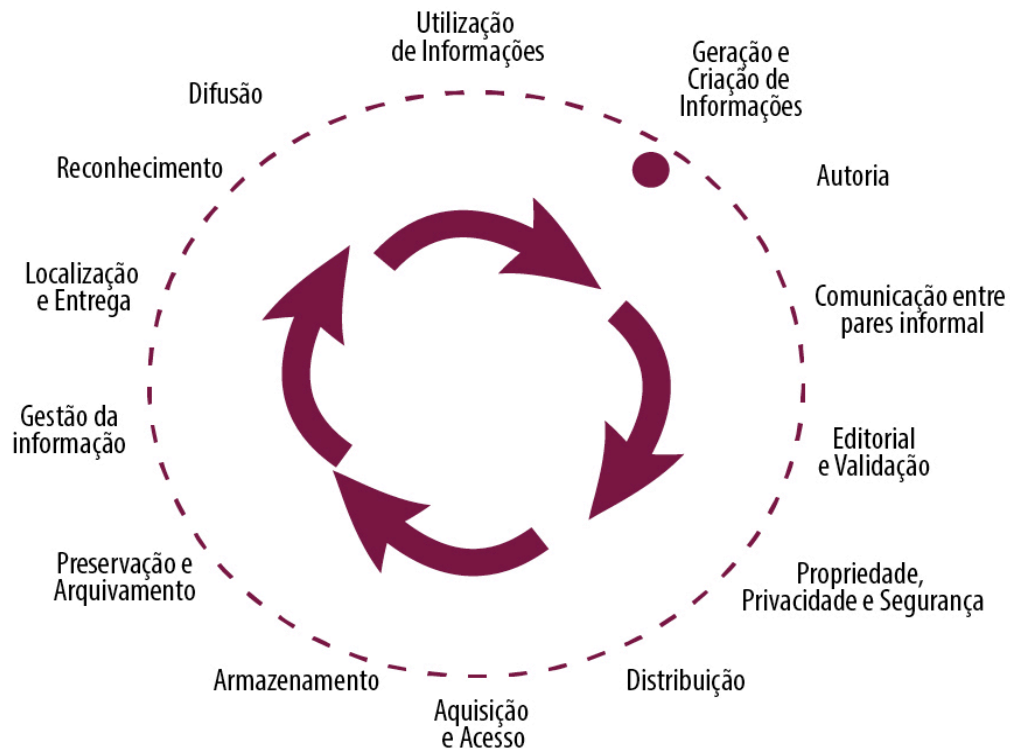


Fonte: Traduzido do modelo de King, Mcdonald e Roderer (1981).

Este modelo tem formato de espiral e inclui 11 (onze) funções que começam com a pesquisa e a criação de outros recursos informacionais e terminam com a geração de pesquisa e informação. Além dos estágios de geração e processamento de informações, o modelo prevê diferentes papéis e regras que estão relacionados aos usuários do sistema. Os usuários podem ser autores, pesquisadores, bibliotecários, bibliotecas e centros de informação, provedores de serviços de indexação e resumos e os usuários. Percebe-se que é um modelo complexo e funcional com regras e funções bem estabelecidas com um grau de eficiência considerável. Este modelo possui semelhança com o atual modelo aplicado aos dados de pesquisa.

O modelo *Scholarly Journal Information Cycle Model* desenvolvido em 1994, pela *Association of American University Research Libraries* trata do ciclo informacional em periódicos acadêmicos e científicos. Este modelo apresenta um esquema de funcionamento em espiral que reflete o processo comunicacional contínuo e regenerativo, conforme demonstrado na Figura 9.

Figura 9 – Modelo de ciclo da informação de periódico acadêmico



Fonte: Traduzido do modelo da *Association of American University Research Libraries Project Report* (1994).

O grupo de trabalho desenvolveu este modelo a partir de uma ampliação das funções desenvolvidas anteriormente para a revista impressa tradicional. O estudo identificou três variações potenciais que eles chamaram de “modelo clássico”, “modelo modernizado” e “modelo emergente”. Todos os modelos possuem fluxos, funções e características similares. Algumas diferenças se configuram na ordem de importância e nas facilidades de desempenho. O “modelo clássico” é direcionado aos recursos impressos, a exemplo dos jornais e revistas impressas, com os elementos de autoria; edição e validação; propriedade, privacidade e proteção; aquisição e acesso; armazenagem e preservação; gerenciamento da informação; reconhecimento; e utilização. As informações referentes ao autor e à instituição são retidas de forma confiável e segura. Este modelo é mais aplicado aos jornais impressos, muito embora possa também ser aplicado ao jornal eletrônico.

O “modelo modernizado” também descreve os mesmos elementos para recursos impressos, mas concentra a entrega sob a demanda de artigos ou revistas e incorpora algum uso de recursos eletrônicos. O “modelo emergente”, por sua vez, é baseado quase na sua totalidade em meios eletrônicos para formatos e produtos de comunicação. A natureza eletrônica das tecnologias do modelo emergente levanta questões relacionadas com a propriedade, privacidade e proteção, porque afeta diretamente a propriedade intelectual, devido à facilidade de cópia e dificuldade no controle de acesso e integridade do trabalho.

Apesar de este modelo ter sido desenvolvido antes da *internet*, dos periódicos eletrônicos e do acesso aberto, ele apresenta questões que são refletidas na atualidade, como a propriedade intelectual. Na seção seguinte serão abordados os modelos relacionados com o atual contexto da *e-Science*.

4.2 MODELOS CONTEMPORÂNEOS DE PUBLICAÇÃO CIENTÍFICA

O ciclo de vida dos dados é o modelo que apoia o processo de coleta, compartilhamento, gerenciamento, curadoria, acesso, uso e reuso. A adoção deste modelo preconiza que “os dados são a infraestrutura da ciência. Dados confiáveis são críticos à medida que eles formam a base para boas decisões científicas, gerenciamento sábio e uso de recursos na tomada de decisão” (TENOPIR et al., 2011, p. 1, tradução nossa).

O compartilhamento e o gerenciamento são imprescindíveis para tornar dados científicos acessíveis e recuperáveis.

4.2.1 Gerenciamento e Compartilhamento dos Dados Científicos

As questões de uso e reuso de pesquisa no processo de compartilhamento e colaboração científica estudadas por Gray (2007) e Ambinder e Marcondes (2012) afirmam que na *e-Science*, os pesquisadores estão focados além de obter a recompensa científica, reusar e reproduzir experimentos de pesquisa semelhantes com o intuito de evitar erros inúteis e aprimorar a investigação. As experiências de pesquisas replicadas e reutilizadas estão na essência da *e-Science*.

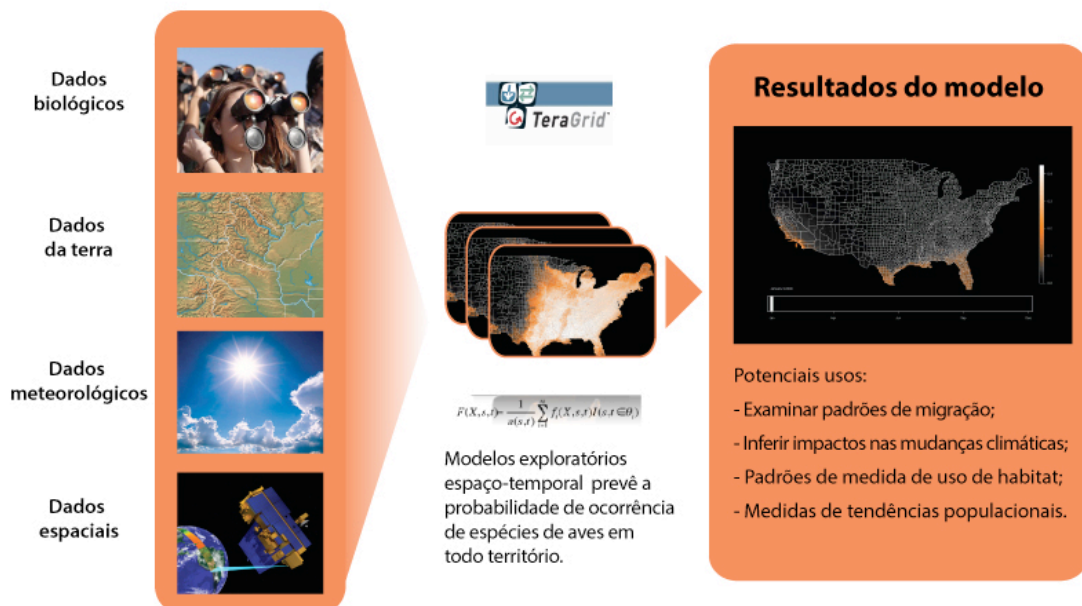
O ciclo de vida dos dados tem uma importância fundamental no plano de gerenciamento e compartilhamento de dados de pesquisa no contexto da *e-Science*. Assim, iniciar com uma visão geral do ciclo de vida, identificando os passos de todo o processo de investigação é o mais recomendável para o pesquisador que tem interesse em adotar este modelo para gerenciamento de toda a sua pesquisa. Instituições de pesquisa e organizações internacionais que utilizam o modelo adotam tal metodologia e recomendam que o processo inicial do compartilhamento de dados já utilize essa sistemática.

Na perspectiva do pesquisador, o gerenciamento de dados apresenta a possibilidade de manter os dados organizados e acessíveis. Acompanhar o processo de reuso e controle da versão de seus dados, para os seus próprios estudos e de outros, tornando-os mais eficientes e com qualidade, evita a perda de dados a partir de um *backup* eficaz e proporciona que a

reutilização dos dados traga maior entendimento, credibilidade e reconhecimento pelo esforço científico. A perspectiva do gerenciamento dos dados favorece o compartilhamento e reuso.

Na perspectiva institucional ou governamental, os dados bem gerenciados e armazenados adequadamente, além de proporcionarem o compartilhamento e reuso, aumentam a visibilidade e impacto da investigação, promovem a inovação, incentivam as colaborações coletivas, maximizam a transparência e o controle dos processos e resultados de investigações, reduzem os custos e duplicação na coleta de dados primários. A representação da Figura 10 exemplifica a perspectiva de que os dados bem geridos podem resultar no reuso, integração e numa nova ciência aberta.

Figura 10 – Gerenciamento e compartilhamento de dados geo-espaciais



Fonte: Adaptado do *DataONE* (2015).

O exemplo apresentado reflete a boa prática de um processo de coleta, gestão e compartilhamento de dados abertos. Neste processo, é possível realizar o acesso, uso e reuso para novas pesquisas e descobertas. Ainda nessa perspectiva, o aspecto do armazenamento e preservação de dados compartilhados em longo prazo é outra nuance desse contexto que deve ser considerada.

Com base nas experiências internacionais acerca do compartilhamento de dados e da sua importância para que ocorra sucesso em um empreendimento desta natureza. Os benefícios almejados devem se estender a um público específico, mas ao mesmo tempo, que atenda ao patrocinador da pesquisa, à agência financiadora, à comunidade da pesquisa e aos

pesquisadores que utilizarão o compartilhamento dos dados para uso e reuso de suas pesquisas. O Quadro 5 apresenta alguns benefícios agrupados e para quem se destinam dentro de um contexto científico e social.

Quadro 5 – Benefícios do compartilhamento de dados

BENEFÍCIOS DO COMPARTILHAMENTO DE DADOS	
Público (<i>Public</i>)	Uma melhor informação dos rendimentos públicos que melhoram a decisão no que diz respeito ao: Planejamento ambiental e econômico; Política federal, estadual e local; Opções de melhoria social e na educação como o uso de dinheiro dos impostos; Melhorar o estilo de vida pessoal e saúde, como recreação e nutrição.
Patrocinador ou agências de financiamento de pesquisa (<i>Research sponsor</i>)	Organizações que patrocinam a pesquisa devem maximizar o valor dos dólares para a pesquisa. O compartilhamento de dados aumenta o valor de investimentos em pesquisa, permitindo: Verificação de métricas de desempenho e resultados; Novas pesquisas e maior retorno sobre o investimento; Avanço da ciência; Redução de gastos evitando a duplicação de dados.
Comunidade Científica (<i>Scientific Community</i>)	Acesso a pesquisas relacionadas permite aos membros da comunidade: Construir sobre o trabalho de outros e adicionar, em vez de repetir, a ciência; Realizar meta-análises que não podem ser realizadas com conjuntos de dados individuais ou laboratoriais; Compartilhar recursos e perspectivas de modo que a compreensão seja expandida e compartilhada; Incremento na transparência, reprodutibilidade e comparabilidade de resultados; Expandir avaliação metodológica, recomendações e aperfeiçoamento; Educar novos pesquisadores quanto aos resultados mais atuais e significativos.
Cientista (<i>Scientist</i>)	Cientistas que compartilham dados ganham o benefício de: Pesquisa patrocinada reconhecida como uma autoridade; Reconhecimento de patrocinador da pesquisa como uma fonte autorizada e sábio investimento; Melhoria da qualidade de dados devido ao uso expandido, verificações de campo e feedback; Maior oportunidade para a troca de dados; Estabelecimento de melhores ligações à rede científica, colegas e potenciais colaboradores.

Fonte: Adaptado do DataONE *Education Module* (2012).

Esse modelo apresenta diferentes perspectivas (do pesquisador e da sociedade) para a comunicação científica. A base desses novos modelos está pautada no gerenciamento de dados científicos, de natureza digital e aberta. O gerenciamento de dados científicos envolve diferentes fluxos de informação com processos e subprocessos.

Além do compartilhamento, é necessário que o processo de gestão dos dados adote os procedimentos, padrões e técnicas em conformidade internacional. Por exemplo, fornecer metadados confiáveis e com qualidade evita o uso inapropriado de termos ou parâmetros, seguridade e confiabilidade de dados sensíveis, falta de créditos e agradecimentos, perda de vantagem competitiva em pesquisas financiadas.

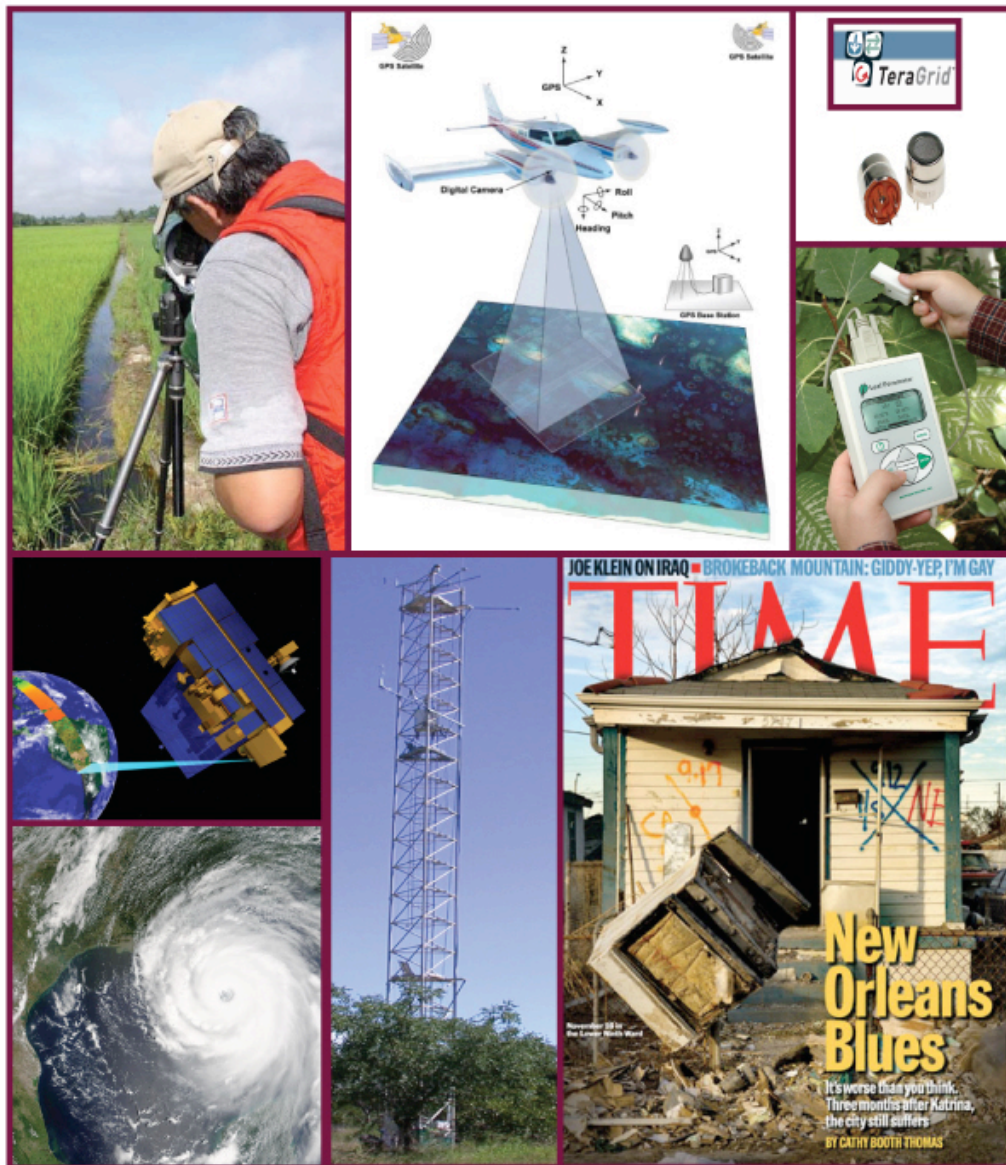
No cenário internacional, as agências financiadoras e instituições de pesquisa *nos* “últimos anos emitiram declarações e políticas ressaltando a necessidade para solicitação de arquivamento de dados, e algumas agências de financiamento começaram a exigir que os projetos de dados que elas financiam sejam depositados em um arquivo público” (ICPSR, 2012, *on-line*). A partir das políticas institucionalizadas no âmbito destas instituições, a maioria dos projetos de pesquisas financiados segue as recomendações para adotar uma prática baseada em ferramentas e modelos para o gerenciamento de dados científicos.

Na conjectura atual o tempo é despendido na geração de dados primários e o valor agregado destes em cada área do conhecimento. Dessa forma, o gerenciamento de dados científicos traz à tona diversas questões que envolvem o uso, reuso, reprodutibilidade, compartilhamento, gerenciamento, armazenamento, curadoria/preservação de dados, além de questões de propriedade intelectual, *software*, *hardware*, servidores, padrões de metadados, protocolos de interoperabilidade, controle de qualidade, bem como pesquisadores qualificados e pessoal especializado.

A sistemática do modelo de ciclo de vida possui diversos estágios, objetivando um processo eficiente e adequado para o compartilhamento, uso e reuso de dados científicos. Em áreas como “as ciências da vida e outros campos que dependem de dados observacionais, a situação está ficando cada vez mais complexa à medida que os dados são usados para modelagem avançada, a qual cria novos conjuntos de dados” (TENOPIR et al., 2011, p. 4, tradução nossa).

Para o tratamento e curadoria de dados científicos, o pesquisador precisa compartilhar seus dados para uso e reuso de longo prazo. Pesquisadores e institutos de pesquisa começaram a identificar que existem desperdícios na duplicação de pesquisas financiadas, tempo demandado para a execução e do descarte na geração do valor agregado aos dados primários em diferentes áreas. O universo que envolve os dados de pesquisa é infinito. Dados de pesquisas são gerados, instantaneamente, em diversos ambientes e áreas do conhecimento. Para o tratamento e curadoria de dados científicos, o pesquisador precisa compartilhar seus dados para uso e reuso de longo prazo. Pesquisadores e institutos de pesquisa começaram a identificar que existem desperdícios na duplicação de pesquisas financiadas, tempo demandado para a execução e do descarte na geração do valor agregado aos dados primários em diferentes áreas. A Figura 11 ilustra os diferentes cenários de coleta de dados primários.

Figura 11 – Cenário dos dados primários



Fonte: Adaptado do DataONE (2015).

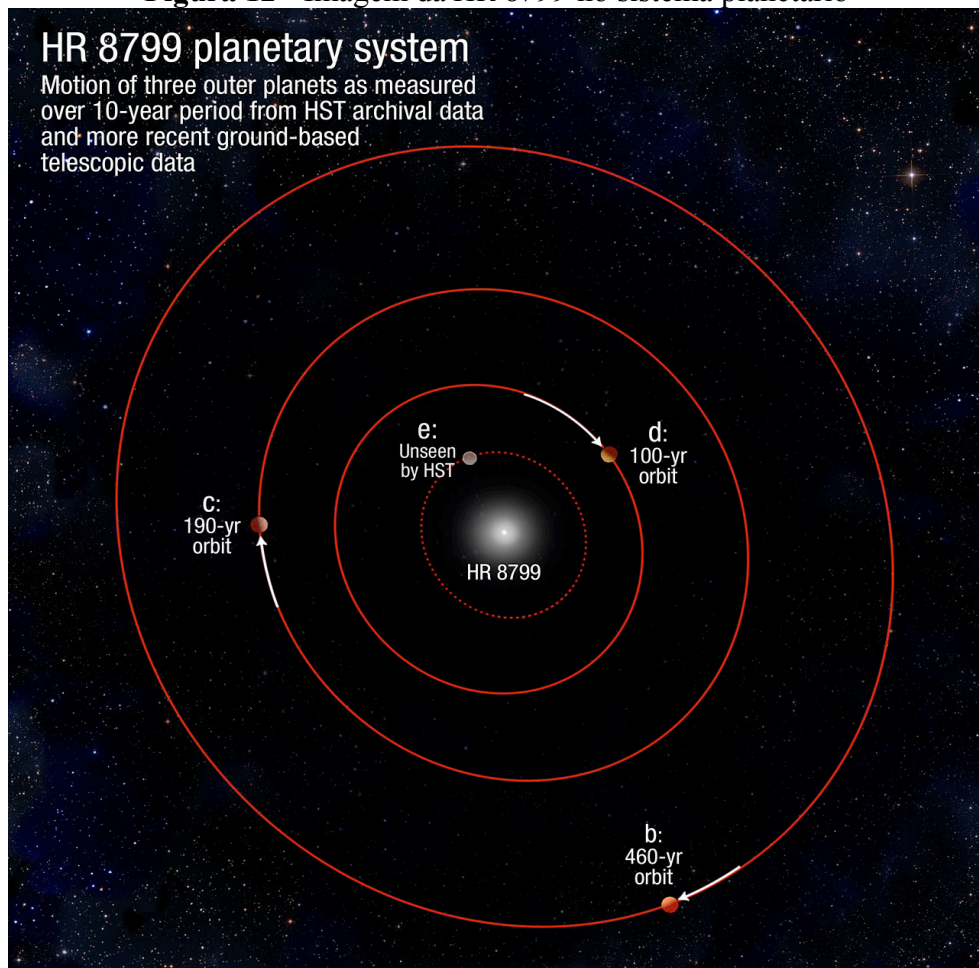
O compartilhamento de dados de pesquisa não é algo recente, “há mais de 40 anos, cientistas da computação já compartilhavam arquivos, de forma anônima, através de *File Transfer Protocol* (FTP), [...] para transferir arquivos de um computador hospedeiro para outro, em uma rede” (MACHADO, 2015, p. 201). O compartilhamento inclui, além dos resultados, os dados primários de pesquisas. As abordagens contemporâneas das publicações científicas que permeiam a ambiência da *e-Science* estão direcionadas para o aproveitamento intenso dos recursos e ferramentas tecnológicas, possibilidades de compartilhamento e colaboração científica e reutilização de pesquisas.

Em 2009, David Lafrenière, da Universidade de Montreal, recuperou uma imagem de HR 8799 do planeta mais externo usando dados do Hubble coletados muito mais cedo, em

1998, com a câmera de infravermelho próximo e o Multiobjeto Espectrômetro (NICMOS). Duas décadas posteriores à descoberta de David Lafrenière, numa nova análise meticulosa dos mesmos dados arquivísticos NICMOS, o pesquisador Rémi Soummer do *Space Telescope Science Institute* em Baltimore encontrou evidências visuais para todos os três planetas externos (NATIONAL AERONAUTICS AND SPACE ADMINISTRATION, 2011).

O quarto planeta não pôde ser visto porque a *coronagraph NICMOS* o obscureceu. Ao encontrar estes planetas mais velhos no *Hubble data*, a equipe demonstrou ter um recurso valioso para determinar o movimento orbital que utiliza dados retrospectivos para comparar à posição anterior do planeta com aqueles vistos em observações mais recentes, como é demonstrada na Figura 12.

Figura 12 - Imagem da *HR 8799* no sistema planetário



Fonte: NASA (2011).

Esse exemplo demonstra a utilidade do arquivo do *Hubble Data* que contém imagens e informações espectrais coletadas em mais de 20 (vinte) anos de observações. A descoberta tem uma base fundada no método de Lafrenière. Lafrenière foi o primeiro pesquisador a

demonstrar o poder de uma nova técnica de processamento de dados para recuperar assinaturas planetárias fracas do brilho de suas estrelas centrais. Ele também reforça uma nova abordagem para a descoberta adicional de planetas que, a partir de dados arquivados no *Hubble data* somados a novos dados e informações trouxeram um avanço para a ciência e perspectivas para novas descobertas.

No cenário nacional, a iniciativa brasileira na área de Biodiversidade é uma iniciativa que desenvolveu um sistema de gestão, integração e monitoramento de dados da Biodiversidade no Brasil. Considera-se que o “processo de compartilhamento e gestão de dados abertos é muito incipiente, muito embora já existam algumas iniciativas que estão sendo direcionadas a partir de um processo integrado de gestão de monitoramento de dados primários da biodiversidade” (CORRÊA et al., 2004, p. 1). O projeto pioneiro com dados da Biodiversidade no Brasil está sendo desenvolvido pela Universidade de São Paulo, instituições de pesquisas ambientais e o Ministério do Meio Ambiente (MMA). Os dados foram adequados à sistemática e à infraestrutura tecnológica baseados na perspectiva da *e-Science*. O gerenciamento de dados de pesquisa é possível a partir do seu compartilhamento.

A importância do compartilhamento de dados no cenário brasileiro requer esforços coletivos, recursos, diretrizes, políticas governamentais e institucionais. A disponibilidade dos pesquisadores para prover seus dados e torná-los abertos e acessíveis para reuso possibilitará benefícios em larga escala e o avanço da ciência aberta no contexto brasileiro. Instituições internacionais, organizações governamentais, acadêmicas e de pesquisas, em especial nos Estados Unidos, adotam o modelo *Data Lifecycle* como uma prática mandatória e rotineira na investigação científica. Esta fase da pesquisa será conduzida dentro de um modelo e da adoção de padrões internacionais validados para o gerenciamento de dados científicos.

Apointar para a importância do compartilhamento de dados no cenário brasileiro requer que esforços, recursos, diretrizes, políticas e disponibilidade dos pesquisadores sejam agregados em uma finalidade coletiva, para que os benefícios sejam em larga escala e em prol do avanço da ciência.

As subseções seguintes serão destinadas ao modelo de ciclo de vida dos dados. Com a intenção de delimitar o foco deste capítulo serão abordados os dados de pesquisa abertos de natureza científica e serão denominados, para efeito desta pesquisa, dados científicos.

4.3 MODELOS DE CICLO DE VIDA DOS DADOS

O ciclo de vida dos dados científicos é um modelo aplicado à gestão de dados abertos e de natureza científica. As etapas do processo de gestão e compartilhamento de dados estão contidas no ciclo de vida dos dados. Muitas iniciativas institucionais e governamentais no contexto internacional foram desenvolvidas a partir desse modelo para a padronização no processo de coleta, gerenciamento, curadoria e preservação desses dados. As seções apresentam modelos de ciclo de vida que são adotados por diferentes instituições.

4.3.1 Modelo de Ciclo de Vida do ICPSR

O modelo de ciclo de vida de dados desenvolvido pelo *Inter-university Consortium for Political and Social Research* (ICPSR) em 2013, tem uma sistemática de com seis estágios. É recomendado que o gerenciamento de dados e os planos de compartilhamento sejam desenvolvidos em conjunto com um arquivo para maximizar a utilidade dos dados e garantir a disponibilidade desses dados no futuro.

O ciclo inicia com a proposta de desenvolvimento do plano de gerenciamento dos dados e finaliza com o depósito dos dados. Conforme o referido modelo, o pesquisador deve cumprir a sistemática contida em cada fase do ciclo da seguinte forma:

Fase 1: Desenvolvimento da proposta e plano de gerenciamento dos dados. Nesta fase, o pesquisador deve ter contato com o arquivo para supervisionar e criar o plano de gerenciamento dos dados para assegurar por longo prazo a avaliação de recursos de dados.

Fase 2: Início do projeto. Este é o momento em que o pesquisador faz as suas decisões sobre a forma de documentação e conteúdo.

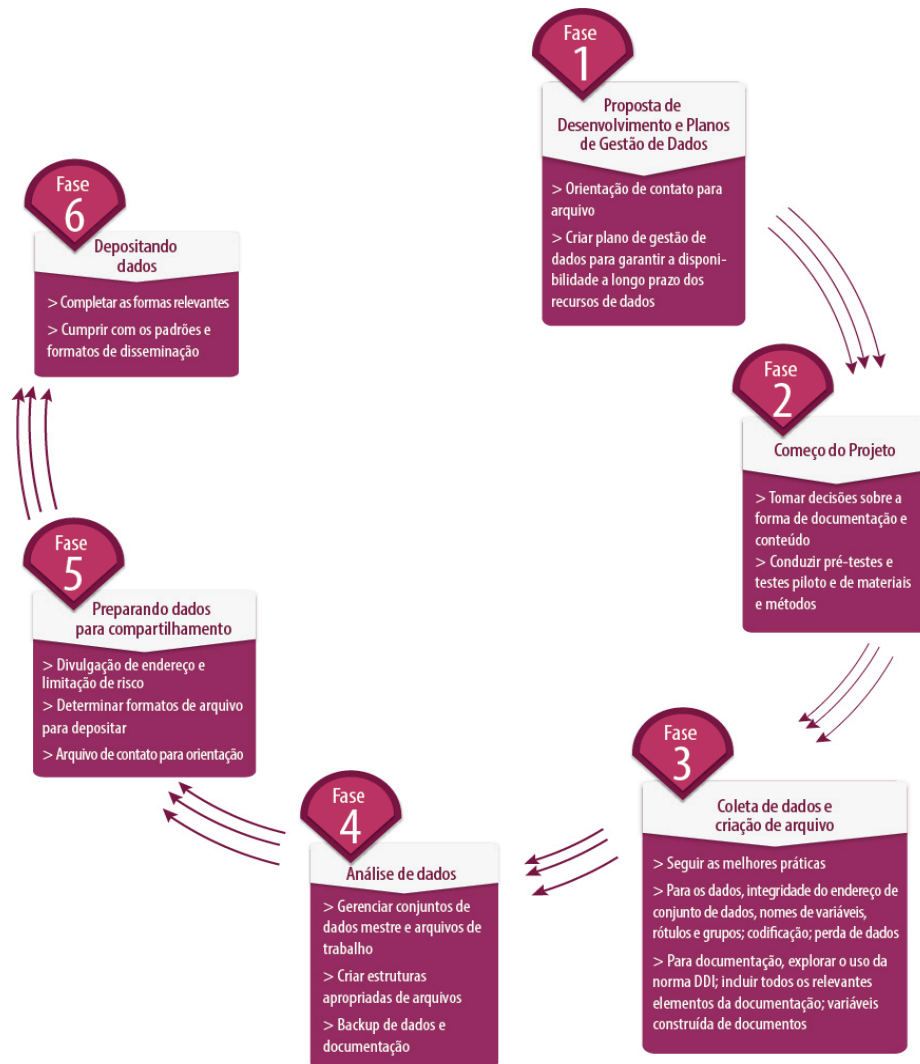
Fase 3: Coleções de dados e criação de arquivos. Nesta fase, o pesquisador já sabe quais os dados farão parte de sua coleção e quais os arquivos deverão ser gerados. A escolha do endereço do *Dataset* com integridade, nomes variáveis, etiquetas e grupos, codificação e dados em falta.

Fase 4: Análise dos dados. É o momento de gerenciar os conjuntos de dados mestre e arquivos de trabalho. Criar estruturas de arquivos apropriados, *backup* de dados e documentação.

Fase 5: Preparação dos dados para compartilhamento. Disponibilizar o endereço de divulgação com risco de limitação, determinar os formatos de arquivo para depositar e o contato do arquivo para aconselhamento.

Fase 6: Depositando dados. É a ocasião de preencher os formulários relevantes e cumprir com os padrões de divulgação e formatos desejáveis.

Figura 13 – Modelo de ciclo de vida dos dados de pesquisa do ICPSR



Fonte: Traduzido do Modelo ICPSR (2013).

Este diagrama apresenta apenas um resumo das fases em que o pesquisador do ICPSR deve adotar durante o processo de compartilhamento de dados. O modelo é aplicável para uma realidade específica institucional e foi enriquecido com padrões, orientações, políticas e regulamentações apropriadas a este contexto.

4.3.2 Modelo de Ciclo de Vida da USGS

A *United States Geological Survey* (USGS) é uma organização científica federal dos Estados Unidos que estuda o ambiente, seus recursos e perigos naturais que ameaçam o país. A USGS tem quatro maiores disciplinas científicas relativas à Biologia, Geografia, Geologia e Hidrologia. Pertence ao *United States Department of The Interior* como uma agência científica única e exclusiva desse departamento. A USGS representa dados corporativos com valor potencial para além de algum uso imediato de pesquisa, e, portanto, esses precisam ser considerados e devidamente gerenciados durante todo o seu ciclo de vida.

Com a devida importância relacionada ao gerenciamento de dados científicos, foi constituída no âmbito da USGS uma equipe que desenvolveu *Scientific Data LifeCycle Model* (SDLM) com uma visão de alto nível dos dados, desde a sua concepção até a preservação e partilha, para ilustrar como as atividades de gerenciamento de dados referem-se aos fluxos de trabalho do projeto e para ajudar na compreensão das expectativas adequadas de gerenciamento de dados.

Em 2009, foi criada pela USGS a *Community For Data Integration* (CDI). A CDI representa uma comunidade dinâmica de práticas focadas no avanço de dados científicos e gerenciamento de informações e recursos transversais de integração dentro da USGS. Assim, a CDI através de parcerias e grupos de trabalhos desenvolveu ferramentas e práticas de gerenciamento de dados, infraestrutura cibernética, ferramentas colaborativas, debates e treinamentos para suporte de cientistas e especialistas tecnológicos.

Em 2010, a CDI estabeleceu o *Data Management Working Group* (DMWG) para desenvolver, ampliar e recomendar as melhores práticas e políticas para assistir a organização efetivamente na manipulação, documentação, preservação e promoção do acesso à agência de dados científicos. O subgrupo de trabalho investigou exclusivamente a existência de modelos de ciclo de vida de dados, que resultou na compilação de aproximadamente 50 (cinquenta) modelos advindos da academia, agências federais e organizações internacionais. Tais modelos diferenciavam largamente na granularidade, complexidade, apresentação e perspectiva da comunidade, bem como diferentes modelos de ciclo, por exemplo, circulares, espirais, lineares e hierárquicos.

O modelo consiste em um fluxo intuitivo (da esquerda para a direita) e um ponto de partida claramente definido que se alinha com o início do projeto de pesquisa. Atenção especial é dada à efetiva tomada de escolhas semânticas, com o firme propósito de prover comunicação clara dos componentes do modelo. O desenho final do modelo foi revisado e aceito em novembro de 2012 pelos patrocinadores executivos do CDI. A finalidade primordial com a aplicação desse modelo às atividades de pesquisa era de que cientistas da

USGS pudessem garantir que os produtos de dados fossem bem descritos, preservados, acessíveis e aptos para reutilização. Aprofundamentos e estudos contínuos se realizaram até a definição em 2011 de um modelo gráfico linear, conforme demonstrado na Figura 14.

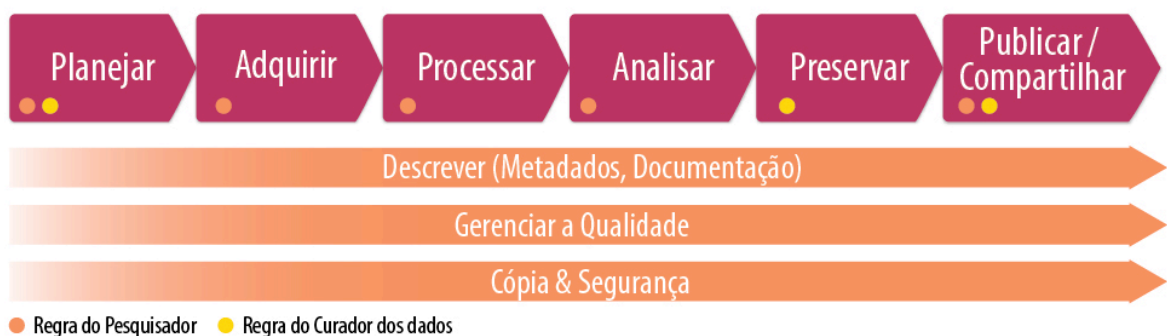
Figura 14 - Modelo de ciclo de vida dos dados de pesquisa da USGS



Fonte: Traduzido do Modelo USGS (2015).

Este modelo define também os diferentes papéis e responsabilidades para serem realizados pelo pesquisador e curador dos dados de pesquisa, que inclui dados e atividades de gestão que exigem conhecimento e habilidades especializados, bem como a educação permanente sobre os métodos e padrões. A Figura 15 apresenta os estágios do *Data Management Roles and Responsibilities in Research Project into USGS Science Data Lifecycle Model*.

Figura 15 – Papéis e responsabilidades do Modelo USGS



Fonte: Traduzido do Modelo USGS (2015).

Os pesquisadores ou grupo de pesquisa são responsáveis pelo planejamento, preservação e publicação/compartilhamento. Os curadores dos dados são responsáveis pelos estágios de coleta, processo, análise e, em conjunto com o pesquisador, realizam a publicação e o compartilhamento. Os elementos primários do modelo que compõem o *Data Life Cycle* da USGS são:

a) **Planejar** (*Plan*): é o primeiro elemento do ciclo que se destina a ajudar os cientistas em planejar e considerar todas as atividades relacionadas ao manejo de ativos de

dados do projeto, da concepção do projeto ao arquivamento. Durante este estágio, todos os elementos do modelo devem ser avaliados, tratados e documentados. A equipe do projeto deve considerar abordagens, recursos necessários (incluindo financiamento e pessoal) e saídas previstas para cada etapa do ciclo de vida dos dados. Um plano de gestão de dados é o resultado almejado neste estágio.

b) **Adquirir** (*Acquire*) ou **Coletar** (*Collect*): é o segundo elemento do ciclo que representam as atividades através das quais dados novos ou já existentes são coletados, gerados ou considerados e evoluem para reuso. Este é um estágio importante que requer dos pesquisadores, habilidades e técnicas de pesquisa para a coleta de dados, considerando as questões éticas e metodológicas, a fim de manter a qualidade, proveniência e integridade destes.

c) **Processar** (*Process*): o terceiro elemento do modelo que representa as várias atividades associadas. É o processamento de entrada de dados que pode implicar a definição de elementos de dados, integração de conjuntos de dados díspares, extração, transformação e carga de operações, e aplicação de calibrações para preparar os dados para análise. Este estágio lembra aos pesquisadores que padrões e ferramentas deverão ser adotados de acordo com os requisitos do projeto e, ao mesmo tempo, construam uma base de dados para a ciência integrada. As saídas deste elemento são conjuntos de dados (*datasets*) que deverão estar prontos para a interoperabilidade e análise.

d) **Analisar** (*Analyze*): é o quarto elemento que representa as atividades relacionadas com a exploração e interpretação do processamento de dados, quando as hipóteses são testadas, as descobertas são feitas e as conclusões são desenhadas. As atividades analíticas nesse estágio do processo incluem sumarização, desenhos, análises estatísticas e espaciais, modelos que são usados para produzir informações e resultados científicos. Nesse elemento, novos dados são gerados, versões são rastreadas e processos são documentados. O gerenciamento de dados durante a análise aprimora a eficiência das atividades, preserva a documentação que é crítica para a integridade científica e cria uma base para futuras pesquisas. A saída de dados são interpretações ou novos *datasets* que, muitas vezes, são publicados em relatórios escritos e outros formatos.

e) **Preservar** (*Preserve*): é o quinto elemento do modelo. Representa as atividades associadas ao armazenamento de dados para o uso e acessibilidade por longo prazo. Preservação muitas vezes não é considerada até o final da etapa de um projeto. Nesse estágio a publicação e o compartilhamento reforça o compromisso de que cientistas financiados pelo governo federal devem planejar a preservação de dados, metadados, produtos auxiliares,

formatos de armazenamento de aplicativos neutros e toda a documentação adicional a longo prazo, garantindo a disponibilidade e reutilização futura.

f) **Publicar** (*Publish*) e **Compartilhar** (*Share*): é o sexto elemento no modelo que envolvem os sistemas de coletas (*Harvesting*) e os conceitos de publicação *peer-reviewed* tradicional que se encontram distribuídos por meio de *sites*, catálogos de dados, mídias sociais e outros locais. A publicação e disseminação de dados são componentes críticos que devem receber especial atenção. Em suma esse elemento lembra aos cientistas que tanto os dados como as publicações tradicionais são produtos de pesquisas.

Além dos elementos principais, o modelo propõe elementos transversais chamados de *Cross-Cutting Elements*. A execução ocorre simultaneamente com a execução dos estágios (processo e análise):

a.1) Descrever (*Describe*): este estágio é o primeiro elemento transversal entre os elementos principais (*Process e Analyze*). É o estágio que se descreve os metadados e a documentação baseados em padrões de metadados, tais como código de *software*, modelos de dados e fluxos de trabalhos. A descrição inicia no plano de gerenciamento dos dados e no detalhamento da documentação em todas as fases. O fluxo descritivo de trabalho facilita a indexação, adesão aos padrões e procedimentos, compreensão de cada fase e uso futuro dos dados. Registrar as informações em cada estágio do ciclo garante a precisão e compreensão dos dados científicos criados, compilados, processados e compartilhados. A finalidade é que outros cientistas possam validar os resultados através da replicação, resultados ou reuso dos dados.

a.2) Gerenciar a qualidade (*Manage Quality*): gerenciar a qualidade é o segundo elemento transversal. Os pesquisadores devem adotar as medidas de garantia de qualidade para os dados no início do projeto e, em seguida, realizar o controle de qualidade durante o ciclo. O monitoramento e ajuste devem ser realizados durante todo o ciclo para que haja alta qualidade nos produtos da ciência. Este elemento alinha com as políticas mandatórias das agências de financiamento que propõe a realização de produtos científicos de alta qualidade.

a.3) Cópia & Segurança (*Backup e Secure*): guardar em segurança é o terceiro elemento transversal. Envolve o gerenciamento físico dos riscos do ciclo de vida dos dados ao mesmo tempo que garante a acessibilidade dos mesmos. Este elemento informa aos cientistas que as rotinas de cópias de segurança são críticas para prevenção de dados, devido a falhas nos *hardware* e *software*, desastres naturais ou erros humanos. Para a USGS os “dados representam ativos corporativos com valor potencial para além de qualquer uso em

investigação imediata e, portanto, precisam ser considerados e devidamente gerenciados durante todo o seu ciclo de vida” (USGS, 2015, *on-line*, tradução nossa).

Há medidas preventivas para evitar perdas, com a cópia do plano de gerenciamento de dados, estratégia de aquisição de dados, processamento de procedimentos, versionamento³, métodos de análises, principalmente quando o trabalho dos cientistas dos projetos tem múltiplas instalações ou comodidades (USGS, 2015). Observa-se que, na maioria dos estágios, o pesquisador possui papeis e responsabilidades que asseguram a validade e especificidade dos dados que são atributos inerentes do pesquisador.

Por fim, o modelo também serve como uma estrutura para ajudar a organização, avaliar e melhorar as políticas e práticas para o gerenciamento de dados científicos e identificar as áreas que necessitam desenvolver novos instrumentos e padrões. Por apresentar uma completude em seu arcabouço conceitual, este modelo será inspirador para a construção do modelo conceitual proposto no que tange aos aspectos autorais.

4.3.3 Modelo *Curation Lifecycle* do *Digital Curation Centre* (DCC)

O *Curation Lifecycle Model* é um modelo de ciclo de vida e curadoria desenvolvido pelo *Digital Curation Centre* (DCC) do Reino Unido. O DCC é um centro líder mundial de *expertise* em informação digital de curadoria com um foco na construção de capacidade e habilidades para a gestão de dados de pesquisa em toda a maior comunidade de pesquisas da educação do Reino Unido.

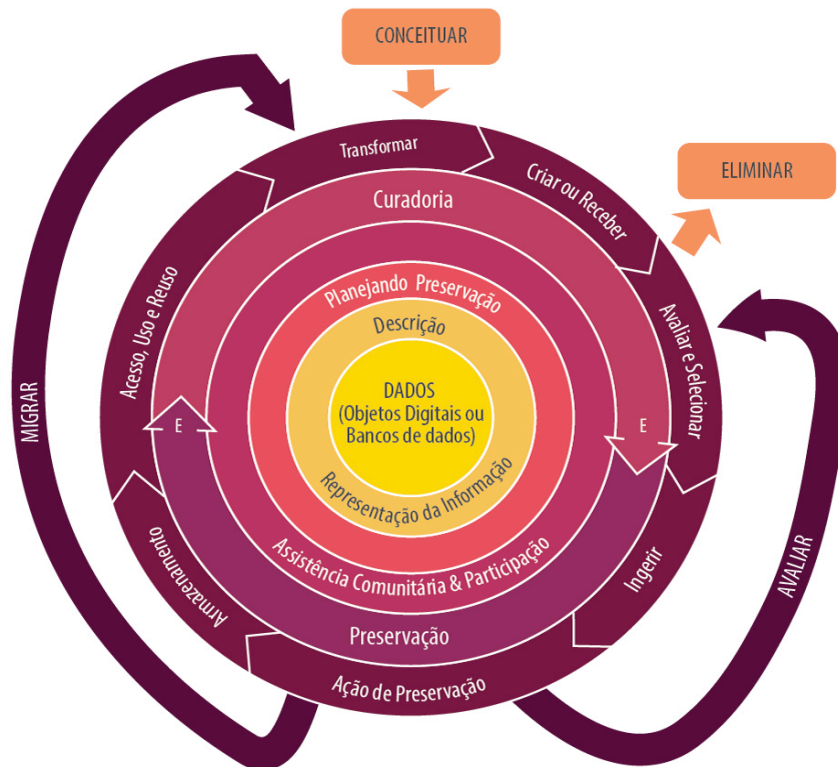
O modelo fornece uma visão geral de alto nível gráfico das etapas necessárias para a curadoria e preservação de dados de conceituação inicial ou recebimento através do ciclo de curadoria iterativo. Apresenta elementos-chave para o modelo de ciclo e curadoria baseado nos dados, prevendo as ações completas, sequenciais e ocasionais. Os dados são definidos no modelo como objetos digitais ou bancos de dados (*databases*). Os objetos digitais podem ser simples ou complexos em formato digital binário.

Objetos digitais simples compreendem itens digitais discretos, tais como: arquivos de texto, arquivos de imagem ou arquivos de som, juntamente com os seus identificadores relacionados e metadados. Os objetos digitais complexos são itens digitais discretos feitos através da combinação de uma série de outros objetos digitais, por exemplo, sítios web. Os bancos de dados incluem coleções estruturadas de registros ou dados armazenados em um

³Tradução de *Versioning*: Versionamento. Controle de versões.

sistema de computador. Todas as fases que representam o modelo *DCC Curation Lifecycle* estão na Figura 16.

Figura 16 – Modelo de ciclo de vida e curadoria



Fonte: Traduzido do *Curation Lifecycle Model* (2012).

O modelo em formato espiral e eixo central é constituído pelos dados que são objetos digitais (simples e complexos) ou armazenados em banco de dados (DCC, 2012), conforme o entendimento a seguir:

a) A descrição (*Description*) e a representação da informação (*Representation information*) é um estágio voltado para a atribuição administrativa, descrição, técnica, metadados estruturais e de preservação. Deve adotar padrões que garanta a descrição e controle por longo prazo, coletando e atribuindo a representação das informações para a compreensão e processamento do material digital e dos metadados associados;

b) Planejando a preservação (*Preservation Planning*) é o estágio relacionado ao plano de preservação da curadoria durante o ciclo de vida do material (dado) digital. Isso pode incluir planos para o direcionamento das ações de ciclo de vida de curadoria;

c) Assistência comunitária (*Community Watch*) e Participação (*Participation*) é o estágio relacionado com a manutenção e prazo das atividades comunitárias apropriadas e a participação do desenvolvimento de normas comuns, ferramentas e *software* adequados;

d) Curadoria (*Curate*) e Preservação (*Preserve*) é o estágio de realização da gestão e ações administrativas planejadas para promover a curadoria e preservação ao longo do ciclo de vida. As demais questões-chave apresentadas no modelo estão relacionadas com as ações ocasionais de dispor, avaliar e migrar os dados;

e) O eixo externo é norteador pelas funções conceituar e eliminar, envolve as ações de criar ou receber, avaliar e selecionar, ingerir, preservar e armazenar. Este eixo norteia o projeto de curadoria e preservação para que o acesso, uso e reuso. Além de gerar possibilidades de migração, avaliação e eliminação.

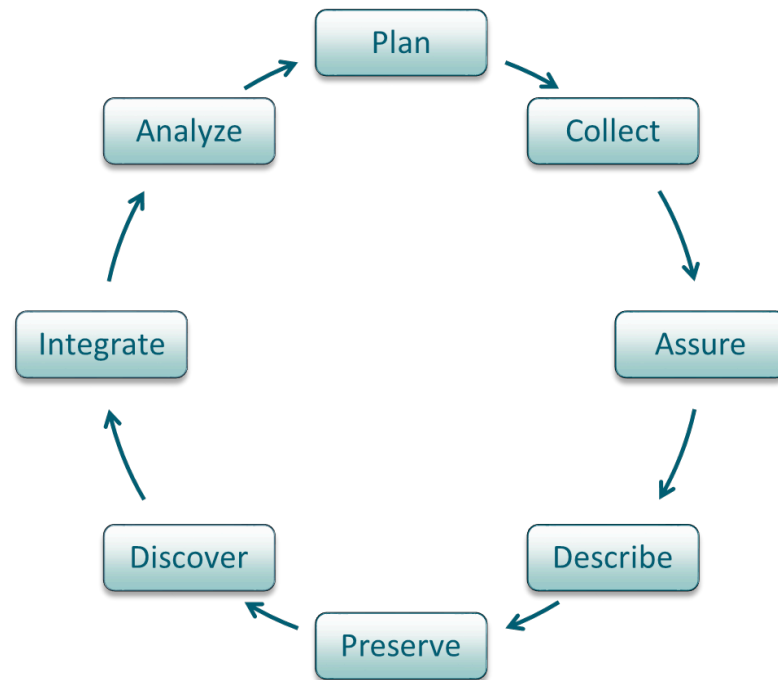
É um modelo amplo e complexo que envolve ações de todo o ciclo de vida e de curadoria dos dados para longo prazo. Oferece ferramentas, repositórios, publicações e serviços de apoio para pesquisadores, acadêmicos e instituições que desejam prover dados de pesquisa.

4.3.4 Modelo *Data Lifecycle* do DataONE

O *Data Lifecycle Model* foi desenvolvido pelo *Data Observation Network for Earth* (DataONE) no âmbito da *University of Tennessee* e outras instituições de pesquisa dos Estados Unidos e colaboradores em outros países como a Austrália, Canadá e Brasil. O DataONE é uma rede de observação de dados para a Terra que tem como fundamento a nova ciência ambiental inovadora. A estrutura é distribuída numa ciberinfraestrutura sustentável e atende as necessidades da ciência e da sociedade, em prol do acesso aberto, persistente, robusto e seguro dos dados científicos de observação da Terra (DATAONE, 2015).

O modelo é constituído por oito estágios que atende todo o ciclo de vida dos dados. Este modelo tem semelhanças com o modelo de ciclo de informação científica citado em capítulo anterior, que trabalha com 11 (onze) funções que se retroalimentam; entretanto, o modelo ora exposto apresenta diferenças, por tratar de forma descritiva e detalhada os dados de pesquisa primários, iniciando com o estágio de planejamento até o de análise.

Inicialmente, propõe-se uma descrição sucinta de cada estágio, a partir da visualização do *Data Life Cycle Model* na Figura 17.

Figura 17 – Modelo de ciclo de vida dos dados

Fonte: DataONE (2015).

Para que toda a cadeia do ciclo de vida dos dados ocorra de forma eficaz e dentro do planejado, o compartilhamento efetivo de dados requer uma cuidadosa reflexão durante cada estágio do processo de desenvolvimento de dados, incluindo:

- Descrição e documentação do processo, conteúdo e características dos dados;
- Depósito e armazenamento dos dados em um local a partir do qual podem ser acessados ou compartilhados, prevê ainda a preservação dos dados usando um formato e mídia que possibilitem o reuso em longo prazo;
- A descoberta dos dados é feita pela divulgação de informações sobre os dados em publicações de investigação, câmaras de compensação de dados e por portais de distribuição de dados (DATAONE, 2012).

O modelo de ciclo de vida do DataONE é chamado de *Data Life Cycle Model*, representado por oito estágios que constitui todo o processo de gerenciamento de dados de pesquisa. Para efeito desta pesquisa o modelo de comunicação científica foi adotado para ilustrar os estágios que são sistematizados em práticas científicas com dados na *e-Science*.

A Figura 18 representa este modelo como uma adaptação da ser adotada do original e aplicada aos dados científicos.

Figura 18 – Modelo de ciclo de vida dos dados



Fonte: Adaptado do modelo DataONE (2015).

a) Estágio Planejar (*Plan*) está relacionado com o plano de pesquisa que será desenvolvido pelo pesquisador que optou pela metodologia baseada no modelo *Data Life Cycle* para o gerenciamento dos seus dados de pesquisa. Para Wiggins e outros (2013, p. 3), “é um processo iterativo do planejamento do projeto, durante o qual todos os aspectos do gerenciamento de dados são analisados e as decisões são tomadas para documentação e implementação em fases posteriores”. Todas as etapas e procedimentos metodológicos sobre o projeto de pesquisa devem ser pensados neste estágio.

b) Estágio Coletar (*Collect*) é direcionado ao processo de coleta dos dados primários, os procedimentos e processamentos em banco de dados, *software*, laboratórios ou repositórios. Geralmente, os dados primários são oriundos de pesquisa de campo (observacionais, experimentais, simulacionais, documentais) desenvolvida pelo pesquisador ou grupo de pesquisa. Para o estágio coletar deve-se “determinar a melhor maneira de obter informações dos participantes em um arquivo de dados utilizável. O resultado final deste processo de decisão é um modelo de dados que descreve a forma como os dados são estruturados” (WIGGINS et al., 2013, p. 3, tradução nossa). É um estágio que apoia toda a coleta e estruturação dos dados desde a sua origem para dar continuidade aos demais.

c) Estágio Assegurar (*Assure*) está relacionado com os procedimentos de garantia e qualidade desenvolvidos pelos pesquisadores que apoiam o processo de gerenciamento. Nesse estágio é necessário adotar “procedimentos e controle que melhoram a qualidade dos dados

(por exemplo, participantes do treinamento, calibração de instrumentos de rotina) e identificar os erros e as técnicas potenciais para abordá-las” (WIGGINS et al., 2013, p. 3, tradução nossa). É nesse estágio que a adoção de critérios e procedimentos que assegurem o gerenciamento com qualidade e a validação dos dados por longo tempo.

d) Estágio Descrever (*Describe*) contempla a adoção de um padrão de metadados que prover a descrição minuciosa dos dados de pesquisa. É uma fase que requer a adoção de padrões, protocolos e repositórios visando o processamento, armazenagem, curadoria, preservação e disponibilidade dos dados em curto e longo prazos. É considerado um dos estágios mais relevantes. A qualidade na descrição dos metadados e a adoção de padrões e protocolos irão contribuir no compartilhamento dos dados para futuro uso e reuso.

e) Estágio Preservar (*Preserve*) deve ser adotado o plano de preservação dos dados por curto e longo prazos. Deve-se analisar que perdas de dados são possíveis por diferentes motivos. A preservação dos dados em curto prazo tem a finalidade de estocar primariamente e preservar a coleta primária. A preservação de dados em longo prazo tem a finalidade de proporcionar a reutilização dos dados no futuro para outros pesquisadores, parceiros e instituições de pesquisa. A preservação de dados inclui “*backups* e arquivamento assim como processos, tais como: conversão, reformatação e salvamento de dados” (DATAONE, 2015, *on-line*, tradução nossa). Esta infraestrutura está relacionada com as adoções de tecnologias e procedimentos que apoiem a preservação ao longo prazo.

f) Estágio Descobrir (*Discover*) está relacionado com a identificação de outros *datasets* e repositórios que podem complementar e agregar valor ao projeto de dados. A importância dos dados e seus metadados estarem disponíveis para serem localizados e potencialmente capazes de serem usados são abordagens que devem ser adotadas neste estágio. Algumas estratégias e parcerias institucionais podem garantir que os dados tenham uma maior visibilidade à pesquisa e proporcionem o compartilhamento e a colaboração científica. A descoberta de dados tem duas faces: “a primeira é encontrar dados existentes para análise em conjunto com outras fontes de informação; o segundo é tornar as informações sobre os dados disponíveis para que outros possam descobrir e acessá-los” (WIGGINS et al., 2013, p. 11, tradução nossa). Este estágio proporciona a visibilidade do projeto e do seu potencial para uso mais amplo e benefício para a investigação científica, de apoio à decisão e desenvolvimento de políticas. O registro do projeto de dados em diretórios e outros repositórios abertos são alternativas que podem ser realizadas para aperfeiçoar o projeto de dados.

g) Estágio Integrar (*Integrate*) está relacionado com a integração do projeto de dados a outros recursos de dados que possam ser capazes de novas análises e investigações. O estágio *Integrate* tem a finalidade de agregar o projeto de dados com dois propósitos: o primeiro é a agregação interna dos diferentes dados ou conjunto de dados gerados na pesquisa científica em um mesmo ambiente; o segundo propósito é o projeto de dados ser integrado com outras fontes externas.

h) Estágio Analisar (*Analyze*) está relacionado com a análise do projeto de dados por pesquisadores, comunidade científica, parceiros e profissionais. Existem ferramentas e *software* que suportam a exploração, análise e visualização. Este é o estágio em que os dados são analisados e se relacionam com a comunidade científica. Todo o processo de pesquisa já está completo e o pesquisador pode fazer derivações dos seus dados através de publicações em jornais, repositórios, base de dados e câmaras de compensação de dados e portais de dados. Proporciona que os dados de pesquisa já possam ser utilizados por outros pesquisadores para uso e reuso.

O projeto DataONE orienta pesquisadores e profissionais envolvidos com o plano de gerenciamento de dados visando a adoção de melhores decisões e práticas. O DataONE define também que com a execução de todo o ciclo seja adotada a aplicação de oito identificadores que possuem atributos vinculados aos Planos de Integração de Dados (PIDs) e representam as suas características. Os atributos de exclusividade, autoridade, opacidade, imutabilidade, resolubilidade, granularidade, estrutura e serialidade são necessários para que os dados de pesquisa compartilhados na ambiência do DataONE atendam o princípio de uma ciência inovadora, robusta e persistente.

O modelo *Data Lifecycle* do DataONE foi utilizado como exemplificação para o gerenciamento dos dados científicos. Entende-se que os modelos de ciclo de vida dos dados, independente de sua composição, proporcionam uma sistemática metodológica, técnica e tecnológicas para o processo de tratamento, gerenciamento e curadoria a longo prazo. Critérios de conformidade internacional devem ser adotados com padrões, protocolos, modelo, procedimentos e tecnologias que favorecem a eficiência e eficácia do compartilhamento e colaboração científicas, com vistas ao acesso, uso e reuso futuros.

5 MODELAGEM TECNOLÓGICA PARA O GERENCIAMENTO DE DADOS CIENTÍFICOS

A infraestrutura tecnológica apoia o gerenciamento de dados científicos dentro da sistemática do ciclo de vida. Essa infraestrutura se constitui em soluções tecnológicas providas por ferramentas, recursos, identificadores, padrões de metadados, protocolos de interoperabilidade, *software* e repositórios que constituem o arcabouço sustentável para as práticas científicas da *e-Science*. O capítulo foi pensado em uma dimensão morfológica com vistas a dar forma e explicitar em cada estágio do ciclo de vida, as funcionalidades tecnológicas que dão sustentabilidade ao processo de gerenciamento dos dados.

5.1 INFRAESTRUTURA BASEADA NO MODELO DE CICLO DE VIDA DOS DADOS

As novas práticas de publicação científica já são discutidas há certo tempo desde o advento do acesso aberto. Vários estudos se concentraram na possibilidade de convergência e entrelaçamento de produtos de *e-pesquisa*, *e-Science*, *open data* e *open science* na comunicação científica. Em todas as áreas do conhecimento mudanças estão ocorrendo de forma globalizada e impulsionadas pelo uso das Tecnologias de Informação e Comunicação (TICs) e *internet*. Na literatura internacional a noção de infraestrutura sustentável se amplia para o termo *cyberInfrastructure* para prover a ciência aberta. Por uma conveniência conceitual o referido termo será utilizado na língua vernácula.

Os estudos de Atkins e outros (2003), Tenopir e outros (2011), Kowalczyk e Shankar (2013) e Borgman (2014) afirmam que: a ciberinfraestrutura abrange *hardware*, *software*, formatos de dados e protocolos, que precisa ser construída, implementada e avaliada para a utilização de dados como produtos de pesquisa. Tenopir e outros (2011, p. 1, tradução nossa) afirmam que “a quantidade de dados coletados, analisados, reanalisados e armazenados têm aumentado enormemente devido à evolução na simulação computacional e modelagem, aquisição de dados automatizados, e tecnologias de comunicação”. É através da ciberinfraestrutura que essa enorme quantidade de dados são estruturados dentro da sistemática do ciclo de vida. A sistematização promove também as possibilidades de reuso e reprodutibilidade dos dados coletados, geridos, analisados e armazenados.

O estudo de Atkins e outros (2003) aponta para uma ciberinfraestrutura constituída em duas camadas: a camada subjacente é composta pelos componentes integrados *eletro-ópticos* da computação que favorecem a armazenagem e comunicação integradas; e a camada superior

é composta por programas de *software*, serviços, instrumentos, dados, informações, conhecimentos e práticas sociais aplicáveis aos projetos específicos, treinamento e comunidades de pesquisa. Entre as duas camadas a ciberinfraestrutura integra os componentes internos - algoritmos, *hardware*, *software*, e o ambiente externo - comunicação científica, pessoas e instituições envolvidas – no processo de pesquisa. As camadas que constituem a ciberinfraestrutura “deve fornecer uma plataforma eficaz e eficiente para o empoderamento das comunidades específicas de pesquisadores e, eventualmente, inovar e revolucionar o que eles fazem, como eles fazem, e quem participa” (ATKINS et al., 2003, p. 5, tradução nossa).

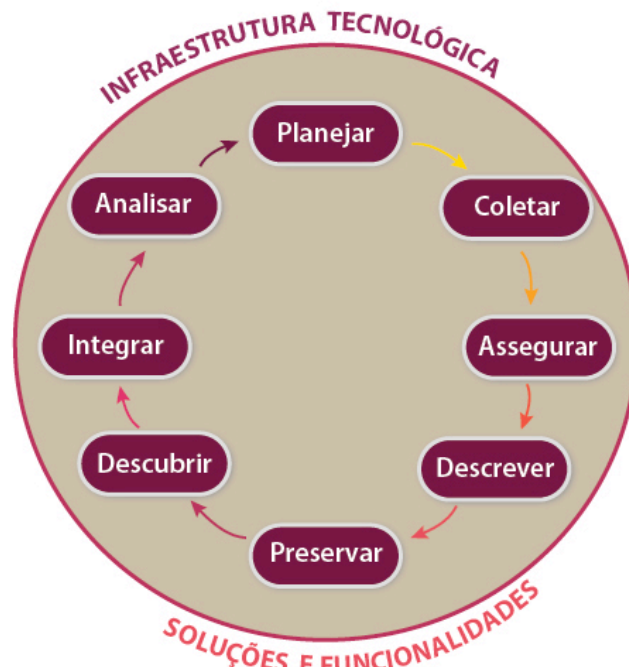
Dessa forma, além da plataforma que sustenta a colaboração e o empoderamento das comunidades envolvidas, a sistemática do processo de gerenciamento do ciclo de vida que dar sustentabilidade à ciberinfraestrutura eficiente está alicerçada em três aspectos intrínsecos:

- a) Os aspectos tecnológicos envolve a extensa e complexa ciberinfraestrutura que dar sustentabilidade ao agrupamento de plataformas abertas, aplicações, padrões de metadados, protocolos de interoperabilidade, identificadores persistentes, linguagens compatíveis *web* e repositórios especializados para tornar dados científicos acessíveis, persistentes, confiáveis, detectáveis e reusáveis. Na *e-Science* para uma ciberinfraestrutura tecnológica sustentável é recomendado que tecnologias e modelos sejam abertos e de código livre.
- b) Os aspectos técnicos estão relacionados aos procedimentos, normas, padrões, vocabulários controlados e outros recursos que proporcionem a qualidade e conformidade internacional para que os dados científicos sejam detectados e interoperados. Reforça-se a qualidade na descrição dos metadados e recursos automáticos para a completa identificação através dos mecanismos de atribuição, citação e o uso de identificadores persistentes e licenças públicas. Para a infraestrutura técnica os autores sugerem adotar três critérios mínimos para o compartilhamento de dados: persistência, longevidade e qualidade. Os dados devem ser confiáveis e acessíveis a partir de um armazenamento seguro e longínquo. São previstos critérios que contemplem alta tecnologia, recursos e computação intensiva para lidar com um grande volume de dados (estruturados e desestruturados) possibilitando o processo de gestão dos dados e para que a ciberinfraestrutura tecnológica apoie todo o ciclo de vida.
- c) Os aspectos metodológicos contemplam os procedimentos científicos. A ciberinfraestrutura suporta a colaboração científica e as diferentes formas de publicação - fontes de dados, informações, pesquisas, *links*, resultados, mídias e

simulações - com os públicos (pesquisador, colaborador, produtor, editor e usuário), possibilitando a convergência, interatividade e o compartilhamento. Além da colaboração e compartilhamento científico os aspectos metodológicos dão sustentabilidade ao processo de pesquisa desde o estágio de planejamento, com a contemplação dos métodos científicos e procedimentos de pesquisa que estão associadas ao domínio específico do conhecimento.

Como exemplificação de uma ciberinfraestrutura modelada para uma pesquisa orientada aos dados científicos como produtos primários, utilizou-se como representação o modelo de ciclo de vida do DataONE para a explicitação de algumas soluções e funcionalidades necessárias em cada estágio do ciclo de vida. A utilização de um modelo de ciclo de vida para o gerenciamento de dados é mandatória. O modelo contempla os aspectos que sustentam a ciberinfraestrutura e tem a finalidade de garantir um padrão internacional com a qualidade e a conformidade que sustentem o compartilhamento, a colaboração e a gestão de dados. Para exemplificar os estágios do ciclo, adotou-se o modelo do DataONE para representação conforme a Figura 19.

Figura 19 - Modelo para infraestrutura tecnológica



Fonte: Adaptado do modelo DataONE (2015).

O modelo *Data Lifecycle* do DataONE é constituído por oito estágios que sistematiza as ações: Planejar (*Plan*); Coletar (*Collect*); Assegurar (*Assure*); Descrever (*Describe*); Preservar (*Preserve*); Descobrir (*Discover*); Integrar (*Integrate*); e, Analisar (*Analyze*). A

seguir, os estágios são descritos com algumas exemplificações de soluções tecnológicas e técnicas.

5.1.1 Primeiro estágio do ciclo: Planejar (*Plan*)

O estágio planejar está relacionado ao planejamento e descreve o mapeamento dos processos metodológicos e técnicos e a ciberinfraestrutura tecnológica necessária para a execução de todos os estágios subsequentes do ciclo de vida. Recomenda-se como uma boa prática adotar uma ferramenta de planejamento para detalhar o projeto de dados. Para apoiar a execução deste estágio, é recomendado o uso de ferramentas tecnológicas para ajudar o pesquisador a desenvolver seu plano com base em orientações, regras estabelecidas e uma metodologia que norteie o processo investigativo.

O *Data Management Plan Tool* (DMPTool) é uma ferramenta aberta e livre que pode ser utilizada para o cumprimento dessa fase e fornece guia de boas práticas. O intuito dessa ferramenta é favorecer a gestão de dados eficiente e proporcionar que os dados sejam usados, compartilhados e reutilizados, bem como que ela seja utilizada como uma plataforma de pré-teste e *open-source* para a interoperabilidade dos dados com sistemas de informação, conjunto de dados e repositórios.

Lançado em 2011 e atualizado em 2014, o *Data Management Plan Tool* (DMPTool) tem uma forte utilização em agências de fomento, instituições de ensino e científica, bibliotecas e pesquisadores em geral que gerenciam seus dados por meio desta ferramenta. É um *open-source*, aplicativo gratuito, *on-line* que auxilia na criação dos planos de gestão de dados. Além de ser uma ferramenta é ao mesmo tempo uma metodologia recomendada nos modelos de ciclo de vida dos dados e que também ajuda ao pesquisador ou ao grupo de pesquisa no planejamento e execução do projeto. É baseada em cinco componentes: informações sobre os dados e formato de dados, conteúdo de metadados e formatos, políticas para acesso, compartilhamento e reuso, armazenagem em longo prazo e gerenciamento, e orçamento. Estes planos ou DMPs fazem parte de uma política mandatória adotada pelas agências de financiamento dos Estados Unidos e de outros países, como parte do processo de apresentação de propostas de financiamentos de pesquisa.

O estágio planejar se constituiu adotar a ferramenta, instalar, customizar, mapear processos e recursos para a execução do plano de gestão baseado nas fases propostas pelo DMP. Pensar no projeto de gestão é idealizar todo o processo e “começar com os objetivos do projeto (saídas desejadas, resultados e impactos) e trabalhar previamente, construindo um

plano de gestão de dados, apoiado nas políticas de dados e planos de sustentabilidade” (WIGGINS et al., 2013, p. 3, tradução nossa).

Recentemente, a ferramenta *Data Management Plan online* (DMPonline) foi atualizada em 2015, na versão 4.2, pelo *Digital Curation Centre* (DCC) do Reino Unido. O objetivo do DCC é ajudar as equipes de pesquisa a responderem a essa exigência das agências de fomento e quaisquer outras expectativas institucionais. A versão atual do DMPonline inclui marca institucional, orientação opcional para financiadores, melhorias na área de perguntas, melhorias na interface de administração, solicitações de extração do *GitHub*, novo arquivo para dados primários e uma atualização dos arquivos de armazenagem (DDC, 2016).

Além da ferramenta de gestão, recomenda-se o uso de um identificador persistente para cada pesquisador envolvido no projeto. Essa recomendação ajuda a garantir que as atribuições de autoria sejam agregadas aos dados de pesquisa. Para a implantação do DMPTool, o pesquisador é orientado a gerar um identificador *Open researcher and Contributor Identification* (ORCID) para a conexão com a sua pesquisa. O identificador ORCID é composto por 16 (dezesesseis) dígitos vinculados ao identificador (ID) do pesquisador e ao e-mail. Oferece um ID ORCID completo e o *link* para o seu registro público cingido à pesquisa e ao pesquisador. A ferramenta é de fácil utilização, gratuita, *on-line* e pode ser realizada por pesquisadores, profissionais e acadêmicos sem restrições e custos. A importância de se adotar estas ferramentas são recomendáveis para o prosseguimento dos próximos estágios do ciclo.

5.1.2 Segundo estágio do ciclo: Coletar (*Collect*)

Este estágio está relacionado com a coleta primária dos dados, apoiando todo o processo do ciclo desde a sua origem. O estágio se apoia na infraestrutura tecnológica que seja capaz de suportar diferentes tipos de dados (sensíveis, sigilosos, estáticos ou dinâmicos), fatores intrínsecos e extrínsecos relacionados com a natureza e com as condições de coleta dos dados. As condições (física, temporal e acondicionamento) que os dados foram coletados e armazenados (preliminarmente), requer atenção especial do pesquisador e dos colaboradores envolvidos no processo de coleta e curadoria.

Além da coleta, o pesquisador deve se preocupar com a curadoria preliminar. Para apoiar esta curadoria, é recomendado ter uma infraestrutura inicial com ferramentas de comunicação geral, a exemplo do *dropbox*, *google drive*, *software* específico ou equipamentos laboratoriais que condicionem e assegurem a proteção dos dados primários. O

pesquisador deve observar os aspectos relacionados com o volume, a extensão, a mídia dos dados, como também o *hardware*, *software* e outros recursos humanos que estejam envolvidos no processo. O volume estimado, por exemplo, para os dados pode variar de *megabyte* (MB), *gigabyte* (GB), *terabyte* (TB) e até *pentabyte* (PB). Esta previsão deve ser prevista na fase em que os dados estão sendo coletados e armazenados preliminarmente.

Para a interface textual, quando necessário, recomenda-se a adoção de formatos legíveis por máquina, tais como: HTML, XML, RDF e PDF-A que também são úteis para manipular textos e documentos não estruturados, baseados em padrões abertos. O formato HTML é uma linguagem de marcação que auxilia na vinculação dos dados, conjuntos de dados (diferentes formatos, mídias e extensões) e documentos utilizando *hyperlinks* e *links* como âncoras para a vinculação das publicações, que podem ser dados, artigos, documentos ou outros suportes digitais. O formato XML (*Extensible Markup Language*) é outra linguagem utilizada para a criação de documentos com dados organizados de forma hierárquica, que geram linguagens de marcação para necessidades especiais com a finalidade de compartilhar informações pela Internet. Tais linguagens favorecem a vinculação de dados ou conjuntos de dados e documentos utilizando *hyperlinks* e *links* como âncoras para a vinculação das entidades.

Os arquivos *Comma-separated Values* (CSV) é outro padrão aberto aplicado para o formato textual que faz uma ordenação de *bytes* ou um formato de terminador de linha. Isso permite que os dados possam ser manipulados por diferentes ferramentas, perdendo o contexto de limitação temporal pela obsolescência de *software* de acesso aos dados. Ele comumente é usado em *software* oficial, tais como: *Microsoft Excel* e *LibreOffice Calc*.

O padrão *Moving Picture Experts Group* (MPEG-4) é utilizado primeiramente para compressão de dados digitais de áudio e vídeo. O acrônimo *Tagged Image File Format* (TIFF) é um formato de arquivo para imagens digitais de arquivos gráficos com elevada definição de cores.

O *Joint Photographic Experts Group* (JPEG) conhecido como JPEG2000 é um padrão aberto para compressão de imagens de alta definição, facilitando a aplicação de diferentes ferramentas computacionais para a manipulação de imagens. Outras referências para padrões sobre recomendações de formato de dados pode ser consultada nas orientações⁴.

⁴ Saiba mais: <http://libraries.mit.edu/guides/subjects/data-management/formats.html>.

Aspectos relevantes são observados quanto aos procedimentos especiais, tais como: licenças, permissões, equipamentos, padrões, protocolos e repositórios que serão atendidos nos próximos estágios.

5.1.3 Terceiro estágio do ciclo: Assegurar (*Assure*)

Realizados os procedimentos preliminares de qualidade, validação e curadoria, ocorre a migração dos dados científicos para uma ferramenta compatível que favoreça o compartilhamento, curadoria e preservação a longo prazo em uma única interface. Relaciona-se com os procedimentos de garantia, qualidade, validação e atribuição autoral preliminar dos envolvidos no projeto.

Estimula-se neste estágio a adoção de identificadores para os dados científicos, a exemplo do *Digital Object Identifier* (DOI). O *Digital Object Identifier* é um identificador persistente que agrega os dados (digitais), objetos (digitais) e documentos relacionados numa mesma pesquisa. É um sistema que além de assegurar a qualidade e identificação, proporciona gerenciar os direitos de propriedade intelectual de dados e objetos digitais. Estabelece a ordem de atribuição realizada para as citações de pesquisador(es), colaborador(es), instituições, agências de fomento e parceiros externos.

No que tange aos aspectos autorais, é relevante adotar procedimentos legais e levantamento de licenças públicas antes de avançar para o estágio de descrição dos metadados. O caráter público do *commons*, *semion*, *open access* e *open science* impulsionaram as mudanças nos instrumentos de propriedade intelectual consolidados. O conceito *open* é da ciência 2.0, inspirado na *Web 2.0* que se baseou numa nova abordagem da ciência baseada no compartilhamento e colaboração em rede. A palavra *open* tem sido cada vez mais utilizada para os “modelos proprietários – com restrições para cópia, distribuição e reusos da informação. As expressões *open science*, *open access*, *open research*, *open education*, *open contents*, *open source*, *open notebook*, *open licenses*, *open courses* etc.” (MACHADO, 2015, p. 210) e são utilizadas atualmente para definir as diferentes iniciativas, novos negócios e modelos contemporâneos científicos.

A definição de procedimentos legais, éticos, normativos e critérios para as condições e restrições de uso e o *status* (aberto, fechado ou embargado) dos dados é recomendado através de diretrizes e políticas que antecedam ao processo inicial de gestão. O Apêndice B - Licenças Livres, Abertas e Públicas complementa estas discussões e apresenta as licenças públicas que estão sendo adotadas no âmbito da ciência aberta. Nesse estágio todos os procedimentos

deverão ser conduzidos com vistas a garantir os procedimentos que promovam a autoria de todos os colaboradores envolvidos na pesquisa.

Além dos procedimentos já descritos, a infraestrutura de *software* ou repositório que sustente os estágios futuros de compartilhamento, curadoria e preservação a longo prazo já é contemplada preliminarmente no estágio assegurar. É também recomendada a implantação de uma plataforma gerenciável de código aberto com possibilidades de agregação, aplicações, configuração de metadados, protocolos de interoperabilidade e ferramentas *web*.

5.1.4 Quarto estágio do ciclo: Descrever (*Describe*)

Esse estágio é de extrema relevância e a escolha de um padrão de metadados adequado para descrição dos dados de pesquisa é essencial. A qualidade na descrição dos metadados e a adoção de padrões irão contribuir para o futuro compartilhamento dos dados para uso e reuso. Inúmeros padrões de metadados específicos foram desenvolvidos para diferentes tipos de dados científicos.

O padrão de metadados define os detalhes e a sistemática de descrição de determinados elementos de dados ou conjunto de dados contidos em um recurso. As “informações sobre dados (metadados) são registradas para que outros possam descobrir, adquirir, interpretar e utilizar os dados de forma eficaz” (WIGGINS et al., 2013, p. 7, tradução nossa). Existem padrões de metadados específicos para diferentes áreas que proporcionam uma descrição minuciosa e compatível com a natureza da pesquisa dos dados primários. A *National Information Standards Organization* (NISO) define que “metadados são informações estruturadas que descrevem, explicam, localizam, ou permitem que um recurso informacional possa ser fácil de recuperar, usar ou gerenciar” (NISO, 2004, *on-line*, tradução nossa). Tem a função de fornecer, descrever e identificar os detalhes sobre um determinado recurso. Esta informação adicional irá proporcionar a oportunidade para a compreensão e o compartilhamento de dados e permite sua integração, uso e reuso.

Os metadados são dados que refletem os elementos que envolvem um processo de investigação e o padrão colabora na descrição sistemática dos elementos. A complexidade da estruturação de dados contidos em um mesmo recurso ou em diferentes conjuntos de dados requer a adoção de padrões, recursos auxiliares, técnicas e vocabulários controlados que apoiem a descrição dos metadados. Descrever os metadados a partir da adoção de um padrão propicia que os metadados se tornem enriquecidos e legíveis por computador e acessíveis pelo usuário final. Um padrão de metadados é usado para realizar a representação adequada e

padronizada de recursos armazenados de modo que eles possam ser subsequentemente permutados e interoperáveis. Os padrões de metadados podem ser divididos em três níveis, conforme a Quadro 6.

Quadro 6 – Níveis dos padrões de metadados

Nível 1	São metadados com padrões simples e consiste de metadados desestruturados que são, em alguns casos, extraídos automaticamente por robôs automatizados. Na maioria dos casos possuem uma redução semântica.
Nível 2	São padrões de metadados estruturados que consistem e definem os metadados em um esquema. Eles são baseados em padrões que fornecem uma descrição mínima do recurso para a identificação, localização e valorização.
Nível 3	São os padrões de metadados ricos, considerados altamente estruturados. Eles são constituídos por metadados complexos e apresentam uma descrição mais detalhada formal e de estrutura. Eles são baseados em normas e códigos de um domínio específico especializado, permitindo a descrição de um recurso de informação individual ou pertencente a uma coleção. Facilitando a localização, recuperação e interoperabilidade de recursos de informação.

Fonte: Alves e outros (2014, p. 155).

Todos os níveis de padrões de metadados realizam uma descrição do recurso. A natureza do recurso e o grau de detalhamento que se precisa alcançar irão determinar o nível a ser adotado. Inicialmente, os padrões foram desenvolvidos para descrever os recursos impressos contidos em bibliotecas, museus, arquivos e centros de documentação. Com a evolução tecnológica, padrões foram sendo adaptados para a descrição de recursos digitais em formato eletrônico e aplicáveis em sistemas automatizados, bases de dados e gerenciadores de periódicos eletrônicos. Os padrões evoluíram para se adequarem à descrição de dados de pesquisa ou conjuntos de dados (*datasets*). Para a descrição dos metadados e a escolha de um padrão que seja compatível com a natureza do experimento deve seguir critérios que assegurem a qualidade da descrição e o preenchimento completo. As questões a partir das perguntas: *Why*, *Who*, *What*, *When*, *Where* e *How* foram adotadas. Estas orientações foram elaboradas a partir do <http://murray.harvard.edu/policies> de gerenciamento de dados para participação pública na pesquisa científica do DataONE e estão sintetizadas no Quadro 7.

Quadro 7 – Orientações para a descrição de metadados

PERGUNTAS	ORIENTAÇÕES
<i>Why</i> (Porquê)	Refere-se ao entendimento dos dados e o contexto que eles foram criados. Os detalhes devem incluir as questões científicas ou proposições para a coleção de dados, razões que envolvem contribuidores voluntários e uma visão geral sobre o conjunto de dados.
<i>Who</i> (Quem)	Trata-se da descrição dos metadados relacionados com quem coletou os dados ou quem patrocinou a coleta. Ele também deve especificar quem deverá ser contactado com perguntas sobre os dados e como deve ser citado o conjunto de dados para garantir que todos os envolvidos obtenham reconhecimento na pesquisa. Este é um espaço que deve descrever os metadados de todos os pesquisadores, colaboradores, agências e instituições que se relacionaram à pesquisa.

<i>What</i> (O quê)	Requer uma descrição apurada dos dados. Foram desenvolvidas algumas categorias que ajudam a nortear a descrição. Com base no <i>Data Management Guide for Public Participation in Scientific Research</i> , as categorias recomendam seis passos para a descrição desta questão, relacionadas ao contexto digital (nomes de arquivos, formatos de arquivos, modificações de datas, arquivos de exemplo, conjuntos de dados relacionados e procedimentos de processamento de dados); <u>Parâmetro/ detalhes variáveis</u> (unidades de medida, formatos de dados, precisão e exatidão); <u>Informações sobre dados</u> (taxonomias, codificação, procedimentos); <u>Conteúdo de arquivos de dados</u> (definições de parâmetros e explicação de seus formatos, avaliação de qualidade notas); <u>Informações taxonômicas adicionais</u> (usando taxonomias padrão quando possível); <u>Organização dos dados</u> (relacionamentos entre entidades de dados, arquivos, diretórios de banco de dados).
<i>When</i> (Quando)	Relaciona-se ao detalhamento dos aspectos temporais, como período (ano, mês, dia), tempo, regularidade e intervalos da coleta de campo.
<i>Where</i> (Onde)	Direciona-se ao detalhamento dos aspectos de cobertura geográfica e espacial que estão relacionadas com o país, a cidade, local, coordenadas espaciais (latitude e longitude) e condições meteorológicas (clima, temperatura).
<i>How</i> (Como)	Estabelece as condições de criação dos dados. Os detalhes são muito importantes para as possibilidades de reuso e interpretação futura dos dados. O detalhamento de protocolos, técnicas, métodos, diagramas, <i>software</i> , modelos, esquemas e estatísticas utilizadas nos procedimentos metodológicos e de coleta de campo que originou os dados são relevantes para uma descrição qualitativa dos metadados. Documentos adicionais aos dados também devem ser informados e se possível vinculados ao dado ou conjunto de dados.

Fonte: Elaborado pela autora (2015).

Importante salientar que obedecer aos procedimentos de qualidade, descrição, confidencialidade, fidedignidade dos metadados e escolher um padrão de metadados adequados são recomendações internacionais que irão assegurar a conformidade para a interoperabilidade, preservação, visualização, análise e acesso com possibilidades de uso e reuso no futuro. Para complementar este tópico, o Apêndice B – Padrões de Metadados com Ênfase no Domínio Específico, foi elaborado para exemplificar padrões adotados em áreas específicas - Ecologia, Biodiversidade, Química, Física, Genoma - da ciência. Por fim, este estágio é de grande relevância para a execução dos procedimentos normativos e legais que promovem a recompensa autoral dos colaboradores envolvidos.

5.1.5 Quinto estágio do ciclo: Preservar (*Preserve*)

A adoção de um plano de preservação para os dados primários a longo prazo é recomendada neste estágio. A preservação tem a finalidade de proporcionar *backup* e guarda segura dos arquivos digitais. A preservação utiliza recursos (manuais, eletrônicos e duplicados), ferramentas de armazenagem em nuvens, guarda em *software* de gerenciamento eletrônico, laboratoriais, estufas, *data centers* e repositórios. Realizar a preservação em repositório especializado permite a curadoria a longo prazo e viabiliza os processos de conversão, reformatação e salvamento dos dados para o uso, reuso e reprodutibilidade futuros.

Esse estágio conduz dois procedimentos importantes no gerenciamento dos dados:

- a) Primeiro está direcionado ao término da coleta dos dados primários. Identificar as tecnologias que podem suportar o *backup* a curto prazo, mas que garantam a preservação e guarda preliminar.
- b) Segundo direciona a armazenagem a longo prazo ressaltando a importância de adotar requerimentos específicos relacionados com a proveniência e tipologia dos dados, bem como licenças autorais e de patentes *on-line* e as versões do *software*.

Recomenda-se reforçar a instalação de licenças públicas abertas que preveem a revisão e aprovação de licenciamento público para diferentes contextos com atribuições para documentos, códigos, dados e não atribuições.

Na vertente do acesso aberto a Rede Brasileira de Serviços de Preservação Digital – CARINIANA promovida pelo IBICT é um excelente exemplo de curadoria e preservação digital. A rede é responsável pela preservação digital compartilhada da informação científica, tecnológica e cultural e “busca disponibilizar serviços de preservação digital para instituições com publicações de acesso livre, além de mecanismos que facilitem a automatização dos processos de identificação, armazenamento, validação e conversão para novos formatos digitais” (IBICT, 2015, *on-line*). Em 2015, a rede ampliou seu escopo para a preservação de dados de pesquisa com a integração ao Repositório *Dataverse*.

O *Dataverse* é um *software* de código aberto publicado pelo *Harvard Institute of Quantitative Social Science* da Universidade de Harvard, nos Estados Unidos. No caso do IBICT a rede *Dataverse* irá atuar como administradora de conteúdos digitais de repositórios nacionais. A rede está aberta para depósito de membros de instituições parceiras e para todos os pesquisadores e bibliotecários via *internet*. Pesquisadores podem customizar, gerenciar, descrever (metadados) e compartilhar seus dados de pesquisa (abertos ou privados) em condições de acesso personalizadas (IBICT, 2015).

A rede disponibiliza políticas com orientações da *Harvard Dataverse* para as condições de acesso e permissões de uso. A política orientada para aumentar a visibilidade dos dados da pesquisa é realizada por meio do reconhecimento e crédito acadêmico através da citação. As permissões e restrições adotadas para o compartilhamento e disponibilidade dos conteúdos e dados de pesquisa no repositório *Murray Archive* são gerenciados por licenças no *DataCite*.

A citação de dados no *Dataverse* é estabelecida com base nos princípios da *Joint Declaration of Data Citation Principles* (2014): crédito e atribuição; identificador persistente; e especificidade e verificabilidade. Para a descrição da citação no *Dataverse* são definidos sete componentes: nome autor, título, ano (data publicado em *Dataverse*), publicador (repositório que publicou o conjunto de dados), identificador persistente (DOI ou Handle),

número da versão e impressão digital numérica universal (Universal Numerical Fingerprint – UNF), adotados para dados tabulares.

Esse estágio não ocorre de forma simples e rápida. Algumas estratégias auxiliam na localização e aquisição de dados potencialmente úteis, tais como:

- a) identificar dados complementares que possam agregar valor aos dados do projeto;
- b) adotar os buscadores genéricos na web não são eficientes para encontrar dados úteis;
- c) os dados são mais facilmente encontrados por meio das redes, diretórios de projetos e repositórios específicos.

Muito embora algumas iniciativas desenvolvidas já apresentem soluções de estruturação, descrição de metadados e agregação de objetos por meio de protocolos de interoperabilidade, a atual infraestrutura de repositórios com base em OAI-PMH, XML, Dublin Core, DIDL ou DDI é capaz de lidar com os padrões mais complexos e ambientes abertos (VEERNOOY-GERRITSEN, 2009) e domínios específicos. O Apêndice C - Repositórios de Dados Digitais apresenta um levantamento de repositórios digitais com ênfase em domínios específicos. O potencial dos dados de pesquisa quando são preservados de forma adequada apresenta perspectivas no desenvolvimento e sustentabilidade econômica, política e social além de promover o progresso científico do país.

A complexidade entre os diferentes tipos de dados requer padrões e protocolos que possam suportar as possibilidades de estruturação, agregação, curadoria e interoperabilidade em uma estrutura complexa para integração e descoberta eficazes.

5.1.6 Sexto estágio do ciclo: Descobrir (*Discover*)

Após a garantia da curadoria e preservação a longo prazo, o estágio descobrir proporciona o completo compartilhamento, integração e visibilidade dos dados. A partir da integração com repositórios especializados é também possível acessar diretórios de dados, *datasets*, buscadores especializados ou meta buscadores para a descoberta de dados. A importância de tornar dados e metadados localizáveis, torna-os detectáveis, reutilizáveis e citáveis. Reforça-se a importância dos metadados estarem disponíveis para serem localizados e potencialmente capazes de serem usados e reproduzidos.

O *CiteBank* é um exemplo de repositório especializado de acesso aberto que faz a agregação de citações para artigos e publicações relacionados à biodiversidade. Ele fornece recursos de pesquisa e navegação para publicações armazenadas em vários repositórios internacionais e proporciona que pesquisadores compartilhem bibliografias especializadas.

O *Data Asset Framework* (DAF) fornece um conjunto de ferramentas para as organizações identificarem seus ativos digitais e avaliar como eles são gerenciados. É uma plataforma útil para instituições, departamentos ou grupos de pesquisa que começam a pensar em gerenciamento de dados e que precisam preparar um registro de seus ativos de dados.

O *DataONE Search* é uma plataforma que agrega Repositórios. Os participantes agregadores são denominados de nós membros. *DataONE Search* é uma implementação personalizada do aplicativo de pesquisa de código aberto *MetacatUI* (<https://github.com/NCEAS/metacatui>), desenvolvido no Centro Nacional de Análise e Síntese Ecológica, e foi lançado em novembro de 2015.

A *Global Change Master Directory* (GCMD) é um diretório que contém metadados de coleção de dados e dados armazenados em repositórios especializados nas Ciências da Terra e Ambiental em nível mundial. O site do GCMD oferece pesquisa facetada e pesquisa de conjuntos de dados por extensão espaciotemporal, resolução, centro de dados, localização, instrumento, plataforma e projeto.

Estimula-se a elaboração de parcerias institucionais para garantir que os dados tenham uma maior visibilidade da pesquisa e proporcione o compartilhamento e a colaboração científica. Algumas estratégias e parcerias institucionais podem garantir que os dados tenham uma maior visibilidade à pesquisa e proporcionem o compartilhamento e a colaboração científica.

5.1.7 Sétimo estágio do ciclo: Integrar (*Integrate*)

Os protocolos de interoperabilidade são responsáveis pela integração dos metadados com a finalidade de prover visibilidade, novas análises e investigações para o reuso e reprodutibilidade futuras. Podem-se adotar as ferramentas disponíveis pelo *Open Archival Information System* (OAIS) com o intuito de preservar e disponibilizar por meio de seus protocolos de interoperabilidade os dados, informações e documentos relacionados ao projeto de pesquisa.

O *Open Archives Initiative* (OAI) é o protocolo de interoperabilidade adotado para tornar eficiente a disseminação dos conteúdos contidos no projeto. E o *Open Archives Initiative Object Reuse and Exchange* (OAI-ORE) é responsável pela definição de padrões, recursos e especificações que designam e auxiliam na troca de agregações de recursos da *web*.

A agregação do projeto de dados possui dois propósitos:

1) agregação interna dos diferentes dados ou conjunto de dados gerados na pesquisa científica em uma mesma interface; 2) projeto de dados ser integrado com outras fontes externas. Este é um estágio que incentiva publicar e compartilhar combinando conceitos publicação tradicional avaliada por pares, com a distribuição de dados por meio de sites, catálogos de dados, mídia social e outros locais (USGS, 2013, p. 3, tradução nossa).

Estas agregações devem ser adotadas para combinar os objetos distribuídos em vários tipos de mídia e recursos. O *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) deve ser adotado como uma opção técnica simples para prover a possibilidade de metadados serem interoperáveis e disponíveis para serviços baseados em padrões abertos HTTP e XML.

No que diz respeito ao padrão de ligação dos objetos digitais, as relações não lineares de compartilhamento e interoperabilidade de dados são aspectos que devem ser considerados. Estas tendências têm reduzido a distância, permitindo aos cientistas compartilhar recursos remotamente, e aumentaram a complexidade e quantidade de dados de investigação (KOWALCZYK; SHANKAR, 2014).

A agregação dos dados ou conjunto de dados e o seu compartilhamento podem ser complexos, muito embora seja possível integrá-los em um repositório, *site* ou *software* que já esteja preparado para esta integração. A partir da publicação dos dados em uma plataforma compatível, já é possível atingir o segundo propósito que está relacionado com a interoperabilidade com outros repositórios especializados e diretórios. A integração e a disponibilidade do projeto propiciam o uso e reuso para a continuidade de outras pesquisas.

5.1.8 Oitavo estágio do ciclo: Analisar (*Analyze*)

Este estágio está voltado para a análise e avaliações que podem ser realizadas entre os dados de pesquisa por meio de *software*, câmaras de compensação de dados e ferramentas analíticas e métricas especializadas. Todo o processo de pesquisa já está completo e o pesquisador pode fazer derivações dos seus dados através de publicações em jornais, repositórios, *datasets* e portais de dados para duas performances:

- a) Utilizar os dados para análises que atendam os objetivos do projeto. Muitas ferramentas de *software* estão disponíveis para apoiar as atividades de exploração, análise e visualização dos dados. A *Data Analyze* são conduzidas por computadores, *grid* ou *cloud computing* que promovem as tarefas estatísticas, execução de modelos, estimativas de parâmetros, gráficos e plotagens;

- b) Proporcionar a completa reprodutibilidade. A reprodutibilidade é a chave para os métodos científicos e processos complexos que são mais difíceis de serem reproduzidos. Recomenda-se que haja a completa e eficiente preservação de uma boa documentação como pré-requisito essencial para a reprodutibilidade eficaz.

Ferramentas que documentam o workflow e facilitam a reprodutibilidade são adotadas nesse estágio, a exemplo do *VisTrails* e do Projeto *Kepler*.

O *VisTrails* é um sistema de código aberto para o gerenciamento de fluxo de trabalho e processos de conteúdos científicos que oferece suporte à exploração e visualização de dados desenvolvido no *Scientific Computing and Imaging Institute* da Universidade de Utah. Além de fornecer uma infraestrutura geral que pode ser combinada com repositórios, sistemas e bibliotecas existentes que fornece suporte para a exploração e visualização de dados, possui versões pré-compiladas para Windows, Mac OS X e Linux.

O Projeto *Kepler* promove e apoia as capacidades, o uso e a conscientização da aplicação de fluxos de trabalho científicos. O *software* Kepler é de código aberto e livre, desenvolvido e mantido pela colaboração entre equipes compostas pela University of California de Davis, Santa Barbara e San Diego e pela *NSF-funded Kepler/CORE team*. Foi projetado para auxiliar cientistas, analistas e programadores de computadores a criar, executar e compartilhar modelos e análises em uma ampla gama de disciplinas científicas e de engenharia. Adotar ferramentas de alta performance para a análise métrica e visualização de dados é uma prática recomendada. Este estágio também apoia iniciativas e abordagens que descrevem e associam todo o conteúdo de uma pesquisa em um mecanismo de leitura óptica para que ele possa ser mais facilmente compartilhado e permutado para além do reuso, oportunizando a reprodutibilidade de dados e códigos abertos. Analisar e visualizar os dados são processos de qualidade e validação da ciência aberta. Este procedimento corrobora com o reuso e a reprodutibilidade de dados científicos.

Por fim, como reflexão deste capítulo é possível concluir que as soluções tecnológicas que acompanham todos os estágios do ciclo de vida são mandatórias para que haja um processo investigativo nos moldes das práticas da *e-Science*. Os procedimentos e as soluções adotadas são muito específicas de cada instituição e grupos colaborativos de pesquisa. As políticas e diretrizes, bem como o modelo de ciclo de vida deve atender o propósito da pesquisa e das instituições envolvidas. Contudo, o princípio básico da ciência aberta está em tornar os dados científicos, documentos, ferramentas tecnológicas e as boas práticas, acessíveis, interoperáveis, reutilizados e reprodutíveis.

6 REPENSAR OS DIREITOS DE PROPRIEDADE INTELECTUAL FRENTE AS TENDÊNCIAS CONTEMPORÂNEAS DA *E-SCIENCE*

A reflexão do capítulo se funda na essência do quarto paradigma científico que apregoa um novo fazer no qual a ciência se propõe ser aberta, gratuita e de livre acesso. A ideia principal a ser explorada no capítulo está direcionada para o repensar dos direitos de propriedade intelectual na perspectiva de uma sociedade globalizada, conectada, digital, colaborativa e compartilhada. Esse repensar se apoia na essência de uma ciência do ‘comum’ e do *commons paradigm* que trazem a colaboração e a coletividade para o espectro da ciência contemporânea. A noção de propriedade exclusivista e individualista dos sistemas autorais vigentes se direcionam para o olhar colaborativo. As dimensões teórica e política são norteadoras para contextualizar essas tendências contemporâneas.

6.1 A CIÊNCIA DO COMUM E COLABORATIVA

As primeiras manifestações dos Direitos de Propriedade Intelectual foram originados na cultura ocidental dos impérios grego e romano. Mas foi a partir da invenção das ‘prensas de tipos móveis’ por Johannes Gutenberg (1450) que questões sobre a autoralidade e a PI começaram a emergir efetivamente a visão coletivista da época medieval que versava por uma propriedade comum.

A invenção de Gutenberg provocou a mudança no estado das coisas, mundialmente, sendo associada ao trio ‘imprensa-pólvora-bússola’ (EISENSTEIN, 1998). Até então, a propriedade não possuía atribuições e considerada um bem privativo das cortes. Com a proliferação da imprensa, problemas de direitos autorais e reivindicações pela proteção e remuneração autoral afluíram (MARTINS FILHO, 1998). Assim, surgiram os sistemas de propriedade intelectual foram condicionados a um regime de propriedade privada e exploração econômica.

A comunicação científica foi incorporada pela lógica capitalista com diferentes formas de apropriação da produção intelectual com os primeiros jornais científicos surgidos na França e Inglaterra. Reforça-se que a preocupação obsessiva pela propriedade intelectual, especialmente nas últimas décadas do século XX, conduziu ao “alargamento dos mecanismos de apropriação privada da produção intelectual e cultural, ampliando e aprofundando relações capitalistas de mercado para áreas que até então constituíam uma reserva social” (ALBAGLI,

2015, p. 10). Esses mecanismos estão baseados em um paradigma proprietário econômico e em relações capitalistas regulando a propriedade intelectual há séculos.

No contexto científico, tendências direcionadas as novas dinâmicas projetadas pela ciência aberta estão aflorando discussões e preocupações também na seara dos direitos de propriedade intelectual. As tendências contemporâneas apontam para uma cultura colaborativa, coletiva e compartilhada como pressuposto para uma ciência aberta e comum. Assim a reflexão emerge no cerne dos DPIs, do paradigma proprietário econômico, arraigado nos sistemas de propriedade intelectual vigentes, para o *Commons Paradigm* que está assentado em uma cultura coletiva e colaborativa de bens intelectuais.

As diferentes legislações nacionais e o equilíbrio dos DPIs nas relações locais, nacionais e internacionais devem ser conduzidas no contexto atual científico para uma ciência aberta. O movimento do acesso aberto impulsionou o entrelaçamento dos modos de comunicação eletrônicos com novas dimensões ao acesso à informação (BORGMAN, 2003). A compreensão fundamenta as características relacionadas à colaboração e ao compartilhamento presente nas práticas da *e-Science* e levantam reflexões para o quadro jurídico existente da propriedade intelectual.

Os DPIs vigentes regem os instrumentos legais regulatórios para o desenvolvimento econômico, tecnológico, industrial, científico, técnico, cultural e social. Albagli (2015, p. 10) afirma que “os instrumentos de propriedade intelectual, em seu atual formato, já não cabem no novo paradigma”. Na atualidade os sistemas DPIs estão sendo repensados para atender as atuais dinâmicas de uma Sociedade da Informação digital. Diante das mudanças e interações almejadas pela era digital o arcabouço jurídico da propriedade intelectual se apresenta ineficaz e não contempla práticas que surgem de ambientes científicos, tecnológicos, sociais e culturais. “As novas formas de comunicação não cancelam o conceito de propriedade intelectual mas, ao contrário, estimulam o legislador a intervir para tutelar os direitos dos autores que fixam seus pensamentos em escritos divulgados pela rede [...]”. (PAESANI, 2012, p. 2). Grande desafio político-ético-legal está em conduzir uma agenda global para o entendimento de que a ciência aberta altera as práticas existentes e interfere nos mecanismos regulatórios da PI, ao mesmo tempo que se busca garantir os direitos de autoralidade aos criadores e inventores.

O âmbito científico é a principal reflexão deste estudo. As tendências contemporâneas científicas em âmbito mundial foram direcionadas inicialmente pelo movimento open access (acesso aberto) e mais recentemente se expande para a ciência aberta e ciência dos dados. Com base nisso, as tendências no ambiente científico estão sendo conduzidas por uma

perspectiva de ciência aberta com práticas científicas conduzidas pelo quarto paradigma da *e-Science* e pelo *Commons Paradigm*.

Assim, o estudo propõe duas interpelações acerca da transversalidade da ciência aberta com os DPIs. A primeira interpelação se manifesta nas dimensões político-legal, enfatizando os DPI como subdomínio de uma política orientada aos dados de pesquisa. E a segunda interpelação está idealizada na dimensão epistemológica, refletindo a cultura do *commons*, para uma sociedade colaborativa, compartilhada, aberta e digital.

A **primeira interpelação** preconiza que a ciência aberta constitui um *locus* de poder e negociação, contradições, ideologias, culturas, saberes e conhecimentos. As relações entre ciência-poder-saber são manifestadas em instâncias, dentro e fora do *corpus* institucional.

Os esforços da ciência aberta envolvem instâncias de “ação e decisão diferenciadas, internas e externas à ciência, que vão desde o pesquisador individual e equipes de pesquisa até o nível macro das políticas públicas e das regulações internacionais, passando pelo nível das instituições científicas e agências de fomento” (ALBAGLI, 2015, p. 21). O empreendimento político voltado para a ciência aberta que utiliza dados de pesquisa nas práticas científicas, deve considerar os níveis hierárquicos e as diferentes instâncias.

Ademais, aspectos culturais, éticos, ideológicos e formas de poder estão enraizados nos valores intrínsecos da nação ou de uma instituição. Pensar nesses valores no sentido *stricto sensu* e relacioná-los aos aspectos internos e institucionais (pessoas, instrumentos, processos). No sentido *lato sensu* eles se voltam para os aspectos externos e amplo, tal como a política governamental pública ou uma área do conhecimento.

Nessa direção pensar nos sentidos *stricto sensu* e *lato sensu* é essencial na condução de políticas de informação. Por política de informação se entende o conjunto de leis, regulações e posições doutrinárias e outras tomadas de decisões e práticas com efeitos constitutivos em toda sociedade envolvendo a produção, processamento, fluxo, acesso e uso (BRAMAN, 2011). Analogamente, conjecturar uma política orientada aos dados de pesquisa inclui todos esses elementos e se acrescentam aspectos de compartilhamento, colaboração, reuso e reprodutibilidade.

A política legitimada por autoridade governamental passa a ser compreendida e nomeada como uma política pública de informação. Expectativa semelhante deve ser direcionada à ciência aberta, norteando a elaboração da política pública orientada aos dados científicos. Igual propósito deve ser conduzido para uma nova constituição dos DPI como um subdomínio dessa política.

A dimensão político-legal que envolve os DPIs e a *e-Science* precisa ser pensada

como uma nova conjectura que promova as adequações legais e contemplem os aspectos colaborativos, coletivos, éticos e culturais. Deve auxiliar na compreensão das políticas e diretrizes que atentem para valores (extrínsecos e intrínsecos) que sustentem o amplo espectro da ciência aberta (MARQUES; PINHEIRO, 2013).

O estudo apresenta ainda as correlações conceituais das políticas voltadas para a informação, nomeadas como subdomínios. O subdomínio ‘Protecionismos da Informação’ em seu arcabouço político está relacionado aos “regulamentos e mecanismos que controlam o acesso à informação e divulgação na esfera pública, (ex.: segredo oficial) e em mercados de informação (ex.: Proteção de dados)” (MARQUES; PINHEIRO, 2013, p. 121). Na dimensão política e com base nesse subdomínio a predominância dos mecanismos de regulação dos mercados de informação, a exemplo da proteção dos dados é predominante.

Assim a reflexão se volta para que o subdomínio dos DPIs seja contemplado numa visão multidimensional, com o olhar voltado para as interações colaborativas, mercados informacionais e culturais, além dos produtos intelectuais provenientes das instituições formais (universidades, indústrias e pesquisa), modelados na cultura do *commons* e do compartilhamento.

Isto posto, a **segunda interpelação** expõe o olhar multidimensional para a cultura do *commons* e da Economia Criativa. Compreender os dados de pesquisa numa perspectiva de bem comum e, por conseguinte, se constituir em um elemento de interesse da coletividade é merecedor de novas reflexões normativas e legais. Logo, é mister compreender essa perspectiva a partir da ótica do objeto de estudo (dados de pesquisa) na noção do *commons*. A ideia preliminar pressupõe que cultura, conhecimento e criatividade podem ser encapsuladas na noção da ciência aberta (FROSIO, 2014).

Na perspectiva do *commons* e da *e-Science*, os DPIs possuem conexão com a Economia, por meio da Teoria do *Commons* e da Economia Criativa. Afirmar Albagli (2015) que a ciência aberta amplia o espectro de possibilidades e espaços de criação do saber quando ocorre uma interlocução da ciência com outros segmentos sociais e outros tipos de saberes.

Lafuente (2015) enfatiza essa perspectiva quando afirma que a ‘ciência do comum’ é sua própria condição. Defende que essa derivação deve ser um dos imperativos da nova gestão da ciência que rejeitam as regulações impostas pelos mecanismos econômicos da propriedade intelectual que transformam o conhecimento em informação codificada. A dinâmica tradicional, gera consequências devastadoras, quando há o incentivo de vários processos de privatização do conhecimento. Assim, a noção de *commons* se apresenta em

duas vertentes: a primeira relacionada com o entendimento de ‘não *propertising*’ e ‘tragédia do *commons*’; e a segunda, com a ideia da Teoria do *Commons* e da Economia Criativa.

Inicialmente deve ser compreendida nessa **primeira vertente** a noção conceitual de recursos, bem intelectual e bem comum. Recursos engloba os componentes (naturais, ecológicos, ambientais, entre outros) dispostos na natureza e à disposição de uma coletividade. Bens intelectuais são resultantes de processos criativos, inovadores, culturais e científicos oriundos de um processamento de conhecimento cognitivo. E por, bem comum entende-se a junção de recursos naturais e de bens intelectuais disponíveis e acessíveis em prol da sustentabilidade de uma determinada coletividade.

Explica Herscovici (2013, p. 18, grifo nosso) que “bem comum, em determinada coletividade (componentes ecológicos, recursos naturais, **conhecimento comum**, etc.), a apropriação privada de tais bens pode ser prejudicial para a comunidade e pode se traduzir pelo esgotamento do estoque disponível”. Certo é que o questionamento se funda entre a apropriação dos bens comuns regidos por DPI e o acesso aberto a favor de um coletivo. Alarga-se o escopo para a classificação de que os bens comuns estão na seara dos direitos difusos e coletivos.

O acesso aos bens intelectuais procedentes da própria ecologia social, deve caminhar para uma nova postura assentada na cultura colaborativa e do compartilhamento entre os envolvidos. A economia criativa e o *commons* visam que novas práticas e espaços interativos e colaborativos promovam a criatividade, inovação e saberes científicos, artísticos e culturais a favor da sociedade.

Albagli (2015) e Krokosz (2015) reforçam esse pensamento na ideia de que o paradigma subjacente do *commons* e a cultura de bens comuns são alicerçados por práticas relacionadas com a autoria colaborativa e coletiva por meio das cocriações e cocolaborações que estão representadas na *e-science*, ciência aberta, inovação aberta, produções *peer-to-peer*, *wiki*, *crowdsourcing*, entre outras, reforçando a noção emergente da cultura colaborativa e do compartilhamento.

O entendimento de ‘não *propertising*’ é defendido pela ideia de que para acesso e uso dos bens comuns é necessária a intervenção de um ente regulador. Os estudiosos afirmam que tais bens precisam ser regidos por DPIs. Afirmam ainda que a ausência de uma intervenção e o uso indiscriminado (recursos naturais e bens intelectuais) podem levar a escassez irreversível desses bens (FROSIO, 2014). A escassez é chamada de tragédia do *commons*.

No que diz respeito aos direitos intelectuais a “open science no que se refere aos DPIs foi corroída pela *propertising* implacável de direitos intelectuais” (FROSIO, 2014, p. 7,

tradução nossa). A *‘propertising’* mantém-se até os dias atuais, quando a indústria editorial e a iniciativa privada que regem os mercados de conhecimento e informação, pressionam por um reforço dos códigos jurídicos que garantam a continuidade dos DPIs para controle desses bens. Tais discussões estão assentadas na **segunda vertente** relacionada ao *commons*.

Avança o movimento da ciência aberta com o surgimento de “mecanismos regulatórios visando reverter assimetrias resultantes da apropriação e distribuição privada dos conhecimentos, sobretudo em áreas sensíveis e com forte apelo social, como saúde, agricultura, alimentação e meio ambiente” (ALBAGLI, 2015, p. 18). Esta é uma acepção generalista da noção do *commons*. É a ideia preliminar de uma ciência do ‘comum’, pressupondo que os dados de pesquisa são encapsulados como bens intelectuais, público e comum na ciência aberta.

A Teoria do *Commons* originada na corrente denominada *Commons Paradigms* foi teorizada por Ostrom⁵ (1990). É o marco teórico no contexto da governança econômica dos recursos intelectuais e naturais e levanta questionamentos acerca da proprietização desses bens de acesso comum. A teoria demonstra que as pessoas comuns são capazes de criar regras e instituições que permitem a gestão sustentável e equitativa dos recursos compartilhados.

Pressuposto essencial da ciência do comum que busca na sustentabilidade e no compartilhamento de bens comuns, benefícios para uma determinada coletividade. Krokosz (2015) e Nielsen (2011) afirmam que a ideia do paradigma do *commons* no que se refere aos conhecimentos científicos estão sendo compartilhados, abertamente, desde o processo de descoberta (dados de pesquisa) por cientistas, ambientalistas, usuários da internet e outros envolvidos na cultura da ciência do comum. Essa é a nova postura da cultura colaborativa e do compartilhamento. Visa que as novas práticas e espaços interativos e colaborativos promovam a criatividade, inovação e saberes científicos, artísticos e culturais.

Diferentes formas de economia solidária e colaborativa apresentam eficiência social na economia digital quando “os sistemas dos *commons* baseados sobre o compartilhamento de informações e de bens culturais são social e economicamente mais eficientes que os sistemas baseados sobre a propriedade privada e sobre a distribuição a partir de suportes materiais individualizados” (HERSCOVICI, 2013, p. 190, tradução nossa).

É neste contexto que os DPIs entram no embate, pois de um lado estão os modelos tradicionais de propriedade arraigados na cultura do *copyright* e das patentes, e do outro a busca pelo acesso aos bens intelectuais e culturais por meio do viés aberto e público visto que:

⁵ Elinor Claire ‘Lin’ Ostrom era uma economista e cientista política da Indiana University, Estados Unidos, que recebeu o Prêmio Nobel de 2009 em Ciências Econômicas pela sua pesquisa inovadora.

A propriedade privada e suas consequências econômicas e jurídicas vão de encontro à tendência de compartilhar, de cocriação e inovação; a regulação ao livre acesso; a vontade de manter determinado bem ou produto em estado de escassez, incluindo-a valor, contrasta com a abundância e a capacidade de renovação de recursos na Economia Criativa – recursos estes, em sua maioria, culturais e intelectuais [...]. (CARVALHO et al., 2015, p. 180).

O conhecimento é um bem intelectual cujo entendimento se apoia como um bem intangível, sendo assim: “conhecimento é diferente. Um produto hoje se torna viável e útil muito mais pelo conhecimento incorporado (pesquisa, *design*, comunicação etc., os chamados intangíveis) do que pela matéria-prima e trabalho físico” (DOWBOR, 2015, p. 2). O conhecimento é um ativo intangível incorpóreo, que necessita ser incorporado e encapsulado para ganhar forma, adquirindo um *status* de bem intelectual e consequentemente, com direitos associados. O conhecimento em si mesmo não é um bem intelectual tangível, precisa ganhar forma, por meio de um processo de trabalho científico ou criativo. É por meio da incorporação e apreensão do saber, que os recursos intelectuais, criativos, artísticos e culturais se manifestam em bens culturais e intelectuais (produtos) e assumem a natureza pública.

Todavia, há de se considerar que os processos exploratórios nas searas científica, econômica, cultural e industrial são conduzidos, na maioria das vezes, por mercados capitalistas que regem a exploração comercial da PI. Bens comuns resultantes dos processos científicos, inovativos e criativos rendem-se aos apelos econômicos do mercado.

O questionamento está na regulação dos bens comuns por mecanismos e instrumentos que envolvem a PI. A reflexão direciona para a atenção do acesso aos bens comuns que promovam uma sustentabilidade para uma determinada coletividade. Tal sustentabilidade deve estimular a inovação e a criatividade para impulsionar soluções e crescimento social, econômico e científico sem que haja uma restrição no acesso e uso.

Trazendo para o contexto da *e-Science* se pressupõe o acesso aberto e público aos dados de pesquisa. Estando os dados de pesquisa encapsulados como um bem comum é necessário considerar aspectos relacionados com a tipologia (observacionais, experimentais e computacionais), natureza (científica, governamental, cultural, artística) e acesso (aberto, fechado, restrito, embargado). Por exemplo, o direito à privacidade (dados pessoais) e conteúdo de segurança nacional devem ser previstos nos instrumentos normativos e legais no escopo da política.

Na noção do *commons* não desconsidera a autoria relacionada ao direito moral, mas busca uma equidade entre o direito de acesso para a coletividade e a paternidade da obra. “O reconhecimento perene da paternidade ou criação original de uma determinada obra, o que de

acordo com a legislação é algo inalienável, ou seja, a paternidade de uma obra é intransferível” (KROKOSZ, 2015, p. 101). A paternidade pela obra é mantida prevalecendo os direitos morais do autor.

Além do *commons* outro destaque dessa vertente é a Economia Criativa (EC). A EC é um conceito novo e não usual. Na seara do Direito é um *locus* de interseção com a propriedade intelectual. A EC possui características que buscam a valorização de produtos culturais e intelectuais, a dinamicidade de seus processos e a tendência do compartilhamento (CARVALHO et al., 2015).

Richard Florida (2010) considerou em seus estudos que o fim do predomínio da economia pautada na segunda Revolução Industrial abriu novos espaços para uma dinâmica que ele chamou de Economia Criativa. É um dos percussores do tema relacionado à EC e aos seus processos criativos. Analogamente, referenciou a mudança de paradigma, referindo-se à teoria de Thomas Kuhn na obra ‘A Estrutura das Revoluções’, para afirmar que a mudança do paradigma proprietário intelectual transitará para o compartilhamento de bens intelectuais e culturais oriundos dos processos criativos.

A EC tem como sustentabilidade o tripé: conhecimento, informação e inovação criativa. Este tripé delineia atividades que requerem conhecimento, informação (como ferramentas para a criatividade) e inovação que se dá através de novos aparatos tecnológicos e com os novos modelos de negócios (CARVALHO et al., 2015). Os processos criativos, recursos tecnológicos e novos modelos de negócios estão em mudança. Apesar da Economia Criativa ser oriunda do conceito gerado na seara econômica, e portanto, sustentada no paradigma proprietário, tem vertente ligada à inovação e ao compartilhamento.

Carvalho e outros (2015) afirmam que a sua natureza permite que atividades ligadas à valorização e à comercialização de produtos intelectuais, tecnológicos e culturais se tornem atividades dinâmicas, fluídas, empreendedoras e compartilhadas. Enquanto que, na visão de Howkins (2013) a EC tem sua base na criatividade. O autor afirma que a criatividade sozinha não se configura como uma atividade econômica, mas quando ideias se tornam um produto criativo (bem ou serviço), passam a ter implicações e valor econômicos.

Howkins (2013), em seus estudos, traçou um limite cronológico para situar o estudo da criatividade por meio de cinco gerações demonstrado no Quadro 8.

Quadro 8 – As cinco gerações do estudo da criatividade

	Denominação	Ênfase	Época
1 ^a . Geração	Pensamento criativo	Desenvolvimento de habilidades	Década de 1950
2 ^a . Geração	Solução criativa de problemas	Produtividade e competitividade	Década de 1960
3 ^a . Geração	O viver criativo	Autotransformação	Década de 1980
4 ^a . Geração	Criatividade como valor social	Solução de problemas sociais, aberta à vida, à juventude, ao cotidiano	Década de 1990
5 ^a . Geração	Economia criativa	Geração e exploração da propriedade intelectual	Década de 2000

Fonte: Machado (2015, p. 40) baseado em Howkins (2013).

O entendimento desta divisão é baseado em uma trajetória de gerações. A 1^a geração centrada no pensamento criativo e no desenvolvimento de comportamentos técnico-especializados. A 2^a geração buscou a solução criativa de problemas, baseada na produtividade e competitividade, ocasionada pelo auge da Revolução Industrial da década de 1960. A 3^a geração desabrochou na década de 1980, na autotransformação, com os estudos e pesquisas sobre a criatividade que estavam mais voltados para a dimensão individual. A 4^a geração se abre para uma busca mais ampla por soluções baseadas nas questões sociais, sustentabilidade e para a formulação de políticas públicas. E por fim, a 5^a geração é baseada na economia criativa e na exploração da propriedade intelectual a partir da década de 2000. O estudo de um lado reflete a exploração econômica dos produtos criativos, do outro enfatiza o compartilhamento de bens intelectuais, prevalecendo o espírito inovador e colaborativo.

A ruptura com os modelos tradicionais de negócios (econômicos, educacionais, científicos) abre espaços para essas dinâmicas. Apesar dos produtos criativos estarem encapsulados por DPIs, eles são bens intelectuais que promovem a inovação e a cultura e, portanto, pressupõem que devem ser acessíveis, públicos e gratuitos. No entendimento de Carvalho e outros (2015, p. 178) a “Economia Criativa parece rimar ainda com as ideias como as de conectividade, interação, compartilhamento, inovação; relaciona-se às novas mídias, novas tecnologias, ao empreendedorismo e à mudança de padrões”.

No que concerne à ciência aberta, os bens intangíveis estão associados a diversos recursos, desde os bens intelectuais (renováveis), a exemplo do conhecimento, até os recursos naturais (escassos), relacionados com os bens comuns. Os recursos intangíveis que se manifestam através da cultura, conhecimento e experiência são infinitos e renováveis e representam uma economia em abundância baseada nos modelos de colaboração (CARVALHO et al., 2015). Nesse ponto converge o maior desafio da ciência aberta que é transformar recursos intangíveis e bens intelectuais, em bens comuns e públicos. A cultura do comum e do compartilhamento devem fortalecer a “busca de uma ciência aberta, colaborativa

e conectada, uma reflexão sobre os desafios da propriedade intelectual e o direito à informação, e as implicações para relações de trabalho colaborativo e compartilhado na construção do conhecimento” (DOWBOR et al., 2015, p. 9).

As questões de PI baseadas na dicotomia entre o público e o privado atestam que:

A dicotomia derivada da propriedade privada baseada nas leis derivadas do modelo econômico tradicional e a proteção dos direitos autorais exacerbada entram em conflito com os direitos inovadores, provocando um hiato entre o que as tecnologias de informação permitem e a lei proíbe, prevalecendo em uma sociedade digital a contrafação. Outra questão se funda nos preceitos da propriedade intelectual de bens intangíveis (*on-line*) e digitais (CARVALHO et al., 2015, p. 182).

A confluência da Economia Criativa com a *e-Science* está no contexto de bens intelectuais intangíveis, como por exemplo: o conhecimento. A colaboração e o acesso são necessários e vitais “para a sobrevivência de todos, o acesso às tecnologias que reduzem o impacto climático, por exemplo, não só não deve ser travado por patentes, como fomentado” (DOWBOR, 2011, p. 3). Tal reflexão se fundamenta em dois entendimentos: o primeiro relaciona-se ao crescente argumento de que dados de pesquisa oriundos de pesquisa financiadas com recursos públicos devem retornar à sociedade com um bem público e acessível a todos. Contrapõe-se com os direitos proprietários garantidos aos autores, pesquisadores e inventores. É sabido que não é um movimento fácil e rápido de acontecer. A PI tem séculos de institucionalização alicerçada em sentimentos individualizados e proprietários; o segundo é a defesa da produção que exploram as ideias de justiça ambiental, no sentido econômico-sócio-cultural com novas vertentes de capitalização do intelectual que antes era pautado em um paradigma econômico proprietário.

Assim, eliminar as práticas atuais que aumentam a segregação e estimulam o monopólio social ou econômico se faz imperativo (ALBAGLI, 2015). As implicações no tocante aos DPIs constituem um dos gargalos da ciência aberta. Grande desafio político-ético-legal está em conduzir uma agenda global para o entendimento de que a ciência aberta quebra com paradigmas de práticas existentes. A promoção de uma ciência que supere a proprietização para uma sustentabilidade mais ampla em áreas como, saúde, meio ambiente, alimentação, agricultura, segurança nacional, fomentando o desenvolvimento econômico, científico e tecnológico do país. A reflexão final deste capítulo se pauta na ampliação dos direitos de propriedade intelectual instituídos em um paradigma proprietário para uma visão multidimensional que direciona a autoralidade para um processo colaborativo, compartilhado e de atribuição na perspectiva da *e-Science*.

6.2 ARCABOUÇO POLÍTICO-LEGAL BRASILEIRO DOS DIREITOS DE PROPRIEDADE INTELECTUAL

A reflexão do capítulo versa sobre os instrumentos legais que regem os Direitos de Propriedade Intelectual (DPIs) no Brasil e as iniciativas que emergiram do movimento de acesso aberto, acesso à informação pública e governo eletrônico. Muito embora essas iniciativas não são direcionadas especificamente aos dados científicos, as reflexões em cima das legislações brasileiras nortearam o entendimento na seara político-legal do contexto nacional. O capítulo está apoiado nas dimensões político e ético-legal.

6.2.1 Normatização Legal dos Direitos de Propriedade Intelectual

Na conjuntura da atual cultura digital, tornou-se possível a qualquer um que tenha acesso à *Internet*, copiar, recortar, apropriar-se de obras alheias sem que haja controle efetivo. Dois aspectos são observados nessa conjuntura: o primeiro está relacionado à facilidade de acesso à informação e ao conhecimento e as legislações vigentes; e o segundo com a conduta ética na sociedade contemporânea.

Acrescenta-se ainda os aspectos culturais e legais que regem as instituições, notadamente as de ensino e pesquisa. Os DPIs regidos por leis, decretos e regulamentos que protegem as produções e inovações e sustentam o desenvolvimento econômico, político, tecnológico, cultural e científico do país. Agora estão movidos para as dinâmicas emergentes para uma ciência aberta.

No Brasil, o direito autoral tem por finalidade regulamentar a proteção de obras (literárias, científicas e artísticas) e os seus direitos conexos, além de reger os direitos (produtores, artistas, intérpretes e executantes) de fonogramas e organismos de radiodifusão (FRAGOSO, 2009). Na legislação autoral há a coexistência de duas naturezas: moral e patrimonial.

A Constituição Federal Brasileira (1988) faz previsão sobre os direitos de propriedade intelectual no Capítulo I que trata dos direitos e deveres individuais e coletivos. O artigo 5º prevê no inciso XXII (garantia do direito de propriedade), inciso XXIII (a propriedade atenderá à sua função social), inciso XXVII (aos autores pertence o direito exclusivo de utilização, publicação ou reprodução de suas obras, transmissível aos herdeiros pelo tempo que a lei fixar), inciso XXVIII (a proteção às participações individuais em obras coletivas e à reprodução da imagem e voz humana, atividades desportivas e o direito de fiscalização para

fins econômicos) e por fim, no inciso XXIX (assegura aos autores de inventos industriais o privilégio temporário de patentes, criações industriais, marcas, nomes de empresas e signos distintivos). Os incisos constituintes pressupõe um direito que rege a proteção autoral baseada nos pressupostos da propriedade e dos direitos individuais. Os processos e regulações funcionam em diferentes segmentos da sociedade e se estende em fronteiras nacionais e internacionais.

O Código Civil não faz menção acerca dos direitos autorais, mas prevê os direitos personalíssimos no artigo 11, do capítulo II. Descreve que “com exceção dos casos previstos em lei, os direitos da personalidade são intransmissíveis e irrenunciáveis, não podendo o seu exercício sofrer limitação voluntária (BRASIL, 2002, *on-line*). Os direitos autorais estando associados aos direitos personalíssimos, no que tange aos direitos morais são por conseguinte, irrenunciáveis. Estando os direitos autorais regidos pela natureza moral, é permitido apenas, a transmissão dos direitos patrimoniais para que terceiros possam usufruir, comercializar e reproduzir a obra por tempo determinado.

A Lei de Direitos Autorais (LDA), sob o nº 9.610, de 19 de fevereiro de 1998 estabelece as matérias acerca do Direito Autoral e seus Direitos Conexos. A LDA possui sua natureza jurídica assentada nos direitos da personalidade como um atributo moral e nos direitos patrimoniais relacionados aos atributos econômicos concernentes à criação materializada. Assim, “o Direito de Autor é composto por um conjunto de normas jurídicas que individualizam, tutelam e disciplinam as criações intelectuais. A ligação entre a criação intelectual e o autor que a criou é indissolúvel e incindível” (PAESANI, 2012, p. 9).

O Artigo 7º é um dos mais extensos e trata da proteção de “criações do espírito, expressas por qualquer meio ou fixadas em qualquer suporte, tangível ou intangível, conhecido ou que se invente no futuro” (BRASIL, 1998, *on-line*). O artigo se desdobra em diversas formas de proteção autoral com destaque para os incisos:

- a) I - textos de obras (literárias, artísticas ou científicas);
- b) II - conferências, alocações, sermões e outras obras (mesma natureza);
- c) III - obras (dramáticas e dramático-musicais);
- d) IV - obras coreográficas e pantomímicas;
- e) V - composições musicais (tenham ou não letra);
- f) VI - obras audiovisuais (sonorizadas ou não) incluindo as cinematográficas;
- g) VII - obras fotográficas e por qualquer processo análogo;
- h) VIII - obras de desenho, pintura, gravura, escultura, litografia e arte cinética;
- i) IX - ilustrações, cartas geográficas e outras obras da mesma natureza;

- j) X - projetos, esboços e obras plásticas concernentes as áreas (geografia, engenharia, topografia, arquitetura, paisagismo, cenografia e ciência);
- k) XI - adaptações, traduções e outras transformações de obras originais, (criação intelectual nova);
- l) XII - programas de computador (que possuem legislação específica);
- m) XIII - coletâneas ou compilações. (BRASIL, 1998).

A **primeira reflexão** destaca o § 2º da LDA, que trata do inciso XIII que refere-se a proteção concedida as coletâneas e compilações que não abrangem “os dados ou materiais em si mesmos e se entende sem prejuízo de quaisquer direitos autorais que subsistam a respeito dos dados ou materiais contidos nas obras” (BRASIL, 1998, *on-line*). Porém, constatou-se a ausência de dispositivos que faça menção a proteção autoral de dados de pesquisas.

A **segunda reflexão** refere-se ao § 3º desse artigo quando descreve “no domínio das ciências, a proteção recairá sobre a forma literária ou artística, não abrangendo o seu conteúdo científico ou técnico, sem prejuízo dos direitos que protegem os demais campos da propriedade imaterial” (BRASIL, 1998, *on-line*). Se considerar a proteção de produções científicas e técnicas advindas dos processos científicos é inferido que as mesmas não estão contempladas no arcabouço da LDA.

A primeira reflexão da lei é pontuada pela incompatibilidade em um mesmo artigo da proteção às obras científicas, com a contradição entre o inciso I e o § 3. Essa proteção está relacionada às obras que emanam do espírito humano e se constitui no *corpus mysticum* (corpo espiritual) e *corpus mechanicum* (corpo material) da obra. Envolve uma questão de análise sobre os efeitos práticos quanto a forma e o modo de expressão do autor. A esta proteção se inclui também as obras científicas (forma e modo) que se fundem e complementam-se em si mesma.

Ademais, a proteção estende-se ao autor como um direito intrínseco relacionado à criação (paternidade) da obra, constituindo um direito de caráter pessoal (personalíssimo) e intransferível. A LDA é categórica quando descreve no artigo 11 que o autor é pessoa física, o criador de uma obra literária, artística ou científica. A obra nasce de cada pessoa, fruto de sua personalidade, denominando-se de direitos morais, intimamente relacionados à elaboração, divulgação e titulação da sua criação.

A abordagem moral é prevista no artigo 24 e seus incisos contemplam os direitos de reivindicação e sua identificação de autoria; conservação do ineditismo da obra; assegura a integridade, modificação e retirada de circulação da obra e de ter acesso a exemplar único e raro da obra é previsto. Os direitos patrimoniais garantem a exclusividade (utilizar, fruir e

dispor) ao autor da sua obra (literária, artística, ou científica) conforme previsto no artigo 28. O artigo 49 garante a transferência (total ou parcial) por meio de três modalidades (licença pública, concessão ou cessão de direitos). A utilização da sua obra depende de autorização prévia para permitir a reprodução (parcial ou integral), edição e tradução, utilização (recitação ou execução musical) e distribuição da obra por qualquer sistema.

Nos direitos patrimoniais, o autor pode decidir como colocará a sua obra à disposição do público deve prever a forma, local, tempo e custo. Cabe ao autor autorizar a uma terceira através da transmissão ou cessão total, transmissão ou cessão parcial, licença ou autorização e sucessão. Os instrumentos contratuais são utilizados para garantir a transferência da titularidade da obra a terceiros mantendo o direito de paternidade do autor criador da obra.

A LDA não prevê as sanções penais quanto à violação autoral de obras (eletrônicas, digitais e impressas). Assim, as sanções legais são previstas no Código Penal brasileiro, regido pela atualização da Lei nº 9.777, de 26 de dezembro de 1998, nos artigos 171 e 184.

A violação autoral (apropriação indébita e plágio) no Artigo 171 se enquadra como crime de estelionato que prevê: “obter, para si ou para outrem, vantagem ilícita, em prejuízo alheio, induzindo ou mantendo alguém em erro, mediante artifício, ardil, ou qualquer outro meio fraudulento: Pena - reclusão, de um a cinco anos, e multa” (BRASIL, 1998, *on-line*). E o Artigo 184, prever que: “violar direitos de autor e os que lhe são conexos, ocasiona uma pena com detenção, de 3 (três) meses a 1 (um) ano, ou multa e pena com reclusão, de 2 (dois) a 4 (quatro) anos, e multa, conforme previstos nos § 1º. e 2º”. (BRASIL, 1998, *on-line*).

Os parágrafos desse artigo, preveem que o indivíduo incorre em violação e tem o intuito de lucro (direto ou indireto), ao distribuir, vender, expor à venda, alugar, introduzir no País, adquirir, ocultar, ter em depósito (original ou cópia) de obra intelectual ou fonograma reproduzido com violação do direito (autor, artista, intérprete ou executante, produtor) ou, ainda, alugar (original ou cópia) de obra intelectual ou fonograma, sem a expressa autorização dos titulares ou de quem os represente, comete crime de violação autoral (BRASIL, 1998).

Vê-se que a legislação é bastante branda quanto à pena aplicada, o que pode ocasionar um estímulo à prática indevida de apropriação e reprodução ilegal, seja pessoa física ou jurídica, com fins comerciais e obtenção de lucro. A LDA “encontra-se desatualizada e não oferece subsídios jurídicos [...], por exemplo, o compartilhamento *on-line* de obras científicas, artigos, livros, prática esta promovida e valorizada no contexto atual” (CARVALHO et al., 2015, p. 182). Isto posto, é mister afirmar que a lei específica brasileira sobre os direitos autorais não abrange as dinâmicas necessárias para dar suporte legal e normativo a uma ciência orientada aos dados de pesquisa abertos.

Sobre a gestão coletiva de direitos autorais no Brasil foi instituído o Decreto nº 8.469, de 22 de Junho de 2015, que regulamenta alguns artigos da Lei nº 9.610, de 19 de fevereiro de 1998, e a Lei nº 12.853, de 14 de agosto de 2013 e decreta ser as associações responsáveis pela gestão e cadastramento.

O parágrafo 1º do decreto, prevê que as associações a que se refere o art. 99 no *caput* deverão centralizar à autoria e às interpretações, além do cadastro no Escritório Central de uma base de dados que contenha todas as informações de titularidade (obras, execuções, fonogramas, participações individuais, interpretação ou execução) e contidas nos contratos, declarações ou outros documentos de qualquer natureza, observado o disposto em ato do Ministério da Cultura.

No Artigo 10 do decreto é previsto que as associações de gestão coletiva de direitos de autor e dos que lhes são conexos deverão manter um cadastro centralizado de todos os contratos, declarações ou documentos de qualquer natureza que comprovem a autoria e a titularidade do autor da obra.

A Instrução Normativa nº 3, de 7 de Julho de 2015, instrui a normatização do inciso II do parágrafo único do art. 87 da Constituição, com base no disposto no artigo 34 do Decreto nº 8.469, de 22 de junho de 2015, e em conformidade com o inciso I do art. 1º do Decreto nº 6.532, de 5 de agosto de 2008, a habilitação, os procedimentos, organização do cadastro, supervisão e aplicação de sanções para a atividade de cobrança de direitos autorais por associações de gestão coletiva e pelo ente arrecadador de que trata a Lei nº 9.610, de 19 de fevereiro de 1998.

Para reger a gestão, a Instrução Normativa nº 4, de 7 de Julho de 2015, aprova o Regulamento de Mediação e Arbitragem no âmbito do Ministério da Cultura, nos termos da Lei nº 9.610, de 19 de fevereiro de 1998, e do Decreto nº 8.469, de 22 de junho de 2015, que regulamenta a Diretoria de Direitos Intelectuais de promoção e mediação entre usuários e titulares de direitos autorais ou seus mandatários, em relação à falta de pagamento, aos critérios de cobrança, às formas de oferecimento de repertório e aos valores de arrecadação, e entre titulares e suas associações, em relação aos valores e critérios de distribuição.

A Lei de Acesso à Informação (LAI) regida sob o nº 12.527, de 18 de novembro de 2011, dispõe sobre os procedimentos a serem observados pela União, Estados, Distrito Federal e Municípios, com o fim de garantir o acesso às informações conforme previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal. Para normatizar a LAI foi instituído o Decreto nº 7.724, de 16 de maio de 2012.

A LAI constitui um “marco para a democratização da informação pública, e preconiza, dentre outros requisitos técnicos, que a informação solicitada pelo cidadão deve seguir critérios tecnológicos alinhados com as ‘3 leis de dados abertos’” (MANUAL..., 2011, p. 8).

O Decreto nº 7.724 dispõe sobre o acesso as informações conforme previsto no inciso XXXIII do *caput* do art. 5º, no inciso II do § 3º do art. 37, art. 84, *caput*, incisos IV e VI, alínea “a” e no § 2º do art. 216 da Constituição, tendo em vista o disposto na Lei nº 12.527, de 18 de novembro de 2011.

O escopo do Artigo 3º do Decreto que disciplina a LAI as definições sobre dados e informação, requisitos para o acesso, que são relevantes para o estudo em tela, nos seguintes incisos:

- a) I, que *informação* são dados, processados ou não, utilizados para produção e transmissão de conhecimento (qualquer meio, suporte ou formato);
- b) III, são *dados processados* por processamento eletrônico ou automatizado com o emprego de tecnologia da informação;
- c) IV, trata do acesso à *informação sigilosa*, que é submetida temporariamente à restrição de acesso público, em razão de sua imprescindibilidade. Para a segurança da sociedade e do Estado, e as hipóteses legais de sigilo;
- d) V, trata da *informação pessoal* definida como informação relacionada à pessoa natural identificada ou identificável, relativa à intimidade, vida privada, honra e imagem;
- e) VII, trata da *disponibilidade* entendida como a qualidade da informação que pode ser conhecida e utilizada (por indivíduos, equipamentos ou sistemas autorizados);
- f) VIII, refere-se à *autenticidade* que diz respeito à qualidade da informação (produzida, expedida, recebida ou modificada) por determinado (indivíduo, equipamento ou sistema);
- g) IX, diz respeito à *integridade* relacionada à qualidade da informação não modificada, inclusive quanto à origem, trânsito e destino;
- h) X, é relacionado à *primariedade* que aborda a qualidade da informação coletada na fonte (com o máximo de detalhamento e sem modificações);
- i) XI, dispõe sobre a *informação atualizada* que se refere à informação que reúne os dados mais recentes sobre o tema, natureza e prazos previstos (BRASIL, 2011).

A LAI tenha sido promulgada para disciplinar o acesso à informação pública, as definições, os princípios norteadores, os requisitos de acesso e sigilosidade são pertinentes para pensar na gestão de dados científicos no ambiente nacional. A lei está norteada pelos princípios de publicidade máxima da administração pública, sendo o sigilo a exceção. Para tal exceção foram estabelecidas três categorias de sigilo e seus respectivos prazos: ultrassecreto (vinte e cinco anos); secreto (quinze anos) e reservado (cinco anos). Os prazos dependerão da natureza e sigilo de dados e informações, por exemplo as informações sobre a segurança nacional.

A abrangência da LAI e o Decreto que a regulamenta esteja direcionada ao escopo das informações públicas e dados abertos, observa-se que analogamente pode ser aplicado a uma realidade de dados científicos (abertos e restritos). Vale ressaltar que toda informação produzida e armazenada pela administração pública possui natureza pública, podendo ter seu acesso restrito ou aberto mediante as restrições legais.

O Projeto de Lei nº. 3396/2015 em tramitação no Congresso Nacional, versa sobre a criação, definições, princípios norteadores e objetivos para a Política Nacional de Incentivo à Economia Criativa. O projeto é embasado pelo entendimento de que o Brasil é um país de enorme diversidade cultural, o que resulta num potencial criativo em múltiplos setores e pode ser atestado pelo reconhecimento internacional da música, cinema, teatro, artesanato, e outras formas de expressão (BRASIL, 2015).

Este projeto de lei e outras regulamentações emergem em consonância com a preocupação generalizada nas organizações internacionais, a exemplo dos acordos TRIPS e da WIPO que discutem “os debates estão no centro das lutas por uma sociedade livre, onde, no século passado, a batalha era em torno da propriedade dos bens de produção, hoje, deslocou-se para a área da economia da criatividade” (DOWBOR, 2011, p. 33). O Projeto de Lei visa estabelecer o marco legal para criação e adequação da Economia Criativa no Brasil.

Assim, iniciativas que minimizem a dicotomia devem ser conduzidas para que possibilite o mínimo de proteção aos titulares de direito sem limitar o acesso, pois

A tensão entre a disseminação livre de informação e os conteúdos apropriados por direitos de propriedade intelectual vêm levando a busca de um consenso entre o direito do autor e o interesse do público no acesso, uma vez que o criador do conteúdo precisa ter sua obra protegida e, ao mesmo tempo, garantir que a obra seja objeto de uso pela sociedade (PERALTA et al., 2011, p. 1).

É notório que a existência de mecanismos legais em conformidade com os preceitos legislativos e jurídicos não garantem que o Direito acompanhe as mudanças e anseios da

sociedade. A exemplo da LDA que não está adequada para acomodar os direitos de propriedade intelectual numa sociedade baseada na cultura aberta e colaborativa.

Tal afirmação baseia-se no reduzido número de produções acadêmico-científicas acerca do assunto, de poucas iniciativas em andamento e da ausência de políticas e diretrizes que guiem estas ações de dados de pesquisa orientadas à ciência aberta. Potencialmente, a coleção de dados pode “capacitar os pesquisadores a formular novos tipos de indagações, hipóteses e a usar métodos analíticos inovadores no estudo de questões críticas para a ciência e para a sociedade” (SAYÃO; SALES, 2015, p. 5). Soma-se a isto, a ausência de uma prática mandatória que potencialize as atividades de pesquisa e tornem os dados abertos e acessíveis.

Em contrapartida, ações governamentais sobre acesso aberto, governo aberto, transparência pública, acesso à informação pública, preservação digital e repositórios institucionais já estão em andamento no Brasil. Iniciativas emergiram no âmbito governamental e em áreas específicas, como Biodiversidade, Agropecuária e Saúde Pública.

Ressalta-se que a abordagem dada a esta seção, traz um recorte que versa pelos direitos autorais, não abrangendo o escopo legal voltado aos direitos de proteção a patente e aos direitos *sui generis*.

6.2.2 Cenário Brasileiro do Acesso Aberto à Informação e aos Dados Governamentais

As iniciativas governamentais e as políticas públicas são fundamentais para impulsionar o desenvolvimento, conduzir processos públicos e regulamentar as mudanças em vários segmentos da sociedade. O movimento em prol do acesso aberto, transparência das informações públicas e abertura de dados abertos são aspectos relacionados ao atual escopo político-normativo brasileiro.

No Brasil, o entendimento para a abertura de dados governamentais tem como um dos princípios a promoção de um Governo Aberto que afirma:

Os órgãos públicos precisam promover ativamente a abertura do governo. As diretrizes de um governo aberto estimulam a criação de processos e procedimentos governamentais mais transparentes. A mudança de uma cultura de sigilo, que muitas vezes está incorporada ao setor público, para uma cultura de abertura é essencial para a promoção do direito à informação (BRASIL, 2013, p. 8).

O termo *open data* que ganhou popularidade no movimento de transparência e governo aberto mundialmente, será conduzido conceitualmente nesta seção como sendo os dados que referem-se as informações geradas por todos os órgãos públicos oriundos de

atividades administrativas governamentais. E por, Dados Abertos Governamentais é definido que: “são a publicação e disseminação do setor público na Web, compartilhadas em formato bruto e aberto, compreensíveis, logicamente, de modo a permitir sua reutilização em aplicações digitais desenvolvidas pela sociedade” (CGI.br, 2015, *on-line*).

Para que os dados ou conjunto de dados atendam ao princípio do governo aberto, é necessário estarem imbuídos de características que os tornem abertos e acessíveis. As características que representam este princípio partem de três premissas: disponibilidade e acesso; reutilização e redistribuição; e, participação universal.

Essas características estão associadas às três leis do direito preconizada por David Eaves⁶: “se o dado não pode ser encontrado ou indexado na web, ele não existe; se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado; Se algum dispositivo legal não permitir sua reaplicação, ele não é útil” (EAVES, 2009, *on-line*). A essência das três leis de Eaves está na premissa: os dados precisam ser encontrados, acessíveis, úteis, interoperáveis, legalizados, abertos, compartilhados e reproduzidos.

O governo tem a responsabilidade e competência para direcionar uma “política de dados abertos tanto por causa da quantidade e da centralidade dos dados que coleta quanto pelo fato de que tais dados são públicos [...] e estão gerando valor, e há potencial para muito mais” (BRASIL, 2011, p. 20), além de que, o acesso aos dados públicos é um direito garantido no artigo 5º da Constituição Federal brasileira.

A partir da promulgação da LAI no âmbito governamental em 2011, o Decreto sem número, de 15 de setembro de 2011, instituiu o Plano de Ação Nacional sobre Governo Aberto em consonância com o art. 84, inciso VI, alínea “a” da Constituição Federal. É destinado à promoção de ações e medidas que visem ao incremento da transparência e do acesso à informação pública, à melhoria na prestação de serviços públicos e ao fortalecimento da integridade pública, que serão pautadas, entre outras, por diretrizes. “A LAI representa uma mudança de paradigma em matéria de transparência pública, pois define que o acesso é a regra e o sigilo, a exceção” (BRASIL, 2013, p. 12).

A Secretaria de Logística e Tecnologia da Informação do Ministério do Planejamento, Orçamento e Gestão criou a Instrução Normativa nº. 4, de 13 de abril de 2012, que instituiu a Infraestrutura Nacional de Dados Abertos – INDA, com base no art. 31 do Decreto no 7.675, de 20 de janeiro de 2012, tendo em vista o disposto no Decreto no 7.579, de 11 de outubro de 2011, e considerando a Parceria para Governo Aberto, celebrada em setembro de 2011 entre o

⁶ David Eaves é um estudioso de políticas públicas e ativista para o governo aberto. Preconizou a *Three Laws of Open Government Data*.

Brasil e sete outros países, cuja co-liderança compete ao Brasil nos anos de 2011 e 2012, bem como o Decreto s/n, de 15 de setembro de 2011, sobre o Plano de Ação Nacional sobre Governo Aberto, o qual estabelece o compromisso do governo de implantar a Infraestrutura Nacional de Dados Abertos.

A instrução normativa regulamenta a política para garantir e facilitar o acesso pelos cidadãos, pela sociedade e, em especial, pelas diversas instâncias do setor público aos dados e informações produzidas ou custodiadas pelo Poder Executivo federal, dentro do escopo do Governo Aberto. Estabelece em seu escopo a definição, estrutura, gestão e as atribuições do comitê gestor.

Essas regulamentações legais impulsionaram a abertura efetiva dos dados governamentais, a sistematização e a transparência das práticas e resultados da gestão pública no Brasil. Estudos estão sendo realizados e apontam para benefícios que favorecem diversas áreas no âmbito do governo.

Na área política e social, as regulamentações conduzem para a transparência, responsabilização e confiança no governo, pela melhoria dos serviços e apoio na tomada de decisão no âmbito da gestão pública. A área econômica, visa o incentivo à inovação, ao crescimento econômico, a competitividade e acesso ao conhecimento coletivo. E nas áreas operacionais e técnicas do próprio governo, favorece a melhoria no relacionamento com usuários dos serviços e redução do volume de interações, reuso de dados, melhoria nos processos administrativos, sustentabilidade dos dados, possibilidade de combinação de dados, fortalecendo a visibilidade de resultados da gestão pública. Para tornar, produzir e publicar dados abertos é necessário preliminarmente, analisar aspectos relacionados com custos e mudanças culturais, institucionais, organizacionais e técnicas (SÃO PAULO, 2016).

O Plano de Ação Nacional sobre Governo Aberto contempla iniciativas, ações, projetos, programas e políticas públicas voltados para os dados e informações do Governo Aberto brasileiro. Em 2011, o Brasil firmou parceria com a *Open Government Partnership* (OGP) para o desenvolvimento de ações do Governo Aberto. Concomitante foi lançado o primeiro Plano de Ação do Brasil para Governo Aberto sob a responsabilidade da Controladoria-Geral da União.

A Controladoria-Geral da União “foi o órgão responsável por liderar a inserção do Brasil na OGP, articulando-se com diversos órgãos públicos e com setores da sociedade civil para construir o Plano de Ação Brasileiro” (BRASIL, 2016, *on-line*). Os compromissos firmados no plano brasileiro foram pensados para refletir diretrizes, desafios e princípios orientadores para os dados abertos no âmbito de um Governo Aberto.

A iniciativa do Plano de Dados Abertos é um exemplo de política pública que visa o definição de padrões, procedimentos para a publicação e escolha de futuras publicações e o fortalecimento da consciência e confiança dos intermediários nos dados abertos. Pode conter ainda informações sobre as realizações do órgão sobre o tema ou mesmo destacar os benefícios dos dados abertos para o governo, empresas e cidadãos (BRASIL, 2016).

Como instrumento de apoio ao Plano de Dados Abertos e as regulamentações legais, foi idealizado em 2011, o Portal Brasileiro de Dados Abertos. O Portal Brasileiro de Dados Abertos foi criado a partir da Lei de Acesso à Informação (LAI), como recurso disponibilizado pelo governo para que todos possam encontrar e utilizar dados abertos e informações públicas do governo brasileiro. O portal tem por objetivo disponibilizar todo e qualquer tipo de dado, por exemplo: dados da saúde suplementar, do sistema de transporte, de segurança pública, indicadores de educação, gastos governamentais, processo eleitoral, etc. A finalidade é promover a interlocução entre cidadãos e o governo para pensar a melhor utilização dos dados em prol de uma sociedade melhor (BRASIL, 2015).

Ainda no âmbito da Administração Pública Federal foi promulgada a política de publicação e compartilhamento de dados do Ministério da Educação (MEC) e do Fundo Nacional de Desenvolvimento da Educação (FNDE). A política possui como pressupostos os princípios do Governo Aberto e Transparência Pública. Percebe-se que o Governo Federal vem avançando lentamente para consolidar as políticas de compartilhamento e abertura das publicações e dados governamentais no cenário nacional.

Na esfera estadual o destaque é para o Governo de São Paulo, que na vertente do Governo Aberto possui o Portal Governo Aberto SP, regulamentado pelo Decreto Estadual nº 58.052, de 16 de maio de 2012.

O portal do Governo Aberto SP é uma iniciativa do Governo do Estado de São Paulo, de disponibilização, através da *internet*, de documentos, informações e dados governamentais de domínio público para a livre utilização pela sociedade, garantindo à mesma, acesso aos dados primários, de forma que possam ser reutilizados produzindo novas informações e aplicações digitais para a sociedade (SÃO PAULO, 2012, *on-line*).

Muitas outras iniciativas estão em andamento no âmbito governamental. A política mandatória no sentido de tornar os dados abertos e a informação pública no âmbito governamental é uma realidade em construção e está sendo conduzida e norteadas por princípios basilares do Governo Aberto, legislações específicas e iniciativas oriundas de instituições governamentais e organizações não governamentais.

6.2.3 Acesso Aberto aos Dados Científicos

No que tange aos dados científicos o processo está evoluindo lentamente. A estruturação e normatização da informação científica aberta foi acolhida em 2004, a partir do movimento de acesso aberto pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), ligado ao Ministério de Ciência, Tecnologia e Inovação (MCTi). Em relação aos dados científicos ainda não existe uma regulamentação que discipline e direcione a política para este contexto.

A partir de 2005, foi lançado o Manifesto brasileiro de apoio ao acesso livre ao conhecimento científico guiado por um conjunto de recomendações para toda a comunidade científica brasileira para a implantação de ações de promoção do acesso aberto. Em 2007, foi submetido à Câmara dos Deputados um projeto de lei com a finalidade de reforçar e respaldar as iniciativas pelo acesso aberto no Brasil, esse projeto já passou por revisões e atualmente encontra-se em tramitação no Congresso Nacional (IBICT, 2015).

Em 2009, o IBICT financiado pela FINEP, iniciou as iniciativas para a criação de repositórios institucionais em universidades públicas e centros de pesquisas financiados com recursos públicos (IBICT, 2012). Nessa direção o IBICT tem desenvolvido ações para oferecer um portfólio do seu conjunto de serviços de acesso aberto e de políticas editoriais para o Brasil. O Diadorim é um Diretório de Políticas Editoriais das Revistas Científicas Brasileiras. Oferece um serviço de informações relativas às autorizações concedidas para o armazenamento e o acesso dos artigos das revistas brasileiras em repositórios digitais de acesso aberto. Tem como objetivo auxiliar na localização de informações sobre a política editorial das revistas e sobre a permissão para o armazenamento de artigos em Repositórios Institucionais (IBICT, 2012).

No que tange as políticas para nortear o acesso e os DPIs são normatizadas em três condições: tipo de acesso, disponibilidade para depósito e licença de conteúdo. O tipo de acesso está relacionado a permissão definida pela editora para os artigos publicados em suas revistas que serão armazenados em repositórios institucionais.

Para a disponibilidade de depósito e armazenamento das versões pré-print e post-print e suas restrições são definidas na codificação por cores. A licença de conteúdo está relacionada as permissões estabelecidas pela editora para acesso e uso dos artigos publicados pela revista. As licenças estabelecidas no Diadorim para o uso dos artigos são baseadas no *Creative Commons*. O Quadro 9 a seguir disponibiliza as políticas de acesso aberto definidas para o diretório Diadorim.

Quadro 9 – Políticas de acesso aberto em repositórios institucionais

POLÍTICA	NATUREZA	DESCRIÇÃO
Tipo de acesso	Acesso aberto	Os artigos armazenados nos repositórios institucionais estarão disponíveis para qualquer usuário acessá-lo.
	Acesso aberto após período de embargo	Os artigos armazenados nos repositórios institucionais só estarão disponíveis para acesso aberto, depois de terminado o tempo de embargo estabelecido pela Editora.
	Acesso restrito	Apenas as informações referenciais (metadados) do artigo estarão acessíveis no repositório institucional.
Disponibilidade para depósito	Amarela: permite o depósito da versão pré-print de um artigo	A permissão será concedida ao repositório institucional para armazenar os artigos publicados pela Revista, somente da versão pré-print. Deve ser mantida a versão originada submetida pelo autor (manuscrito digital).
	Azul: permite o depósito da versão pós-print de um artigo.	A concessão para armazenar no repositório é permitida para os artigos publicados pela Revista, somente da versão pós-print. A versão final da publicação do artigo (avaliada pelos pares e corrigida pelos autores).
	Branca: apresenta restrições para o depósito das versões pré-print e pós-print de um artigo.	O armazenamento e acesso para disponibilizar os artigos da revista no repositório não é permitido.
	Verde: permite o depósito das versões pré-print e pós-print de um artigo	É concedida ao repositório institucional a permissão de armazenar os artigos publicados pela Revista na versão pré-print (versão original do manuscrito digital) ou a versão pós-print (versão de publicação do artigo avaliada pelos pares e corrigida pelos autores).
Licença de conteúdo	Não permite que copiem a obra publicada por sua Revista	A editora não permite que copiem a obra publicada por sua Revista.
	Não permite que distribuam a obra publicada por sua Revista	A editora não permite que distribuam a obra publicada por sua Revista.
	Permite cópia e distribuição da obra sem uma licença idêntica a do Editor	Uso não comercial. A Editora permite que outras pessoas copiem e distribuam as obras publicadas por sua Revista, mas somente para fins não comerciais.
	Permite distribuição e cópias exatas, mas não obras derivadas.	Não à obras derivadas. A editora permite que outras pessoas copiem e distribuam somente cópias exatas da sua obra, mas não obras derivadas (tais como tradução).

Fonte: IBICT (2015).

As políticas apresentadas são adotadas para o armazenamento e acesso de artigos de periódicos que estão sob a responsabilidade editorial de revistas para o depósito em repositórios institucionais. No entanto, mesmo que estejam sob a regência de uma política editorial deve se considerar as diferentes possibilidades para o acesso aberto no diretório. As permissões que são definidas pelo autor e editora são respeitadas pelo diretório. O Diadorim é um diretório que atende apenas artigos de periódicos para disponibilidade em repositórios institucionais não abrangendo os dados de pesquisa.

O Portal da Biodiversidade é outra iniciativa brasileira que disponibiliza os dados de biodiversidade mantidos pelo Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio) e Ministério do Meio Ambiente (MMA), com apoio da *Deutsche Gesellschaft für Internationale Zusammenarbeit* (GIZ), como parte do Projeto Monitoramento da Biodiversidade com Relevância para o Clima como parte integrante da Cooperação Brasil-Alemanha para o Desenvolvimento Sustentável.

O sistema foi desenvolvido a partir do *Atlas of Living Australia* (ALA), em parceria com pesquisadores da Escola Politécnica da Universidade de São Paulo (EPUSP), Escola de Artes, Ciência e Humanidades da Universidade de São Paulo (EACH/USP) e do Instituto de Geociências e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho” (IGCE/UNESP), com apoio do Núcleo BioComp/USP e da Fundação de Desenvolvimento Tecnológico da Engenharia (FDTE) da USP.

O portal é uma iniciativa conjunta lançada em 2015 a partir de bases de dados de Biodiversidade do ICMBio integradas. Possibilita o “acesso público via *internet* e a utilização de dados de biodiversidade atualmente disponíveis nos bancos de dados do ICMBio, a partir de buscas textuais e geoespaciais, visualização e *download* de registros de ocorrência de espécies” (BRASIL, 2015, *on-line*).

O ICMBio instituiu a Política de Dados e Informações sobre Biodiversidade do Instituto Chico Mendes de Conservação da Biodiversidade através da Instrução Normativa nº 2, de 25 de novembro de 2015. A política regulamenta a disponibilidade, o acesso e uso de dados e informações custodiadas em bases e sistemas de informação. Nessa instrução são considerados os autores (pessoa física e jurídica), dados de biodiversidade e o período de carência para disponibilidade dos dados (até cinco anos). O período de carência ou embargo obedece aos critérios e prazos estabelecidos na instrução.

O acesso é aberto as informações e documentos que são oriundos do planejamento de ações das unidades de conservação, publicações (técnicas e científicas) que envolvem análises e sínteses de informação sobre animais e plantas (níveis taxonômicos) e dados sensíveis. A regulamentação para o período e formas de restrição dos dados sensíveis são regulamentados por ato administrativo específico. Institui ainda que os dados e informações disponibilizados, devem ser identificadas para reconhecimento da autoria (autor e instituição) e fonte original (sistemas de informação).

A incidência maior de repositórios institucionais estão associados às universidades como um instrumento para disponibilizar a sua produção intelectual. Além das universidades alguns institutos de pesquisa possuem iniciativas de promover o acesso aberto às publicações

científicas e oferecem serviços de apoio. Ainda não há uma realidade consolidada para repositórios especializados em dados científicos.

Este capítulo identificou as iniciativas nacionais acerca das políticas e legislações que estão relacionados com os direitos autorais e dados abertos no âmbito brasileiro.

A primeira observação refere-se a LDA e a incompatibilidade com as práticas de acesso aberto e da *e-Science*. Já a LAI, com seus Decretos e Instruções Normativas avançaram e possibilitaram a ampliação de tornar a informação pública aberta e acessível. A informação pública e os dados abertos governamentais estão regidos pela LAI e por decretos específicos que apregoam iniciativas para o plano nacional do governo aberto e a transparência pública. A consolidação dos repositórios institucionais apresentam um avanço para a disponibilidade da produção científica aberta no Brasil.

A segunda observação refere-se as iniciativas institucionais e governamentais emergentes para a abertura de dados científicos, a exemplo do ICMBio. O engajamento entre governo, institutos de pesquisa e universidades aponta para um caminho promissor em prol de uma institucionalização da ciência aberta no Brasil. Muito embora não foi identificada uma legislação específica que aponte para direitos e atribuições autorais para dados científicos no cenário brasileiro. A existência de políticas editoriais que estão sendo conduzidas no escopo dos projetos capitaneados pelo IBICT já aponta para uma orientação nesse sentido.

7 DIRETRIZES INTERNACIONAIS POLÍTICO-NORMATIVAS PARA UMA CIÊNCIA ORIENTADA AOS DADOS CIENTÍFICOS

O capítulo dar continuidade às reflexões políticas e teóricas da categoria relacionada à Propriedade Intelectual (PI). Nessa direção as dimensões política e teórica norteiam o capítulo na busca de identificar as diretrizes internacionais político-normativas estabelecidas para os Direitos de Propriedade Intelectual na *e-Science*.

No cenário internacional as diretrizes político-normativas que contemplam uma ciência orientada para os dados científicos, já estão sendo concretizadas. Diretrizes institucionais, arcabouços normativos, legais e orçamentos são direcionados como instrumentos norteadores pelas agências de fomento, institutos de pesquisa, instituições de ensino e organizações governamentais para a condução de iniciativas e ações da *e-Science*. As diretrizes institucionais orientam também para uma nova dinâmica dos DPIs nesse contexto.

O foco da comunicação científica antes concentrado na divulgação dos resultados validados em meios formais de publicação, atualmente se concentra nos dados oriundos da investigação científica. Assim, qualquer mudança desse porte importa pensar nos impactos políticos, econômicos, legais, éticos, culturais e sociais.

7.1 ESCOPO INTERNACIONAL POLÍTICO-NORMATIVO

A seção parte da premissa que a informação científica e tecnológica é um bem público e, por conseguinte, se constitui em um elemento de interesse da coletividade e merecedor de formulações normativas e legais. Isto posto, é mister compreender o arcabouço político, normativo e legal voltado à regulação de dados de pesquisa no cenário internacional.

No que tange aos DPIs as diretrizes estão inseridas nas políticas orientadas por iniciativas desenvolvidas por agências de fomento e/ou financiamento, institutos de pesquisa, universidades, instituições governamentais e organizações não governamentais. A finalidade é gerar um escopo comum, balizador e norteador que direcione aplicações políticas, legais, normativas e éticas para a garantia de autoralidade de obras intelectuais, artísticas e culturais que envolvem os DPIs. Além das legislações específicas de cada país, os instrumentos internacionais apoiam e convencionam assuntos comuns por meio de convenções, tratados, acordos, diretrizes e mecanismos diretivos.

Independente dos instrumentos e legislações específicas de cada país, as novas dinâmicas da ciência ultrapassam as barreiras geográficas e legais. Avançam nos mais

diferentes segmentos da sociedade gerando novos fenômenos para a seara do Direito. A exemplo das práticas científicas que trouxeram mudanças no arcabouço jurídico provocadas pelo movimento de acesso aberto, proteção de direitos digitais, licenças públicas e, recentemente a alteração no novo fazer da pesquisa científica que têm provocado mudanças no escopo normativo-legal. As novas práticas ocasionam:

[...] uma quebra de paradigma entre os bens tangíveis e intangíveis. O conhecimento é um ativo intangível, impulsionador de desenvolvimento, que cada vez mais necessita ser codificado, apropriado para ser negociado. E, é nesse contexto, que os novos acordos internacionais sobre propriedade intelectual estruturam as suas diretrizes, dependentes das entidades que os administram e dos seus consequentes interesses (PERALTA et al., 2011, p. 126).

Os ativos tangíveis e intangíveis são de algum modo protegidos por DPIs. O conhecimento, como mencionado anteriormente, é um ativo intangível que impulsiona o desenvolvimento econômico, tecnológico e científico de um país. Dados de pesquisa também é considerado um ativo, possui valor agregado quando compartilhado e direcionado para além do desenvolvimento científico, impulsionar aspectos econômicos, tecnológicos e de outras áreas de sustentabilidade do país. Dessa forma há uma necessidade de aprimoramento dos instrumentos de DPIs para acomodar na sociedade da informação um outro componente da era digital, os dados de pesquisa.

Os atuais mecanismos normativos internacionais vigentes em acordos, tratados e convenções abordam assuntos comuns e buscam balizar para questões comuns e complexas dos DPIs. Muito embora, ainda seja predominante a expropriação econômica e privada nos sistemas vigentes. Em sentido oposto, várias iniciativas têm emergido em prol da ciência aberta e da condução de novas práticas para apoiar a ciência aberta no âmbito das agências de fomento, institutos de pesquisa, universidades e instituições governamentais. O estudo traz uma reflexão sobre os principais instrumentos regulatórios internacionais, com o intuito de levantar as questões direcionadas as novas práticas científicas.

7.1.1 Convenção Unionística de Berna (CUB)

A Convenção de Berna é considerada uma das convenções mais importantes e antigas que abrange os DPIs no âmbito internacional. As discussões iniciais se deram na Convenção de Paris, aprovada em 1883, direcionada à propriedade industrial no sentido mais amplo, incluindo patentes, marcas (logotipos), desenhos industriais, modelos de utilidade, marcas de

serviço, nomes comerciais, indicações geográficas e repressão à concorrência desleal.

Posteriormente, foi instituída a Convenção Unionística de Berna (CUB) em 1886 com a fusão da Convenção de Berna e a Convenção de Paris. A Convenção de Berna foi assinada de forma multilateral e definiu o escopo e abrangência da proteção relativa ao direito de autor, que até então era voltada para os direitos de patente. A CUB é baseada em três princípios básicos e contém uma série de disposições que determinam a proteção mínima a ser concedida, bem como disposições especiais disponíveis para os países em desenvolvimento que desejam fazer uso deles.

Inicialmente, a Convenção de Berna procurava garantir a proteção das obras denominadas de Belas Artes (pintura, escultura, gravura, arquitetura), obras literárias (livros em toda a generalidade de seus temas e peças teatrais) e depois obras científicas que foram incorporadas. Atualmente a convenção no Artigo 2 estabelece a proteção das produções de domínio literário, científico e artístico (qualquer que seja o modo ou forma de expressão), obras multimídias e produções tecnológicos. A base do sistema autoralista é o princípio norteador da convenção, onde:

A proteção dos direitos autorais que recai sobre todas as criações do domínio literário e artístico que possam ser tangibilizadas de alguma forma. Este é o princípio norteador advindo da Convenção da União de Berna, a primeira Convenção Internacional a estabelecer diretrizes e padrões mínimos de proteção às obras autorais para todos os partícipes da União (PERALTA et al., 2011, p. 119).

Muito embora a convenção seja uma diretriz internacional a sua adoção não é mandatória, visto que cada país possui suas legislações específicas e podem, ou não, optar pela adoção das recomendações. A Convenção de Berna é um dos principais instrumentos de balizamento das questões autorais no âmbito internacional. No Brasil a convenção foi promulgada pelo Decreto nº 75.699/1975. Muito embora seja considerada um instrumento importante para os DPIs no âmbito internacional, a convenção ainda não prevê diretrizes contemporâneas em conformidade com a *e-Science*.

7.1.2 *Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS)*

Destaque para outro instrumento importante na esfera internacional denominado de *Trade Related Aspects of Intellectual Property Rights (TRIPS)*, administrado pela *World Trade Organization (WTO)* conhecida como Organização Mundial do Comércio (OMC).

A WTC é a única organização internacional global que trata das regras do comércio entre as nações. O objetivo é ajudar os produtores de bens e serviços, exportadores, importadores na proteção de suas produções e na condução de seus negócios. Define que os DPIs são os direitos concedidos as pessoas pelas criações de suas mentes. Eles costumam dar ao criador um direito exclusivo pela criação por tempo determinado. O acordo aborda duas áreas: direito de autor e conexos e propriedade industrial (WTC, 2015).

O Acordo TRIPS objetiva o estabelecimento de padrões e princípios adequados à existência, abrangência e exercício dos DPIs relacionados ao comércio. Visa a proteção da PI para a contribuição da inovação técnica e transferência de tecnologia. Busca reduzir as lacunas na forma como esses direitos são protegidos mundialmente e harmonizar as regras comuns internacionalmente. Estabelece também padrões mínimos de proteção para todos os campos tecnológicos. É destinado aos estados-membros e tem caráter facultativo.

Paesani (2012) afirma que o Acordo TRIPS estabelece padrões mínimos regulatórios para aplicação às legislações nacionais de cada país. É um acordo de normas resultantes de discussões entre membros que permitem sua aplicação racional e livre para que possam estar de acordo com os seus próprios ordenamentos jurídicos.

O direito de autor engloba as obras literárias e artísticas (como livros e outros escritos, composições musicais, pinturas, esculturas, programas de computador e filmes) por um período mínimo de 50 (cinquenta) anos após a morte do autor.

Os direitos conexos são reservados aos artistas (atores, cantores e músicos), produtores de fonogramas (registros de som) e organismos de radiodifusão. Nesta seara, a proteção tem a finalidade de proteger os direitos conexos como incentivo e recompensa ao trabalho criativo. Abrange o escopo do Direito do Autor e *Copyright*.

A propriedade industrial envolve várias áreas: patentes, sinais distintivos, marcas comerciais (distingue os produtos e serviços de uma empresa) e indicações geográficas (identificam um produto como originário de uma localidade com características e atributos à sua origem geográfica).

A proteção de sinais distintivos visa estimular e assegurar a concorrência leal e proteger os consumidores, permitindo-lhes fazer escolhas entre vários produtos e serviços. A proteção pode durar indefinidamente desde que o sinal em causa continue a ser distinto.

Outros tipos de propriedade industrial são protegidos principalmente para estimular a inovação, *design* e criação de tecnologia. Recaem nesta tipologia, as invenções (patentes e desenhos industriais) e segredos comerciais.

Além das proteções acima mencionadas, o acordo prevê que o regime de PI em vigência, facilite a transferência de tecnologias sob a forma de investimento direto estrangeiro, *joint ventures* e licenciamento com a finalidade de equilibrar o interesse legítimo dos titulares de direitos e utilizadores.

No contexto contemporâneo já se discute formas práticas para os desafios existentes nos processos de inovação e acesso aberto. Reconhece a complexidade entre a relação comercial entre áreas como, tecnologias médicas, patentes, inovação, acesso e o uso de licenças compulsórias e licenciamento voluntário, por exemplo. Porém, ainda não disponibiliza direcionamentos para as modalidades emergentes de publicações eletrônicas e dados de pesquisa.

7.1.3 *World Intellectual Property Organization (WIPO)*

Em 1967 foi criada a *World Intellectual Property Organization* (WIPO) conhecida como Organização Mundial de Propriedade Intelectual (OMPI). A WIPO é o fórum global para serviços de propriedade intelectual, política, informação e cooperação. É agência de autofinanciamento das Nações Unidas com 188 estados-membros. Tem como missão liderar o desenvolvimento de um sistema internacional de Propriedade Intelectual equilibrado e eficaz que permite a inovação e a criatividade para benefício de um coletivo (WIPO, 2015, *on-line*).

Define que a Propriedade Intelectual (PI) refere-se às criações da mente, tais como: invenções, obras literárias e artísticas, desenhos e símbolos, nomes e imagens usadas no comércio (patentes, direitos autorais e marcas comerciais) que permitem que pessoas ganhem reconhecimento ou benefício financeiro a partir de suas criações e invenções (WIPO, 2015).

Abrange diretrizes, serviços e sistemas de apoio à institucionalização dos DPIs nos países-membros. Busca proporcionar uma base harmônica entre o justo equilíbrio do interesse privado dos inventores e o interesse público. O sistema da PI fomenta um ambiente para que criatividade e inovação floresçam. No que tange ao *copyright* (ou direito de autor) é definido com termo jurídico usado para descrever os direitos que os criadores têm sobre as suas obras literárias e artísticas. A WIPO apresenta flexibilidade ao acomodar diferentes tipos de direitos de diferentes formatos de publicações.

A WIPO criou em 1996 dois tratados que apoiam o ambiente digital tentando adequar a legislação atual às novas formas de uso para as obras protegidas, são eles: *WIPO Performances and Phonograms Treaty* (WPPT) e *WIPO Copyright Treaty* (WCT).

O *WIPO Performances and Phonograms Treaty* (WPPT) aborda alterações referentes aos direitos vizinhos e conexos do autor. O arcabouço legal presente no WPPT trata da regulação do aluguel de material protegido pelos direitos conexos, complementando o Tratado de *Copyright* da Organização Mundial da Propriedade Intelectual (WTC). Faz referências às disposições encontradas na Convenção de Roma que abriga a proteção aos artistas, intérpretes, produtores de fonograma, organismos de radiodifusão, além de programas de computador. A WIPO oferece alguns serviços de suporte através de banco de dados conforme demonstrado no Quadro 10:

Quadro 10 – Banco de dados da WIPO

SISTEMAS	DESCRIÇÃO
PATENTSCOPE	<i>Patent Cooperation Treaty (PCT)</i> é um banco de dados que fornece acesso a aplicativos internacionais em formato de texto completo no dia da publicação, bem como aos documentos de patentes de participação de escritórios nacionais e regionais de patentes. (Website: http://www.wipo.int/patentscope/en/).
Global Brand Database	É o banco de dados da marca global que permite busca fácil em milhões de registros relacionados com marcas protegidas internacionalmente, denominações de origem e rolamentos heráldicos, bandeiras e outros emblemas de Estado, bem como os nomes, siglas e emblemas de organizações intergovernamentais. (Website: http://www.wipo.int/reference/en/branddb/).
MADRID	O <i>International Trademark Research</i> é um banco de dados que permite a busca de informações detalhadas sobre todas as marcas internacionais registradas ao abrigo do sistema de MADRID da WIPO. (Website: http://www.wipo.int/reference/en/branddb/).
HAGUE	Através de uma única interface, intuitiva, o <i>Global Design Database</i> permite buscas livres e simultâneas de projetos industriais registrados no âmbito do sistema de HAGUE, conhecido internacionalmente como sistema de Haia. (Website: http://www.wipo.int/ipdl/en/6ter/search-struct.jsp).

Fonte: Elaborado pela autora a partir do website WIPO (2016).

Diversos países são beneficiados pelos sistemas da WIPO e possuem escritórios que promovem os serviços internacionais de patentes, marcas, desenhos, direitos autorais, além de oportunizar a resolução e arbitragem alternativa de litígios em PI. Os escritórios externos ainda se beneficiam de bancos de dados, atividades de investigação, desenvolvimento, divulgação e capacitação além da visibilidade de processos de inovação e criatividade no país.

No Brasil o escritório foi inaugurado em 2009 e trabalha com instituições locais, com o intuito de criar valor duradouro de ativos. Atualmente o órgão responsável no Brasil é o Instituto Nacional de Propriedade Intelectual (INPI).

No que se refere aos direitos relacionados com a ambiência digital as recomendações da WIPO estão no entendimento e expansão de domínio público. A *WIPO Development Agenda* discorre as recomendações 16 e 20. A recomendação 16 orienta a preservação do domínio público dentro do processo normativo da WIPO e sugere aos Estados-Membros

aprofundar a análise das implicações e benefícios ao acesso às publicações de domínio público. A Recomendação 20 busca a promoção de atividades normativas do domínio público e a PI, incluindo a possibilidade de preparar diretrizes para dar assistência aos Estados-Membros e identificar matérias que caíram em domínio público das respectivas jurisdições.

As políticas tanto nacionais como internacionais sempre concentraram seus ativos (tangíveis e intangíveis) regidos pelos direitos exclusivos da propriedade intelectual. Agora a atenção está mudando para as limitações da PI e promoção de bens intelectuais, não-exclusivos, como por exemplo as produções de domínio público (DUSOLLIER, 2010).

O estudo de Dusollier (2010) intitulado *Scoping Study on Copyright and Related Rights and the Public Domain* apresenta uma compilação de pesquisas realizadas sobre o paradoxo do domínio público e os DPIs. O referido documento reflete a intenção da WIPO e de inúmeros países acerca do assunto. Demonstra que a vertente do domínio público inicialmente reforçado com os movimentos *open access* e a cultura do *remix*, por exemplo, causaram um estremecimento nos sistemas autorais vigentes. As organizações enfrentam a necessidade de ajustar os mecanismos regulatórios, diante do valor intrínseco do domínio público como matéria-prima para o processos de criação, inovação e desenvolvimento (científico, literário e artístico).

O estudo também observou a perda econômica de produtos e serviços regidos por DPIs, os custos econômicos provocados pelo domínio público e o impacto na sociedade mercantilista. No que tange aos dados, o documento traz as considerações voltadas para os metadados de obras já publicadas e que podem ser utilizadas para acesso em domínio público. Alternativas particulares de proteção de direitos digitais, a exemplo dos sistemas de gestão de direitos digitais e do sistema de informação ccREL do projeto da *Creative Commons* estão cada vez mais, sendo utilizados para informar o *status* de domínio público de uma obra e suas condições de licenciamento aberto por meio dos metadados.

No âmbito da WIPO as discussões giram em torno da promoção de domínio público de obras não-exclusivas. Mesmo diante de tais movimentos prevalece indubitavelmente, o esforço em manter os mecanismos reguladores que promovem a duração e a extensão dos direitos (autorais e inventivos) com o aprimoramento de medidas tecnológicas de proteção e a continuidade da expropriação econômica dos ativos (tangíveis e intangíveis).

Em sentido oposto, como mencionado anteriormente, iniciativas emergem como prática mandatória no âmbito internacional para tornar os dados de pesquisa, acessíveis e públicos. Além do compartilhamento e gerenciamento de dados de pesquisa é necessário que os DPIs se acomodem nessa nova dinâmica de empreendimento global.

Ademais, em alguns países já existe uma compreensão por parte dos pesquisadores, de agências de fomento e instituições de pesquisa que os dados de pesquisa possuem valor quando são devidamente tratados, preservados e gerenciados. São potenciais ativos para a ciência (SAYÃO, SALES, 2015).

A adoção de práticas mandatórias e rotineiras são iniciativas que vão na contramão dos DPIs no âmbito privado. Agências de financiamento, institutos de pesquisa, universidades e instituições governamentais estão exigindo que os dados de projetos de pesquisa financiados por eles estejam em domínio público (ICPSR, 2015). As diretrizes por vezes mandatórias têm sido estabelecidas no âmbito internacional para direcionar a ciência aberta como guarda-chuva no arcabouço político-normativo-legal e técnico-tecnológico para dar sustentabilidade às práticas da *e-Science*.

7.2 DIRETRIZES POLÍTICO-NORMATIVAS PARA AS PRÁTICAS CIENTÍFICAS DA E-SCIENCE

O movimento da ciência aberta está em expansão e avança no sentido de promover mecanismos e instrumentos na seara político-normativo-legal e dos aspectos técnicos e tecnológicos. Nos Estados Unidos, o documento endereçado aos chefes de departamentos e agências executivas do governo americano, expedido pelo Departamento Executivo de Política de Tecnologia e Ciência da Casa Branca, denominado de *Memorandum for the Heads of Executive Departments and Agencies*⁷, atualizado em 04 de maio de 2016 e afirma que a investigação científica apoiada pelo Governo Federal catalisa avanços inovadores que impulsionam a economia e se tornam combustível para novas investigações e inovações.

O *Memorandum for the Heads of Executive Departments and Agencies* estabelece a política de dados abertos como estratégia de gestão dos ativos intangíveis. Isso se constitui em uma política mandatória na qual estabelece os critérios mínimos para a interoperabilidade, plataformas abertas, investimentos necessários para adequar os sistemas existentes, responsabilidades dos atores envolvidos em todo o ciclo de vida da informação e dos dados e mecanismos de responsabilização. Os resultados dessa pesquisa se tornam o combustível para decisões governamentais e são ativos para o progresso em áreas como saúde, energia, ambiente, agricultura e segurança nacional.

⁷ Ver: <https://www.whitehouse.gov/sites/default/files/omb/memoranda/2016/m-16-11.pdf>

A *National Science Foundation* (NSF) é a principal agência de fomento americana. Financia as iniciativas que objetivam abordar aspectos essenciais vinculados ao tratamento, curadoria e compartilhamento de dados de pesquisa para uso e reuso. Adiciona-se outras agências de fomento no escopo americano: *National Academic of Science* (NAS), *National Academic of Engineering* (NAE), *National Research Council* (NRC) e a *National Science Board* (NSB). Tais agências já disponibilizam que iniciativas que contemplam a ciência aberta deve seguir a política mandatória governamental.

A importância das agências de fomento para o desenvolvimento da ciência se refere a todas as entidades - locais, nacionais e internacionais; governo, sem fins lucrativos, e entidades com fins lucrativos - que fornecem apoio financeiro para a produção de dados, arquivamento, gestão e utilização (NSB, 2005). Além do apoio financeiro, essas agências sustentam as compilações de dados que residem em organizações de investigação e de ensino (como é típico para pesquisas financiadas pelo NSF) e aquelas que dão suporte as pesquisas internas dentro das agências financiadoras.

Criada pelo Congresso Nacional em 1950, a *National Science Foundation* (NSF) é uma agência de fomento à ciência e tornou-se a principal financiadora de incentivo à pesquisa. Abrange cerca de 20% (vinte por cento) dos financiamentos de projetos de pesquisa de universidades e institutos de pesquisa americanos.

A *National Academic of Science* (NAS) foi fundada em 1863, por uma lei do Congresso Nacional, assinada pelo presidente Abraham Lincoln no auge da Guerra Civil. É uma sociedade privada, sem fins lucrativos e composta por pesquisadores ilustres. Sua principal missão é fornecer suporte para a pesquisa sobre assuntos relacionados à ciência e tecnologia.

Fundadas em 1964 a *National Academic of Engineering* (NAE) e em 1970, a *National Academy of Medicine* (NAM) são instituições privadas, independentes, sem fins lucrativos que fornecem um escopo para as áreas de pesquisa da Engenharia e Medicina, respectivamente. A missão de ambas é realizar o avanço em prol do bem-estar da nação, promovendo a profissão e mobilizando os conhecimentos e *insights* de profissionais que fornecem aconselhamento ao governo federal de questões que envolvem as áreas de Engenharia, Tecnologia e Medicina.

Além de atender as suas especificidades, as academias juntas e de forma objetiva e independente, possuem a finalidade de fornecer subsídios científicos e aconselhamento para a resolução de problemas complexos que substanciem as decisões de políticas públicas do governo e da comunidade científica americana.

Acrescenta-se a este rol a *National Research Council* (NRC) fundada em 1916. É uma instituição privada, sem fins lucrativos, que produz relatórios sobre as políticas e opiniões públicas para a promoção da Ciência em áreas como Engenharia, Medicina e Tecnologia no âmbito das agências de fomento e no bojo das academias.

No que corresponde ao contexto contemporâneo na era digital, existem dois documentos elaborados em parceria com as referidas academias e publicado pela NAS. Em 2012 foi publicado o documento intitulado *For Attribution - Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop*. Este documento foi resultante de um seminário realizado em agosto de 2011 na Califórnia. Reuniu diversos especialistas engajados em prol da *e-Science* e dos dados de pesquisa na era digital. Abrange além dessas questões as novas dinâmicas que envolvem os DPIs. Orienta para as novas abordagens baseadas na atribuição e citação de autoria.

O segundo documento publicado em 2013 intitulado de *Committee on the Impact of Copyright Policy on Innovation in the Digital Era* apresenta a abordagem política de direitos autorais orientada à infraestrutura de dados e os seus dilemas na era digital. Fruto de políticas e orientações estabelecidas por agências de fomento e relatórios resultantes de conferências acerca da temática, os documentos mencionados serão norteadores para o aprofundamento do capítulo da análise de conteúdo.

No âmbito da Comunidade Europeia excelentes iniciativas estão em condução para a *e-Science*. As iniciativas são realizadas em parcerias institucionais e governamentais apoiadas por institutos de pesquisa, universidades e agências financiadoras. Instrumentos regulatórios e normativos para a condução das práticas de *e-Science*, tem por finalidade, além tratar da infraestrutura (metodológica, técnica e tecnológica), atender aos aspectos de proteção autoral, privacidade e assuntos governamentais.

Destaque inicial à *Organization for Economic Co-operation and Development* (OECD). É uma instituição não governamental que reúne diversas nações por meio de um fórum único, representado por trinta governos que trabalham em conjunto para enfrentar os desafios econômicos, sociais e ambientais globalizados. Tem como missão unir esforços para compreender e ajudar as nações-membros a lidarem com assuntos importantes, tais como: governança corporativa, economia da informação e os desafios do envelhecimento populacional.

Em 2007, a OECD publicou o documento intitulado *OECD Principles and Guidelines for Access to Research Data from Public Funding* com o objetivo de promover uma cultura de abertura e compartilhamento de dados de pesquisa entre as comunidades científicas dos

países membros da OECD. O referido documento reúne as primeiras iniciativas para a pesquisa aberta global como resultado de uma reunião ocorrida em Paris (2004) com a participação de trinta ministros de Ciência e Tecnologia e representantes de quatro países. A reunião deflagrou o desenho de um documento com um conjunto de princípios e orientações para o empreendimento de políticas de pesquisas institucionais.

As diretrizes foram definidas para orientar as instituições que precisam de políticas e cooperação de pesquisas internacionais. Esse documento contempla a investigação científica inovadora como um papel crucial na resolução dos desafios globais, tais como: saúde, mudança climática, energias renováveis e gestão de recursos naturais. Aborda a velocidade e a profundidade da pesquisa inovadora e o intercâmbio de colaboração entre as diferentes comunidades, assegurando assim a ampla divulgação na vertente da ciência aberta.

Na direção de uma ciência aberta, a OECD assumiu essa liderança para o desenvolvimento de princípios e diretrizes para facilitar o acesso e disponibilidade de dados de pesquisa gerados com financiamento público. Os países filiados a OECD não tem obrigação legal de seguir suas diretrizes, mas por sua adesão aos princípios da instituição e devido a interesses mútuos, acabam por aplicar em diferentes graus, tais diretrizes em seus países. Os princípios se constituem: abertura, flexibilidade, transparência, conformidade legal, proteção da propriedade intelectual, profissionalismo, interoperabilidade, qualidade, segurança, eficiência, responsabilização e sustentabilidade.

Considera-se que as recomendações contidas no referido documento contemplam um conjunto de princípios, padrões e objetivos que precisam ser implementados para uma política orientada aos dados de pesquisa. Possui efeito de um instrumento legal que aponta para os princípios éticos e morais que devem ser considerados.

O referido documento contempla orientações como um pressuposto para admissibilidade de política de ciência aberta e a acessibilidade aos dados de pesquisa abertos, que conduzem ao investimento público e criação de forte valor de inovação, reforçando os inquéritos na pesquisa aberta e encorajamento à diversidade de análise e opiniões. Busca incentivar a promoção de novas pesquisas, com a realização de possibilidades que gerem novas hipóteses ou alternativas de métodos de análises. Contempla as possibilidades de estudos que suportem métodos e mensuração das publicações e conjuntos de dados, com permissão da criação de novos conjuntos de dados quando questões múltiplas são combinadas. Além, de promover a educação de novos pesquisadores para a pesquisa com dados científicos (OECD, 2007).

Esses itens fazem parte de uma estrutura mínima que deve ser considerada para a elaboração de uma política orientada aos dados científicos. Visa oportunizar o acesso aberto, o compartilhamento e a colaboração científica entre pesquisadores que obtiveram financiamento público, cuja expectativa gira em torno de uma política de dados para a ciência aberta. O intuito é sempre a busca por novos investimentos científicos, econômicos e sociais em prol de uma sustentabilidade para a sociedade, além de ajudar a maximizar o potencial colaborativo das novas tecnologias digitais e redes de pesquisa colaborativas.

Além dos pressupostos acima apresentados o documento contempla doze princípios norteadores que direcionam para diferentes aplicações de investigações científicas com dados abertos. No que tange aos DPIs três princípios destacam-se para aprofundamento, a saber: conformidade legal, proteção da propriedade intelectual e a responsabilidade formal.

D. Conformidade Legal

Regime de acesso de dados deve respeitar os direitos legais e os interesses legítimos de todas as partes interessadas na empresa pública de pesquisa.

E. Proteção da propriedade intelectual

Acordos de acesso de dados devem considerar a aplicabilidade do direito de autor ou de outras leis de propriedade intelectual que possam ser relevantes para os bancos de dados da investigação financiada por fundos públicos.

F. Responsabilidade Formal

Os acordos de acesso devem promover práticas explícitas, formais e institucionais, tais como: o desenvolvimento de regras e regulamentos, a respeito das responsabilidades das várias partes envolvidas em atividades relacionadas a dados. Estas práticas devem dizer respeito à autoria, créditos de produção, posse, divulgação, restrições de uso, disposições financeiras, regras éticas, termos de licenciamento, a responsabilidade e arquivamento sustentável (OECD, 2007, p. 16, grifo nosso).

Esses três princípios conduzem as políticas e normativas, além de direcionar para aspectos éticos e legais acerca das garantias autorais de dados científicos. Além disso, os princípios contidos neste documento buscam orientar as nações, agências de financiamento e pesquisadores a lidarem com as barreiras e desafios do compartilhamento e acesso aos dados de científicos abertos. Questões jurídicas e políticas, tais como: leis nacionais, acordos internacionais dos DPIs e a proteção à privacidade, afetam diretamente o acesso aos dados e as práticas de compartilhamento. As práticas estão sendo pensadas para atender questões relacionadas com: autoria, créditos de produção, posse, divulgação, restrições de uso, disposições financeiras, regras éticas, termos de licenciamento, responsabilidade e arquivamento sustentável.

Ainda no âmbito europeu, o governo do Reino Unido publicou o *Data Protection Act* 1998, que se constitui em um ato legal que regula o processamento de informações relacionadas com indivíduos, incluindo a obtenção, detenção, utilização e divulgação dos dados abertos.

Nessa direção a Comissão Europeia por meio de sua *Directorate General for Research & Innovation* publicou o *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*. Utilizou das declarações: Budapest Declaration (2002) e Berlin Declaration (2003) para definir o acesso como o direito de ler, baixar, imprimir, copiar, distribuir, buscar, rastrear e minerar informação e dados. E assim, propuseram o *Model Grant Agreement*, o qual estabelece no Artigo 29.2, requisitos legais sobre o acesso aberto às publicações científicas, a exemplo do depósito em repositórios de todos os artigos que passaram pelo processo *peer-review*. Também está sendo proposto nesse documento o *Open Research Data Pilot* como forma de estimular pesquisadores a aderirem ao depósito e compartilhamento de dados de pesquisa no âmbito dos países europeus.

A parceria entre o Reino Unido e a Inglaterra é outra iniciativa que através das instituições *Libraries and Archives Copyright Alliance* (LACA) e *Chartered Institute of Library and Information Professionals* (CILIP) está promovendo a defesa e difusão dos direitos autorais de forma justa e equilibrada nesses países. A LACA e a CILIP apresentam algumas iniciativas e um grande corpo de evidências para apoiar os litígios acerca da propriedade intelectual e consultas da União Europeia sobre direitos de autor (LACA, 2015).

A LACA é uma aliança de membros de bibliotecas e arquivos que fiscaliza os direitos autorais e os direitos conexos, no Reino Unido e na Europa em nome de seus membros, usuários e organizações. É um membro que gerencia o *copyright* para as produções criativas e iniciativas de uma base ampla representada pela sociedade civil europeia, bibliotecas, indústria e criadores que procuram debater sobre como os direitos autorais podem apoiar a inovação, o acesso e a criatividade para todos os europeus de forma eficaz. A aliança é membro colaborador da *International Federation of Library Associations and Institutions* (IFLA) e do *European Bureau of Library, Information and Documentation Associations on Copyright Issues* (EBLIDA).

Ainda no Reino Unido, outra iniciativa que faz parte de serviços oferecidos pela *University of Nottingham* é chamada de SHERPA/RoMEO. O projeto RoMEO é fruto de um trabalho colaborativo com parceiros internacionais que contribuem para o desenvolvimento e manutenção de um serviço que promove políticas de informação e dados. O projeto é atualmente financiado pelo *Joint Information Systems Committee* (JISC), uma organização

sem fins lucrativos vinculada as instituições de ensino superior do Reino Unido que promovem educação e competências para os serviços de informação e soluções digitais. O projeto RoMEO dar sustentabilidade ao portal SHERPA/RoMEO que reúne políticas para o autoarquivamento de pesquisas científicas em repositórios institucionais.

As informações sobre propriedade intelectual são disponibilizadas pelo JISC. No escopo do *JISC Legal Information* inúmeras diretrizes e políticas de *copyright* e autoarquivamento são disponibilizadas para autores e editores. Uma peculiaridade da política é o gerenciamento dos direitos autorais por meio de um sistema de codificação por cores. A codificação de cores RoMEO obedece as permissões fornecidas pelos editores e atendem os critérios da via *green* (SHERPA/RoMEO, 2015).

A codificação por cores são definidas pelo projeto RoMEO e representam a política de autoarquivamento e acesso aberto que orientam os autores cadastrados. As políticas de acesso aberto e de *copyright* são apresentadas no sistema. O sistema permite a busca de informações em uma base de dados que identifica o *copyright* dos editores e orienta o *status* para o autoarquivamento das publicações (*pré-print* e *post-print*), estabelecendo as condições ou restrições impostas pelo editor acerca dos direitos autorais. O Quadro 11 apresenta um extrato da política de autoarquivamento.

Quadro 11 – Políticas de *copyright* e de autoarquivo de editores

Cores RoMEO	Política de arquivo
Verde	Pode arquivar a versão <i>pré-print</i> e <i>post-print</i> ou Versão/PDF do editor
Azul	Pode arquivar a versão <i>post-print</i> (i.e. o rascunho final após o <i>peer-review</i>) ou Versão/PDF do editor
Amarelo	Pode arquivar a versão <i>pré-print</i> (i.e. antes do <i>peer-review</i>)
Branco	O arquivo não é suportado formalmente

Fonte: SHERPA-RoMEO (2015).

As restrições são gerenciadas pelo sistema que bloqueia o acesso às publicações (*e-print*) nos casos de embargo (divulgação pública ou acesso controlado por senha) e direitos de acesso com *status* de arquivamento (parcial e total).

A política de arquivamento e de acesso aos dados de pesquisa já é contemplada no SHERPA. Na política é estabelecido o período de embargo para a disponibilidade da publicação e abertura para acesso aos dados. O período de embargo pode variar de seis meses a cinco anos. Os cumprimentos de tais prazos variam de acordo com o domínio específico da disciplina. Por exemplo, “o período de **seis meses** poderia ser mais adequado para um projeto de química, mas uma escavação arqueológica poderia exigir vários anos, o que temos fixado em **cinco anos**” (SHERPA/RoMEO, 2015, *on-line*, grifo nosso). O embargo (tempo de

carência) garante a exclusividade ao(s) pesquisador(es) dos dados científicos (originários).

Ainda no escopo do SHERPA/RoMEO existe o projeto denominado JULIET que tem a finalidade de orientar administradores e autores para o uso de repositórios institucionais. JULIET apoia e suporta a pesquisa de acesso aberto, gerencia os acordos de transferência de direitos autorais estabelecidos por editores, financiadores e a relação com as políticas de arquivamento nos repositórios.

O JISC disponibiliza diretrizes não mandatórias para a normatização de DPIs, definindo que “são direitos concedidos aos criadores e proprietários de obras que resultam da criatividade intelectual humana. Estas obras advêm do domínio industrial, científico, literário e artístico” (JISC, 2016, *on-line*, tradução nossa). As recomendações do JISC abrangem os direitos de *copyright*, patentes, marcas, *design* e banco de dados. Quanto aos direitos de uso e reuso, o JISC fornece diversos recursos digitais e orientações para aplicação aos diferentes tipos de publicações, tais como: dados de pesquisa, mapas, filmes, imagens, jornais, *e-books*, repositórios e recursos educacionais e de descobertas. Disponibiliza o compartilhamento de dados geoespaciais para reuso. No tocante aos dados abertos o JISC apoia diferentes recomendações para uma melhor aplicação da PI com orientações, *blogs*, eventos, produtos, serviços, tecnologias, licenças e recomendações que norteiam a política de acesso aberto aos dados de pesquisa. Para a proteção de dados em *databases* as recomendações são:

Um banco de dados, que é uma coleção de dados ou outro material que seja organizado de tal maneira que os itens sejam individualmente acessíveis, pode ser protegido por direitos de autor como uma obra literária e /ou para direito de banco de dados. Esta proteção pode aplicar-se tanto para o papel como para as bases de dados eletrônicas (JISC, 2015, *on-line*, tradução nossa).

Defende firmemente uma “PI mais liberal ‘aberta’ que possa suportar utilizações inovadoras de ativos digitais. Para universidades e faculdades, a orientação é permitir e incentivar a digitalização, preservação, acesso e reutilização de conteúdos digitais [...]” (JISC, 2015, *on-line*, tradução nossa). Recomenda que a proteção contratual de direitos autorais não impeçam a inovação e o crescimento educacional e científico.

Na Holanda o *Dutch Ministry of Education, Culture and Science* apoia a *Netherlands Organisation for Scientific Research* (NWO) que é o Conselho Nacional de Pesquisa e financiador de orçamento de 625 milhões de euros por ano. A NWO direciona a ciência holandesa por meio de programas e gestão dos processos de conhecimento, promovendo a qualidade e a inovação.

No escopo político, a NWO como agência de fomento e de financiamento promove a ciência aberta para tornar o livre acesso aos resultados de pesquisa pagos por fundos públicos em todo o mundo. Se aplica as publicações e outras formas de produção científica, inclusive o compartilhamento de dados de pesquisa. No âmbito governamental holandês, foi instituída pela Secretaria de Estado da Educação, Cultura e Ciência, regulamentações que determinam que até 2018, 60% e até 2024, 100% das publicações científicas financiadas com dinheiro público devem ser enquadradas na classificação *Gold Open Access*. A classificação assegura a “qualidade e acesso aberto de publicações e dados científicos. As orientações governamentais instituídas a partir de 1º de Janeiro de 2015 é no sentido de que todas as publicações oriundas de financiamento público devem ser de acesso livre e gratuito” (NWO, 2016, *on-line*, tradução nossa).

Dentre os oito institutos de pesquisa que o NWO administra o *Data Archiving and Networked Services* (DANS), afirma que dados científicos comportam diversos formatos, como, gravações de voz, questionários preenchidos e amostras de solo, desde que provenientes de resultados de pesquisa científica com muita informação valiosa. A rede DANS em parceria com a *Royal Netherlands Academy of Arts and Sciences* (KNAW) e a organização *Netherlands eScience Center* (NLeSC) desenvolveram o *National Academic Research and Collaborations Information System* (NARCIS) para apoiar o arquivamento sustentável de publicações e dados de pesquisa da comunidade científica holandesa.

O NARCIS foi constituído em 2007 e está sob a responsabilidade da DANS desde 2011. O objetivo da DANS é promover o acesso sustentável aos dados de pesquisas digitais e incentivar pesquisadores científicos para arquivar e reutilizá-los de forma sustentável e por meio do arquivamento *on-line*. É um repositório nacional que alberga publicações científicas de pesquisadores holandeses, que incluem acesso aberto aos repositórios institucionais das universidades holandesas, institutos de pesquisa e academias de ciência. Disponibiliza também dados de pesquisa, publicações ampliadas, *datasets* e documentos das descrições de projetos de pesquisa holandesas.

Questões de PI e acesso são direcionadas pela NWO como um desafio para tornar os dados científicos abertos e públicos. Orienta que os dados científicos sejam abertos, bem descritos (metadados), armazenados de forma inequívoca, sustentáveis e acessíveis para outros pesquisadores. A NWO está ciente de que informações confidenciais devem ser protegidas. Por esta razão, o critério aplica-se para os dados, em particular: “**aberto, onde possível, protegida onde necessário**” (NWO, 2016, *on-line*, grifo nosso).

Como preparação para uma política efetiva de promoção à ciência aberta e gerenciamento de dados, a NWO iniciou em Janeiro de 2015 um projeto piloto intitulado *Start Pilot Data Management*. O projeto reflete a preocupação sobre os direitos digitais que afetam pesquisadores, professores e estudantes, sob dois aspectos: o primeiro é sobre a obtenção dos direitos para os próprios dados (primários) e segundo é o reuso dos dados de outros pesquisadores. Para dirimir essas questões a NWO conta com o desenvolvimento e adoção de ferramentas, serviços e licenças em parceria como a iniciativa SURF.

A SURF é uma organização colaborativa de Tecnologias de Informação e Comunicação de apoio à educação superior e pesquisa holandesa, que oferece aos alunos, professores, cientistas e publicadores o acesso para as melhores facilidades das TIC e da *internet* e “apoia o gerenciamento dos direitos digitais no escopo dos direitos autorais, licenças, acesso aberto, reutilização de dados, imagens e áudio e aspectos legais dos materiais utilizados em repositórios, ensino e pesquisa” (SURF, 2015, *on-line*, tradução nossa). Oferece um portfólio de serviços, ferramentas abertas e identificadores persistentes para apoiar pesquisadores, educadores e alunos.

O OpenAIRE é outra importante organização que representa uma rede de repositórios de acesso livre, arquivos e periódicos de apoio às políticas de acesso aberto da comunidade europeia. Tem o papel de promover a agregação de agências financiadoras, provedores da pesquisa científica com as publicações e seus dados de pesquisa. Fornece uma ampla gama de serviços, depósitos e estatísticas. A política para a promoção da ciência aberta é contemplada pelo projeto *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020* da OpenAIRE.

As diretrizes tem a finalidade de promover a Política da Comissão Europeia sobre investigação de dados de pesquisa. A premissa básica do projeto é que a informação já paga pelo erário público não deve ser paga novamente cada vez que ela for acessada ou usada, e que deve beneficiar no máximo as empresas e os cidadãos europeus. Isso significa fazer com “financiamento público, informações científicas disponíveis *on-line*, sem custo adicional, para os investigadores europeus, indústrias inovadoras e cidadãos, assegurando ao mesmo tempo a preservação em longo prazo” (OpenAIRE, 2016, *on-line*, tradução nossa).

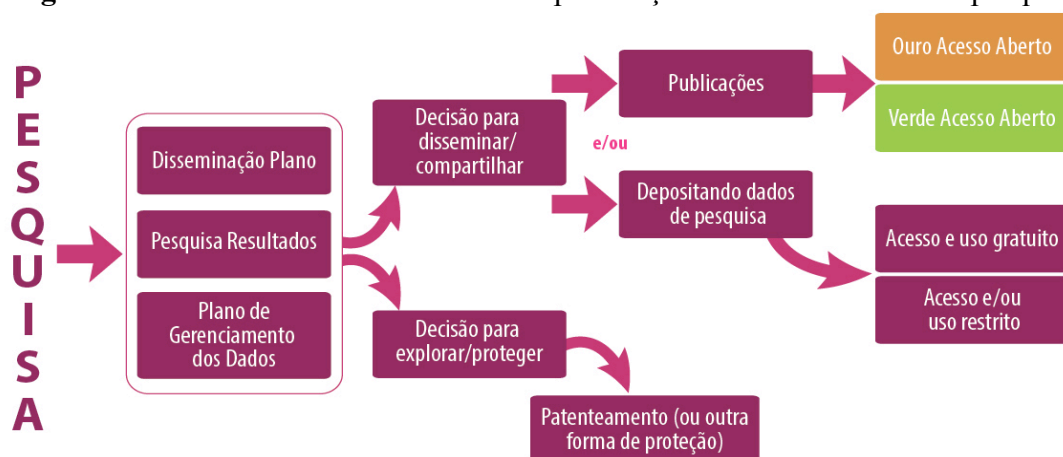
As orientações propostas fornecem regras com base em dois pressupostos: o primeiro é que artigos de investigação científica revisada por pares (publicado em revistas acadêmicas), e segundo, que os dados de pesquisa (dados de publicações subjacentes, dados curadoria e/ou de dados brutos) sejam provenientes de projetos financiados. Dessa forma as publicações e dados de pesquisa devem ser aplicáveis para acesso aberto.

Recentemente, o OpenAIRE publicou em Fevereiro de 2016 como parte das estratégias para dados de pesquisa, o projeto piloto intitulado *Open access to research data: the open research data pilot*. O projeto tem a finalidade de melhorar e maximizar o acesso e reutilização dos dados de pesquisa. Sob o aspecto jurídico o “acesso aberto refere-se ao direito de acesso e reutilização de dados digitais de investigação nos termos e condições estabelecidos no contrato de concessão” (OpenAIRE, 2016, *on-line*, tradução nossa).

Em um contexto científico, os dados da pesquisa se referem à informação, em fatos ou números específicos, recolhida para ser examinada e considerada como base para o raciocínio, discussão ou cálculo. Por exemplo, os dados incluem estatísticas, resultados de experimentos, medidas e observações resultantes do trabalho de campo, gravações de entrevistas e imagens. O foco está na abertura dos dados em formato digital com possibilidades de ser acessados, extraídos, explorados, reproduzidos e divulgados de forma gratuita para o usuário (OPENAIRE, 2016).

A Figura 20 representa o processo de compartilhamento e livre acesso à publicação e aos dados de pesquisa no contexto mais amplo de divulgação e exploração. No primeiro direcionamento, o ato de decisão abrange a exploração por meio da proteção de DPIs (direito autoral e patente). No segundo direcionamento a decisão condiciona a disponibilidade e compartilhamento de publicações e dados científicos.

Figura 20 – *Framework* do livre acesso à publicação científica e dados de pesquisa



Fonte: Traduzido do projeto *Horizon 2020* (OpenAIRE, 2016, p. 3).

No caso das publicações sugere adotar a classificação de acesso aberto: *Gold OA* (Acesso Aberto Ouro) e *Green OA* (Acesso Aberto Verde). E no compartilhamento dos dados de pesquisa estabelece duas premissas: acesso aberto e gratuito (via verde) e acesso e uso restrito (via vermelha).

O *framework* representa uma solução democrática que orienta pesquisadores e envolvidos nos projetos de pesquisa a conduzirem a disponibilidade e compartilhamento dos produtos de pesquisa de forma que atenda aspectos relacionados com os DPIs, ao mesmo tempo que prevê o acesso aberto e gratuito na perspectiva da *e-Science*.

Além das diretrizes apresentadas pelo OpenAIRE para a comunidade europeia, a iniciativa denominada Projeto Blimunda é uma das iniciativas no contexto de Portugal. Foi originado para dar suporte ao contexto do Repositório Científico de Acesso Aberto de Portugal (RCAAP). É uma iniciativa apoiada e financiada pela Fundação para a Computação Científica Nacional (FCCN) e está sendo desenvolvido pela Divisão de Biblioteca e Documentação da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa. O Blimunda é um projeto piloto de gestão de dados científicos para enfrentamento dos desafios e formulação de estratégias com base nas diretrizes do OpenAIRE e apoiado pela Fundação para a Ciência e Tecnologia (FCT) de Portugal. A meta inicial do Blimunda foi conduzida para que houvesse a integração de publicações resultantes de projetos e bolsas financiadas pela FCT e arquivadas nos repositórios da rede RCAAP ao portal OpenAIRE.

Na América Latina o destaque é a rede *Red de Revistas Científicas de América Latina y el Caribe, España y Portugal* da *Universidad Autónoma del Estado de México*. É um projeto promovido pela Universidade Autônoma do Estado do México, em colaboração com centenas de instituições de ensino superior, centros de investigação, associações profissionais e editores latino-americanos. Tem a finalidade de ser uma plataforma líder para serviços de informação científica de acesso aberto a nível internacional, visando atender as necessidades de informação especializada de estudantes, pesquisadores e tomadores de decisão no desenvolvimento científico e tecnológico. Busca a equidade e neutralidade no acesso à informação, democratização do conhecimento e acesso aberto ao conhecimento científico.

Este capítulo refletiu algumas iniciativas internacionais direcionadas por políticas, diretrizes e boas práticas que orientam a gestão de publicações e dados científicos. As reflexões emergidas do capítulo resultam em duas considerações:

A **primeira consideração** refere-se ao fato de que algumas iniciativas se encaixam na noção de *soft law*⁸, que se baseia em normativas que não são obrigatórias e não possuem natureza legal, mas servem como princípios para nortear políticas, processos e

⁸ O termo “*soft law*” refere-se aos instrumentos de quase-legais que não têm qualquer força juridicamente vinculativa, ou cuja força de ligação é um pouco “mais fraca” do que a força obrigatória da lei tradicional, muitas vezes contrastado com *soft law* por ser referido como “*hard lei*”. Tradicionalmente, o termo “*soft law*” está associada com o direito internacional, embora mais recentemente, foi transferido para outros ramos de direito interno também.

comportamentos. Tais como, os princípios elaborados pela OECD, os estabelecidos pela Comissão Europeia e as diretrizes da NWO na Holanda.

A **segunda consideração** está condicionada a defesa em prol do movimento da ciência aberta com a abertura dos dados científicos. Fortalecendo o argumento que esse movimento incentiva e acelera a produção científica, diminui custos e possibilita o desenvolvimento econômico, científico e social.

Diante disso, o estudo entende que é imperioso o engajamento do governo e da comunidade científica na elaboração de diretrizes e princípios para a ciência aberta e a gestão de dados científicos como estratégia de impulsionar o desenvolvimento social e econômico de um país. Considerando para tal empreitada as conjugações político-normativas e ético-legais.

8 STATUS DA AUTORALIDADE NA *E-SCIENCE*

Este capítulo versa sobre a reflexão da autoria dos dados científicos com vistas a apreender onde reside a autoralidade colaborativa. O capítulo foi norteado pelas dimensões epistemológica e ético-legais. Parte do pressuposto que o fazer científico da *e-Science* está baseado no trabalho colaborativo e na preocupação com a postura ética de pesquisadores e colaboradores nas práticas científicas. Apregoa-se também que a sensibilização e o processo de educação deve guiar a conduta e o comportamento ético dos envolvidos. No que diz respeito aos direitos de propriedade intelectual vigentes as questões autorais estão sendo reconfiguradas para um entendimento contemporâneo da autoralidade colaborativa.

Tradicionalmente, o entendimento doutrinário sobre a autoralidade, envolve duas acepções nas definições de Aulete e Valente (2016) e Krokosz (2015):

- a) A primeira acepção refere-se a noção de ‘autor’. O autor é o criador da obra literária, artística, científica e inventiva. É o sujeito responsável pela ação que pode ser um objeto, uma ideia ou um fato, uma invenção ou descoberta. Tal ação pode ser empreendida em diferentes propósitos e contextos.
- b) A segunda acepção está direcionada a noção de ‘autoria’. A autoria está relacionada com a função que o autor exerce. É o caráter morfológico da função de autoria, ou seja, é a condição em que o autor exerce sua prática. Assim, a autoria é exercida pelo autor para a realização, execução, produção do trabalho intelectual em diversas formas, tais como: obra literária, música, teoria, simulação, escultura, peça teatral, marca, desenho, dado; ou seja, o resultado da criação científica, inventiva ou artística fruto da mente humana.

Então, o autor é o sujeito da ação responsável pela produção, geração, fundação e invenção da obra, e a autoria é a ação, condição ou função desempenhada pelo autor.

Teoricamente, a simples criação da mente do criador (*corpus misticum*) serve, tão somente, para seu próprio deleite, como o próprio ato de criar. “A criação que por qualquer forma, meio ou processo não for exteriorizada, não pode ser considerada obra, por não ser perceptível no mundo físico, e, por isso, simplesmente não existe” (FRAGOSO, 2009, p. 40).

Um contraponto nesse sentido afirma que o ato criativo não é um ato de criação absoluto “ele não cria algo do nada; ele descobre, seleciona, remodela, combina, sintetiza fatos, ideias, faculdades, competências já existentes” (KOESTLER, 1967, p. 120, tradução nossa). A referida afirmação aponta para requisitos da originalidade, sob o ponto de vista da descoberta científica e de uma produção coletiva e colaborativa chamada de pensamento

bissociativo. O autor não é uma figura indissociável da criação e processo intelectual, mas também não é absoluto.

É o que afirma Barthes (2004) no que ele atribui como **Autor-Deus**, mas ao que corresponde o resultado de uma obra intelectual como uma atividade criativa que opera em mais de um plano [...] (KROCOSZ, 2015). Esses planos são concebidos em vários níveis de pensamento e conhecimento e ocasionando as descobertas originais, são aquelas que não resultam de um conjunto de regras e envolvem processos inconscientes nos vários níveis de profundidade. Sendo assim, a originalidade do autor é oriunda de uma coletânea de conhecimentos e pensamentos, adquiridos, internalizados a partir de níveis de profundidade e que resultam em um novo *corpus misticum*.

Seguindo o entendimento de Fragoso (2009), pode-se compreender que o *corpus mechanicum* refere-se à forma de expressão, que constitui o processo intelectual escolhido para a expressão do subjetivismo do criador (intelectual, artística, musical, literária) e, ao mesmo tempo, como parte de um mesmo todo, o seu revestimento material (a forma partitura, a forma disco, a forma livro, a forma escultura, etc., o seu *corpus mechanicum*), enfim, é a forma como se diz. As discussões em torno do que constituem o *corpus misticum* e *corpus mechanicum*, envolvem os efeitos práticos entre a forma dada a ideia e o modo de expressão do autor.

Ainda sob a égide da noção de ‘autor’ o entendimento trazido por Foucault (2001) aflora uma reflexão contemporânea no sentido de que “o autor não é exatamente nem o proprietário, nem o responsável por seus textos; não é nem o produtor, nem o inventor deles. Esta é a visão de que o autor não é o proprietário absoluto de seus textos, escritos ou obras, mas um compilador de ideias, textos e agrupamentos de conhecimentos que geram um novo sentido e, por conseguinte, uma nova criação” (FOUCAULT, 2001, p. 265).

Já o modo de expressão representa o tratamento que o criador confere ao conteúdo intangível (*corpus misticum*) de tal forma, ou o seu significado: é o modo como se diz o que se pretende dizer. Ainda faz referência ao meio de expressão que representa os processos ou sistemas, enfim, os veículos pelos quais a obra é dada a conhecer ao público (reproduzida em cópias, transmitida por radiodifusão, representada cenicamente, etc.), ou seja, através do que se diz (FRAGOSO, 2009).

Assim, pensando em um contexto direcionado para as práticas científicas abertas, a reflexão de Foucault apresenta uma nova compreensão sobre o conceito de autor. Visto que o autor na contemporaneidade, já é partícipe de um processo de construção colaborativa e

compartilhada. O que leva a uma dedução preliminar que o caráter proprietário e exclusivo autoral está baseado em um *constructo* coletivo e colaborativo.

Ainda na perspectiva autoral, por séculos o entendimento sobre o autor esteve arraigado intrinsecamente à sua natureza proprietária e exclusivista, ainda presente nos sistemas de propriedade intelectual vigentes. O que leva a outra perspectiva centrada no seu reconhecimento, no sentido de atribuição: “o nome do autor funciona para caracterizar um certo modo de ele ser do discurso: para um discurso, o fato de haver um nome de autor, o fato de que se possa dizer ‘isso foi escrito por tal pessoa’, ou ‘tal pessoa é o autor disso’ [...]”, recebida de uma certa maneira e que deve, em uma dada cultura, receber um certo *status* (FOUCAULT, 2001, p. 275).

O termo *status* o qual Foucault está ligado a noção de ‘apropriação’. O discurso que o autor emite através de sua obra é expressado por diferentes formas e manifestações autorais. O autor recebe um *status* e uma posição diferenciando-o pela titularidade, escritos e até por indicadores de mensuração presentes na comunicação científica. Atribui-se ao autor um *status* como forma de reconhecimento da sua função autoral. Tal *status* posiciona a função do autor em uma situação de destaque.

Dessa forma, essa posição de destaque traz nova reflexão sobre as diferentes configurações de autoria (autor, coautor e colaborador). Então, o que é ‘autor’ e o que é ‘autoria’ continuam a constituir a indagação inicial sobre a autoralidade contemporânea. Por ora, será visto na seção seguinte as configurações sob o ponto de vista da legislação e doutrina brasileira e em seguida as várias concepções sobre autoria.

8.1 PONTO DE VISTA LEGAL

A previsão legal do Direito Autoral na legislação brasileira coexiste de duas naturezas básicas, a saber: natureza moral e patrimonial.

Na Constituição Federal brasileira o Artigo 5º trata dos direitos morais do autor. Tais direitos também são chamados de direitos da personalidade. O artigo prevê que são consideradas invioláveis a intimidade, a vida privada, a honra e a imagem das pessoas e assegurado o direito a indenização por dano material ou moral decorrente da sua violação. Assim, o autor ao realizar uma criação intelectual, realça o vínculo que estabelece com a obra, de modo que se manifesta através de expressões que o individualizam, podendo estas ser identificadas a partir de elementos ligados à sua personalidade (CARBONI, 2009).

No Código Civil é previsto que os direitos de personalidade são intransmissíveis e irrenunciáveis. Entretanto, as características que regem os direitos de personalidade não podem ser atribuídas aos direitos autorais.

A Lei de Direito Autoral (LDA) nº. 9.610, de 19 de fevereiro de 1998, prevê que o autor goza de direitos morais e patrimoniais. A LDA estabelece no seu artigo 24 e incisos, que o autor pode reivindicar a autoria; ser identificado como autor; conservar a obra inédita; assegurar a integridade da obra; modificar a obra; retirar a obra de circulação; e, ter acesso ao exemplar único e raro da obra; tudo isso se configura como direitos morais do autor.

A LDA faz previsão legal nos artigos de 11 a 13 sobre a autoria das obras intelectuais, estabelecendo no artigo 11 que: autor é a pessoa física criadora de obra literária, artística ou científica. No parágrafo único, a proteção concedida ao autor se aplica à pessoa jurídica. O artigo 12, prevê a identificação do autor como criador da obra. No artigo 13, a lei prevê que o autor da obra intelectual, estando em conformidade com o artigo anterior, goza do seu uso e utilização, desde que não haja prova em contrário. O autor de qualquer obra intelectual, artística e cultural com base na legislação brasileira a rigor é uma pessoa física.

A Lei autoral brasileira faz uma confusão entre autoria e titularidade, abrigando sobre o manto da autoria, ora a pessoa física, ora a pessoa jurídica. O artigo 27 dispõe que os direitos morais do autor são inalienáveis e irrenunciáveis. Neste entendimento, o autor possui o seu direito intrínseco de caráter personalíssimo e intransferível, cujos direitos são atribuídos à pessoa humana e, por conseguinte, inerentes à pessoa física.

Sobre os Direitos Patrimoniais, conforme a LDA, o capítulo II dos artigos 28 ao 45 descrevem que esses direitos são objeto de contrato e perduram por um prazo determinado. A titularidade é transferida do autor para terceiros. O terceiro passa a ser o titular do direito autoral com a sua propriedade limitada nos termos da lei. Pois, cabe ainda ao autor, como proprietário da obra, o direito exclusivo de utilizar, fruir e dispor de sua obra.

A LDA faz previsão legal dos artigos 49 a 52 da transferência dos direitos autorais. Esses artigos estabelecem a transferência para terceiros, total ou parcial, a título universal ou singular, por meio de licenciamento, concessão, cessão ou por outros meios admitidos no Direito, pelo autor ou por seus sucessores, a transmissão e a cessão da obra intelectual. Conforme o artigo 49, o autor pode transferir os seus direitos, total ou parcial, por meio de três modalidades: licença pública, concessão ou cessão de direitos.

Em virtude de uma obra ser fruto da criatividade intelectual humana, no tocante à pessoa jurídica, esta usufrui apenas dos direitos de titularidade para uso comercial. Estamos a falar dos direitos patrimoniais, pois os direitos autorais são intransferíveis, mantendo a

originalidade de autor atribuída à pessoa física. Os direitos patrimoniais podem ser transferidos para utilização e reprodução da obra. É a titularidade que pode ser exercida por outra pessoa física ou pessoa jurídica. A pessoa jurídica pode ser uma entidade, empresa, instituição ou organização.

Considera-se que para compreender estas nomenclaturas e sua função no processo de criação e invenção científica, literária, artística ou cultural é necessário entender suas características etimológicas, morfológicas e a previsão legal. Adiciona-se a esta compreensão os dados científicos oriundos de investigações no contexto da *e-Science*.

A partir do entendimento do direito moral e patrimonial, voltam-se às discussões das acepções de autor e autoria, à distinção entre pessoa física e jurídica e à aplicação de apropriação e atribuição. Questões relacionadas à autoralidade originária, derivada e científica, bem como a aplicação de apropriação e atribuição autoral carecem de discussões.

8.2 AUTORIA ORIGINÁRIA E DERIVADA: UMA REFLEXÃO SOB O PONTO DE VISTA DA PROPRIEDADE

Na ambiência da *e-Science* os produtos e invenções oriundas da pesquisa científica estão relacionados com o compartilhamento, gerenciamento e reuso de dados de pesquisa. As questões relacionadas à autoralidade (originária, derivada, científica, coletiva e colaborativa), bem como a aplicação de apropriação e atribuição autoral carecem de discussões e levantam proposições acerca da propriedade autoral nos atuais sistemas legais.

Considerando o requisito da originalidade, o autor tem por definição a autoria primária de sua criação. A proteção autoral, bem como os limites da autoria primária, derivada, coletiva e colaborativa na perspectiva do compartilhamento de dados de pesquisa e informações são inovadoras e evidenciadas. O atual entendimento está baseado nas múltiplas autorias e na pluralidade de colaborações.

A legislação específica brasileira declara que a autoria é uma propriedade exclusiva do autor, criador ou inventor. Assim a criação originária é albergada pelos requisitos de originalidade e ineditismo e que são exclusivas do autor originário. O artigo 5º da LDA, inciso VIII abrangem o entendimento dos tipos de autorias previstas e que serão pertinentes para as reflexões contemporâneas:

- a) A obra originária é a criação primígena;
- b) A obra derivada é a que, constituindo criação intelectual nova, resulta da transformação de obra originária;

- c) A obra coletiva, é a criada por iniciativa, organização e responsabilidade de uma pessoa física ou jurídica, que publica sob seu nome ou marca e que é constituída pela participação de diferentes autores, cujas contribuições se fundem numa criação autônoma.

Desta maneira, a proteção autoral é alicerçada no requisito primordial da originalidade do autor. Assim, as obras originárias e derivadas se distinguem na sua constituição. No primeiro caso, é uma obra originária, que são “as obras intelectuais que podem nascer de forma autônoma, sem qualquer vinculação com obras anteriores, ou destas se originarem” (CARBONI, 2009, p. 14). A obra originária é autônoma e nasce de uma ideia da mente humana. É possível atribuir que o autor primário é sempre uma pessoa física e que goza dos direitos de propriedade sobre seus dados ou recursos informacionais oriundos do processo investigativo. E no segundo, pelo processo de transformação, trata-se de “obra derivada, que tem como base outra preexistente e depende de prévia aprovação do criador da obra anterior. O direito de autor protege tanto a obra originária como a derivada” (CARBONI, 2009, p. 14). A obra derivada é um processo de transformação da obra originária, utilizando-a esta como base para o aprimoramento, acréscimo e novas hipóteses. Pode ser oriunda do autor originário ou de terceiros.

De acordo com Paesani (2012, p. 27), a “obra derivada é a obra que resulta da transformação de obra originária e constitui uma criação intelectual nova”, por conseguinte, a nova criação. Para que outro autor possa transformar a obra originária, requer a autorização prévia do autor originário. Este é o viés ético da pesquisa científica.

A legislação brasileira vigente mantém sua abordagem tradicional na autoralidade convencional (originária, derivada e coletiva). A partir dessa abordagem, dois direcionamentos serão adotados para evidenciar as reflexões no contexto da *e-Science*: obra originária pertencente ao autor primário e a obra derivada atribuída ao autor secundário.

A primeira abordagem considera que mesmo em um ambiente que prevaleça o ambiente de colaboração e coletivo, mantém-se o *status* de autor primário para a criação primígena. A segunda abordagem destina-se a autoria derivada, com base na previsão legal em vigência, pode transferir os direitos de uso, reprodução e fruição da obra originária, mediante prévia permissão expressa do autor primário. Frente a segunda abordagem, repousam as reflexões acerca do compartilhamento e reuso de pesquisas científicas, mais notadamente, aos dados científicos.

A distinção conceitual entre autoria originária e derivada foi conduzida com maior ênfase, pela necessidade de aprofundar o entendimento epistemológico da autoralidade e a

transversalidade com os diferentes papéis que o autor passa a assumir nas configurações postas pelas práticas da *e-Science*.

8.3 RESSIGNIFICANDO A AUTORIA CIENTÍFICA

A autoria científica é uma função já presente no ambiente acadêmico-científico. É evidenciada como uma função de destaque no âmbito da investigação científica. A especificidade da autoria científica não é assunto recente, mas tomou outros contornos a partir do que os autores chamam de “condição pós-moderna da autoria científica” do século XXI. Krokosz (2015, p. 97) “chama de noção compartilhada [...] o surgimento de uma corrente de reflexão que vem ressignificando a concepção tida sobre o conhecimento e suas implicações”.

Esta ressignificação está atingindo todos os campos dos saberes, primordialmente no ambiente científico quando se busca resultados válidos e aceitáveis para a ciência. A autoria científica é portanto, “uma modalidade autoral, típica dos processos de produção textual acadêmica, a qual pode ser exemplificada como editoriais, resumos, pôsteres, artigos, capítulos de livro e livros inteiros” (KROKOSZ, 2015, p. 90). Assim, prevalece suas características de criação, colaboração, produção e invenção nos diferentes campos dos saberes e se propaga por meio do processo de comunicação científica vigente.

A autoria patrimonial é centrada no direito privado, propriedade individual e transferível, e por conseguinte, recebe a recompensa pecuniária. E a autoria científica está centrada em um direito público, recebe por recompensa autoral: merecimento, aceitação, reconhecimento dos pares e a atribuição. Geralmente, a autoria científica não possui natureza pecuniária e a propriedade é inalienável. A recompensa é pautada no reconhecimento científico sendo a atribuição e os créditos as formas usuais.

Ainda no limiar da autoria científica a propriedade está associada ao direito imaterial centrado na diferença que a propriedade moral entendida como *propriety* voltada para o reconhecimento autoral por meio de um sistema de créditos autorais e a inexistência econômica. Enquanto que a noção de *property* está direcionada a uma recompensa que se pauta nos sistemas de direitos de propriedade intelectual, com base em um direito privado e na remuneração econômica. Essa dualidade, muito embora não seja tão evidente na prática da comunicação científica, reflete de forma latente duas compreensões: o conhecimento científico mediado sempre por uma dicotomia entre o público e o privado; e, a expropriação econômica dos feitos científicos como forma de recompensa autoral, por meio da remuneração.

Baracat e Rigolin (2012, p. 350) afirmam que “as contradições de âmbito social, econômico e cultural são pontos centrais nas discussões sobre propriedade intelectual, pois é a lei de proteção da propriedade intelectual que regula a produção, a circulação e o acesso ao conhecimento estão voltadas para uma vertente privada em um contexto contemporâneo de acesso aberto e público”. No entanto, o que se leva em consideração é o comportamento moral e ético do autor científico, devendo prevalecer sempre o bem comum e público para o progresso da ciência. O reconhecimento autoral é baseado na reputação, prestígio e promoção científica.

8.4 AUTORIA COLETIVA OU COLABORATIVA: ONDE RESIDE A AUTORALIDADE NA *E-SCIENCE*?

Na ambiência contemporânea baseada na *e-Science*, o processo da comunicação científica é validado pelo modelo ciclo de vida dos dados, aplicado ao gerenciamento dos dados científicos compartilhados. Assim, no espaço científico além da autoria científica, outras duas condições autorais se destacam: autoria coletiva e autoria colaborativa. Os ambientes que perpassam as práticas da *e-Science* é realizado por diferentes pesquisadores, colaboradores e instituições.

Na doutrina, perdura ainda a diferença conceitual entre a autoria coletiva e colaborativa. A obra coletiva, “é criada por iniciativa, organização e responsabilidade de uma pessoa física ou jurídica, que a publica sob seu nome ou marca, e que é constituída pela participação de diferentes autores, cujas contribuições que se fundem numa criação autônoma” (PAESANI, 2012, p. 27). Assim, os direitos autorais também são coletivos, indivisíveis e destinam-se a múltiplos autores.

No sistema legal brasileiro, as obras colaborativas não estão previstas na LDA, uma vez que a previsão legal só prevê as obras em coautoria. A lei atual vigente consagrou o “princípio da coautoria, que englobaria, parcialmente, o de colaboração, definindo a obra em coautoria como aquela criada em comum por dois ou mais autores” (FRAGOSO, 2009, p. 162). A previsão legal reforça que o coautor tem a sua proteção autoral preservada. No que se refere a autoria, é o autor primário, o responsável pela pesquisa e gerador dos dados primários oriundos de um processo de investigação científica que goza do *status* de autoria originária.

Cada contribuição individual na criação colaborativa representa um “aporte intelectual de verdadeira criação, individualizada e identificada, resultando numa obra divisível, onde cada colaborador conserva o direito individual sobre seu trabalho [...], com interdependência

na sua obra, estamos a falar de uma obra colaborativa” (FRAGOSO, 2009, p. 162). É o caso dos editores, produtores, revisores e curadores que participam como auxiliares com contribuições secundárias ou acessórias, mas que seu papel complementa a obra intelectual em sua totalidade.

Porém, outra reflexão é imprescindível no atual contexto, considerando as novas vertentes de colaboração científica em uma ambiência compartilhada. Ressaltamos a figura do autor na contemporaneidade e que na obra de Barthes, ele aponta para ‘a morte do autor’, onde ocorre “o afastamento do Autor [...] (que seria a «mensagem» do Autor-Deus), para um espaço de dimensões múltiplas, onde se casam e se contestam escritas variadas, nenhuma das quais é original: o texto é um tecido de citações, saldas dos mil focos da cultura [...]” (BARTHES, 2004, p. 3). Esta reflexão é bastante atual no sentido de mostrar que a figura do autor deixa de ter uma função centrada na sua propriedade individualista para uma função colaborativa. A dimensão teológica que Barthes posiciona como Autor-Deus cede espaço para as posições e escritas múltiplas e compartilhadas, saídas de várias culturas e fundem-se em diálogos. O foco não está na obra e no autor, mas na multiplicidade e na coletividade de diferentes colaboradores.

Na contemporaneidade, a versão colaborativa é exemplificada pela Wikipédia. Krokosz (2015, p. 79) afirma que “é a Wikipédia, uma enciclopédia eletrônica que vem sendo perenemente construída e permanentemente editada na *internet* desde 2001”. Essa obra colaborativa permite que qualquer pessoa possa publicar, criar, modificar e distribuir o conteúdo, seja para fins ou não comerciais, desde que obedeça aos parâmetros das licenças *Creative Commons Attribution-Share Alike* e a condição da *GNU Free Documentation License*. Ambas são licenças para documentos e textos livres, apesar de “permitirem o uso livre de conteúdos, exigem a condição de que os autores originais sejam creditados e o direito de uso livre por outras pessoas seja mantido” (KROKOSZ, 2015, p. 80). As dimensões geográficas e linguísticas abrem espaço para os ambientes digitais colaborativos, conectados, interativos e sem fronteiras, proporcionados pelo uso intenso de tecnologias e através da *internet*. Assim, é mister, também, entender as autorias já mencionadas e ter a compreensão do processo autoral na seara compartilhada - *strict sensu* - da autoria colaborativa.

8.5 AUTORIA COLABORATIVA E ATRIBUIÇÃO COMPARTILHADA

Considerando o entendimento que a autoria colaborativa está em evidência na atual era digital, proporcionada pelas tecnologias digitais de informação e comunicação, redes

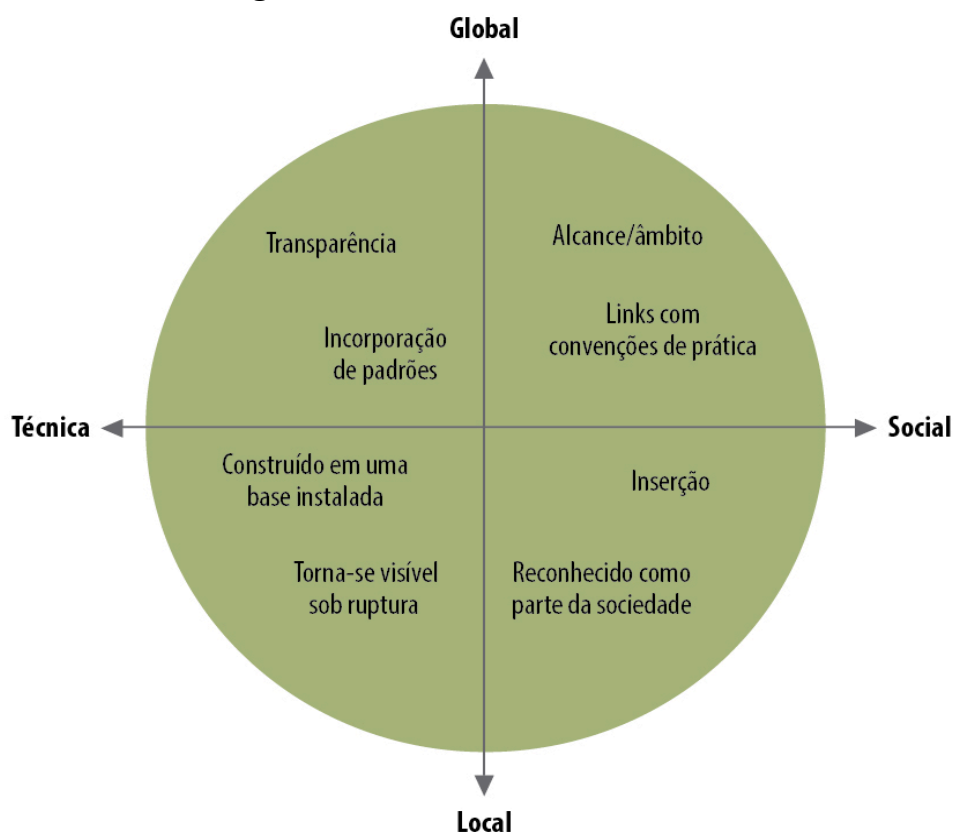
sociais colaborativas e pela *internet*. No caso da *e-Science* há de se considerar que a autoria é colaborativa em virtude de que os dados científicos são compartilhados e gerenciados por diferentes autores e propósitos distintos, que configuram a multiplicidade autoral e colaborativa. A autoria colaborativa é diferenciada pela proporcionalidade de sua colaboração à obra, partindo do princípio que a pesquisa envolveu a participação individualizada e proporcional de diferentes atores no processo criativo ou inventivo.

Desta forma, o autor científico pode ser compreendido sob a condição de autoria científica em um ambiente colaborativo. Além dos pesquisadores-autores configuram também neste cenário outros participantes, tais como: curadores, bibliotecários, indexadores, especialistas tecnológicos, dentre outros profissionais envolvidos com a área-fim. A condição autoral científica passa a ser mais um requisito que valida os preceitos metodológicos, éticos e científicos no processo de investigação científica, mas não é o centro da autoralidade na *e-Science*.

Assim, sob o ponto de vista moral permanece o entendimento de autoria primária (originária) e sob a ótica do compartilhamento dos dados de pesquisa, os autores são derivados e passam a gozar de uma condição de autoria colaborativa. É mister ressaltar que o cerne das atuais discussões concentra-se não mais na propriedade e autoralidade individual e privada, mas na recompensa autoral. O Direito de Autor no cenário de compartilhamento deve ser considerado numa perspectiva onde,

A capacidade de estabelecer rapidamente se uma ideia anteriormente foi levada adiante - e creditá-la corretamente, se assim for - é importante para os cientistas. Completa identificação do autor e o conteúdo pesquisável de forma central irão simplificar este processo e facilitarão a atribuição e o reconhecimento (SERINGHAUS; GERSTEIN, 2007, p. 2, tradução nossa).

Acerca da atribuição de dados, o estudo de Bowker e outros (2010) foi desenvolvido um modelo a partir de dimensões relacionadas às práticas que envolvem a citação e atribuição de dados. A infraestrutura no modelo exemplificado convencionou características que permitem as práticas para alcance global, inserção social, construção de uma base local e transparência técnica a partir da incorporação de padrões e procedimentos. Cada dimensão, “ela é um ponto de partida útil para avaliar como essas características de infraestrutura são aplicadas à citação de dados e atribuição: a prática social, usabilidade, identidade, persistência, descoberta, procedência, relacionamentos, propriedade intelectual e política” (BOWKER et al., 2010). O modelo é constituído em quatro partes chamadas de dimensões. A representação dessa infraestrutura dimensional está demonstrada na Figura 21.

Figura 21 - Dimensões de infraestrutura

Fonte: Traduzido de BOWKER e outros (2010).

A dimensão técnica do modelo está relacionada com a infraestrutura e incorporação de padrões e executada em uma base local. Entre as bases instaladas que influenciam a citação de dados estão os protocolos de *Internet*, práticas de publicação e métodos de catalogação da biblioteca.

Na dimensão social, a infraestrutura está ligada a convenções de prática – quer seja catalogação ou gestão de dados – e é reconhecida como parte de membros em uma comunidade (por exemplo, bibliotecários ou astrônomos). Essa dimensão é apoiada pelo modelo de gerenciamento de dados e associada ao sistema de recompensa autoral, e um tópico social de particular interesse é a relação dos sistemas de recompensa para a citação de dados (BOWKER et al., 2010).

Na dimensão local do modelo está contemplado o contexto com menor amplitude, que pode estar focado em uma instituição, uma biblioteca ou um laboratório de pesquisa, e onde encontram-se práticas individuais para gerenciar dados e práticas de biblioteca para administração de dados (BOWKER et al., 2010). As práticas científicas preferencialmente devem ser conduzidas pelo modelo de ciclo de vida para o gerenciamento de dados. A

realização de práticas e procedimentos ocorre em um ambiente local, onde o projeto de pesquisa está sendo desenvolvido.

A dimensão global, por sua vez, é a que representa o caráter inerentemente internacional do projeto de pesquisa científico. O contexto internacional oportuniza as práticas colaborativas, compartilhamento e intercâmbio de dados. É uma dimensão importante, onde a infraestrutura já deve estar garantida de modo que a citação e atribuição devem funcionar de forma eficaz através das fronteiras disciplinares, institucionais e políticas (BOWKER et al., 2010). No quadrante técnico-global, os dados são mais passíveis de captura automática, gerenciamento e descoberta. Tecnicamente, a infraestrutura deve estar preparada para que os dados sejam capturados, interoperáveis, gerenciáveis e que possam ser descobertos e reusados.

O modelo se aplica ao compartilhamento de diferentes tipologias de dados. Dados recolhidos de telescópios baseados no espaço, e associados com as estruturas estabelecidas de dados, ferramentas analíticas e repositórios, por exemplo, são mais facilmente citados. Por outro lado, os dados relacionados às dimensões do quadrante local-social tendem a ser mais heterogêneos em forma e conteúdo, mais artesanais em métodos de coleta de dados, e mais variados em práticas de gestão, utilização e reutilização (BOWKER et al., 2010). Devido a essa heterogeneidade, contexto local e procedimentos metodológicos ineficazes poderão dificultar a captura, gerenciamento, citação, e atribuição de dados em contextos locais e sociais.

Isto posto, o modelo exemplificado colaborou no entendimento de uma infraestrutura sustentável que envolve os aspectos técnicos e práticas locais, sociais e internacionais. Tem a finalidade de dar visibilidade aos dados científicos a partir do reconhecimento da autoridade, utilizando a atribuição e citação como meio técnico.

A atribuição e o crédito autoral devem ser fixados dentro de sua proporcionalidade. A identificação pode ser realizada aos autores primários às suas publicações ou dados de pesquisa, com identificadores tais como: ORCID e DOI. Além da figura do autor, outros colaboradores devem ser destacados a exemplo, curadores e gestores de dados. Por vezes, essas atribuições podem ser contribuições primárias, secundárias ou acessórias, que exercem a função de colaboração típica, requerendo que as atribuições dos créditos sejam feitas na proporcionalidade de participação de cada envolvido, pois a,

Atribuição é feita à parte responsável. Atribuição pode, portanto, ser dada a um investigador individual, a uma equipe de pesquisa, para uma universidade, para uma agência de financiamento, para um repositório de dados, para uma biblioteca ou para

outra parte responsável pela coleta, montagem, curadoria, ou, por outro lado, contribuindo para a disponibilidade de dados para outros usarem. (BORGMAN, 2012, p. 5, tradução nossa).

Atribuir a cada colaborador o reconhecimento autoral é proporcionalmente dar o crédito individual pela contribuição e responsabilidade no processo de pesquisa. A distinção entre a atribuição e a citação: o reconhecimento atribuído é realizado tanto à pessoa física como pessoa jurídica e a citação utiliza-se de mecanismos para realizar as referências a cada atribuição. Na linguagem bibliométrica, referências são feitas e citações são recebidas (BORGMAN, 2012). Assim é por meio da citação que se atribui o reconhecimento autoral. Atualmente existem sistemas de citação, identificação e *software* que ajudam na atribuição e mensuração autoral.

Não se trata de propriedade e transferência dos dados científicos (primários e derivados), mas, sim de reconhecimento sob os aspectos de atribuição e licenciamento dos direitos autorais de bens intelectuais, científicos e artísticos. A citação e a atribuição autoral devem ser consideradas no contexto da utilidade dos dados digitais como objetos reusáveis e de bem comum.

Nas práticas da *e-Science* é percebido claramente que as diferentes atribuições são dadas na proporcionalidade da responsabilidade e participação de cada colaborador. A criação não mais goza apenas de um autor, coautores ou pesquisadores, mas de contribuídores que colaboram no processo de gerenciamento e de responsabilidade científica mútuas. Portanto, a

Atribuição é mais estreitamente associada à noção de contribuição ou colaborador do que com o autor, que está entre as diferenças entre a manipulação de dados e manipulação de publicações. Como publicações, no entanto, a atribuição implica responsabilidade social para dar crédito onde o crédito é devido. Quando escrevemos artigos de periódicos e livros, nós fazemos referência a outras publicações e às provas em que se baseiam para atribuir as nossas fontes (BORGMAN, 2012, p. 5, tradução nossa).

Preliminarmente é entendido que a noção de contribuidor ou colaborador direciona a diferentes envolvidos, que pode ser pesquisador, autor, co-autor, curador, fotografo, agência, universidade, organizações governamentais e não governamentais e institutos científicos sob a condição de autoria colaborativa, cuja atribuição é proporcional à sua responsabilidade e contribuição. A responsabilidade se fixa na conduta ética e profissional que cada envolvido contribui no processo de investigação científica. Essa contribuição ela não possui uma hierarquia que se sobrepõe entre os diferentes envolvidos, pois a relação é conduzida pela colaboração mútua. A citação é distinta para cada colaborador. A atribuição é realizada por meio de identificadores persistentes e créditos autorais são fixados individualizados. A

condição autoral passa a ser um requisito que valida os preceitos metodológicos, éticos e científicos no processo de investigação científica.

Do ponto de vista da autoria originária - que cabe ao autor, co-autor, colaborador ou inventor - a primariedade da criação, invenção e participação considera a condição de autoria originária se estende a todos os envolvidos na pesquisa. No caso do reuso de dados primários a condição de autoria tem caráter colaborativo e assume a autoria derivada.

O capítulo trouxe uma visão distinta dos processos científicos tradicionais, visto que as práticas da *e-Science* são realizadas pelos envolvidos que atuam com atribuições e finalidades distintas para um propósito único. A individualidade de atribuição e a citação proporcional a cada colaborador não distingue a importância, pois a autoria é constituída em processo coletivo, pluralista, e por conseguinte, colaborativo.

9 NOVAS ACEPÇÕES DO MODELO EMERGENTE DA *E-SCIENCE*: AUTORALIDADE COLABORATIVA E RECOMPENSA AUTORAL

Este capítulo reflete a análise de conteúdo que emergiram com as conexões inferenciais e os capítulos descritos. As categorias da tese foram norteadas pelas temáticas ciência aberta, Propriedade Intelectual e a transversalidade com a Economia e a tecnologia. A análise buscou responder a questão-problema e direcionou as inferências para o objeto de estudo focado na recompensa autorial na *e-Science* e o processo de ressignificação dos Direitos de Propriedade Intelectual na *e-Science*.

9.1 CONCEPÇÕES CONTEMPORÂNEAS DA CIÊNCIA ABERTA E *E-SCIENCE*

A seção foi conduzida para as concepções conceituais da ciência aberta e *e-Science*. As reflexões buscam entender a distinção conceitual entre essas concepções que estão reestruturando as práticas científicas contemporâneas.

No que tange a ciência aberta reforça-se que é uma concepção ampla, complexa e diversificada do pensar e fazer a ciência. Engloba espaços comuns do fazer intelectual de maneira que exista sempre a colaboração, o compartilhamento, a coletividade e a promoção do bem intelectual como um bem comum. Então, pensar na ciência aberta é representá-la como um conceito guarda-chuva em contextos multidisciplinares e multidimensionais.

O entendimento de um conceito guarda-chuva confirma que o escopo da ciência aberta traz em seu bojo múltiplas dimensões que abrangem aspectos, tais como: político, legal, normativo, econômico, científico, técnico, tecnológico, humano, ético, social e cultural. Muito embora o conceito alargado da ciência aberta foi recortado na pesquisa apenas para dar visibilidade aos aspectos que estão ligados com a comunicação científica.

Assim foi considerada que a ciência aberta possui um escopo multidimensional dentro de um novo pensar da pesquisa científica colaborativa aberta a partir das seguintes reflexões:

- a) Na Sociedade da Informação, a abertura do conhecimento humano acumulado passa a ser uma realidade possível em busca de políticas sustentáveis que visem a melhoria das condições de vida dos cidadãos e a redução das grandes desigualdades globais;
- b) O propósito primordial é beneficiar a sociedade com transparência, sustentabilidade e confiança a partir dos resultados de pesquisa;

- c) A abertura dos dados de investigação e da ciência tem um papel decisivo em vários segmentos da sociedade e economia global;
- d) A pesquisa aberta colaborativa tem como premissa que as práticas científicas gerem dados, documentos, resultados, métodos, modelos, publicações e tecnologias abertas e gratuitas;
- e) Todos os resultados e criações oriundos da pesquisa colaborativa além de abertos, devem ser curados e preservados em repositórios específicos, permanentemente, considerando os princípios políticos, normativos, éticos e legais;
- f) As iniciativas internacionais levam ao entendimento de que a economia do futuro aponta para a investigação (transparente e confiável), inovação (competitiva) e experiências científicas (qualitativas e abertas);
- g) Abordagens de projetos de pesquisa abertos e colaborativos realçam a dimensão de novidade, inovação e mudança na Sociedade Digital contemporânea;
- h) Aspectos legais vigentes são considerados e analisados previamente antes da abertura dos dados. Entretanto, a primordial premissa da ciência aberta é que os produtos de pesquisa estejam acessíveis, abertos, interoperáveis e gratuitos;

Isto posto, as constatações apresentadas direcionam a ciência aberta para duas vertentes: a formal e a *práxis*. A vertente formal é constituída por instrumentos formais, como as políticas, diretrizes, normativas, legislações e mecanismos regulatórios. E a *práxis* é constituída pela convergência de teorias, modelos, métodos e práticas (vigentes e emergentes) que ocorrem no processo de comunicação científica.

No tocante a vertente formal, as dinâmicas da ciência aberta ocorrem em diferentes contextos, variando entre países, regiões e locais. Essa variação se acentua a partir de economias, culturas, instituições e espaços distintos. Daí a importância de pensar a pesquisa aberta e colaborativa numa perspectiva global e local, observando diferentes propósitos, políticas, culturas e contextos nacionais, regionais e institucionais. Mesmo que haja uma política formal que norteie a pesquisa com parcerias internacionais é importante que questões nacionais e locais sejam observadas e previstas, dentro do possível, com vistas a estabelecer uma conformidade mínima que diminua as diversidades presentes nesses contextos. Essas diversidades influenciam dinâmicas, procedimentos e comportamentos no decorrer do processo de pesquisa. A exemplo, legislações específicas, documentos normativos institucionais e códigos de conduta ética e comportamental que regem práticas científicas locais. Portanto, a *e-Science* é conduzida por um arcabouço formal de políticas, diretrizes, normativas, procedimentos e instruções que regulamentam e norteiam as práticas. Os aspectos

formais estão direcionados por escopos: político-normativo, ético-legal-cultural e os práticos pelo teórico, técnico e morfológico.

A vertente prática direciona-se ao ambiente científico com as novas configurações de uma ciência colaborativa e aberta que representa a *práxis* da *e-Science*. A *práxis* da *e-Science* tem como fio condutor a colaboração científica e o compartilhamento baseados em dados de pesquisa. O escopo morfológico é a infraestrutura sustentável representa seu espaço morfológico onde procedimentos, práticas, métodos, instrumentos, tecnologias, ferramentas e recursos são providos, compartilhados, acessíveis e abertos.

Considerando as vertentes formais e práticas que apoiam a *práxis* da *e-Science*, buscou-se no *constructo* da pesquisa entender os aspectos transversais da ciência aberta e a conexão com a propriedade intelectual. Daí as discussões contemporâneas acerca do *Commons*, Economia Criativa e dos bens intelectuais, como bens comuns que direcionam o entendimento da colaboração, compartilhamento e coletividade da ciência aberta nesta tese. Os bens intelectuais e culturais, nessa perspectiva são de natureza comum, pública e universal. O acesso a tais bens, idealmente, deveria ser democrático e aberto.

Assim, esta seção que reflete a primeira categoria da análise de conteúdo inferiu três reflexões:

- a) A **primeira reflexão** refere-se aos dados científicos está no quarto paradigma científico como eixo norteador da *práxis*. A premissa básica do quarto paradigma está associada ao uso intensivo dos dados na pesquisa científica. Isso pressupõe que todos os dados primários gerados e coletados a partir dos processos de investigações científicas deverão ser aproveitados como produtos primários de pesquisa;
- b) A **segunda reflexão** se volta para os dados científicos como produtos primários na pesquisa científica. A comunicação científica vigente conduz que os resultados da investigação científica sejam publicados como produtos finais. Na *e-Science* as publicações já são realizadas a partir dos dados científicos e se constituem em produtos primários de processos compartilhados e colaborativos de pesquisa. Entendeu-se assim que, a autoralidade antes centrada na figura individualista do pesquisador direciona-se para a perspectiva colaborativa. Reflete também nos DPIs vigentes baseado no paradigma propriedade econômico para abordagens pautadas no *commons paradigm*, na ciência do comum e na colaboração científica. Esses assuntos serão aprofundados em seção específica.

- c) A **terceira reflexão** levou ao entendimento que o quarto paradigma tem uma confluência com os paradigmas anteriores da ciência. Considerou-se que o surgimento de um novo paradigma não rompe com os paradigmas anteriores, mas provoca mudanças no modo de fazer e publicar os feitos científicos. O novo fazer científico ocorre a partir da convergência, incremento, ressignificados e aprimoramentos de modelos, métodos, técnicas e tecnologias que dão sustentabilidade às práticas colaborativas baseadas nos dados científicos.

Em suma, as reflexões acima conduzem a ciência aberta para um ambiente que promove a colaboração, compartilhamento e a coletividade da ciência moderna e a *e-Science* para uma reconfiguração de práticas científicas com dados científicos. Por ora, a categoria contribuiu para a construção do entendimento epistemológico e teórico que norteará os padrões do modelo conceitual.

9.2 ENTENDIMENTO DOS DADOS CIENTÍFICOS SOB A PERSPECTIVA DO *COMMONS*

A seção foi conduzida para entender os dados científicos sob a perspectiva do *commons*, contudo a compreensão preliminar dos instrumentos regulatórios no âmbito internacional, movimentos do domínio público e acesso aberto emergiram na análise e conduziram as novas reflexões para o bem intelectual público e comum. A análise considerou aspectos políticos e normativos vigentes nos instrumentos regulatórios, iniciativas institucionais internacionais e legislações nacionais. As diretrizes que envolvem os aspectos colaborativos a partir do *commons* e as normativas voltadas para o escopo da ciência aberta foram analisadas.

9.2.1 Instrumentos Internacionais Regulatórios

A análise foi conduzida na busca de identificar nos instrumentos regulatórios internacionais contidos em tratados, acordos e convenções direcionamentos à Ciência Aberta e às práticas científicas da *e-Science*. Visto que tais instrumentos são assentados no paradigma proprietário econômico que norteia os sistemas de direitos de propriedade intelectual vigentes. Para auxiliar a análise buscou-se fontes bibliográficas pautadas na doutrina jurídica para dar respaldo a essa subcategoria.

Inicialmente, observou-se que os DPIs na seara científica são utilizados para a proteção autoral e a exploração comercial. A informação ao longo dos séculos assumiu o papel na cadeia informacional e cada vez mais foi sendo filtrada, editada e certificada por uma classe de “peritos” (especialistas em todas as áreas do conhecimento), a qual culminou com a centralização e a corporativização da informação e do acesso e, conseqüentemente, a erosão da esfera pública (CARBONI, 2003). Ao assumir o controle da informação na tomada de decisões a influência sobre as formas de exercício de poder e, por conseguinte, a propriedade dos bens intelectuais foi sendo massificada.

No que tange aos instrumentos regulatórios internacionais, as questões contemporâneas ainda são exordiais. A Convenção de Berna, um dos instrumentos mais antigos no cenário internacional, responsável pelo estabelecimento de acordos e diretrizes para a proteção intelectual, ainda não contempla as questões baseadas na *e-Science*, mas a cúpula de membros já reconhece as tendências contemporâneas. Em 2002, a Organização Mundial do Comércio (OMC) apresentou uma série de apontamentos importantes relativos aos princípios modernos da propriedade intelectual e tendências para o desenvolvimento do país, concedendo privilégios temporários para o uso de criações, tendo em vista o interesse social e o desenvolvimento tecnológico e econômico. O entendimento enaltece que mesmo havendo inovações e descobertas científicas que possibilitem o desenvolvimento econômico, tecnológico e social, na maioria dos países, o imperialismo econômico proprietário se sobrepõe ao bem público e interesse social coletivo, porém as tendências contemporâneas da ciência colaborativa e aberta não são abordadas pela OMC.

No âmbito dos Acordos TRIPS, pode-se dizer que a função social do direito de autor, expressa no artigo 7º, dispõe sobre a proteção e aplicação de normas de proteção dos direitos de propriedade intelectual, a partir da contribuição na promoção da inovação, transferência e difusão tecnológica. Estabelece que haja um benefício mútuo de produtores e usuários de conhecimento tecnológico e, conseqüentemente, propicie bem-estar social, econômico e um equilíbrio entre direitos e obrigações nessa seara. Mais uma vez, é notória a importância dada por esses instrumentos em prol do desenvolvimento econômico e tecnológico e da inovação para o bem-estar social como princípio da função social, mas ainda prevalece o controle do paradigma proprietário econômico.

A orientação de membros que compõem a cúpula da WIPO, por meio de relações bilaterais, negociações multilaterais, como a ACTA, TPP e arranjos institucionais, impõe ajustes às normas globais, tais como as diretrizes da WIPO e WTO, direcionando as políticas de pesquisa aberta (NAS, 2015).

Nos Estados Unidos e na Europa, o entendimento está na direção do princípio do acesso aberto e nas políticas moldadas por evidências baseadas em pesquisa moldadas para alterações no escopo dos direitos autorais. No Reino Unido, a recomendação é que pesquisadores e autores participem dos debates recentes de direitos autorais em especial no âmbito da pesquisa “na elaboração de questões de investigação, bem como na identificação de oportunidades para estudos comparativos a nível internacional” (NAS, 2015, p. 10, tradução nossa). A Diretiva 96/9/CE, de 11 de março de 1996, do Parlamento Europeu e do Conselho, já previa a proteção *sui generis* das bases de dados. As discussões estão se ampliando para que as legislações autorais e de patentes sejam repensadas para a sociedade digital.

Em suma, a subcategoria aponta para duas reflexões:

- a) A primeira reflexão centra-se no âmbito dos acordos, tratados e convenções que ensejam a predominância do sistema econômico proprietário da propriedade intelectual.
- b) A segunda reflexão está no entendimento que, mesmo lentamente, já existe um diálogo sendo conduzido entre as cúpulas da OMC, WIPO e TRIPS.

Isto posto, observou-se que os sistemas DPIs em âmbito internacional ainda estão arraigados no paradigma proprietário econômico. Reforça-se a garantia protetiva aos criadores (autores e inventores), prevalecendo o caráter individualista e privatístico do direito autoral. A exploração comercial ainda é uma moeda de valor econômico que remunera autores e inventores, e a expropriação de bens intelectuais com a prevalência do monopólio industrial e editorial é imperiosa nos instrumentos regulatórios internacionais.

9.2.2 Domínio Público

A análise dessa subseção foi conduzida para entender o domínio público como um movimento que antecedeu ao *open access*. A noção de domínio público nos moldes tradicionais é definida como os elementos intelectuais que não são protegidos por direitos autorais devido à expiração do prazo de proteção. Por vezes, tal noção é mais rigorosa, incidindo apenas sobre as obras cuja proteção dos direitos autorais já foi extinta, ou mais ampla, dando boas-vindas ao uso, mas legitimando que se trata de uma exceção ou licença exclusiva. A extensão ou restrição do domínio público inclui qualquer recurso cuja utilização é livre do requisito de exclusividade.

A frequência de palavras resultantes da análise de conteúdo gerou a terceira subcategoria Domínio Público. A incidência dos termos ‘data’ e ‘informação’ foi destacada

no material analisado, seguidos pelas palavras *copyright*, *open*, *public*, acesso, propriedade, conhecimento e pesquisa (Figura 22). Levando inicialmente à compreensão que já existem discussões sendo conduzidas para o contexto emergente da *e-Science*, muito embora os termos informação e *copyright* ainda estejam presentes no contexto temático do domínio público.

Figura 22 – Subcategoria domínio público



Fonte: Resultados de Pesquisa do *Software* NVivo (2016).

A visão tradicional do domínio público soa negativamente no cenário da comunicação científica. A abordagem negativa do domínio público prevalece na maioria dos regimes de direitos de autor, mas nem sempre é explícita na legislação. “Os elementos do próprio domínio público em geral não estão sujeitos a quaisquer regras ou proteção: os termos ‘domínio público’ raramente aparecem nas disposições da lei” (DUSOLLIER, 2010, p. 7, tradução nossa). Restrições que são vigentes nas práticas científicas e nos sistemas de propriedade intelectual.

Numa acepção contemporânea, “a terminologia de ‘commons’, ‘commons intelectual’ ou ‘conteúdo aberto’ começou a substituir os termos de ‘domínio público’, de modo a insistir no uso aberto ou livre de materiais de domínio público e da natureza coletiva e compartilhada de cada uso” (DUSOLLIER, 2010, p. 7, tradução nossa). Considera-se esta como a definição mais apropriada do domínio público. A influência do domínio público nos movimentos contemporâneos emergentes, a exemplo dos termos *data*, *open*, *commons* e da ciência aberta, não é formalizada em sistemas legais. Mesmo para publicações científicas

tradicionais, as leis autorais nem sempre preveem a promoção do domínio público em seu escopo.

O domínio público a dois requisitos intrínsecos (exclusividade e gratuidade). Disto isto, a subcategoria apontou para as seguintes considerações:

- a) O requisito **exclusividade** é um dos pontos que devem ser desconsiderado quando se tratar de acesso público. A principal característica ressaltada nas obras de domínio público é a ausência ou expiração dos direitos de autor de um conteúdo, ou seja, de qualquer exclusividade. A exclusividade dada pelo direito de autor implica apenas o controle de atos de exploração pública de obras, que abrange reprodução da obra, comunicação, distribuição e outros atos públicos de difusão do trabalho, tais como aluguel ou comodato. O controle está na difusão pública e reprodução da obra a um público. Pode-se traduzir esta ideia, dizendo que o centro das atenções dos direitos de autor é a exploração de uma obra, definida como a difusão pública. A não exclusividade determina o acesso livre sem restrições legais à publicação e, atualmente, aos dados de pesquisa.
- b) O requisito **gratuidade** é relacionado quando uma obra está sob o regime do domínio público deve ser assegurada a utilização gratuita. Esse é o entendimento para as políticas de dados científicos no âmbito da ciência aberta, a qual norteia esta tese. Em primeiro lugar, salienta-se que a proteção intrínseca de direitos autorais, em certa medida e representa uma liberdade relativa quanto ao acesso.

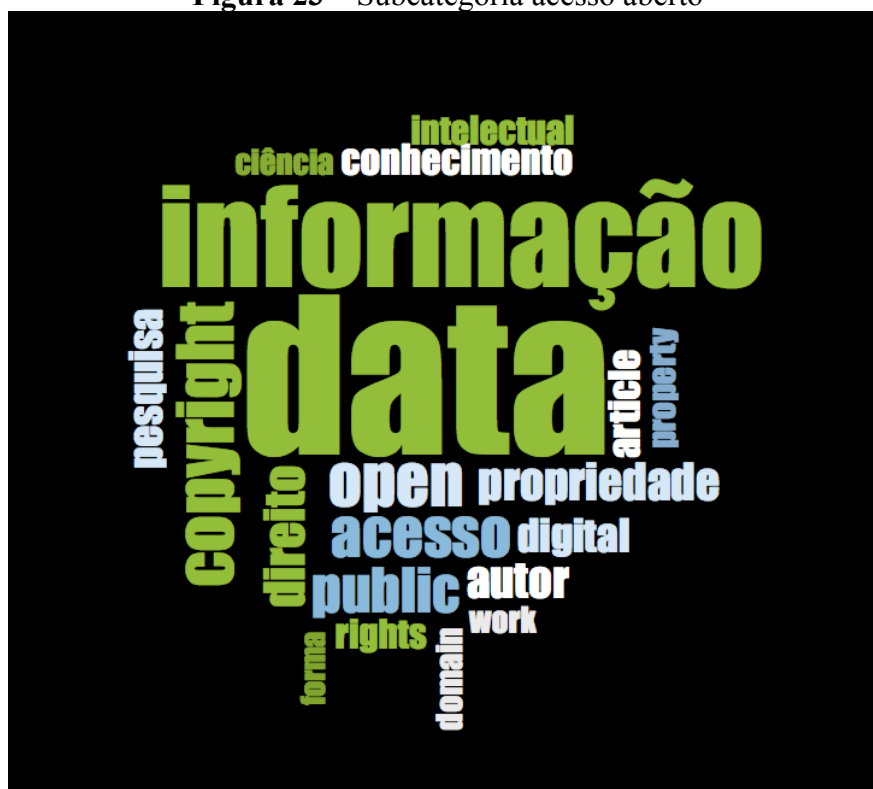
De fato, a possibilidade de desfrutar intelectualmente o conteúdo de um trabalho, ter conhecimento do seu significado, busca proporcionar a qualquer cidadão o acesso livre e gratuito aos resultados de pesquisa. O domínio público e gratuito não isenta que garantias de atribuição e citação sejam fixadas e asseguradas para as fontes de informação, sejam estas, dados de pesquisa e resultados científicos. Pois, um dos elementos que compõem o princípio norteador dos movimentos emergentes abertos são a gratuidade e a não exclusividade. Assim, a subcategoria ajudou na reflexão dos requisitos que irão contribuir para os padrões político-normativo e ético-legal.

9.2.3 Acesso Aberto

A análise dessa subcategoria foi conduzida para entender o acesso aberto como um movimento emergente que conduz as tendências atuais para a ciência aberta. O movimento do acesso aberto denominado de *Open Access* (OA) surgiu a favor da abertura gratuita à

informação e ao conhecimento e como uma vertente que se contrapõe aos sistemas vigentes de propriedade intelectual. Do *corpus* de análise emergiram as discussões contemporâneas evidenciadas pelos termos: *data*, informação, *open*, *public*, *copyright* e acesso, demonstradas na Figura 23.

Figura 23 – Subcategoria acesso aberto



Fonte: Resultados de Pesquisa do *Software NVivo* (2016).

Inicialmente, observou-se que os fatores restritivos vão além das questões tecnológicas e técnicas, mas se dão principalmente nos âmbitos políticos, legais, econômicos e culturais. As TICs e a *internet* criam a infraestrutura e dão sustentabilidade à disseminação da comunicação científica. Todavia, não resolvem as assimetrias no acesso à informação ou dados científicos, se não houver o respaldo político-normativo que preze pelas condições e regulações relativas à manutenção do ciclo de produção e distribuição de informação e dados.

O acesso aberto é constituído como uma prática do acesso *on-line* à informação científica, livre de encargos para o usuário final e favorável para reuso no contexto da investigação e da inovação. Os resultados da pesquisa científica, em grande parte, são compostos pelos artigos científicos revisados por pares (publicado em revistas acadêmicas, geralmente, eletrônicas) e, mais recentemente, por dados de pesquisa (publicações de dados subjacentes, dados curados e/ou dados brutos) baseados nas práticas científicas da *e-Science* que conduziu a três reflexões:

- a) A primeira reflexão direciona o acesso gratuito, público e *on-line* a qualquer cidadão. O **acesso aberto** dar o direito de ler, baixar, imprimir, além de permitir a cópia, distribuição, busca, conexão e rastreamento.
- b) A segunda forma é direcionada aos dados de pesquisa com o direito de acesso, uso e reutilização. O direito de acesso e reutilização de dados de pesquisa sob os termos e condições estabelecidos no contrato de concessão, permite que os mesmos sejam acessíveis, extraídos, explorados, reproduzidos e disseminados a título gratuito para qualquer cidadão.
- c) A terceira reflexão foi compreendida no âmbito brasileiro, com base na lei de acesso para o acesso público. No Brasil, o acesso aberto foi regulamentado por meio da Lei de Acesso à Informação Pública (LAI) que parte do princípio da Transparência Pública do Estado brasileiro pressupostos para a abertura dos Portais de Transparência Pública, Governo Eletrônico e Governo Aberto no Brasil. Tal princípio conduz as políticas públicas de acesso à informação governamental e arquivísticas. e acesso às informações governamentais.

O artigo 3 da LAI busca assegurar o direito fundamental de acesso à informação a todos os cidadãos brasileiros como princípio básico da administração pública e a publicidade e divulgação de informações como interesse público. Os meios de comunicação e tecnologias de informação devem ser utilizados para tornar claros a transparência e o desenvolvimento da administração pública, considerando o sigilo como uma exceção.

Percebeu-se também, que a LAI trouxe uma mudança na dinâmica de comunicação estatal, bem como o desenvolvimento de uma cultura para o acesso aberto junto à sociedade. Embora, não trate de uma legislação específica aos dados científicos, a reflexão serviu de parâmetro para discussões futuras em torno da ciência aberta.

É considerável o avanço na seara governamental com a disponibilidade e acesso à informação pública brasileira. Tal avanço também inspirou os programas na seara eletrônica governamental em prol de uma política voltada para o governo eletrônico e aberto, embora ainda sejam iniciativas lentas e fragmentadas, em comparação ao cenário internacional.

O foco da comunicação científica, antes concentrado na divulgação dos resultados validados, atualmente se preocupa com os dados primários da investigação científica. Ressalta-se que o acesso aberto as publicações científicas não necessariamente é livre de custo e remuneração. Entretanto, é possível observar que a pesquisa moderna, mesmo que lentamente, está caminhando para a abertura e gratuidade às publicações científicas, notadamente, às que estão disponíveis em meio eletrônico. O movimento do acesso aberto

conseguiu uma adesão significativa da comunidade científica, muito embora ainda enfrente desafios para a sua plena adoção nos âmbitos institucionais, legais, governamentais e acadêmicos, visto que a cultura e a política científica de cada país podem representar um fator restritivo ou estimulador da adoção de tais ideais. Assim, qualquer proposição de mudança de paradigma enfrentará contradições, confluências e divergências de interpretação acerca dos benefícios e prejuízos.

Quanto aos dados científicos, ainda é um empreendimento que irá percorrer um longo caminho. Pensar no acesso aberto e domínio público em torno do bem intelectual como um produto final de pesquisa é desafiador. Pois, qualquer mudança desse porte estimula pensar nos impactos políticos, econômicos, legais, sociais e culturais e nos embates que serão reforçados pelos mecanismos jurídicos da propriedade intelectual e mercados editoriais vigentes.

9.2.4 Bem Intelectual Público e Comum

A subcategoria em tela buscou refletir sobre o contexto científico contemporâneo que conduz a uma nova acepção da autoria e do bem intelectual. Nesse sentido, dois aspectos emergem e conduzem a análise:

- a) O primeiro se concentra na proteção da base de dados e sua abrangência de proteção aos conteúdos (publicações e dados) que são dotados de originalidade. O segundo reflete sobre os conteúdos que receberam consideráveis investimentos qualitativos e quantitativos no seu processo de criação são considerados uma publicação *sui generis*.

No primeiro aspecto, é considerada a proteção baseada na originalidade. O segundo aspecto, por analogia ao direito *sui generis* de pesquisas duplicadas e transformadas, pode ser compreendido como um reuso de pesquisa, considerando a existência de uma contribuição (qualitativa ou quantitativa) ao conteúdo que gerou uma contribuição adicional.

Os bens oriundos de criações intelectuais, culturais, criativas e inventivas no escopo da propriedade intelectual sofrem a influência do exercício de poder na esfera pública, prevalecendo a exploração mais predominante na esfera privada, mesmo em ambientes colaborativos e abertos.

Numa abordagem jurídica, **bens públicos** têm dupla acepção: “podem designar bens de uso comum do povo, como bens pertencentes às pessoas jurídicas de direito público interno” (BARBOSA, 2009, p. 10) e para a Economia “são aqueles não rivais e não

excludentes, que podem ser consumidos sem que terceiros se privem dos mesmos, os quais é impossível evitar que terceiros tenham acesso” (BARBOSA, 2009, p. 11). A informação e o conhecimento são considerados bens públicos, e, portanto, comuns a todos. Na perspectiva econômica tais bens são acessados por várias pessoas e utilizados sem que suas características primárias se alterem.

Na era digital, perspectivas colaborativas e abertas direcionam as novas formas de recompensa autoral. Buscam, pois, proporcionar através de bens públicos e comuns benefícios que tragam bem-estar e melhorias para uma determinada coletividade. A NAS tem promovido o debate e apoio às políticas em prol do acesso à pesquisa aberta, ao equilíbrio dos direitos e à recompensa autoral, pois as pesquisas podem “alterar o cálculo incentivo para vários atores no sistema de direitos autorais, impactar os custos das transações voluntárias de direitos autorais, colocar novos desafios de execução e alterar o equilíbrio ideal entre a proteção de direitos autorais e as exceções” (NAS, 2013, p. 10). Os novos contornos na pesquisa científica invocam que essa preocupação de governos, agências de fomento à pesquisa, instituições de ensino e organizações regulatórias sejam repensados.

No que tange a recompensa do direito de autor,

Historicamente, é interessante observar que o direito de autor não nasceu como um direito de propriedade, mas como uma política do governo para concessão de um monopólio bastante restrito, voltado à comercialização temporária de determinadas criações intelectuais. Com o passar do tempo, os tratados internacionais, juntamente com as legislações de diversos países e decisões dos tribunais, passaram a considerar o direito de autor como algo próximo a um direito de propriedade, ou ainda, como um direito natural do criador intelectual (CARBONI, 2009, p. 20).

Assim, a recompensa autoral de criações intelectuais e inventivas foi sendo encapsulada gradativamente nos moldes econômicos e comerciais. Não obstante a essas forças dominantes, as instituições de ensino, bem como as científicas também foram absorvidas no processo de encapsulamento econômico, pois “o interesse da indústria dos bens intelectuais move todo um processo de alargamento do objeto de proteção do direito de autor e de prolongamento do seu prazo de proteção. [...]” (CARBONI, 2009, p. 21).

A figura do sujeito-autor tem condição mandatória de abrir mão de seus direitos em prol de um monopólio editorial ou industrial que possa auferir lucro através da comercialização da obra. No entanto, tal condição não é opressora ao sujeito-autor, pois mesmo em ambientes científicos, a exploração comercial das criações intelectuais do pesquisador (sujeito-autor) é deliberada.

- b) A segunda direciona o entendimento do acesso aos bens intelectuais e culturais, procedentes da própria ecologia social, caminha para uma nova postura assentada na cultura colaborativa e do compartilhamento entre os envolvidos. A economia criativa e o *commons* visam que novas práticas e espaços interativos e colaborativos promovam a criatividade, inovação e saberes científicos, artísticos e culturais a favor e para a sociedade. É a ideia preliminar de uma ciência do ‘comum’.

Assim, buscou-se em disciplinas afins um alargamento para a compreensão desse escopo com os seguintes pressupostos:

b.1) A cultura do ‘comum’ reflete a noção do *commons law*, a teoria do *commons* e da economia criativa para uma ciência aberta.

b.2) Sugere que o ‘comum’ está associado à abertura e gratuidade de recursos intelectuais e criativos, adicionando-se os bens de natureza científica, como dados, informação e conhecimento.

b.3) A economia criativa é um paradigma de mudança que enfatiza o acesso aos bens intelectuais, frutos do processo de criatividade humana, estando enraizados na inovação e, portanto, devem ser de acesso aberto.

A noção do *commons* ou ‘comum’ é enfatizada pela ideia da ‘ciência do comum’ e sua própria condição. Essa derivação é defendida como um dos imperativos da nova gestão da ciência que rejeitam as regulações impostas pelos mecanismos econômicos de propriedade intelectual. A ciência do comum estabelece novas condições para que o conhecimento como um bem público deixe de ser codificado e privado.

O conceito de *commons* é uma “expressão britânica que se referia aos campos comuns onde todos podiam pastar os seus rebanhos – está ligado diretamente ao direito de acesso, uso e controle sobre um determinado bem” (GATTI; CORTIZ; DOWBOR, 2014, p. 149). Isso conduz à primeira inferência de que produção por pares está diretamente associada ao processo de colaboração, compartilhamento e coletividade de bens intelectuais.

Os bens intelectuais são compreendidos como “resultado universalista (humanista) da produção do intelecto de todas as artes ou em terminologia contemporânea, compreende as criações artísticas, técnicas e científicas” (BARBOSA, 2009, p. 13). A informação e o conhecimento entram como sua função utilitária. Assume uma condição de bem intelectual, na perspectiva de ser um bem público e de acesso a todos.

Esse é o entendimento dos bens intelectuais associado à ideia de cultura, conhecimento e criatividade. Tais bens são recursos não comercializáveis no mercado aberto

(FROSIO, 2014). A corrente do *commons* estabelece que “conhecimento deve fluir pelos indivíduos da sociedade sem qualquer barreira, para que novos modelos de criação possam surgir de maneira espontânea” (GATTI; CORTIZ; DOWBOR, 2014, p. 149). O que torna as comunidades inovadoras, a exemplo do *software* livre, que não objetiva produzir para todos, mas construir através da colaboração e do compartilhamento um bem comum. Assim, “bens comuns não são garantidos apenas pelo acesso gratuito, mas não excluem qualquer forma de colaboração para melhorar o resultado. [...] não se referem apenas aos indivíduos, mas também às culturas” (LAFUENTE; ESTALELLA, 2015, p. 40, tradução nossa).

Os bens comuns, de alguma forma, vêm sendo regulados por mecanismos econômicos, legais e políticos. A Economia tem na sua missão principal a alocação racional de recursos escassos: alocação de bens que utilizados em um produto, não estarão disponíveis para outros (DOWBOR, 2011), diferindo do bem intelectual – conhecimento –, gerado principalmente por investigações científicas é considerado um recurso econômico, mas que pode ser consumido simultaneamente por diferentes pessoas. Isto, contrapõe-se ao monopólio dos mercados de conhecimento e informação. Sugere-se como argumento que – o conhecimento e a informação – não devem ser privatizados e restritos, considerando as seguintes premissas: a escassez e o bem não rival.

O bem não rival é “aquele que admite usos simultâneos que não competem entre si. [...] como as ideias, os programas de computador, as obras artísticas, científicas e culturais – são, em geral, não rivais” (SIMON; VIEIRA, 2010, p. 70), e, portanto, não podem ser restringidos. Sendo um bem não rival, se compreende que, são bens públicos, e portanto, irrestritos.

A característica da escassez não afeta o conhecimento por duas razões: primeiro, porque é um bem não rival, pois quem comunica uma ideia, não deixa de tê-la. Segundo, porque é um bem que pode ser transmitido eletronicamente, infinitamente e simultaneamente, por meios digitais e pela *internet*. Esse é um dos argumentos que se contrapõe aos atuais sistemas de DPIs, pois como o mercado editorial e científico encapsulam o conhecimento, tornando-o uma moeda de valor econômico e “[a]s informações que costumavam ser “livres” estão sendo cada vez mais privatizadas, monitoradas, criptografadas e restritas” (FROSIO, 2014, p. 24, tradução nossa) entra em conflito com a Sociedade Digital. Tanto o conhecimento com a informação representam esse paradoxo, por um lado se concentram os conflitos e restrições das leis de propriedade intelectual, e do outro estão as potencialidades das tecnologias e da *internet*.

Nesta tese realizou-se um recorte na abrangência da Teoria do *Commons* com ênfase na cultura do comum e do bem intelectual. A ciência do comum possui distintos direcionamentos (epistemológicos e práticos), “mas ambos compartilham o mesmo **projeto epistêmico** [...], com práticas sustentáveis e protocolos que enfatizam o compromisso com uma cultura aberta, experimental, inalienável, horizontal e distribuída” (LAFUENTE; ESTALELLA, 2015, p. 39, grifo nosso), que sejam conduzidos e destinados à sociedade.

O direcionamento no bojo da Teoria do *Commons* para identificar o equilíbrio entre o acesso livre e público e a garantia de autoralidade. A chave para garantir que os direitos autorais sirvam ao bem-estar público está focada no equilíbrio entre a limitação do *copyright* e os benefícios do acesso aberto aos bens intelectuais. Considerar as exceções e naturezas desses bens intelectuais, visando estabelecer garantias de direitos autorais em algumas circunstâncias, a exemplo de licenças e restrições, é o equilíbrio que se propõe na ciência do comum (FROSIO, 2014). Portanto, o principal requisito baseado no *commons* “é garantir que não haja qualquer indivíduo com **autoridade exclusiva** sobre o recurso, com o objetivo de que qualquer pessoa possa utilizá-lo dentro de algumas regras, determinadas pelos próprios usuários” (GATTI; CORTIZ; DOWBOR, 2014, p. 149, grifo nosso). No contexto científico, a generalização do movimento do acesso aberto direciona para as regras autorais no sistema de comunicação científica diferenciadas.

A WIPO tem conduzido discussões similares nas últimas décadas. Os estudos estão sendo liderados em vários segmentos (indústria e ensino) da sociedade no sentido de identificar e quantificar o alcance econômico dos direitos de autor e suas exceções, seguindo métodos recomendados pela WIPO (2016). Tais estudos visam mensurar o impacto do acesso aberto e domínio público de criações intelectuais e inventivas. Levam ao entendimento que o acesso ao conhecimento e aos recursos naturais, por exemplo, conduzidos pela ideia do *commons* e por movimentos, tais como *open access* e *open science*, estão alargando o espectro e a necessidade de repensar o Instituto da Propriedade Intelectual.

Por fim, observando a natureza peculiar do *commons* e a sua ligação com a ciência, Frosio (2014) propôs a criação de um *commons* intelectual através do acesso aberto. Para tanto, sugere a revisão das leis de direito autoral com a finalidade de:

- a) Abranger o alargamento e a proteção do domínio público revertendo extensões nos prazos de direitos autorais, acrescentando-se a isto as novas modalidades de pesquisa científica.
- b) Garantir que a lei de *copyright* estabeleça contratos ou licenciamento adequando prazos mais restritos.

- c) Estabelecer na doutrina as regras de venda de conteúdo digital.
- d) Restaurar a negociação de uso justo de direitos negados por medidas tecnológicas de proteção.

Tal propositura não é tarefa fácil de ser adotada mundialmente. Pensar na reformulação de legislações autorais é uma medida que deve ser conduzida em âmbito nacional. Adaptar os modelos vigentes de propriedade intelectual e de comunicação científica para uma cultura colaborativa, compartilhada e digital é um longo caminho a percorrer. As questões relativas à colaboração, compartilhamento e acesso aberto nesse contexto do *commons* deve ser pensado com cautela e relatividade, pois cada país tem um contexto econômico, político, científico e sociocultural distinto. Entretanto, a busca pelo equilíbrio entre os sistemas tradicionais e os movimentos de acesso aberto dos modelos emergentes é um imprescindível para conduzir as políticas orientadas à ciência aberta. É inconcebível no século XXI a continuidade de regimes proprietários que encapsulem os bens intelectuais – conhecimento e informação, e recentemente dados de pesquisa – como bens privativos e exclusivistas.

O resultado, obviamente, não é um produto pronto e acabado, mas uma forma de entender a nossa relação com os novos modelos científicos, com as tecnologias abertas e com outros seres humanos, baseando-se no princípio da comunicação eletrônica, colaborativa e aberta, o intuito é que comunidades e pares dissolvam as fronteiras artificiais e imaginárias e compartilhem livre e gratuitamente os bens intelectuais (LAFUENTE; ESTALELLA, 2015).

Na mesma direção, a Economia Criativa avança como um paradigma da inovação orientado aos bens intelectuais provenientes da criatividade humana. A atividade derivada da criatividade não somente gera empregos e riqueza, como também incrementa o bem-estar da população em geral, promove a expressão e participação dos cidadãos politicamente, e ajuda na manutenção de espaços culturais e de conhecimento popular (NOGUEIRA; GUIMARÃES; XAVIER, 2016). Essa ideia está na essência de pesquisas que brotam da ciência do *commons*, ciência aberta e ciência-cidadã.

A criatividade, nesse contexto, surge como um conceito abrangente e correlato à Cultura. Parte-se da premissa de que todo o desenvolvimento industrial e econômico tem por base indivíduos que criam processos a partir de ideias que acabam por gerar uma riqueza dificilmente mensurável pelos padrões econômicos tradicionais. “Unir essas duas palavras, “Economia” e “Criatividade” – que aqui entendemos como cultura –, no entanto, não foi uma tarefa fácil” (NOGUEIRA; GUIMARÃES; XAVIER, 2016, p. 12).

Preliminarmente, entende-se que a Economia Criativa está relacionada com processos produtivos que envolvem a criatividade, conhecimento e capital intelectual como fontes de recursos e resultam em produtos e serviços com valor econômico. O produto criativo é fruto do labor intelectual que também é apropriado por mecanismos de DPIs vigentes.

Mais uma vez, a discussão se apoia na exploração dos direitos de propriedade intelectual da criação humana retidos por uma minoria. A ideia precípua é que o *commons paradigm* proporcione um equilíbrio entre os direitos básicos da sociedade e a garantia da autoralidade do processo criativo. A Economia Criativa somente se torna viável num contexto de desenvolvimento sustentável com liberdade, abundância e proteção dos direitos básicos, garantindo o desenvolvimento pleno da personalidade humana, sem limitar o pensamento intelectual e criativo.

A mudança de paradigma da inovação traz consigo o desafio de delimitar e proteger a criação intelectual, “e é nessa discussão que encontramos a função central do Direito como meio de regulação dessas interações e dessa nova forma de fazer Economia” (NOGUEIRA; GUIMARÃES; XAVIER, 2016, p. 17). É nisso que reside a maior discussão no bojo da Economia Criativa. À medida que, o paradigma da inovação se torna mais orientado ao uso colaborativo, empresas precisam ajustar a sua estratégia de gerenciamento de direitos de propriedade intelectual e desenvolver ferramentas para gestão aberta, “é nessa relação entre economia e conhecimento que nasce o conceito de Economia Criativa, dentro da concepção da sociedade da informação” (WACHOWICZ; SILVA, 2011, p. 562, tradução nossa).

Ademais, “essa revolução criativa, que coloca o criador no centro de convergência econômica em contraposição ao produto/criação, apresenta ainda uma faceta de fundamental relevância no que tange à tutela dos direitos autorais” (NOGUEIRA; GUIMARÃES; XAVIER, 2016, p. 18). Por vezes, o lucro e a comercialização desse produto intelectual são realizados por terceiros. Os direitos autorais nem sempre são creditados aos produtores.

Na contramão do que historicamente vem sendo conduzido, organizações não governamentais e instituições governamentais já estão reconhecendo o valor da Economia Criativa como um paradigma de inovação e, analogamente, como uma cultura do comum.

Em setembro de 2013, o Parlamento Europeu reconheceu a necessidade de reformar regras de direitos autorais da *European Union* (EU) no âmbito da sua resolução (T7-0368/2013) que trata da promoção dos setores culturais e criativos europeus como fontes de crescimento econômico e de emprego. Nessa resolução, o Parlamento Europeu:

52. Salaria que a existência de 27 diferentes sistemas de gestão de direitos de propriedade intelectual é uma carga excessiva para setores culturais e criativos da Europa, e que o regime fragmentado atual precisa ser reformado para facilitar o acesso e aumentar a circulação (global) de conteúdos, e de tal maneira a permitir que artistas, criadores, consumidores, empresas e público se beneficiem da evolução digital, novos canais de distribuição, novos modelos de negócios e outras oportunidades (EU, 2013, *on-line*, tradução nossa).

Observou-se nessa cláusula que a quantidade de sistemas de gestão de direitos autorais dificulta o acesso e a circulação de bens culturais e criativos na Europa, além do excesso de restrições ao uso. A orientação é que novos modelos e canais de distribuição possam beneficiar todos os envolvidos no processo criativo, além de favorecer a sociedade com o acesso ao conhecimento gerado. A cláusula seguinte sugere que mudanças nos sistemas atuais poderão trazer equilíbrio entre a proteção e o acesso.

53. Considera que, na era digital, um sistema moderno e equilibrado para proteger os direitos de propriedade intelectual (DPI), que torna possível tanto para garantir uma remuneração adequada para todas as categorias de titulares de direitos e para garantir que os consumidores tenham acesso fácil a diversos conteúdos legais e uma verdadeira escolha em termos de diversidade linguística e cultural, é uma condição essencial para assegurar que as CCS são competitivas (EU, 2013, *on-line*, tradução nossa).

Sugere também que a remuneração de autores e criadores possa ser garantida sem limitar o acesso ao conteúdo. Estimula-se a competitividade dos setores culturais e criativos, o que eleva a produção e consumo do conhecimento.

Isto posto, o escopo da análise apresentou em sua essência o entendimento a partir da perspectiva colaborativa e coletiva da Teoria do *Commons* e da Economia Criativa. Tal essência foi observada a partir de um olhar diferenciado para o bem intelectual, a exemplo dos produtos intelectuais, criativos e culturais, e mais especificamente a informação e conhecimento e, por conseguinte, os dados de pesquisa, como bens públicos. Assim, como bem público, assume a obrigatoriedade de uma essência do comum, que visa proporcionar além do acesso gratuito, benefícios a uma determinada coletividade.

Analogamente, buscando trazer o que foi inferido para o escopo da ciência aberta e das práticas da *e-Science*, algumas constatações emergiram:

- a) A constituição e o acesso aos dados científicos deverão considerar a essência colaborativa e do comum, tal qual um bem intelectual e público.
- b) O equilíbrio entre os direitos de propriedade intelectual e o acesso aberto deve ser repensado no bojo de políticas e sistemas de proteção, isto é, a garantia da autoralidade primária considerando o requisito da originalidade.

- c) O reconhecimento de que os dados científicos são um produto intelectual fruto de um processo científico e, portanto, possuem importância para a coletividade, devendo ter seu acesso livre e gratuito, principalmente quando se tratar de áreas de sustentabilidade para a sociedade.

A orientação é que novos modelos e canais de distribuição possam beneficiar todos os envolvidos na criação, assim como a sociedade para acesso aos produtos intelectuais gerados. Assim sendo, a análise continuará a ser conduzida pelas recomendações político-normativas inspiradas nos movimentos do acesso aberto, domínio público e iniciativas em prol da ciência aberta em expansão mundial.

Nos Estados Unidos, a *National Science Foundation* estabeleceu, em 2013, requisitos para o compartilhamento e planos de gestão de dados a partir de uma política federal. O *U.S. Federal Policy* determinou nessa política o acesso aberto às publicações e aos dados de pesquisas financiadas com recursos públicos nos Estados Unidos.

A União Europeia também instituiu recomendações para políticas de acesso livre aos dados de investigação. O Reino Unido, desde 2005, proporciona ao público um direito de acesso à informação mantida por uma autoridade pública, que inclui a maioria das universidades, faculdades ou instituições de investigação com financiamento público. Essas informações solicitadas podem incluir seus dados de pesquisa, e devem ser fornecidas, salvo se uma isenção ou exceção de direitos autorais indique que não pode divulgar a informação.

Preocupações e recomendações voltadas para a política de informação científica são norteadoras e servem de componentes para o desenvolvimento de políticas direcionadas aos dados de pesquisa. No âmbito das agências de fomento é entendido que:

A citação de dados e a atribuição têm componentes políticos. Muitos especialistas estão preocupados com a política de informação acadêmica, incluindo agências de financiamento, editores, repositórios de dados, universidades, pesquisadores e estudantes. Cada um tem preocupações políticas, portanto, devemos perguntar qual é a política, que tipos de política, e de quem é a política? Os requisitos de planos de gerenciamento de dados e políticas de compartilhamento de dados são bons exemplos (NSF, 2012, p. 7, tradução nossa).

Estes requisitos serão explicitados com mais detalhes em seção específica. Por ora, o entendimento acerca do domínio público e do acesso aberto precede as políticas autorais.

9.3 DIRETRIZES POLÍTICO-NORMATIVAS ORIENTADAS AOS DADOS CIENTÍFICOS

Ainda sobre a égide político-normativa, a quinta subcategoria emergiu do escopo das iniciativas contemporâneas voltadas para diretrizes que norteiam os dados científicos no bojo da ciência aberta e das práticas da *e-Science*. As diretrizes reforçam que o movimento do acesso aberto promove a economia inteligente e sustentável, constrói resultados a partir de pesquisas anteriores com o reuso de dados, incentiva e acelera a inovação, envolve cidadãos comuns no contexto da ciência e reforça a busca por melhores práticas científicas (EU, 2015).

A pesquisa moderna se baseia no diálogo científico extenso e avança para um trabalho que antecipa a divulgação dos resultados por meio dos dados de pesquisa. O conceito vigente de publicação evoluiu no contexto da era digital. A política para os dados está alicerçada em uma base normativa que estabelece termos e condições convencionadas ao contexto do acesso aberto e da preservação de dados de pesquisa. No contexto da política orientada aos dados de pesquisa, a recomendação é levar em consideração interesses institucionais e de parceiros, pesquisadores, autores e colaboradores.

O compartilhamento e o gerenciamento de dados oriundos de investigações científicas já são considerados um empreendimento global. Muitas iniciativas internacionais foram ou estão em desenvolvimento para conduzir as práticas científicas da *e-Science*. Tais iniciativas pressupõem a adoção de práticas mandatórias e rotineiras na investigação científica em instituições de financiamento, governamentais, acadêmicas e de pesquisas, em especial nos Estados Unidos, Holanda, Reino Unido, Inglaterra e Austrália.

A análise foi norteada por iniciativas internacionais baseadas nas políticas e diretrizes para a gestão e compartilhamento de dados de pesquisa a fim de entender como os países estão estruturando-as. O estudo de Oliveira e Silva (2016) levanta duas observações:

- a) A primeira observação diz respeito ao fato que algumas iniciativas se encaixam na noção de *soft law*, a qual se refere a regras que não são estritamente obrigatórias e nem contêm uma significação legal, mas servem como princípios para nortear processos, ações e comportamentos.
- b) A segunda observação refere-se ao fato da defesa da abertura dos dados como uma forma de acelerar a produção científica, diminuir custos e possibilitar o desenvolvimento social.

Inicialmente, já se observa que é imperioso o engajamento do governo e da comunidade científica na elaboração de diretrizes e princípios para que a ciência aberta e a

gestão de dados de pesquisa sejam uma estratégia de impulsionamento para o desenvolvimento científico, econômico, social e cultural do país. Os princípios elaborados pela OECD, Comissão Europeia e NWO exemplificam em suas recomendações a noção de *soft law*. Para tal empreitada, entendeu-se que conjugações político-normativas e ético-legais e culturais estão agregadas às diretrizes e políticas.

A Comissão Europeia por meio de sua *Directorate General for Research & Innovation* também publicou o *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*. Utilizou das declarações: *Budapest Declaration* (2002) e *Berlin Declaration* (2003) para definir o acesso como o direito de ler, baixar, imprimir, copiar, distribuir, buscar, rastrear e minerar informação e dados. E assim, propuseram o *Model Grant Agreement: H2020 General MGA* (2015).

Observou-se que essa iniciativa é uma estratégia da Comunidade Europeia para promover a busca por uma economia inteligente, sustentável e inclusiva na geração do conhecimento e da inovação para o desenvolvimento dos países envolvidos. O direcionamento da EU (2015) sugere que a ampliação do acesso às publicações científicas, resultados de pesquisa e dados subjacentes da investigação busca:

- a) Construir sobre os resultados de pesquisas anteriores a melhoria da qualidade dos resultados baseados no reuso dos dados científicos;
- b) Fomentar a colaboração e evitar a duplicação de esforços para promover uma maior eficiência;
- c) Acelerar a inovação para estimular o crescimento mais rapidamente do mercado;
- d) Envolver os cidadãos e a sociedade para propiciar a melhoria da transparência do processo científico na ciência;
- e) Aumentar os benefícios do investimento público na investigação científica e, consequentemente, a visibilidade institucional.

Por estas razões, a European Commission se esforça para melhorar o acesso à informação científica e financiada no âmbito do Programa-Quadro da UE para a Investigação e Inovação Horizonte 2020 (2014-2020). Criado pela Comissão Europeia, o projeto foi denominado de *Open Research Data Pilot* como forma de estimular pesquisadores a aderirem ao depósito e ao compartilhamento de dados de pesquisa (EC, 2015).

O projeto também prevê diferentes fontes de dados, a exemplo, “o oleoduto exige como entrada uma combinação de dados fenotípicos, genéticos e de expressão que necessitam ser processados para determinar uma lista de genes associados a doenças” (EC, 2015, *on-line*, tradução nossa). O piloto sobre a *Open Data Research* será monitorado durante todo o

Horizonte 2020 com vista a desenvolver a política da Comunidade Europeia para investigação aberta.

Os requisitos legais que visam melhorar e maximizar o acesso e reutilização de dados de pesquisa gerados por projetos financiados estão contidos no Artigo 29.3. Esse visa estimular os pesquisadores a compartilharem seus dados e resultados preliminares da investigação. O *caput* do artigo estabelece que os beneficiários que participam do projeto-piloto de dados abertos devem seguir ações preestabelecidas. Ressalta-se que no entendimento do projeto, o beneficiário pode ser: pesquisador, grupo de pesquisa e instituição que obteve financiamento com recursos públicos para a execução do projeto.

A alínea (a) do Artigo 29.3 prevê que o beneficiário deve “depositar em um repositório de dados de pesquisa e tomar medidas para tornar possível a terceiros para acessar, explorar, produzir e divulgar - de forma gratuita para qualquer usuário”.

Duas condições foram estabelecidas: “(i) os dados, incluindo os metadados associados, necessários para validar os resultados apresentados nas publicações científicas o mais rapidamente possível; (ii) outros dados, incluindo os metadados associados, como especificado e dentro dos prazos estabelecidos no ‘plano de gestão de dados’” (EC, 2015, p. 66, tradução nossa).

Os dados científicos e os seus metadados devem ser disponibilizados abertamente e acessíveis para o compartilhamento, exploração e reuso por terceiros.

A alínea (b) traz a recomendação ao beneficiário (pesquisador ou curador) de que forneça informações “através do repositório - sobre ferramentas e instrumentos à disposição dos beneficiários e necessários para validar os resultados (e - se possível - forneça as ferramentas e os instrumentos a eles próprios)”.

Por ora, identificou-se que na visão da Comissão Europeia é que a pesquisa financiada pelo erário público não deve ser paga novamente a cada vez que é acessada ou utilizada. Preconiza os benefícios às empresas, cidadãos e instituições públicas. Isso significa que a informação científica disponível *on-line*, sem custo adicional, para os investigadores, indústrias inovadoras e cidadãos deve ser preservada em longo prazo. Enfatiza que acesso aberto não interfere na decisão de explorar comercialmente os resultados da investigação, por exemplo, através de patenteamento.

Nos Estados Unidos, a política é mandatória e estabelece critérios mínimos para a gestão de dados de pesquisa, interoperabilidade, plataformas abertas, investimentos necessários para adequar os sistemas existentes, responsabilidades dos atores envolvidos em todo o ciclo de vida da informação e dos dados e mecanismos de responsabilização. Visa,

ainda, que resultados preliminares de pesquisa e os dados de pesquisa se tornem o combustível para novas percepções. Ela considera que os dados de pesquisa são ativos para o progresso e sustentabilidade em áreas como saúde, energia, ambiente, agricultura e segurança nacional.

A NSF financia projetos que objetivam abordar aspectos essenciais vinculados ao tratamento, curadoria e compartilhamento de dados científicos para uso e reuso. Como é uma instituição de grande representação científica, ela estabeleceu algumas recomendações para o contexto da *e-Science*:

- a) A criação de uma infraestrutura sustentável de apoio à investigação e inovação de dados em nível mundial;
- b) Essa infraestrutura é essencial para manter a liderança de longo prazo dos EUA na investigação científica e um legado que pode conduzir futuras descobertas, inovação e prosperidade nacional;
- c) Para realizar este potencial, os desafios e oportunidades precisam de investimentos focados e sustentáveis.

Nesse sentido, a infraestrutura sustentável requer o desenvolvimento de políticas nacionais e institucionais, diretrizes que conduzam às práticas da *e-Science*.

Na Holanda, a NWO defende que os resultados de pesquisa financiados por fundos públicos devem ser de livre acesso em todo o mundo. Por isso, um conjunto de regras mais rígidas foi elaborado para guiar a publicação científica dessas pesquisas e dos seus respectivos dados em periódicos abertos e repositórios de dados que, de alguma forma, dão acesso gratuito aos resultados de pesquisa.

O atual cenário da pesquisa aberta baseia-se em um novo diálogo científico extenso e complexo quebrando a hegemonia de publicar apenas os resultados finais da investigação, mas se apresenta com a antecipação preliminar de tais resultados através do compartilhamento de dados. A condução da análise dessa subcategoria e com base nos estudos de Oliveira e Silva et al. (2016, no prelo) permitiu inferir alguns pressupostos norteadores para a formulação de políticas orientadas aos dados científicos:

- a) As políticas devem permitir a compreensão do paradigma científico e de suas práticas vigentes, visto que o movimento em prol do compartilhamento dos dados de pesquisa representa um somatório ao processo de produção e comunicação científica estabelecido, a exemplo, dos periódicos científicos abertos;
- b) As políticas ulteriores devem ser consideradas na concepção de novas políticas e no desenho de arcabouços normativos, legais e institucionais. Políticas científicas,

tecnológicas e econômicas são transversais às políticas de informação e, por conseguinte, às políticas orientadas aos dados;

- c) As peculiaridades das distintas áreas de conhecimento, suas prioridades, metodologias e práticas investigativas, bem como a natureza dos dados de pesquisa deverão ser consideradas no escopo político-normativo;
- d) É recomendável estimular a iniciativa de uso, reuso e preservação dos dados de pesquisa como empreendimento institucional e social, encorajando assim a pactuação entre pesquisadores, instituições científicas, acadêmicas, instituições governamentais e não governamentais, setor privado, bibliotecas e demais profissionais envolvidos nesse cenário;
- e) Fomentar as condições materiais, tecnológicas, humanas e legais necessárias ao compartilhamento, uso, reuso e reprodutibilidade de dados de pesquisa como estratégia para alavancar a colaboração científica internacional e o desenvolvimento científico, tecnológico e social do país.

Neste ponto, convergem os movimentos de defesa ao acesso aberto à informação pública, tecnologias e protocolos abertos, educação aberta e outros sistemas abertos e colaborativos. Esses pressupostos servem como guias para nortear a reformulação ou elaboração de legislações, orçamentos, financiamento de projetos de pesquisa e reformas institucionais no contexto da ciência aberta e no direcionamento das práticas científicas da *e-Science*.

No plano político, as novas práticas de colaboração e compartilhamento têm um papel relevante para a difusão de informação, cultura e conhecimento. Muito embora, reformulações e mudanças serão sempre permeadas por conflitos de interesses dos atores envolvidos, bem como pela lógica do capitalismo cognitivo que representa, paradoxalmente, um impeditivo às atualizações legais e institucionais que dão suporte à ciência aberta e, por outro lado, um estimulador de práticas de compartilhamento e colaboração como estratégia de barateamento e agilização da produção científica.

Os esforços da ciência aberta envolvem instâncias de “ação e decisão diferenciadas, internas e externas à ciência, que vão desde o pesquisador individual e equipes de pesquisa até o nível macro das políticas públicas e das regulações internacionais, passando pelo nível meso das instituições científicas e agências de fomento” (ALBAGLI, 2015, p. 21). Assim, a ciência aberta se constitui um *locus* de poder e negociação, contradições, ideologias, culturas, saberes e conhecimentos.

9.3.1 Princípios Normativo-legais para Dados Científicos

Ainda sobre a égide da categoria diretrizes político-normativas emergiu essa subcategoria que foi norteada pelo documento intitulado *Principles and Guidelines for Access to Research Data from Public Funding* da *Organisation for Economic Co-operation and Development*. O referido documento é uma das iniciativas pioneiras e considerado nesta análise uma das principais diretrizes para a conformidade normativo-legal na *e-Science*. As diretrizes orientam ações estruturantes de projetos para modelos colaborativos e abertos e promovem, também, uma cultura de abertura e compartilhamento de dados de pesquisa entre as comunidades científicas dos países membros.

Tendo em conta que “a natureza do financiamento público da investigação varia significativamente de um país para o outro, estes Princípios e Diretrizes reconhecem que tais diferenças exigem uma abordagem flexível para o acesso aos dados da pesquisa” (OECD, 2007, p. 14, tradução nossa). As recomendações visam que governos se preparem para assumir um compromisso político na implementação de princípios e diretrizes governamentais, respeitando as diferenças em contextos legais, culturais, econômicos e sociais buscando praticar a flexibilidade dos países membros sem ter caráter mandatório. A implementação pode ser alcançada através de medidas regulamentares ou ação coordenada entre as partes nacionais interessadas e apoiar as organizações à pesquisa e ao financiamento. Além de incentivar o intercâmbio internacional e o acesso aos dados de investigação científica.

No que tange ao contexto científico as orientações da OECD estão direcionadas a conduzir o acesso aos dados científicos, sob diversos aspectos:

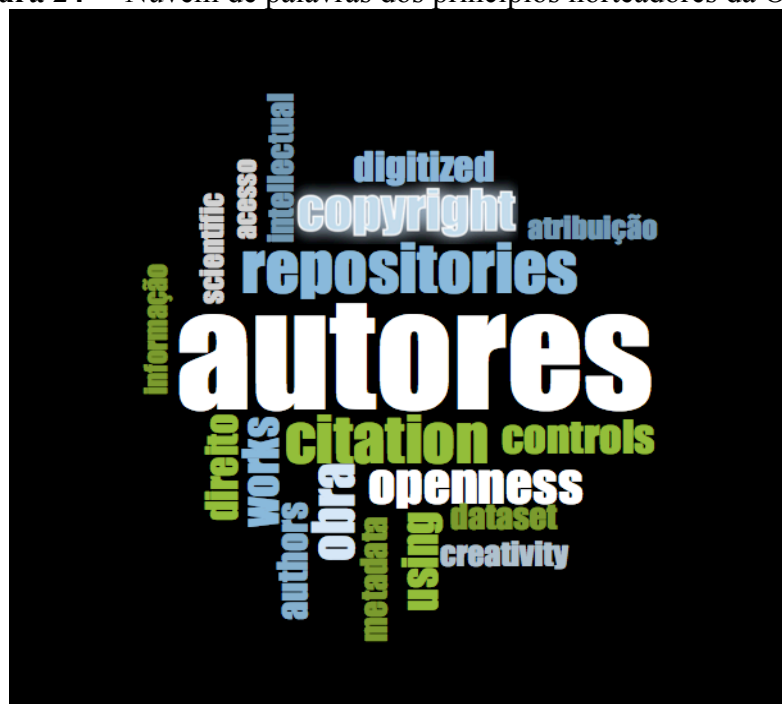
- a) Questões tecnológicas, o acesso aos dados da investigação e a sua exploração requerem uma infraestrutura tecnológica adequada, amplo acordo internacional sobre a interoperabilidade e controles eficazes de qualidade e proveniência.
- b) Questões gerenciais e institucionais são relacionadas a amplitude do acesso às comunidades científicas. A diversidade do empreendimento científico sugere que uma variedade de modelos institucionais e abordagens de gerenciamento de dados personalizados tornarão os projetos mais eficazes na satisfação das necessidades dos investigadores.
- c) Questões financeiras e orçamentárias estão vinculadas à infraestrutura de dados científicos que requer um excelente e contínuo planejamento orçamentário e de apoio financeiro adequado. O uso de dados de pesquisa não será maximizado se o

acesso, o gerenciamento e os custos de preservação forem um suplemento ou reflexão tardia nos projetos de investigação. É importante notar, no entanto, que o custo de armazenamento e gerenciamento de dados diminui drasticamente nos últimos anos e a falta de conhecimento sobre tais mudanças podem, por si só, ser uma barreira para a manutenção e avanço científico. Por isso, a recomendação de adotar recursos, ferramentas e *software* de código aberto e gratuito.

- d) Questões legais e políticas referem-se às leis nacionais e aos acordos internacionais, particularmente em áreas como direitos de propriedade intelectual e direito de privacidade, os quais afetam diretamente as condições e restrições do acesso e uso das práticas de compartilhamento dos dados.
- e) Quanto às questões culturais e comportamentais, referem-se às estruturas educacionais e de remuneração adequadas. As considerações aplicam-se a quem irá financiar, produzir, administrar e usar os dados de pesquisa, sendo um componente necessário para promover o acesso e o compartilhamento.

Os esforços para um projeto de abertura de dados científicos precisam ser realizados dentro das limitações financeiras existentes e de políticas preestabelecidas nacionalmente. Os princípios estabelecidos pela OECD são abrangentes e constituem-se como norteadores para políticas de dados no contexto da *e-Science*. A representação gráfica na Figura 24 ilustra a predominância dos termos contidos nos princípios norteadores.

Figura 24 – Nuvem de palavras dos princípios norteadores da OECD



Fonte: Resultados de Pesquisa do *Software* NVivo (2016).

Um recorte foi realizado no referido documento para que a reflexão seja específica no que diz respeito para os aspectos legais dessa subcategoria, emergindo três princípios: Proteção da Propriedade Intelectual, Conformidade Legal e Responsabilidade Formal.

9.3.1.1 Princípio Norteador 1: Proteção da Propriedade Intelectual

A visão tradicional da proteção autoral está associada aos critérios protetivos de obras intelectuais (originárias e derivadas). As obras intelectuais (científicas, literárias, artísticas e culturais) nascem de forma autônoma, sem qualquer vinculação com obras anteriores e geram derivações. Assim, a obra originária é a primeira criação e a obra derivada é o processo de transformação, que parte de uma obra preexistente. Porém, a obra derivada depende da autorização prévia do criador original da obra. O direito de autor protege tanto a obra originária quanto a que dela se derivou. Tal analogia deve ser considerada para as possibilidades de reuso dos dados primários (originários) no contexto da *e-Science*. Os critérios que tragam a garantia da originalidade da obra ao seu criador primário não é impedimento para o acesso. Mesmo em uma era digital com tecnologias e recursos abertos, é necessário reconhecer que a originalidade pela primariedade das criações primárias é importante e necessita de adaptação para novas dinâmicas e produtos científicos provenientes da *e-Science*.

Nessa direção os elementos que constituem uma criação autoral são descritos por Carboni (2009, p. 14) como componentes fundamentais para a garantia da primariedade:

- a) **Esteticidade:** na *e-Science*, o valor estético é associado à utilidade prática de tecnologias abertas destinadas à sustentabilidade da ciberinfraestrutura que apoia todo o ciclo de vida dos dados. Pois na visão tradicional a esteticidade está voltada para “as obras protegidas pelo direito de autor são as que possuem valor estético autônomo, que se encerram em si mesmo, independentemente da sua origem, destinação ou utilidade prática” (CARBONI, 2009, p. 14), a exemplo dos *software* e base de dados. Portanto, é por meio de ferramentas, protocolos, padrões, *software*, repositórios e aplicações abertas e digitais que a esteticidade da primariedade é conduzida e promove a originalidade primária dos dados primários e a sua derivação a partir do reuso.
- b) **Aporte trazido pelo autor:** na *e-Science* é o reuso dos dados que gera uma criação derivada e o aporte intelectual do criador originário é garantido. Aos pesquisadores e colaboradores envolvidos na criação derivada é promovida a

atribuição pela sua responsabilidade proporcional no processo intelectual de pesquisa. Este é um componente-chave que representa o labor intelectual autoral, tão defendido nos sistemas de direitos autorais. É o que representa a criação primígena da autoria primária. Para o reuso dos dados se entende “a criação intelectual resultante de uma atividade intelectual própria, que acrescente algo de novo à realidade do mundo” (CARBONI, 2009, p. 14). É a criação derivada possui, sobremaneira, aportes intelectuais a partir de acréscimos, melhorias e transformações que foram feitas na utilização de dados científicos reusados em um novo processo de pesquisa.

- c) **Forma:** este componente não está associado diretamente ao conteúdo ou à forma estética da criação. É um componente revestido de subjetividade (*corpus mysticum*), mas dotado de caráter estético intrínseco (CARBONI, 2009). Este é um componente que por si só já expressa a subjetividade e, ao mesmo tempo, se materializa na exterioridade. Tal subjetividade é a essência do conteúdo representada por qualquer formato externo (físico ou digital). Está relacionada aos componentes intrínsecos da criação. Essa relação subjetiva e intrínseca é composta pela especificidade e natureza que compõem o conteúdo, adicionada às referências de estilo, pessoais, éticas, culturais ou institucionais. São elementos peculiares ao processo mental da criação antes que a mesma assuma a sua exterioridade. Mesmo se tratando do reuso de dados científicos, estes por si só já possuem sua forma original digital e legível por máquinas. O que difere é o valor intrínseco que cada forma foi derivada, por elementos de ressignificação, recriação e transformação relacionados ao novo domínio específico da pesquisa a partir da criação original.
- d) **Inserção em suporte:** tradicionalmente é o *corpus mechanicum* que assume a forma exteriorizada que recai sobre a proteção autoral. Por vezes, esta forma exteriorizada pode ser instantânea, como é o caso de comunicação oral ou estilos de expressão corporal. O suporte é denominado de *corpus mechanicum*, ou seja, é a exterioridade da criação a partir por meio de uma forma física ou digital, que “para que haja proteção autoral, a ideia precisa ser materializada em um determinado suporte” (CARBONI, 2009, p. 14). Contudo, ao assumir a exterioridade, a criação é revestida de proteção legal autoral. No caso do **suporte** aos dados, a sua natureza e as condições extrínsecas podem ser utilizadas para representar a forma expressa digitalmente. Os diferentes tipos de dados

científicos são oriundos de pesquisas (observacionais, experimentais, simulacionais) que trazem consigo suas características intrínsecas e possuem características diferenciadas nos mais diversos dados científicos. Porém, a garantia de proteção da autoralidade original e derivada deverá ser realizada. Considera-se que tal garantia, no caso de dados científicos, é realizada pelos procedimentos de curadoria (armazenagem e preservação) adotados na sistemática de um modelo de ciclo de vida.

- e) **Originalidade:** a criação original é princípio de todo processo intelectual. É composta por elementos que tornem tal criação única e inconfundível, revestindo-se de traços e caracteres próprios. “Essa originalidade deve ter caráter relativo, uma vez que é inevitável o aproveitamento, até inconsciente, de conteúdos comuns de outros criadores” (CARBONI, 2009, p. 14). Paira nesse componente a originalidade da criação que, por conseguinte, configura a autoria originária. A doutrina para a desconsideração da originalidade absoluta, pois a composição de uma obra parte de uma junção de concepções intelectuais, retirando o caráter absoluto e passando a configurar a originalidade relativa.

Isto posto, a análise desses componentes foi utilizada para o entendimento de uma nova acepção da proteção autoral direcionada ao foco da originalidade relativa da autoria na *e-Science*.

Outra expressão emergida da análise foi a temática *repositories*. Sua relação está vinculada à expansão em que as questões protetivas com outras configurações, além dos direitos autorais e de *copyright*, como, por exemplo, quando se trata da mineração de dados (*data mining*) e a ideia do alcance sem fronteiras geográficas. A partir da abertura do acesso aberto eletrônico de publicações em bases de dados, a proteção autoral mais contundente no *copyright* teve um direcionamento para a introdução da liberação de cópias temporárias (custo e/ou autorizações necessárias) em serviços *on-line* de bibliotecas tradicionais e recentemente para os repositórios, por exemplo.

Essa é uma discussão reiterada pela *International Federation of Library Associations* (IFLA) como uma preocupação referente à criação de novas proteções de direitos de autor. Tais discussões remetem à observância que, através das funções essenciais da *internet*, as obras que não possuem liberação aberta de uso, sem as legislações vigentes em cada país, levam a uma possibilidade de violação por parte dos usuários IFLA (2014). A ideia preliminar da é que ao identificar e controlar as publicações e os dados, estes estejam habilitados para o consumo.

A utilização de mecanismos automáticos de indexação, *cache* e outras funções de *internet* se possa minimizar “que esses novos usos ocorram sem a regulação de direitos autorais trazendo um efeito colateral às publicações que já estejam garantidas por *copyright* anteriormente” (IFLA, 2014, *on-line*, tradução nossa). Esse é um ponto de reflexão que mostra ainda ser uma prática predominante o fato de que os sistemas de propriedade intelectual utilizam recursos de proteção baseados no modelo proprietário. Considera que se não houver uma flexibilização dos critérios nos termos de licenças e condições (limites e restrições), a exemplo das medidas tecnológicas utilizadas para a proteção em bases de dados, serviços de bibliotecas, e atualmente nos repositórios, o monopólio do mercado editorial continuará a ser vigente. Este efeito colateral está se estendendo ao controle dos dados subjacentes que emergem de processos investigativos em provedores de bases de dados e *dataset* (IFLA, 2014).

Corroborar com as discussões o Parlamento Europeu quando afirma que não há previsão legal para os dados subjacentes de pesquisas, mas existem normativas que estão disponíveis nas políticas de repositórios e bases de dados. Nesse caso, consideram que as funções da *internet*, tais como as ferramentas de indexação e *cache*⁹ sejam mecanismos de identificação e controle.

Nos Estados Unidos, o uso da indexação e armazenamento em *cache* e a reprodução transitória relacionada à obra protegida não são consideradas infração, pois prevalece o princípio do *fair use* (uso justo) (IFLA, 2014). Contudo, há de considerar que tal princípio nos Estados Unidos não se aplica a todas as obras protegidas por *copyright*, sendo a sua aplicação relativa. No Brasil, tal princípio não é aplicado nos sistemas de proteção vigentes, e não há lei específica que regule o uso justo.

A *Australian Law Reform Commission* publicou o relatório intitulado *Copyright and the Digital Economy* (2013) que fez uma revisão dos direitos autorais na economia digital apresentando as exceções e limitações da proteção autoral vigente nos sistemas de *copyright*. A recomendação traz a adoção do dispositivo legal do *fair use*, mas prevê a utilização das funções de *internet*, com os mecanismos de *cache* e indexação. A flexibilidade para as possibilidades de aplicar o uso justo às criações protegidas, utilizando mecanismos no sentido de controle, mas amenizando aspectos de restrições no acesso.

Ainda no bojo da proteção autoral, os bens intelectuais que estão relacionados com a promoção da Economia Criativa sugerem a ligação ao termo ‘interconexão’. A interconexão e a proteção autoral estão associadas às funcionalidades da *internet* que permitem ações como:

⁹ Armazenamento temporário de dados.

linking, *hyperlink* e *browsing* (IFLA, 2014). A disponibilidade de conteúdo *on-line*, que traz, em sua essência, a navegabilidade, funcionalidade e visibilidade à publicação eletrônica traz possibilidades de conexão entre usuários e recursos.

Tais recursos são entendidos no contexto científico, como publicações científicas eletrônicas, entidades, dados de pesquisa e objetos digitais. Essas funcionalidades tecnológicas impactam diretamente na garantia de proteção autoral, validadas por ações de ligação e interconexão (*links* e *hiperlinks*). Por vezes, a ação de interconexão prova a ilicitude e, por consequência, a violação autoral.

Nesse sentido, a necessidade de trabalhar com a arquitetura da *web* e prudência no uso de *hiperlinks* é uma das recomendações em práticas da *e-Science*. Sugere-se que *links*, como referências (chamadas de unidades de informação), tenham mecanismos que solicitem autorização do titular dos dados de pesquisa antes do reuso e que obedeçam aos limites da jurisdição (local, nacional e internacional).

O princípio norteador de proteção da propriedade intelectual sugere que os “acordos de acesso aos dados devem considerar a aplicabilidade do direito de autor ou de outras leis de propriedade intelectual que podem ser relevantes para financiamento público de bases de dados de pesquisa” (OECD, 2007, p. 16, tradução nossa). Observa-se que as iniciativas têm sugerido que os sistemas de direitos de propriedade intelectual, inclusive as leis que regem o *copyright*, sejam repensados e flexibilizados. Esta reflexão se pauta no excesso de proteção que gera desvantagens e impede a competitividade na Economia do Conhecimento em nível mundial.

Do ponto de vista de governos, instituições e agências de fomento e financiamento, as diretrizes estabelecem que alguns fatores devem ser considerados quando se trata de políticas para proteção de dados científicos. A seguir, tais diretrizes serão analisadas na condução desse entendimento:

- a) Visto que as parcerias públicas/privadas no financiamento de pesquisa e a produção de dados estão aumentando, os acordos públicos/privados devem facilitar o acesso amplo aos dados de pesquisa, quando apropriado (OECD, 2007, p. 16, tradução nossa).

Neste caso, algumas parcerias visam também à exploração comercial das pesquisas financiadas por entes privados. A recomendação é que o acesso aos dados de pesquisa possa ser aberto, considerando as possibilidades de prazos de embargo e as atribuições tanto institucionais quanto as dos colaboradores.

- b) O envolvimento do setor privado na coleta de dados não deve, por si só, ser utilizado como razão para restringir o acesso aos dados. Deve ser dada atenção às medidas que promovam o acesso e uso não comerciais e, ao mesmo tempo, que protejam os interesses comerciais, tais como atraso ou liberação parcial desses dados, ou a adoção voluntária de mecanismos de licenciamento. Tais medidas podem permitir que os participantes primários explorem plenamente os dados de pesquisa, sem desnecessariamente impedir o acesso (OECD, 2007, p. 17, tradução nossa).

A reflexão traz uma questão sutil quanto à recomendação de dados científicos que foram gerados com a parceria da iniciativa privada. Tal recomendação sugere medidas apoiadas pelo uso de licenças não comerciais como alternativa de proteger o interesse comercial. A liberação parcial ou restrita desses dados pode ser conduzida pela descrição de metadados, que deve informar o prazo de liberação (total e parcial), as condições do embargo e prazo de duração; já em caso de pesquisas que envolvam uma proteção específica para segredos comerciais, segurança nacional e descobertas que gerem patentes, essas características devem ser consideradas por instruções normativas internas ou por legislações específicas.

- c) Naquelas jurisdições em que os dados de pesquisa do governo e informações são protegidos por direitos de propriedade intelectual, os detentores desses direitos devem, no entanto, facilitar o acesso a tais dados em particular para outros fins de interesse público ou investigação pública (OECD, 2007, p. 17, tradução nossa).

Todas as jurisdições internacionais possuem suas legislações específicas de direito autoral, contudo em países que já possuem práticas estabelecidas para os dados, é recomendável que o acesso seja aberto, especialmente quando as pesquisas forem financiadas com fundos públicos ou se tratar de áreas que proporcionem melhorias e benefícios à sociedade. Cabe nessa diretriz que os detentores originais já estejam cientes da abertura de seus dados. Ademais, dados provenientes da seara governamental devem ser regidos por legislações específicas, que são baseadas na transparência pública e, por natureza, já devem ser obrigatoriamente abertos.

Assim, o movimento da ciência aberta e das práticas científicas orientadas aos dados conduz para que as legislações, jurisdições, políticas e diretrizes possam flexibilizar e adotar procedimentos normativos e legais ainda vigentes nos sistemas de proteção autoral para uma compreensão baseada na atribuição autoral. Os requisitos de propriedade e exclusividade são barreiras à abertura dos dados, necessitando de flexibilidade, negociação e aprimoramento das práticas tradicionais para absorver as dinâmicas da *e-Science*.

9.3.1.2 Princípio Norteador 2: Conformidade Legal

Em relação ao princípio Conformidade Legal na análise observou-se a incidência do termo ‘*copyright*’ seguida por: *work*, autores, obras, acessos, direito, formas, *domain*, intelectual e informação. O termo *copyright* reflete o direito autoral, direitos e restrições dos dados científicos. O princípio observa que os “acordos de acesso de dados devem respeitar os direitos legais e interesses legítimos de todas as partes interessadas no empreendimento de pesquisa pública” (OECD, 2007, p. 16, tradução nossa). Além dos acordos a conformidade legal prevê restrições e limitações a serem consideradas no escopo das diretrizes legais aplicadas em observância aos interesses dos envolvidos no processo de pesquisa.

As restrições podem ser determinantes e impeditivas, então pensar a política nacional e institucional de acordo com as diretrizes da OECD (2007) ajuda a nortear as diretrizes e os aspectos legais. Em especial, sob aspectos que requerem critérios específicos tais como:

- a) **Segurança nacional:** os dados relativos à inteligência, atividades militares, ou tomada de decisão política podem ser classificados e, portanto, sujeitas a acesso restrito (OECD, 2007, p. 16, tradução nossa).

Os dados de pesquisa que possuem conteúdos relacionados à segurança nacional requer adotar critérios especiais de acesso e condições de uso. A definição de acesso (aberto, restrito e parcial) deve ser prevista em uma política adequada à natureza intrínseca dos dados. Tal restrição deve considerar aspectos políticos, institucionais, éticos e legislações específicas de cada país. É recomendado que a restrição seja informada em documentos instituídos em âmbito governamental (federal) por se considerar conteúdos de alta complexidade e sigilosidade. Também deve ser previsto o *status* do acesso durante a descrição dos metadados e da atribuição de licença.

A privacidade e confiabilidade são dois aspectos muito relevantes quando se trata de acesso aos dados de pesquisa. Geralmente esses aspectos são previstos em legislações específicas, contudo as condições e restrições de uso são previstas pela OECD:

- b) **Privacidade e confidencialidade:** os dados sobre seres humanos e outros dados pessoais estão sujeitos a acesso restrito sob as leis nacionais e políticas para proteger a confidencialidade e privacidade. No entanto, os procedimentos de anonimização ou de confidencialidade que garantem um nível satisfatório de confidencialidade devem ser considerados pelos depositários de tais dados para preservar o máximo de utilidade de dados possível para os investigadores (OECD, 2007, p. 16, tradução nossa).

O direito à privacidade está relacionado a uma legislação específica e ao código de conduta ética. Os procedimentos de anonimização e confidencialidade geralmente são conduzidos no âmbito da instituição e são adotados por pesquisadores no momento da intervenção de pesquisa. As regras, documentos e autorizações são previstos, acordados e aprovados entre os pesquisadores, sujeitos e devem ser aprovados nas comissões de ética institucionais. É recomendado que restrições sobre a confiabilidade sejam informadas em metadados, além das documentações, a fim de assegurar informações e os conteúdos dos dados pessoais e sigilosos estejam protegidos.

Acerca dos conteúdos relacionados com segredos comerciais e direitos de propriedade intelectual vigentes, geralmente as informações e dados de pesquisa estão sob a égide dos sistemas autorais e de patentes ou contratuais. Essa diretriz é considerada a mais complexa, por estar relacionada com fatores restritivos de acesso e uso de dados confidenciais, pois os:

- c) **Segredos comerciais e direitos de propriedade intelectual:** os dados sobre, ou de, empresas ou outras partes que contenham informações confidenciais podem não ser acessíveis para a investigação (OECD, 2007, p. 16, tradução nossa).

Os segredos comerciais, na grande maioria das vezes são advindos de descobertas ou invenções e direcionados ao patenteamento. É recomendado que dados de pesquisa (científicos e governamentais) atendam prazos estabelecidos nas legislações específicas em cada país e informe o período de embargo e a restrição de acesso (total ou parcial).

A proteção de espécies da biodiversidade é outro aspecto considerado nessa diretriz. Essa restrição está bem direcionada aos dados científicos de biodiversidade e da natureza (água, terra e solo) que considera a:

- d) **Proteção de espécies raras, ameaçadas ou em perigo de extinção:** em certos casos, pode haver razões legítimas para restringir o acesso aos dados sobre a localização dos recursos biológicos em prol da conservação (OECD, 2007, p. 16, tradução nossa).

As orientações legais e normativas da legislação específica e da instituição devem ser consideradas para a abertura desses dados específicos. Recomenda-se também que as informações de restrição sejam descritas nos metadados, quando parte ou o conjunto de dados relacionados ao projeto maior de pesquisa possua restrições, parciais ou totais.

Quanto aos dados de pesquisa e documentos que estão sob a proteção judicial, a recomendação prevê que o:

- e) **Processo legal:** dados considerados em ações legais (*sub judice*) podem não ser acessíveis. Inscrever-se em códigos de conduta profissional pode facilitar o cumprimento dos requisitos legais (OECD, 2007, p. 16, tradução nossa).

Essa restrição está associada às pesquisas que resultam em dados com restrições judiciais de acesso e que, por vezes, possuem sua natureza vinculada a pesquisas que envolvem patentes e benefícios à sociedade, por exemplo, litígios que envolvem pesquisas na indústria farmacêutica e que tais descobertas podem trazer benefício imediato a uma determinada comunidade. Os códigos profissionais e/ou regulamentações que tratam dessas restrições são adotados no âmbito institucional ou governamental.

Alargando o entendimento das diretrizes da OECD, o documento intitulado *Model Grant Agreement: H2020 General MGA* (2015), elaborado pela comunidade europeia contribui para as recomendações acerca da obrigatoriedade de proteção de dados confidenciais e de segurança. O Artigo 27 prevê a obrigação de proteção da confiabilidade dos dados científicos, sendo desmembrado em quatro cláusulas descritas no Quadro 12:

Quadro 12 - Recomendações para a proteção de dados científicos

ARTIGOS	DESCRIÇÃO
27.1	Cada beneficiário deve examinar a possibilidade de proteger os seus resultados e deve adequadamente protegê-los - por um período adequado e com a cobertura territorial, se: (a) os resultados puderem ser razoavelmente esperados para ser comercialmente ou industrialmente explorados, e (b) protegendo-os é possível, razoável e justificado (dadas as circunstâncias).
27.2	Se um beneficiário não tencione proteger os seus resultados, para parar de protegê-los ou não, busquem uma extensão da proteção, [a UE] [Euratom] [Agência] pode - sob certas condições (ver artigo 26.4) - assumir a propriedade para garantir a sua (continuação) proteção.
27.3	Os pedidos de proteção dos resultados (incluindo pedidos de patentes) movidos por ou em nome de um beneficiário devem - a menos que a [Comissão] [agência] solicite ou concorde, ao contrário, quando é impossível - incluir o seguinte: "O projeto levando a esta aplicação tem recebido financiamento do [programa de investigação e inovação da União Europeia Horizonte 2020] [programa de investigação e formação Euratom 2014-2018] sob o contrato de concessão nº [número]".
27.4	Consequências de não conformidade Se um beneficiário violar qualquer das suas obrigações deste artigo, o subsídio pode ser reduzido (ver artigo 43). Tal quebra também pode levar a qualquer uma das outras medidas descritas no Capítulo 6.

Fonte: Adaptado do H2020 Programme: Multi-Beneficiary General Model Grant Agreement (2015).

A cláusula 27.1 prevê a obrigação de proteger os resultados e os dados de pesquisa, nos quais o pesquisador (beneficiário) tem que observar os prazos, cobertura territorial e extensão de cobertura. O prazo deve ser determinado em observância aos resultados da pesquisa, incluindo seus dados que apresentam potencial de exploração comercial. Sobretudo, deve o pesquisador considerar que os dados de pesquisa, em algum momento, precisam ser liberados para o acesso aberto. Para tanto, devem ser informados prazos de embargo e a

cobertura territorial. Os metadados devem prever as informações referentes à delimitação de acessibilidade e ao tempo de restrição.

É da responsabilidade do pesquisador decidir sobre a proteção, que deve considerar seus próprios interesses legítimos e de terceiros (especialmente comerciais). Assim, é nesse momento que os resultados e dados a serem compartilhados devem ser considerados quanto ao acesso (aberto ou restrito), período de embargo (se houver) e à cobertura de abrangência, por exemplo, se poderá ter abrangência local, nacional e internacional.

A cláusula 27.2 trata da ausência de responsabilidade de proteção que o proprietário (beneficiário) dos dados deveria assumir. Nesse caso, a recomendação prevê que a agência de fomento ou instituição responsável estabeleça as garantias de proteção, conforme previsto no artigo 26.4. As cláusulas previstas no artigo 26.4 tratam:

26.4.1 [A UE] [Euratom] [a agência] pode - com o consentimento do beneficiário em causa - assumir a propriedade desses resultados para protegê-los, se um beneficiário pretende - até quatro anos após o período estabelecido no artigo 3 - para divulgar os seus resultados sem protegê-los, a não ser em qualquer dos seguintes casos: (a) a falta de proteção é porque proteger os resultados não é possível, razoável ou justificado (dadas as circunstâncias); (b) a falta de proteção é porque há uma falta de potencial para exploração comercial ou industrial, ou (c) o beneficiário tem a intenção de transferir os resultados para outro beneficiário ou terceiros estabelecidos num Estado-Membro da UE ou país associado, que irá protegê-los (EU, 2015, *online*, tradução nossa).

Esta cláusula está direcionada à responsabilidade que terceiros, Estado ou instituição assumem as garantias protetivas provisórias que deverão ser atribuídas aos resultados e dados subjacentes da pesquisa.

A cláusula 27.3 faz menção sobre os pedidos de patentes, oriundos de processos de investigação e projetos de inovação que foram financiados por agência de financiamento europeia. Tais pedidos devem ser identificados pelo contrato de concessão de financiamento concedido. Contudo, prevalece ainda a recomendação de tornar resultados e dados científicos acessíveis abertamente, estabelecendo o prazo previsto para o término da restrição.

E, por fim, a cláusula 27.4 trata do não cumprimento das recomendações de conformidade legal previstas no Artigo 27. Sanções e medidas restritivas podem ser aplicadas aos beneficiários (pesquisadores e colaboradores) de financiamentos concedidos, conforme previstos no Artigo 43 e medidas do Capítulo 6, respectivamente.

As recomendações previstas nessa diretriz aponta para requisitos de conformidade legal que além de assegurar critérios que atendam a flexibilidade entre os direitos existentes e o acesso aberto aos resultados de pesquisa e projetos de inovação.

9.3.1.3 Princípio Norteador 3: Responsabilidade Formal

Em relação ao princípio Responsabilidade Formal essa diretriz é uma das mais importantes por se tratar das responsabilidades legais e éticas que envolvem as políticas orientadas aos dados científicos. As práticas de uma política para dados têm como pressuposto a elaboração de normativas que determinem regras, regulamentos e instrumentos que irão conduzir as ações, responsabilidades e comportamentos.

Isto envolve as particularidades e os instrumentos formais os quais promovem:

Acordo de acesso deve promover práticas formais, institucionais e explícitas, tais como o desenvolvimento de regras e regulamentos, a respeito das responsabilidades das várias partes envolvidas em atividades relacionadas aos dados. Estas práticas devem dizer respeito à autoria, créditos de produção, à posse, divulgação, restrições de uso, acordos financeiros, regras éticas, termos de licenciamento, à responsabilidade e ao arquivamento sustentável (OECD, 2007, p. 17, tradução nossa).

Do ponto de vista dos dados científicos além dos aspectos formais e legais da política, questões que tratem das práticas científicas ligadas à operacionalização do processo devem ser previstas formalmente. Sobre a autoria dos dados e os créditos pela produção, esses deverão ser previstos por atribuição e sistemas de recompensa. Porém sobre este assunto específico, a questão sobre autoralidade e atribuição será contemplada em seção específica.

Acordos relacionados com as restrições de uso devem ser previstos em documentação, pois o acesso pode ser limitado dependendo dos tipos de dados e sua origem. Contudo, estas restrições também são pensadas na diretriz que trata da proteção dos dados. Nesse contexto, também se considera se os dados são sigilosos, sensíveis ou privativos como fatores restritivos de acesso. Variações e parâmetros que são originados ou pelos tipos de dados devem ser considerados quando se estabelece o regime de acesso a dados (OECD, 2007).

As práticas científicas da *e-Science* são regidas por acordos que garantam o acesso, difusão e compartilhamento dos dados científicos amplamente, portanto a recomendação prevê que:

- a) Muitos dos problemas relacionados ao acesso, difusão e partilha de dados resultam da falta de acordos institucionais explícitos sobre as condições de acesso e uso. Com o gerenciamento de dados se tornando cada vez mais complexo em determinadas áreas de pesquisa, arranjos informais tradicionais entre pesquisadores podem não ser mais adequados e podem precisar ser complementados por práticas e procedimentos formalmente acordados (OECD, 2007, p. 18, tradução nossa).

Esses acordos devem ser normatizados e explícitos para que todo o processo de compartilhamento e utilização seja assegurado. As iniciativas internacionais disponibilizam documentações, instruções normativas, modelos, padrões e ferramentas abertas que auxiliam na condução das práticas, além de assegurar uma conformidade e normatização nos padrões de condução das mesmas. Adotam a recomendação de observar as ‘boas práticas’ institucionais e o compartilhamento de educação para a atualização.

As práticas da *e-Science* requer respaldo institucional e normativo, a fim de evitar desvios nos procedimentos de pesquisa e na condução do gerenciamento de dados. Qualquer instituição que optar por trabalhar a partir dos dados científicos, deve primordialmente adotar a postura de que tudo deve ser facilmente acessível e aberto, utilizando-se assim das facilidades e recursos disponíveis pelas tecnologias e *internet*.

Os instrumentos formais adotados na política ou diretriz institucional devem apresentar as responsabilidades (internas e externas) dos envolvidos no projeto. Daí a importância de desenvolver documentos que estabeleçam responsabilidades e tarefas formais internas, bem como aqueles que tratem de acordos, contratos e licenças com parceiros externos.

- b) A responsabilidade para os vários aspectos de acesso e gerenciamento de dados deve ser estabelecida em documentos relevantes, tais como as descrições das tarefas formais das instituições, os pedidos de apoio, contratos de investigação, os acordos de publicação e licenças (OECD, 2007, p. 18, tradução nossa).

Esta diretriz está muito associada ao termo denominado *stakeholders*, o qual é muito utilizado nas práticas institucionais para definir acordos e níveis de parcerias de especialistas, grupo de pesquisa ou instituição com o público externo. Está relacionado diretamente ao entendimento de parceria, colaboração, interesse e responsabilidade mútuos.

O arquivamento sustentável é contemplado desde o projeto de pesquisa inicial, pois envolve a responsabilidade formal tanto de instituições de ensino e pesquisa quanto de organizações governamentais, a fim de garantir a:

- c) A sustentabilidade em longo prazo da infraestrutura necessária para o acesso de dados é particularmente importante. Instituições de pesquisa e organizações governamentais devem assumir a responsabilidade formal para garantir que os dados da pesquisa sejam efetivamente preservados, geridos e acessíveis, a fim de que eles possam ser aplicados de forma eficiente e adequada em longo prazo (OECD, 2007, p. 18, tradução nossa).

Para que o arquivamento e a curadoria sejam sustentáveis, prever os custos de manutenção e o uso de tecnologias robustas que apoiem a infraestrutura necessária a longo prazo é recomendado para o sucesso do projeto de pesquisa aberto.

Dito isto, conclui-se que no âmbito internacional as agências financiadoras e instituições de pesquisa, nos últimos anos, emitiram declarações e políticas ressaltando a necessidade de arquivamento de dados provenientes de pesquisa, e algumas agências de financiamento começaram a exigir que os projetos financiados sejam depositados em um arquivo público (ICPSR, 2015). Daí a importância de observar os princípios norteadores promulgados pela OECD e outras organizações que são importantes para projetos de pesquisa nos moldes da ciência aberta. A prática de critérios mandatórios potencializa para que os dados de pesquisa se tornem acessíveis e públicos, e o seu processo de gestão é uma realidade ainda em construção e desafiante, porque depende de outras variáveis referentes às práticas culturais da comunidade científica e às peculiaridades das áreas de conhecimento.

Do ponto de vista político, a falta de visão estratégica, ausência de políticas nacionais e a conjectura política e econômica brasileira são algumas das facetas e desafios para a institucionalização da ciência aberta no Brasil. Aliado a isto, a “falta de visão política tanto dos políticos quanto da própria comunidade científica impede ou, no mínimo, retarda o aproveitamento dos benefícios do compartilhamento de dados de pesquisa” (OLIVEIRA; SILVA, 2016, no prelo).

Outro grande desafio está na dimensão ético-legal no que diz respeito à condução de uma agenda global para o entendimento de que a ciência aberta rompe com paradigmas e práticas de proprietização existentes. A sua promoção vai além do espectro científico-técnico-tecnológico, mas atinge direitos sociais e difusos no sentido de uma sustentabilidade mais ampla, movendo-se para áreas essenciais, como saúde, meio ambiente, água, alimentação, agricultura, segurança nacional, entre outras (OLIVEIRA; SILVA, 2016, no prelo).

Estabelecer um diálogo entre os modelos vigentes de proprietização de bens intelectuais e o acesso aberto, gratuito e público é também outra dificuldade que contribui e dificulta a institucionalização de um empreendimento inovador como a ciência aberta e das práticas científicas baseadas na *e-Science* no contexto nacional.

No contexto nacional, a reflexão aponta para as peculiaridades brasileiras que dificultam ações governamentais abrangentes, tais como suas características territoriais continentais e regionais, universidades públicas e agências de fomento com baixos investimentos, concentrando o potencial de financiamento em poucos Estados, como, por exemplo, os que se localizam na região Sudeste. Tais peculiaridades provocam uma

desproporcionalidade de investimentos financeiros no âmbito científico e, por conseguinte, uma heterogeneidade regional e subutilização dos potenciais de pesquisa em âmbito nacional. Então, pensar em uma política nacional nesse momento se torna um desafio.

Dito isto, a análise foi conduzida com vistas ao cenário brasileiro confirmou uma realidade incipiente quanto à regulamentação político-normativo-legal para abertura de dados científicos. Porém, apesar da ausência de uma política nacional que esteja nos moldes guarda-chuva da ciência aberta, que envolva as dimensões estratégica, gerencial, metodológica, tecnológica, legal e normativa necessárias à condução de práticas da *e-Science*, já existem algumas iniciativas nas áreas da Biodiversidade e dos dados governamentais sendo desenvolvidas.

9.3.2 Diretrizes político-legais brasileiras

Essa subcategoria foi emergida a partir da análise da legislações específicas e sob o olhar do acesso aberto e de iniciativas preliminares da ciencia aberta.

Em 2011, foi instituído o Plano de Ação Nacional do Governo Aberto, o qual estabelece o compromisso do governo federal de implantar a Infraestrutura Nacional de Dados Abertos. A política foi institucionalizada por decretos e instruções normativas para disciplinar escopo, definição, estrutura, gestão e atribuições do comitê gestor. Essa regulamenta ainda as regras que garantem e facilitam o acesso dos cidadãos, da sociedade e, em especial, das diversas instâncias do setor público aos dados e informações produzidos ou custodiados pelo Poder Executivo federal dentro do escopo do Governo Aberto.

O documento intitulado Manual de Dados Abertos definiu responsabilidades e competências para o direcionamento da Política de Dados Abertos no Brasil. Esse documento traz em seu bojo pressupostos constitucionais e do movimento de acesso aberto que conduziram às seguintes reflexões:

- a) A grande quantidade de dados produzida pela gestão pública é enorme, e esses, na maioria das vezes, são centralizados e inutilizados;
- b) As informações e dados gerados no escopo governamental não são potencializados e valorados como potencial para o crescimento e inovação no país;
- c) Os dados são públicos e podem, portanto, gerar valor e serem potencializados, quando se tornam acessíveis e úteis para a sociedade em geral.

Essas reflexões coadunam com as três leis jurídicas de David Eaves quando afirma que os dados governamentais devem ser abertos, públicos, acessíveis, interoperáveis, úteis e replicados. Isto posto, algumas constatações no âmbito brasileiro foram estabelecidas:

- a) A partir da promulgação da LAI e de instrumentos regulatórios e normativos, o Governo Federal vem avançando lentamente para consolidar políticas de compartilhamento e abertura das publicações e dados governamentais abertos no cenário nacional.
- b) Na esfera estadual, o exemplo demonstrado do Estado de São Paulo é uma concretização de que a abertura e compartilhamento de dados governamentais é uma tendência mundial que já se tornou realidade no Brasil.
- c) Acerca dos direitos de propriedade intelectual vigentes no Brasil, quando se trata de dados governamentais, não são percebidas uma rigorosa proteção e restrição para o acesso, uso e reuso dos dados governamentais. Atribui-se a isso, a natureza pública intrínseca.
- d) Ainda sobre os aspectos autorais, observaram-se as recomendações relativas aos critérios de sigilosidade, privacidade, sensibilidade e segurança, prazos de embargo, se necessário, e as regras preestabelecidas nas políticas.

Em suma, o escopo de dados governamentais não se encontrar no foco da tese, a reflexão foi trazida para mostrar que a legislação específica de acesso aberto, políticas e instrumentos normativos estão na vanguarda das iniciativas do Governo Aberto e podem ajudar nos propósitos futuros para a Política Nacional de Ciência Aberta brasileira.

A ausência de uma política nacional que realize iniciativas para uma ciência orientada aos dados científicos ainda é um fator restritivo na condução plena de práticas científicas da *e-Science* no cenário brasileiro. Porém, algumas iniciativas já estão sendo desenvolvidas em parcerias com universidades, institutos de pesquisa, agências de fomento e instituições internacionais, a exemplo do Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio) e do Ministério do Meio Ambiente (MMA).

A Política de Dados e Informações do Instituto Chico Mendes de Conservação da Biodiversidade foi instituída por decreto e instrução normativa que regulamentam a disponibilidade, o acesso e uso de dados e informações custodiadas em bases e sistemas de informação. Essa política prevê o acesso aberto aos dados, informações, documentos e publicações de cunho técnico-científico.

Os dados de Biodiversidade são bem heterogêneos em sua natureza e por si só possuem peculiaridades distintas, como, por exemplo, os dados científicos que foram

provenientes de análises e sínteses de informação sobre animais e plantas (níveis taxonômicos) e que, por sua especificidade, podem ser considerados dados sensíveis. Dessa forma, a política prevê a regulamentação para os períodos e formas de restrição dos dados sensíveis. Para essas particularidades, as regulamentações são feitas por ato administrativo específico no bojo da política e em consonância com as diferentes peculiaridades. Assim, a instrução normativa buscou minimizar a heterogeneidade desse contexto com a denominação dos diferentes colaboradores, participantes institucionais, tipos de dados, períodos de carência e embargos.

No que tange às questões de direitos autorais, tanto a política quanto a instrução normativa possuem algumas cláusulas que estão alicerçadas sobre a influência da Lei de Direitos Autoral (LDA) e da Lei de Acesso à Informação (LAI) vigentes no Brasil. Porém, percebe-se que isto é conduzido no escopo da instrução normativa como uma orientação para que os dados e informações disponibilizados devam ser identificados para reconhecimento da autoria (autor e instituição) e fonte original (sistemas de informação). Isso leva também a uma reflexão mais atualizada do reconhecimento da autoralidade pautada na forma de atribuição relacionada com as responsabilidades de cada colaborador e entidade envolvida.

Em 2015, o Portal da Biodiversidade foi reconhecido como uma iniciativa brasileira que disponibiliza dados científicos de Biodiversidade – com espécies da fauna e flora. Esforço conjunto que só foi possível ser concretizado a partir de uma institucionalização político-normativa e de parcerias nacionais e internacionais. Este esforço político e institucional foi realizado por instituições de pesquisa, universidades e órgãos governamentais e não governamentais, além do apoio federal do Ministério do Meio Ambiente (MMA) e da Universidade de São Paulo, para que a integração, disponibilidade e interoperabilidade fossem realizadas a partir de uma arquitetura complexa que integrou sistemas de informação, bases de dados locais e repositórios institucionais.

Em síntese, é possível inferir que o Brasil possui potencialidades para o desenvolvimento político, econômico, científico e sustentável para uma Política de Ciência Aberta orientada aos dados científicos. Inferências foram realizadas a partir do exposto:

- a) O descompasso do sistema de direitos de propriedade intelectual vigente, notadamente a LDA e a Lei de Patentes, com as práticas de acesso aberto e da *e-Science* já conduz à necessidade urgente de discutir a reformulação e adequações legais no escopo jurídico;

- b) A informação pública e os dados abertos governamentais que estão regidos pela LAI e por decretos específicos apregoaram iniciativas que conduziram para o plano nacional do governo aberto e a transparência pública;
- c) As iniciativas que surgiram a partir da promulgação da LAI, com Políticas, Decretos e Instruções Normativas, avançaram e ampliaram o escopo do acesso à informação e aos dados governamentais, tornando-os públicos, abertos e acessíveis;
- d) Quanto às iniciativas institucionais e governamentais emergentes para a abertura de dados científicos, há um descompasso muito grande entre a LAI, a LDA e os pressupostos que regem a Ciência Aberta e a *e-Science*;
- e) A iniciativa do ICMBio e de instituições parceiras é um marco inicial para a abertura do acesso e compartilhamento de dados científicos. O engajamento entre governo, institutos de pesquisa e universidades aponta para um caminho promissor em prol de uma institucionalização da ciência aberta no Brasil;
- f) A prática mandatória no contexto científico nacional é um fator que deve ser considerado nas discussões junto às agências de financiamento, universidades, instituições governamentais e institutos de pesquisa, quanto à condução de políticas e diretrizes que tornem obrigatório o compartilhamento e acesso aos dados científicos de projetos de pesquisas custeados com fundos públicos.

As reflexões acima tiveram a intenção de demonstrar que apesar da inexistência de uma política norteadora que conduz à abertura e o compartilhamento dos dados científicos consubstanciam novos empreendimentos no cenário brasileiro.

Acerca da LAI e das diretrizes da OECD (2007) foi possível perceber uma interface entre as iniciativas governamentais brasileiras e os princípios norteadores internacionais:

- a) Os instrumentos político-normativos e jurídicos, instituídos a partir da promulgação da LAI e do Governo Aberto, possuem uma conexão com os princípios que regem a abertura, flexibilidade, transparência, interoperabilidade, segurança e sustentabilidade para os dados governamentais. Entretanto, o estudo não inferiu se as demais iniciativas brasileiras estão em consonância com os referidos princípios em virtude do tempo exíguo;
- b) Percebeu-se também que os princípios que tratam da qualidade, eficiência e profissionalismo tem natureza prática e aplicável ao contexto em que os dados (governamentais ou científicos) estão inseridos;

- c) Acerca dos princípios que tratam da conformidade legal, proteção da propriedade intelectual e responsabilidade formal, estes foram refletidos separadamente devido aos direcionamentos específicos para o objeto de estudo desta tese.

Em síntese, reflexões foram inferidas a partir das subcategorias identificaram a ausência de políticas específicas é perceptível, bem como a necessidade de continuidade das iniciativas na esfera governamental para novos diálogos junto à comunidade científica, agências de fomento, universidades, organismos governamentais e privados. No que diz respeito à regulamentação da propriedade intelectual para dados científicos inexiste políticas, diretrizes e instrumentos regulatórios brasileiros que orientem questões legais e jurídicas direcionadas ao gerenciamento de dados científicos. As reflexões levantadas na seção irão nortear o padrão político-normativo e ético-legal.

9.4 SISTEMATIZANDO E GERENCIANDO OS DADOS CIENTÍFICOS

A categoria emergida direcionou as reflexões para os aspectos metodológicos, técnicos e tecnológicos com vistas à apreensão das questões autorais que perpassam pela sistemática da gestão de dados científicos. Essa categoria teve como fio condutor as dimensões morfológicas e técnicas, e a ênfase foi dada aos estágios que apontam para garantias autorais. O ciclo de vida faz parte da ciberinfraestrutura da *e-Science* que é composta por padrões de metadados, protocolos de interoperabilidade e repositórios especializados para a descrição, processamento, integração, preservação e disponibilidade dos dados em curto e longo prazos.

A gestão eficiente dos dados é fundamental para o desenvolvimento de pesquisas de alta qualidade e excelência na *e-Science*. No cenário internacional, existem diversos modelos de ciclo de vida de dados de pesquisa que atendem a diferentes particularidades, concepções e realidades institucionais. Nas práticas da *e-Science*, “a gestão de dados cobre todos os aspectos relativos à manipulação, organização, documentação e agregação de valor, e tem um papel crucial como facilitador nos processos de compartilhamento dos dados, na garantia da sustentabilidade e acessibilidade dos dados em longo prazo” (SAYÃO; SALES, 2015, p. 5). Ela se configura como um conjunto instrumental, político e normativo que aprimora e faz com que as práticas no bojo da ciência possam ser conduzidas por dados científicos como produtos de pesquisa.

Planejar o projeto de pesquisa é pré-requisito obrigatório para as investigações que visam ao compartilhamento para uso e reuso de dados, especialmente quando são financiadas com recursos públicos. O planejamento do projeto de gestão de dados é considerado por

agências de fomento e instituições científicas internacionais um instrumento oficial e mandatório. No projeto de pesquisa é estabelecido o compromisso público de pesquisadores, grupos de pesquisa e instituições de ensino e científica de que os resultados, dados de pesquisa e conjuntos de documentos e métodos subjacentes sejam estabelecidos, previstos e atendidos, garantindo que eles possam ter acessibilidade, usabilidade, reuso e reprodutibilidade otimizados e acessíveis abertamente. O estabelecimento de métodos, metas, ações, compromissos e soluções é previsto neste estágio.

Um plano de gerenciamento é um documento que descreve como os dados coletados ou gerados deverão ser tratados durante um projeto de pesquisa. Após a sua conclusão, esse plano deve apontar quais dados serão descritos, gerados e quais metodologias e padrões estarão seguindo; além disso, se e de que forma esses dados serão compartilhados e/ou garantidos abertamente, e como eles serão curados e preservados (EC, 2016).

Assim, o planejamento visa dar continuidade à pesquisa e à preservação em longo prazo, em virtude dos dados e dos conjuntos de dados de pesquisa serem acondicionados para um tempo de vida maior que os projetos de pesquisa que os criaram. Isso significa que pesquisadores, professores, estudantes e outros profissionais podem continuar a trabalhar sobre esses dados após os projetos e financiamentos terem sido finalizados. “As ações e comprometimentos promovidos pela gestão, coletivamente, permitem que os dados de valor possam ser reusados em outros projetos ao longo do tempo e do espaço” (SAYÃO; SALES, 2015, p. 5). Novos projetos de pesquisa podem conduzir, analisar ou adicionar novos elementos a esses dados de forma que eles possam ser reusados por outros pesquisadores, reiniciando um novo ciclo de pesquisa.

Duas questões críticas devem ser consideradas no momento do planejamento da gestão de dados, especialmente se a política institucional estiver conduzindo para que os mesmos sejam de acesso aberto e público. Nessa direção dois pontos são considerados no planejamento, quais sejam:

Proteção dos dados: Conjunto de dados onde pessoas, agregados familiares ou empresas são identificados. Nesse caso, via de regra, não é possível publicar os dados abertamente, mas em algumas situações, versões dos dados que passem por processos de anonimização podem ser passíveis de disseminação.

Database copyright: Quando os dados de pesquisa são derivados e/ou elaborados a partir de uma base de dados comercial preexistente. Nesse caso, normalmente não é possível republicar as extrações significantes devido a questões de copyright e termos legais de uso. É importante observar que mesmo quando os dados são disponibilizados abertamente, nem sempre é possível republicá-los livremente. (SAYÃO; SALES, 2015, p. 22, grifo nosso).

Essas questões críticas são conduzidas para duas reflexões: proteção de dados e reuso dos dados.

- a) A primeira está relacionada com a proteção dos dados (sensíveis e sigilosos) que devem ser considerados quando o planejamento estiver sendo elaborado. Os dados sensíveis estão relacionados com as informações privativas de pessoas ou instituições. O direito de privacidade aos dados pessoais deve ser levado em consideração e, se possível, a autorização prévia deve ser solicitada.

No estágio de descrição dos metadados desses dados, devem constar informações sobre o status (embargado ou com prazo de liberação). Esse “*status* de outras barreiras ao acesso, tais como o preço e a permissão para reutilização, é indicado pelos termos grátis e livre. A distinção tem sido feita popularmente por Peter Suber, que emprestou a linguagem grátis/livre do mundo do *software*” (FROSIO, 2014, p. 53, tradução nossa).

- b) A segunda questão crítica é sobre como os dados foram gerados ou derivados a partir de pesquisas que estão em bases de dados comerciais ou vinculadas a um contrato de *copyright*. Em países onde os direitos autorais são baseados em *copyright*, os contratos com empresas provedoras de informação ou direto com os autores são muito rigorosos.

Isso leva a uma constatação que abrir os dados científicos para acesso e reuso que estejam sob a proteção contratual, levará a restrição do acesso aberto e gratuito. É possível considerar as possibilidades: domínio público, término do contrato e finalização do período de embargo. Essas são situações conduzem para os casos de modelos híbridos de acesso aberto. Os modelos híbridos de arquivos têm sido distinguidos, por exemplo, como *Open Access* (opcional, retroativo, atrasado ou parcial). Tais modelos são definidos como “*open access* concedidos mediante decisão dos autores após o pagamento de uma taxa, para arquivos retrodigitalizados, tais como volumes de periódicos antigos, depois de um período de embargo ou partes de periódicos” (FROSIO, 2014, p. 53, grifo nosso). As publicações com dados de pesquisa subjacentes que estão arquivadas em bases de dados ou repositórios institucionais podem ser republicadas, desde que haja permissão concedida de autores ou empresas provedoras.

A utilização de ferramentas de planejamento que orienta para a execução de todo o gerenciamento metodológico e tecnológico que deverá ser seguido durante o processo de pesquisa. Os procedimentos metodológicos da pesquisa devem seguir o mesmo rigor dos critérios estabelecidos pela ciência. De acordo com as iniciativas internacionais, sugere-se a adoção da ferramenta aberta DMPTool para o planejamento do projeto de dados ou a

ferramenta de fluxo de trabalho chamado *Taverna* “que ajuda a documentar os processos de dados e permite que o fluxo de trabalho seja revalidado” (NAS, 2015, p. 132, tradução nossa).

A coleta é realizada no estágio *coletar* e deve observar os fatores extrínsecos e intrínsecos relacionados com a tipologia dos dados e as condições de coleta, acondicionamento, integridade e informações sobre a proveniência dos mesmos. A partir do momento em que os dados científicos forem coletados, eles devem ser acondicionados, provisoriamente, em ferramentas de comunicação aberta, tais como *Dropbox* e *Google drive*, ou em ambientes específicos ao domínio da disciplina, como *software*, laboratórios, câmaras, tubos de ensaio, sistemas computacionais ou instrumentos.

É um processo de curadoria preliminar (podendo ser definitivo) que embora alguns arquivamentos sejam realizados em formatos comuns, outros têm considerável variedade. A curadoria de dados “consiste na validação de dados básicos para garantir a integridade e integralidade dos dados [...]. Essa etapa assegura que o formato dos dados seja compreendido, e os metadados solicitados estejam presentes” (NAS, 2015, p. 132, tradução nossa).

A curadoria de dados é necessária como um primeiro passo para fazer a validação de dados básicos para garantir a integridade e integralidade, bem como os metadados necessários. Com a garantia mínima de curadoria, é possível que “os dados de diferentes fontes possam ser descritos, compartilhados, usados, e tornar o processo de descoberta mais fácil” (NAS, 2015, p. 133, tradução nossa). Tal processo é complexo apresentando desafios para o reuso, reprodutibilidade e citação.

Em alguns modelos de ciclo de vida, a partir da preliminar curadoria, os dados passam por um processo de validação com critérios específicos da disciplina, com a finalidade de serem diferenciados em sua natureza (experimental, computacional), por exemplo, e a saída normalizada para os próximos estágios. É evidente que alguns procedimentos em cada estágio do modelo de ciclo de vida dos dados são realizados de acordo com normas institucionais ou por critérios metodológicos de cada pesquisa em uma disciplina específica.

Concomitantemente a essa curadoria, é recomendado o **uso de identificadores persistentes**, que servem para a identificação dos dados primários, bem como de pesquisadores, colaboradores e instituições envolvidos no projeto. Geralmente, essa recomendação é obedecida no estágio *assegurar*, que, além de consolidar a curadoria inicial dos dados, assegura a identificação dos envolvidos no projeto e no fluxo de trabalho.

Os identificadores persistentes, denominados na literatura internacional *Persistent Identifiers (PIDs)*, são vinculados aos contribuidores, dados e publicações para garantir que a indexação, citação e interoperabilidade sejam viáveis. Eles funcionam como *links* que

identificam os objetos dentro do sistema, pois “com a vinculação do identificador persistente, todos os dados, metadados e objetos de mapa de recursos no DataONE têm um identificador único” (DATAONE, 2015, *on-line*). Ele permite ligar uma versão autorizada da publicação, por exemplo, OpenAIRE que fornece meios para verificar o cumprimento de recolhimento dos metadados do repositório.

Identificar e atribuir possibilitam que dados sejam citáveis e atribuíveis de forma exclusiva (única) e em longo prazo. A **identificação apropriada** permite que dados sejam preservados, recuperados, citados e compartilhados, pois, por meio de “identificadores persistentes na qualidade de objetos digitais, as coleções de dados devem ser identificadas por esquemas de identificação persistentes, padronizados e globalmente únicos” (SAYÃO; SALES, 2015, p. 23).

Estimula-se neste estágio a adoção de identificadores para os dados digitais, a exemplo do DOI e do ORCID que são os identificadores persistentes padronizados e recomendados nos modelos de ciclo de vida, como “uma medida concreta para atribuir um identificador persistente aos seus dados” (SAYÃO; SALES, 2015, p. 23). Um sistema de identificação de objetos digitais *on-line* que, além de assegurar a qualidade e identificação, proporciona gerenciar os direitos de propriedade intelectual de dados e objetos digitais. Esses identificadores asseguram a vinculação dos créditos aos dados e aos pesquisadores, colaboradores, entidades e instituições. O fluxo de trabalho também pode ser registrado com um DOI. “Capturar o fluxo de trabalho e atribuir um identificador suporta melhor citação porque o recurso citado é mais reutilizável, e fortalece a reprodutibilidade e validação da pesquisa” (NAS, 2015, p. 133, tradução nossa).

Para a identificação do contribuidor, a iniciativa recomendada é o ORCID, cujo objetivo é estabelecer um identificador único para cada pesquisador no processo de comunicação científica. Este projeto está intimamente relacionado com o padrão *International Standard Name Identifier* (ISNI) da norma (ISO 27729). Esta norma foi recentemente aprovada para publicação e define um identificador para quaisquer “partes” envolvidas no processo de criação de conteúdo em todas as mídias e proporcional à responsabilidade.

Atualmente existem identificadores para todos os elementos de uma citação, a exemplo: por nomes (ORCID ou ISNI), afiliações (*Institutional Identification*, I2), publicações (ISBN ou ISSN), coleções (ISCI), URLs persistentes (DOI, ARC, ou PURL), e Dados (NAS, 2015). Cada um desses padrões pode ser incorporado em URIs acionáveis e os URIs, juntamente com os metadados associados, podem ser servidos à comunidade como parte de armazenamento e visualização dos dados vinculados.

Como exemplo, a identificação adequada com os respectivos metadados vinculados ao criador (autor, pesquisador), colaboradores, curadores e outros responsáveis evita ambiguidade, vincula documentos, publicações ou outras URIs que estão associadas aos dados. Essa identificação colabora ainda para informar o acesso aberto, condições e licenças de uso, bem como o período de embargo, se for o caso. Ademais, usar um serviço de indexação que utiliza os metadados dos dados e publicações, para registrar seus identificadores, como, por exemplo, o *CrossRef* e o *DataCite*, faz parte do conjunto de boas práticas recomendáveis no contexto internacional.

O objetivo da exigência em **metadados** é maximizar a divulgação das publicações e assegurar o reconhecimento do financiamento da European Union (EU). A mineração de dados bibliográfica é mais eficiente do que a mineração de textos completos. A inclusão de informações relativas ao financiamento da EU, como parte do metadados bibliográficos, é necessária para um controle adequado à produção de estatísticas e à avaliação do impacto do Horizonte 2020.

A descrição dos metadados de dados científicos é outra etapa relevante no gerenciamento de dados. Para que os dados sejam citados, é necessário o mínimo de metadados robustos, acessíveis e detectáveis. Os objetivos mais comuns de coleta de metadados são: “(1) fornecer a documentação de dados internos; (2) melhorar o compartilhamento atual dos dados; e (3) aumentar a utilidade futura de arquivos de dados” (RUGG, 2004, p. 1, tradução nossa).

É na descrição dos metadados que a proveniência e a atribuição dos dados são asseguradas. A proveniência é um dos atributos associados à origem e procedência dos dados.

Para a **atribuição** ser realizada, os responsáveis pela pesquisa, como os autores, colaboradores, instituição e informações do conjunto de dados e da pesquisa, são constituídos na descrição dos metadados de elementos. Para que a descrição seja realizada, é obrigatório adotar um padrão de metadados.

Mediante a importância dos metadados para a pesquisa orientada aos dados científicos, algumas recomendações foram observadas:

- a) A construção de um conjunto de metadados robustos para que os dados científicos sejam detectáveis, acessíveis e compartilhados;
- b) A escolha de um padrão internacional de metadados para a descrição minuciosa de elementos que constituem os dados científicos é mandatória. Tais padrões garantem a identificação da atribuição, proveniência e citação dos envolvidos;

- c) Existem padrões gerais e outros para disciplinas especializadas que auxiliam na descrição de recursos e dados científicos. A escolha do padrão é uma decisão relevante e deve conduzir para que a descrição seja apropriada ao domínio da disciplina;
- d) A utilização de instrumentos que auxiliem na padronização, como a *International Organisation for Standardization* (ISO), vocabulários controlados, *thesaurus*, ontologias e palavras-chaves enriquecem a descrição dos metadados;
- e) A descrição de informações referentes aos documentos do projeto de pesquisa original garantirá a preservação das informações que estão relacionadas com a origem dos dados científicos;
- f) A adoção de ferramentas automáticas ou de *software* que gerem os *codes* automáticos de metadados para a interoperabilidade em repositórios especializados é necessária para o acesso, visualização e reuso dos dados científicos;
- g) A inclusão de *links* que possam apresentar outras experiências de *sites*, documentar as boas práticas e a inclusão de informações relativas ao financiamento também são recomendáveis.

É por meio dos metadados que o projeto de dados ganha visibilidade, e eles podem ser analisados e mensurados nos estágios subsequentes. A descrição de metadados é uma tarefa muito relevante do ciclo de vida dos dados. Os metadados referentes aos créditos devem ser atribuídos com a finalidade de identificar os responsáveis principais e colaboradores individuais de dados ou do conjunto de dados.

Na comunidade europeia, a exigência de metadados visa maximizar a divulgação das publicações e assegurar o reconhecimento do financiamento. A inclusão de informações relativas ao financiamento como parte dos metadados é necessária para um controle adequado à produção de estatísticas e avaliação do impacto do programa (EU, 2016). Um conjunto de dados pode ser atribuído a um projeto, ou, em analogia com créditos a um filme, os indivíduos podem ser atribuídos com base em suas contribuições individuais para um projeto.

As informações relacionadas aos envolvidos no projeto de pesquisa devem ser descritas nos campos respectivos aos elementos de identificação (autoria, instituição, licenças e outros). A respeito da atribuição, uma seção será destinada para o seu aprofundamento.

No que tange às políticas para nortear o acesso aos DPIs, essas são normatizadas em três condições: tipo de acesso, disponibilidade para depósito e licença de conteúdo. O tipo de acesso está relacionado à permissão definida pela editora para os artigos publicados em suas revistas, que serão armazenados em repositórios institucionais. Para a disponibilidade de

depósito e armazenamento das versões *pré-print* e *post-print*, suas restrições são definidas na codificação por cores. A licença de conteúdo está relacionada às permissões estabelecidas pela editora para acesso e uso dos artigos publicados pela revista. As licenças estabelecidas no Diadorim para o uso dos artigos são baseadas nas licenças *Creative Commons*.

A importância de adotar o padrão de metadados é importante para as descrições necessárias que garantam a interoperabilidade com o repositório nos estágios futuros, que permite que os metadados possam ser coletados automaticamente por outros sistemas chamados de provedores de serviço. Observou-se que o estabelecimento de normas mínimas para os metadados deve atender a um domínio específico, utilizando um padrão de metadados que esteja em consonância com a natureza do conjunto de dados.

Essa categoria levantou as reflexões que dão sustentabilidade a infraestrutura técnica e tecnológica que substanciará o padrão morfológico-técnico. A importância de adotar critérios baseados no modelo de ciclo de vida e todos os recursos tecnológicos

9.5 AUTORALIDADE COLABORATIVA E RECOMPENSA NA *E-SCIENCE*

Essa seção conduz as reflexões sobre as novas formas de autoria no modelo emergente da *e-Science*. O intuito é refletir sobre o novo fazer científico que pressupõe o trabalho colaborativo, a preocupação com a partilha de dados de pesquisa e a postura ética dos pesquisadores e envolvidos nas novas dinâmicas científicas.

A categorização e codificação da análise de conteúdo dessa seção foram norteadas pelas dimensões ética e legal e a relação dos direitos autorais vigentes e as práticas contemporâneas. A conduta ética, honestidade acadêmica e a postura investigativa são aspectos importantes que devem ser vigiados no processo de compartilhamento e colaboração científica. Assim, dois espectros ético-legais conduziram a reflexão:

- a) a transição de um paradigma econômico proprietário que reforça os direitos privados da propriedade intelectual para o quarto paradigma baseado no compartilhamento e colaboração científica.
- b) a ênfase do novo paradigma científico que direciona a recompensa autoral por meio da atribuição e citação.

O compartilhamento de dados de pesquisa é um fator preponderante na decisão dos pesquisadores. Mesmo no contexto internacional, onde as práticas científicas já estão sendo conduzidas, os “pesquisadores são relutantes para compartilhar dados com grandes preocupações em questões legais, uso indevido de dados e tipos de dados incompatíveis”

(TENOPIR et al., 2011, p. 2, tradução nossa). Por esse comportamento, perpassam a preocupação com a privacidade de dados pessoais, publicações *post-print* e a exclusividade sobre os dados científicos. Muitos pesquisadores alegam que dados e resultados científicos são decorrentes de muitos anos de investigação e que, por merecimento, os direitos de *copyright* são exclusivistas.

O *modus operandi* da pesquisa científica tradicional é direcionado por um comportamento individual, intrínseco do pesquisador-autor e restrito ao ambiente local e nacional entre pares. A comunicação formal através do periódico científico foi, por séculos, o principal meio de publicização de resultados de pesquisa. Com a intensificação das TICs, *internet* e o movimento de acesso aberto, a propagação de publicações eletrônicas, bases de dados e repositórios institucionais abriu o espectro do acesso e uso.

Entretanto, o comportamento do pesquisador-autor se mantém numa postura intrínseca e individualista. A concepção tradicional de que os resultados de pesquisa são de propriedade exclusiva do autor, oriunda de seu labor intelectual, merece ser reconhecida como um mérito acadêmico, de ordem individual e proprietária (OLIVEIRA; SILVA, 2016, no prelo), acrescenta-se a essa concepção a expropriação econômica praticada junto ao mercado editorial, trazendo as nuances contemporâneas sobre a autoralidade na *e-Science*.

9.5.1 Autoralidade na *e-Science*

O *corpus* de análise nessa categoria emergiram temáticas relacionadas com o *copyright*, acesso, *agreement*, obras, intelectual, informação, *articles* e *citations* e, intrinsecamente, os termos relacionados aos conteúdos que tratam da autoria, atribuição, direitos, ética e plágio.

A representação gráfica demonstra a predominância das palavras no escopo da ‘autoralidade’, conforme a Figura 25. A frequência evidencia a vinculação aos referenciais teóricos e à compreensão dos aspectos autorais que tratam da atribuição autoral nas práticas científicas contemporâneas.

Assim, percebeu-se no referencial teórico que as concepções dadas ao autor priorizam o reconhecimento de sua criação, idealização e geração de novos significados e conhecimentos. Porém, exerce a função de autor a partir de concepções, pesquisas e outras formas de conhecimento geradas por outros autores. Ou seja, o texto não é tecido por linhas e palavras, as quais não são exclusivas de um único autor, mas construído em *quilts* de citação com diferentes focos, contextos, culturas e sentidos. É a maturação e influência do autor acerca de conhecimentos já formulados e sua responsabilidade epistemológica e ética que acréscimos e ressignificação são realizados, reorientados, transformados e modificados para novos sentidos e significados nos diferentes campos discursivos. Essa é uma nova percepção da autorialidade sem os requisitos de exclusividade e individualidade predominantes na visão tradicionalista.

Na abordagem tradicional Harvey (2006) e Krokosz (2015) afirmam que a autoria está revestida de características individualista e exclusivista, voltadas para a visão capitalista que se caracterizou pela fragmentação, superficialidade, relatividade e efemeridade do modelo proprietário econômico. Prevalecendo, além disso, a inclinação para a estética da obra, ao invés da ética do conteúdo e da criação.

Na visão contemporânea, observa-se que a ‘autoria’ está sendo configurada na pós-modernidade com uma nova significação da criação individualista, absoluta e una que revestia uma determinada obra. Há de se considerar a “necessidade de superação do paradigma da titularidade estável da autoria, em prol de uma função social mais colaborativa e adequada à noção de economia da informação em rede [...]” (KROKOSZ, 2015, p. 85).

Dessa forma, a predominância dos requisitos exclusividade e individualidade está sendo permutada pela coletividade e colaboratividade. Assim,

apesar de a autoria partir de uma pessoa, esta pessoa se vê, nas atuais circunstâncias, incluída em um mosaico de situações que não mais remetem ao individualismo, e sim à cultura em que ela se insere [...] Se o autor só se reconhece no outro, não se pode mais imaginá-lo ‘uno’ (BARROS, 2010, p. 154).

Paira ainda nessa reflexão, os requisitos da originalidade e novidade a serem considerados. Na visão foucaultiana, Krokosz (2015) afirma que a originalidade da contribuição sobre a função do autor está na percepção da diferença entre autor e autoria. O autor está relacionado à pessoa, como mencionado anteriormente, e, tal qual, ainda prevalece nos sistemas de propriedade intelectual vigentes. Por outro lado, a originalidade da autoria

reflete o feito da obra, sua construção e função que possibilitam a criação, recriação, reuso e ressignificação.

Assim, o requisito da originalidade mantém-se como a ideia da criação primígena. Além disso, na originalidade existem duas distinções: obra original e obra originária. A obra original, entendida como novidade subjetiva, própria do criador, em que “o aspecto personalidade é central na formação da obra original. Em outras palavras, se a obra existe, é porque foi concebida, e esta concepção se deve ao gênio criador” (BITTAR, 2009, p. 145).

A obra originária se define pela “obra primígena, aquela pela primeira vez composta ou criada por seu autor” (FRAGOSO, 2009, p. 161). Trata-se, portanto, de obra originária absoluta. Contudo, no bojo desta análise, há de ser considerada a obra originária no sentido relativo, em que “não se cuida da novidade absoluta, mas de concepção diversa das existentes” (BITTAR, 2000, p. 60).

Logo, na perspectiva da *e-Science*, analogamente o reuso de dados científicos terá como requisito a originalidade relativa. Tal distinção é considerada necessária na doutrina para estabelecer a obra criada original sem preexistência da obra que foi originária a partir de uma obra que a antecedeu, mas que sofreu influências de contexto, cultura e variáveis extrínsecas ao criador.

Sobre a originalidade absoluta e relativa, Barbosa (2007) e Carboni (2009) discutem que a originalidade absoluta não há de ser considerada no labor intelectual e científico, pois tal distinção é apenas para diferenciar a obra criada de outra preexistente. “Essa originalidade deve ter caráter relativo, uma vez que é inevitável o aproveitamento, até inconsciente, do acervo cultural comum” (CARBONI, 2009, p. 16). Para que a originalidade seja absoluta, a criação não pode ter absorvido influências culturais, sociais e científicas do ambiente em que o criador está inserido.

Desta forma, não há o que se questionar quanto ao contexto da *e-Science*, pois as práticas científicas são baseadas em um ambiente de colaboração e compartilhamento para o uso, reuso e reprodução de dados primários, resultados preliminares e documentos relacionados à investigação. Ocorre, portanto, que a originalidade será sempre relativa.

Trazendo para o entendimento da *e-Science*, os dados científicos podem ser revestidos de requisitos de originalidade relativa e de novidade temática. A originalidade relativa está atribuída à natureza e ao contexto do domínio específico, em que os dados estão sendo coletados, gerados, reusados, ressignificados e reproduzidos. A novidade temática, no entanto, consiste na geração de novos sentidos, significados e aplicações pelos quais os dados científicos passaram durante os processos de reuso e reprodutibilidade. A partir do momento

em que os dados científicos são reusados em um novo contexto temático, diferente daquele que originou os dados primários, novas abordagens, hipóteses e métodos são gerados e, por conseguinte, novos direcionamentos investigativos e novidades temáticas surgem.

Assim, “a obra pode ser **novidade** quanto a um determinado tema jamais abordado [...] e **original** quanto ao seu modo de expressão – caso em que a novidade e originalidade se fundem – e também sendo indiferente à forma material assumida” (FRAGOSO, 2009, p. 120, grifo nosso). Além disso, a forma de expressão e os meios de transmissão dos dados científicos não interessam para a garantia dos requisitos descritos.

Até o momento, considerou-se a análise baseada nos requisitos de originalidade relativa e novidade temática dos dados científicos primários. Contudo, no bojo desta análise, as obras derivadas, coletivas e colaborativas devem ser refletidas no contexto da *e-Science*.

Na acepção tradicional, a obra derivada é caracterizada pela “utilização, adaptação e outras transformações de uma obra originária” (FRAGOSO, 2009, p. 171). No escopo do direito autoral tradicional e de acordo com a Convenção de Berna, qualquer transformação na obra original ocasiona uma obra derivada e requer autorização prévia do autor (criador) original. A derivação da obra é revestida de características que denotam ser uma nova criação, podendo ou não ter uma inovação na forma de expressão. Essa deve possuir um vínculo direto de ligação com a obra que a antecede, utilizando-se de processos de transformação, incorporação, complementação, redução ou adição da obra original.

Na compreensão contemporânea, o reuso de dados científicos já caracteriza uma derivação de dados primários originários. No entendimento doutrinário, obras derivadas transformadas por traduções, arranjos, revisões, condensações e outras formas de derivações são consideradas novas criações. Assim, dados científicos reusados ou reproduzidos assumem um caráter de criação derivada ou secundária na concepção tradicional. Complementa-se ainda que pesquisadores, grupos de pesquisa ou instituições são denominados de “usuários secundários – têm que obter liberação do *copyright* do detentor dos direitos sobre os dados, antes dos dados serem reproduzidos” (SAYÃO; SALES, 2015, p. 73). Esta é uma visão tradicionalista que permanece nas práticas contemporâneas da *e-Science*.

No que diz respeito às criações ou obras em colaboração ou colaborativas, essas são definidas pelas diferentes participações que “obviamente, pressupõem a diferença da obra singular, i.e., a obra criada por um só autor, em razão do número de autores que dela participam” (FRAGOSO, 2009, p. 162). Na legislação de direito autoral brasileira, há certa confusão legislativa entre a obra colaborativa e coautoria.

A noção de coautoria não isenta o autor da sua responsabilidade da função-autor. Ocorre uma limitação nas atividades intelectuais em relação à obra, com sua responsabilidade proporcional. Nesse sentido, a atribuição autoral é também proporcional à sua responsabilidade. Ambas as modalidades são protegidas pela Lei nº 9.610/98.

Contudo, para a análise em tela, será considerada a criação colaborativa. Em princípio, esse tipo de criação pressupõe contribuições diversas, de naturezas e graus diferentes. Cada contribuição é representada por um “aporte intelectual de verdadeira criação, individualizada e identificada, resultando numa obra divisível, onde cada colaborador conserva o direito individual sobre seu trabalho” (FRAGOSO, 2009, p. 162). O aporte individual e a criação são perfeitamente distinguidos, independentes e atribuídos no corpo geral da obra. Tal definição, apesar de ser tradicional, pode ser contextualizada na concepção contemporânea da pesquisa científica baseada na *e-Science*.

Na abordagem da *e-Science*, há de se compreender que o compartilhamento de dados científicos pressupõe um trabalho colaborativo de pesquisa. Diferentes colaboradores em contextos e culturas distintas podem estar integrados em um mesmo propósito; razão pela qual a forma de recompensa autoral também se dá de forma compartilhada. É nesse limiar que a autoria na *e-Science* é revestida de uma característica singular, a colaboratividade.

A colaboratividade se apresenta como um elemento que busca trazer a tona um ambiente compartilhado com múltiplos colaboradores, pluralidade de autorias e diferentes instituições. O formato da colaboração na *e-Science* é uma vertente da pesquisa colaborativa aberta. Contudo, ainda é uma prática que precisa ser assentada por condutas, comportamentos e acordos que envolvem os pesquisadores, grupos de pesquisa, instituições e organizações envolvidas.

9.5.2 Autoria Colaborativa

Essa subcategoria emergiu com o intuito de ampliar o entendimento da colaboratividade na *e-Science*, autoria colaborativa e recompensa autoral. No momento, a reflexão está sob a égide da ‘autoria’. Tomando por base as concepções tradicionais, analogamente refletiu-se o alargamento da noção de autor a partir da resignificação do seu papel que adquiriu novos contornos da sua função-autor, relacionando a criação intelectual em espaço colaborativo.

Do ponto de vista morfológico, a autoria se relaciona às diferentes funções e significados que o labor intelectual pode materializar. Do ponto de vista político, assume um

poder legal de autoridade de forma institucionalizada que comanda e impõe a obediência sem coação. O processo científico assume dois posicionamentos, baseados em Japiassú e Marcondes (2006) e Krokosz (2015):

- a) O argumento de autoridade é uma proposição fundada no valor moral ou intelectual de alguém que exerce prestígio ou poder;
- b) Quando revestida de caráter científico, a autoria assume uma modalidade autoral típica dos processos tradicionais de produção acadêmica que variam de acordo com as disciplinas, áreas do conhecimento e instituições.

A percepção preliminar é que na dinâmica das práticas científicas da *e-Science*, a autoria colaborativa possa superar essas relações de autoridade, poder e reconhecimento individual centradas na figura do autor. Considera-se que em decorrência do modelo proprietário dos sistemas de propriedade intelectual, condutas arraigadas dessa relação de poder e autoridade na comunicação científica ainda são vigentes.

Na modalidade autoral colaborativa independente do domínio específico, a nova concepção de autoralidade parte da colaboração científica, divisão de tarefas, coletividade entre pares e espaços interdisciplinares. É predominante o requisito da originalidade e se exclui a exclusividade. Na cultura colaborativa e compartilhamento científico o autor não é o proprietário de seus dados, textos, escritos ou obras, mas sobretudo, um interlocutor em um espaço coletivo e colaborativo de criação que perde o seu *status* de proprietário e passa a figurar como colaborador.

Assim, optou-se por denominar como abordagem conceitual **pesquisador-autor**, isto é, o sujeito criador, gerador ou inventor no contexto científico. Já a condição de autoria é determinada pelo contexto (científico, governamental, cultural, artística), do qual os dados científicos são procedentes. O compartilhamento científico oportuniza ao pesquisador-autor a condição de autoria associada à sua função autoral. O dueto condição de autoria e função autoral determina a autoralidade.

Tradicionalmente, na ciência, a recompensa autoral possui um direcionamento mais sutil do que os impostos pelos sistemas de regulação autoral e se “caracteriza pela recompensa moral e não monetária e observa ‘como uma recompensa não é dada por uma nação específica (de acordo com suas leis), mas por uma comunidade internacional de pares’[...]” (KROKOSZ, 2015, p. 92). Estamos a tratar da recompensa autoral pela atribuição.

Considerou-se que, nessa dinâmica, a análise se voltou apenas para o ambiente científico, muito embora seja possível perceber a amplitude da ciência aberta e das práticas científicas que constituem a *e-Science* extrapolar o escopo da ciência. Emergem assim duas

abordagens autorais dessa nova dinâmica relacionada com a autoria originária e autoria colaborativa. Para entender a dinâmica atual, as acepções tradicionais foram utilizadas:

- a) A primeira abordagem confere o *status* de **pesquisador-autor** ao criador, gerador ou responsável pela criação ou geração de dados primários, dispondo do prestígio da **autoria originária**.

Prescreve a garantia da paternidade dos dados primários através da atribuição ao pesquisador-autor original. Promove os requisitos da originalidade (quem gerou os dados primários) e da proveniência (de onde eles foram originados). No bojo do compartilhamento, é enfatizada a troca de saberes com diferentes autores e propósitos, não configurando uma propriedade autoral, mas a recompensa autoral. Assim, o pesquisador-autor que originou primeiramente a pesquisa e, por conseguinte, os dados primários é o **autor primário**.

- b) A segunda abordagem pressupõe que dados, informações e resultados de pesquisa podem ser usados, reusados, ressignificados e, assim, passam a gerar novas hipóteses, resultados e derivações dos feitos científicos por um outro pesquisador e colaborador, configurando-se a **autoria colaborativa**.

Na acepção moderna, entendeu-se que a abordagem conceitual centrada na autoria colaborativa é a mais adequada. A **autoria colaborativa** versa com duas vertentes: a autoria colaborativa abrange, além dos pesquisadores-autores, outros participantes, tais como: curadores, bibliotecários, indexadores, especialistas tecnológicos, dentre outros profissionais envolvidos com os meios e fins do processo; e a obra derivada, na visão tradicional, é resultante do processo de acréscimo e transformação de uma obra original. Tal entendimento considera que o autor da obra derivada mantém seu *status* de autor secundário (derivado), que, em um processo colaborativo, outros atores-colaboradores figuram também com sua participação e responsabilidade proporcionais.

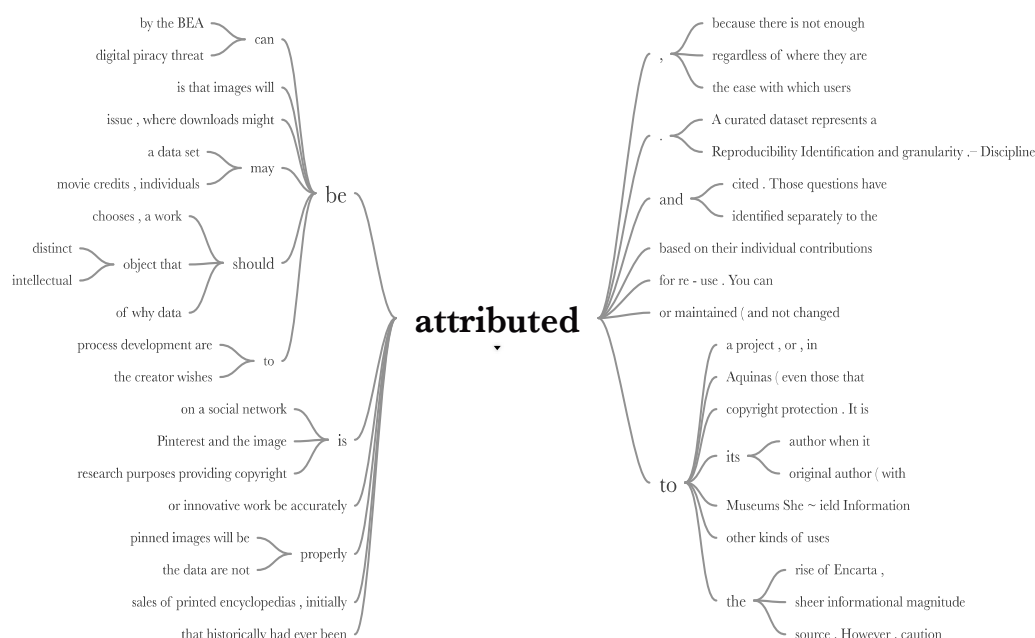
Apregoa-se, assim, que o entendimento hodierno da autoralidade está na autoria colaborativa e, por conseguinte, possui uma função compartilhada e colaborativa. Mantém-se a figura da autoria originária considerando o pesquisador-autor como o criador original. A função de autoria se configura na condição em que o pesquisador-autor se posiciona no processo científico, ou seja, se a função é de pesquisador principal e criador dos dados primários, sua autoria primária é constituída dessa forma. Se a condição parte do reuso de pesquisas, dentre elas os dados científicos, o pesquisador-colaborador e outros colaboradores assumem cada um sua função baseada na responsabilização proporcional. Figura assim, o pesquisador-colaborador na função de **autoralidade colaborativa**.

Posto isto, a análise se conduzirá para as discussões finais concentradas nas atuais discussões da recompensa autoral alicerçada no tripé: atribuição, citação e responsabilização.

9.5.3 Atribuição e Citação como Recompensa Autoral

As diretrizes internacionais e as recomendações estabelecidas pelas instituições que direcionam as práticas científicas da *e-Science* é conduzida pela noção de **atribuição e citação**. A partir da *word tree*, as palavras mais frequentes geraram codificações que foram adicionadas à subcategoria, conforme a Figura 26.

Figura 26 – *Word Tree* da subcategoria Atribuição e Citação



Fonte: Resultados de Pesquisa do *Software* NVivo (2016).

Na frequência de palavras, alguns termos como: *data set*, *movie credits*, *individuals*, *creator*, *cited*, *identified*, *individual contributions* e *original author* emergiram do *corpus* de análise. As palavras em destaque estão relacionadas aos conteúdos relacionados à atribuição e citação ligadas à recompensa autoral. A recomendação internacional é que a recompensa autoral seja feita pela atribuição e reconhecimento. A inexistência de legislações específicas para sustentabilidade legal e jurídica, redireciona para políticas, diretrizes e mecanismos de que orientam para a modalidade de reconhecimento autoral.

São mantidos os pressupostos que dados científicos compartilhados estejam acessíveis para uso e reuso, ao mesmo tempo que fixe a citação. A recompensa autoral é reconhecida e

atribuída imediatamente na devida proporcionalidade. Assim, atribuição autoral é designada a cada responsável de acordo com a sua proporcionalidade. É uma característica de singularidade em um ambiente colaborativo sem a predominância da posse e propriedade individualizada.

Seringhaus e Gerstein (2007) e Borgman (2012) afirmam que a atribuição e o reconhecimento devem ser estabelecidos na sua proporcionalidade. Atribuição é feita à parte responsável. A relação de atribuição “é, sem dúvida, aquele a quem se pode atribuir o que foi dito ou escrito. Mas a atribuição - mesmo quando se trata de um autor conhecido - é o resultado de operações críticas complexas e raramente justificadas” (FOUCAULT, 2001, p.2). Embora seja complexo dar atribuição à parte interessada, considerando a tangibilidade e intangibilidade dos atos de criação, é através da atribuição e crédito aos responsáveis que o reconhecimento autoral é garantido.

O documento denominado de *For Attribution - Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop*, publicado pela *National Academy of Sciences* (2012) elaborou diretrizes com padrões de citação de dados e boas práticas que visam formar a base para o aumento dos incentivos, reconhecimento e recompensas para atividades com dados científicos que, em muitos campos de pesquisa, estão inexistentes ou precários.

O ato de atribuir e citar não é temática nova. Desde o início dos primeiros sistemas de informação automatizados que os padrões de catalogação já estabeleciam os elementos para a atribuição. Na atualidade, as práticas científicas mantêm-se com os mesmos procedimentos, contudo a forma de descrição e as diferentes modalidades de pesquisa promoveram mudanças nas formas de atribuição e citação. Ao longo das últimas décadas, editores de periódicos, catalogadores e os próprios pesquisadores não atentaram para a importância de atribuição e citação como reconhecimento autoral. O comportamento foi mudando a partir do surgimento em escala das publicações científicas eletrônicas e a disponibilidade *on-line* pela *internet*. Em um estudo realizado pela *National Academic Science* dos Estados Unidos, que:

Da mesma forma, os arquivos de dados vêm promovendo práticas de citação de dados por várias décadas. No entanto, durante este mesmo período, muito poucos editores de periódicos exigiram citações de dados, disciplinas não inculcaram citação de dados como uma prática fundamental da boa pesquisa, concessão de agências não recompensaram as citações de dados dos responsáveis, de posse e comissões de recompensa não reconheceram citações de dados em avaliações de desempenho anuais, e os pesquisadores não assumiram a responsabilidade pela citação de fontes de dados (NAS, 2012, p. ix, tradução nossa).

A partir da era digital em rede, a **recompensa autoral** assumiu novos contornos. Sutilmente, é percebida que no contexto internacional, está associada ao reconhecimento intelectual, mérito, premiação e, por vezes, remuneração financeira. No contexto nacional, por sua vez, está firmada no sistema proprietário intelectual, associa-se também às questões éticas. O problema autoral “no campo científico é de ordem ética por estar mais diretamente relacionado à possibilidade de ocorrência de fraude do que jurídica, no que se refere ao direito de apropriação” (KROKOSZ, 2015, p. 86). O uso livre da obra desde que tenha a atribuição de créditos ao autor original é uma prática vigente. Dessa forma, entender essas nuances não basta que seja apenas sob o aspecto legal, mas perpassa pelas questões éticas e culturais que envolvem o contexto da pesquisa.

A abrangência da atribuição é mais ampla e se estende a todos os responsáveis. Assim, a acepção moderna do termo atribuição é diretamente vinculada à contribuição e colaboração. A atribuição e o crédito autoral devem ser fixados dentro de sua proporcionalidade (BORGMAN, 2012; SERINGHAUS; GERSTEIN, 2007). As responsabilidades atribuídas são creditadas a cada colaborador. Em visto disso, cada colaborador que faz parte do projeto como um todo, possui responsabilidades pela parte que lhe cabe.

A atribuição pode, portanto, ser designada a um pesquisador individual, a uma equipe de pesquisa, a uma universidade, agência de financiamento, repositório de dados, biblioteca ou a outra parte responsável pela coleta, montagem e curadoria (BORGMAN, 2012; SERINGHAUS; GERSTEIN, 2007). Destarte, a atribuição é realizada por meio dos créditos que não se limitam à pessoa física, tampouco à figura do pesquisador individual.

Com isso, a responsabilização e méritos são proporcionais. Para que o crédito seja realizado às partes responsáveis, deve ser garantida a citação da autoria originária, considerando o critério de originalidade.

No que tange ao reconhecimento autoral por meio da citação, de fato, há uma série de iniciativas em diferentes organizações e países que estão em andamento. A cultura científica está mudando com,

A citação de dados corrobora o reconhecimento de dados como uma saída de pesquisa primária e não como um subproduto da pesquisa. Quando conjuntos de dados são citados, eles alcançam uma validade e importância dentro do ciclo de comunicação científica. A citação de dados permite o reconhecimento do esforço acadêmico em disciplinas e organizações que querem reconhecer e recompensar as saídas de dados (ANDS, 2016, *on-line*, tradução nossa).

Ademais, a **atribuição e citação** requerem créditos, identificadores persistentes, aplicação de licenças, descrição de metadados, interoperabilidade e definição de políticas de garantia de atribuição para os dados científicos. O importante é estabelecer o crédito corretamente. Realizar a completa identificação do autor e colaborador do conteúdo, pesquisado e reusado, facilitará a atribuição e o reconhecimento autoral dos envolvidos na pesquisa. A citação deve ser expressa, explícita e direcionada ao responsável pela autoria da fonte consultada, reusada e reproduzida. Os produtos finais da citação estabelecem prazos e devem concentrar as referências das publicações e as indicações de autoria (BOLLEN, 2011). A identificação individualizada e proporcional a cada responsável deverá ser descrita nos metadados e realizada por meio dos sistemas de citação.

As diretrizes internacionais afirmam que dados científicos atribuídos e citados influenciam o padrão de excelência da avaliação acadêmica, o reconhecimento de pesquisadores e instituições, bem como as avaliações bibliométricas de materiais de pesquisas já publicados e dados de pesquisa utilizados como fontes primárias (BOLLEN, 2011). É através da citação que são realizados mecanismos de medição do fator de impacto da qualidade de publicações científicas em nível mundial.

Por isso, a citação na *e-Science* refere-se à prática de fornecer uma referência aos dados científicos da mesma forma como pesquisadores rotineiramente fornecem uma referência para recursos impressos e eletrônicos. Contudo, o modo de realização dessa referência está condicionado a ser feito dentro de estágios do ciclo de vida que está sendo adotado no processo de pesquisa. Geralmente, o estágio *descrever* prevê que os metadados relacionados aos pesquisadores, colaboradores e instituições sejam descritos com todas as informações pertinentes para a atribuição e completa citação.

No contexto internacional as práticas da *e-Science* já estejam bem evoluídas e a importância da citação completa para que dados científicos sejam considerados produtos primários está em consolidação. Pois, a atribuição e citação ainda são partes do processo de gerenciamento de dados que não é entendida como relevante por parte dos pesquisadores, colaboradores ou instituições. Muitas vezes, a citação não é realizada de forma eficiente devido ao desconhecimento de regras, ausência de conscientização do pesquisador ou grupo de pesquisa e complexidade de atribuir e descrever corretamente, os dados primários, partes dos dados isolados ou conjuntos de dados. Tarefa esta que, muitas vezes, é realizada por outro colaborador que não seja o pesquisador principal. Na maioria das vezes, no processo de gerenciamento de dados, quem executa esta função é o curador ou bibliotecário envolvidos no processo de ciclo de vida.

Porém, pela obrigatoriedade de as agências de financiamento exigirem, já no processo inicial de submissão, informações que estabeleçam normas, recompensas e incentivos relacionados à recompensa autoral (atribuição e citação), na elaboração dos planos de gerenciamento de dados científicos, muitas vezes, tais procedimentos não são atendidos com eficiência (NAS, 2012). Esta é uma prática mandatória que agências de fomento internacionais estão aplicando em diversos países. Assim, a ausência de compreensão de todo o processo e, muitas vezes, a inexperiência do pesquisador ou equipe de pesquisa acarretam a ineficiência no processo de reconhecimento autoral.

A esta constatação, acrescentam-se as regras de *copyright* vigentes que prevalecem e, por vezes, podem ser aplicadas também à citação de dados científicos. Quando dados científicos são compartilhados ou arquivados para preservação a longo prazo o autor original do *copyright* retenha seus direitos. Quando os dados científicos são oriundos de pesquisas já em andamento, é necessário que haja uma autorização para que os mesmos possam ser arquivados em repositórios ou *data sets*. Os detentores dos direitos sobre esses dados precisam ser identificados e conceder explícita autorização para que os dados sejam compartilhados. Por sua vez, os autores secundários também precisam obter liberação do *copyright* antes que os dados sejam compartilhados para reuso (SAYÃO; SALES, 2015). Acrescentam-se também as questões associadas ao *fair use*, quando não houver interesse comercial.

Claramente é percebida a existência de um processo híbrido de reconhecimento autoral, mesmo no cenário internacional. Essa combinação de atribuição de *copyright* é compreendida em pesquisas ou publicações que estejam sob o regime da propriedade intelectual vigente. Porém, as regras institucionais ou políticas nacionais devem prever tais restrições para que na perspectiva de pesquisa baseada em dados científicos, o acesso aberto, gratuito e público seja garantido.

A NISO e a NAS lançaram um documento que recomendam o uso de identificadores persistentes e interconexão *on-line* para que a citação seja atribuída corretamente e de forma eficiente. Ademais, os sistemas de citação e repositórios de dados conectados estão desenvolvendo as normativas para a atribuição e citação automáticas. A seguir, alguns exemplos de iniciativas internacionais que estão em andamento com propósitos de atribuição e os devidos créditos para os pesquisadores-autores, colaboradores, instituições e agências de fomento.

Na perspectiva internacional, algumas dessas iniciativas estão desenvolvendo ou adotando sistemas de crédito para dados científicos. Recomenda-se também atribuir o crédito

à instituição a qual os dados estão vinculados. Para tanto, a confiabilidade deve ser assegurada “porque as instituições de pesquisa acadêmica dependem de registros confiáveis de realizações acadêmicas para as principais decisões sobre a contratação, promoção e gestão” (SMITH, 2012, p. 77, tradução nossa).

Os sistemas de citação foram evoluindo, ao longo das décadas, do material impresso ao eletrônico. Introduziram-se as avaliações por pares e passou-se a adotar mecanismos eletrônicos conectados aos sistemas de gerenciamento para a completa citação; “além disso, uma série de serviços surgiu para tornar a avaliação do registro mais fácil para a administração. Isso inclui fatores de impacto, análises acadêmicas e outros métodos” (SMITH, 2012, p. 77, tradução nossa).

Na era digital e mais recentemente em práticas que envolvem a *e-Science* a representação institucional é um fator importante no processo de comunicação científica. Especialmente no caso de universidades que são avaliadas por *rankings* (nacionais e internacionais) de qualidade, como, por exemplo, o fator de impacto e avaliação de desempenho de pesquisadores e professores, considerados extremamente importantes para o processo decisório da administração universitária.

Outro aspecto no âmbito institucional está relacionado com o desenvolvimento econômico movido pela inovação e tecnologia. É a vertente da inteligência dos negócios acadêmicos que

muitas universidades agora têm grandes programas de ligação industrial e escritórios de licenciamento de tecnologia. Eles estão sempre tentando descobrir o que os acadêmicos estão produzindo, o que pode ser comercializado ou explorado, tanto para benefício da universidade como do pesquisador (SMITH, 2012, p. 77, tradução nossa).

A maioria das universidades internacionais é privada e depende dos seus *rankings* de classificação para criar estratégias de captação de alunos, corpo docente qualificado e de excelente reputação acadêmica, além de captação de recursos de financiamento de pesquisa em agências de fomento, instituições governamentais e doações de entes privados.

Considera-se, também, no arcabouço institucional que as questões de propriedade intelectual dos dados científicos ainda estão presentes. Ao mesmo tempo em que a geração dos dados científicos é considerada um produto principal de pesquisa, ela possui potencial econômico e comercial. Dessa forma, a atribuição à instituição ou às várias instituições envolvidas com o projeto de dados deve ser citada. Mesmo que pesquisadores suponham que

eles controlam os seus dados e tenham os direitos de propriedade intelectual assegurados, muitas vezes, esses direitos não pertencem a eles.

A recomendação é que a citação seja expressa, pois se espera que os pesquisadores e as instituições compartilhem seus dados científicos, sem custo adicional e dentro de um prazo razoável. Os dados primários, amostras, coleções físicas e outros materiais de apoio criados ou reunidos no curso do trabalho devem ser compartilhados (NAS, 2012). As agências de fomento e institutos de pesquisa entendem que as pesquisas financiadas com fundos públicos já pressupõem seu compartilhamento aberto.

Além das recomendações político-normativas estabelecidas na perspectiva institucional, a infraestrutura tecnológica para citação de dados deve incluir e prever, em seus projetos iniciais, a manutenção de serviços e custos de investimentos. Nesse sentido, o estudo de Smith (2012) levantou algumas recomendações pertinentes:

- a) **Localização persistente ou detectável:** deve funcionar mesmo que os dados se movam ou se houver várias cópias. Primeiro, uma citação tem que ser persistente ou fornecer um local de descoberta. Segundo, mecanismo de descoberta deve detectar a citação independente da localização;
- b) **Conteúdo verificável:** está relacionado com a autenticidade, isto é, a citação deve ser inalterada. Requer a descrição e qualidade de metadados para assegurar a proveniência e a descoberta. É preciso provar a autenticidade dos dados. Em outras palavras, a URI que referencia o trabalho de pesquisa ou conjunto de dados deve ser confiável e verificável a qualquer momento;
- c) **Padronização:** requer adotar identificadores de dados, por exemplo, *DataCite*, DOIs e Identificadores de Pessoas, como o Registro no ORCID para os pesquisadores e colaboradores;
- d) **Identificadores Institucionais:** podem ser adotados para aumentar a visibilidade por meio de padrões de catalogação (OCLC e NISO I2);
- e) **Viabilidade financeira:** prever os custos financeiros e em longo prazo do projeto de pesquisa, pois adotar identificadores e produzi-los requer cessão e manutenção dos custos de serviços.

No que tange à perspectiva do pesquisador, é recomendado que a segurança da atribuição e crédito, bem como as questões relacionadas com os prazos sejam observadas. Nos sistemas tradicionais de propriedade intelectual, a grande maioria não assegura a proteção de ideias. No contexto dos dados, essa possibilidade pode ser prevista.

No estudo de Harley (2012, p. 83, tradução nossa), percebeu-se que em determinados campos do conhecimento, pesquisadores “recebem atribuição para as suas ideias. Em domínios altamente competitivos, como a biologia molecular, que possuem uma corrida para publicar e medo de ser violado, a partilha precoce de ideias, dados e documentos de trabalho é quase inédita”. Sendo que nos atuais sistemas de DPIs a proteção às ideias é vedada.

Contudo, em uma escala menor, emergente e em domínios específicos com baixa concorrência interna, os mecanismos informais de gestão de reputação podem impor atribuição porque, “as comunidades acadêmicas são organizadas e mantidas através da interação face a face (muitas vezes através de conferências e *workshops*) centralmente. Esta cultura de compartilhamento [...] está se movendo em direção a práticas menores de partilha” (HARLEY, 2012, p. 83, tradução nossa).

Como alguns domínios específicos são de alta complexidade e podem gerar patentes, considera-se a possibilidade de atribuição das ideias. A decisão de realizar a atribuição preliminar das ideias deve obedecer a um período de embargo. Decidir também pelo não compartilhamento também é uma opção em áreas do conhecimento com alta complexidade e competitividade.

Outra perspectiva que foi levantada está no crédito pessoal para os pesquisadores e colaboradores para progressão de carreira ou captação de recursos de financiamento. Além disso, “a capacidade de estabelecer rapidamente se uma ideia anteriormente foi levada adiante - e a creditá-la corretamente, se assim for - é importante para os cientistas” (SERINGHAUS; GERSTEIN, 2007, p. 4, tradução nossa).

A identificação do autor e o completo conteúdo central pesquisável irão simplificar este processo e facilitar a atribuição e reconhecimento. Assim, o compartilhamento dos dados pode ser uma vantagem, por conseguinte, quanto mais cedo os mesmos forem disponibilizados, creditados e acessíveis, favorecem as possibilidades de reconhecimento.

Além das perspectivas institucionais e pessoais, considera-se que a infraestrutura tecnológica dos sistemas de citação e repositórios de dados pode ser viabilizada por iniciativas institucionais, empresas privadas e parcerias de consórcios.

O consórcio *DataCite Consortium* é um exemplo de projeto que tem como pano de fundo a ciência aberta global e, portanto, precisa de padrões globais. Pensar local e mundial é um dos pressupostos de um projeto dessa natureza. Brase (2012, p. 95, tradução nossa), em seu estudo, afirma que é necessário desenvolver “fluxos de trabalho e cooperação globais entre os agentes globais, tais como centros de dados globais e editores. E claro, a ciência também é realizada localmente”. Pesquisadores, bibliotecários e inúmeros colaboradores

fazem ciência global, mas são incorporados em suas instituições locais, bibliotecas e agências de financiamento.

Então, esse é o paradoxo do projeto de dados: “queremos ter respostas globais, mas ao mesmo tempo queremos atuar em uma base muito local” (BRASE, 2012, p. 95, tradução nossa). Esse foi o principal motivo por trás do projeto *DataCite*: um consórcio global de instituições locais. Esse funciona como um parceiro central para trabalhar com o fornecimento de citações de dados de inúmeros centros de dados que compartilham individualmente, mas possuem visibilidade e acesso mundial.

A Rede *Data Citation in the Dataverse Network*® é outra iniciativa que dá suporte ao compartilhamento e citação de dados. É um *software* de fonte aberta publicado pelo *Harvard Institute of Quantitative Social Science* da Universidade de *Harvard* nos Estados Unidos. O projeto *Dataverse Network*® (DVN) é um aplicativo de código aberto para a descoberta, publicação, citação e preservação de dados de pesquisa. É referenciado como a infraestrutura prática para o compartilhamento de dados com código-fonte aberto (ALTMAN, 2012).

O *Dataverse* pode atuar como um arquivo virtual para uma revista, um departamento, um grupo de pesquisa, um indivíduo, ou uma biblioteca. O servidor DVN permite uma instituição para fornecer serviços de *backup* unificados, geração de citação, de descoberta, preservação e outros tipos de serviços básicos enquanto permitem que os proprietários dos arquivos virtuais possam gerenciar o depósito de conteúdo, difusão, termos de uso e licenças. A rede *Dataverse* pode ser usada como hospedeira para metadados, texto, imagens e dados científicos de disciplinas diferentes.

No contexto nacional, em 2015, o IBICT passou a utilizar o Repositório *Dataverse* para dar suporte à Rede Brasileira de Serviços de Preservação Digital – CARINIANA como uma administradora de conteúdo. A rede CARINIANA é responsável pela preservação digital compartilhada da informação científica, tecnológica e cultural.

No que diz respeito às políticas de acesso e permissões, são adotadas as orientações da *Harvard Dataverse*. A política orientada para aumentar a visibilidade dos dados da pesquisa é realizada por meio do reconhecimento e crédito acadêmicos através da citação. As permissões e restrições adotadas para o compartilhamento e disponibilidade dos conteúdos e dados de pesquisa no repositório *Murray Archive* são gerenciadas por licenças no *DataCite*.

A citação de dados no *Dataverse* é estabelecida com base nos quatro princípios da *Joint Declaration of Data Citation Principles* (2014): crédito e atribuição; identificador persistente; especificidade; e verificabilidade. Para a descrição da citação no *Dataverse* são definidos oito componentes: nome do autor, título, ano, data de publicação, publicador

(repositório que publicou o conjunto de dados), identificador persistente (DOI ou *Handle*), número da versão e impressão digital numérica universal (*Universal Numerical Fingerprint – UNF*), este último quando tratar de dados tabulares.

Em suma, as iniciativas apresentadas exemplificaram os sistemas de atribuição e citação para os dados científicos. A atribuição e citação constituem um processo complexo e requerem investimentos e políticas institucionais que apoiem o empreendimento local e mundial. O compartilhamento de dados é dificultado pela escassez de tempo dos envolvidos (pesquisadores, curadores e bibliotecários), preparação pessoal para descrever os metadados necessários com qualidade para futura curadoria (que incluem os esforços, por vezes, de conversão de dados analógicos para formatos digitais ou padronização de dados desorganizados) e acesso a longo prazo. A extrema heterogeneidade e as formas idiossincráticas, como no caso da sequenciamento de genes, por exemplo, (HARLEY, 2012) são aspectos que dificultam e, por vezes, inviabilizam o compartilhamento e a atribuição.

No limiar dessas dinâmicas contemporâneas aspectos éticos também são observados. Estimula-se a competitividade sob uma dinâmica ético-cultural que tem como premissa o respeito à originalidade dos dados de pesquisa disponíveis para reuso e reprodutibilidade. No que tange aos aspectos éticos, a atribuição falsa de autores é considerada um desvio de conduta ética. Essa recai no plágio autoral relacionado à fraude, violação ou desonestidade autoral. A desonestidade autoral no meio acadêmico é muito comum, levando à apropriação indébita e fraude de conteúdos alheios, bem como à expropriação econômica por terceiros.

Os sistemas de propriedade intelectual preconizam que a proteção autoral possa ser realizada por meio de legislações específicas e que, por conseguinte, garantam contra o plágio e a violação de conteúdos por terceiros. A violação autoral, no caso brasileiro, é prevista no código penal; já no âmbito internacional, é prevista nos acordos, tratados e convenções vigentes.

Entretanto, após o advento da *internet*, com a facilidade de disseminação e acesso virtual, os riscos de violação aumentaram, e a desonestidade ética prevalece. “Possíveis violações dos direitos de propriedade intelectual incluem a utilização não autorizada de materiais protegidos por direitos autorais, pirataria digital e plágio” (ILKA; RUSSELL, 2014, p. 234). Constata-se que os instrumentos regulatórios da propriedade intelectual vigentes estão em desconformidade como o novo paradigma, muito embora ainda sejam utilizados plenamente como protetores do capital intelectual.

Acrescentam-se, ainda, as preocupações com a ética nos processos científicos que “correspondem à necessidade de reflexão e estabelecimento de parâmetros sobre fraude,

conflito de interesses, atribuição de autoria, degenerações autorais (autoria presenteada, fantasma, coagida) e publicação duplicada” (KROSCOCZ, 2015, p. 80).

Em virtude de a legislação autoral não dar conta das violações, as questões éticas estão sendo conduzidas por códigos de ética ou conduta recomendados pelas agências de fomento, universidades e instituições de pesquisa. Os aspectos éticos que envolvem questões de privacidade dos seus dados devem ser corretamente atribuídos, “se os seus dados envolverem, por exemplo, seres humanos, espécies em risco, ou habitats sensíveis, você deve tomar medidas especiais quando do compartilhamento dos dados” (SAYÃO; SALES, 2015, p. 23). É por meio dos códigos de conduta que parâmetros mínimos preveem os princípios éticos para nortear também as pesquisas científicas vigentes da *e-Science*.

O *Committee on Publication Ethics* (COPE) é um fórum de editores e editoras de periódicos, revisados por pares, para discutir todos os aspectos da ética na publicação. Ele também aconselha editores sobre como lidar com os casos de má conduta na pesquisa. O Código de Conduta foi reorganizado e atualizado em março de 2008 e revisado em 2011 como um “padrão-ouro” das melhores práticas desenvolvidas. Tornou-se uma referência internacional com diretrizes que orientam para: adoção de processos que estimulem a exatidão, integridade, confidencialidade e clareza nas pesquisas; sistemas de autoria e contribuição que promovam as boas práticas; códigos de conduta que desencorajem a autoria fantasma, convidada ou coagida; incentivo à originalidade das submissões e adoção de sistemas de detecção de plágio (COPE, 2011).

O documento da *EU-Directive* 1995/46/EC foi elaborado para conduzir as questões éticas na execução do projeto *Foster* na Comunidade Europeia. O documento abrange os deveres de confiabilidade, dados pessoais, dados sensíveis e exceções legais para o compartilhamento e reuso. Acerca do dever de confiabilidade, é necessário o consentimento prévio dos participantes antes do compartilhamento para que não haja violação da confidencialidade. Nesse caso, se houver interesse público, este pode substituir o dever de confidencialidade. Quando se tratar de dados que estão sob processo judicial, pesquisadores devem restringir o compartilhamento dentro do prazo previsto. E, por fim, devem ser adotadas documentações de consentimento para a liberação dos dados.

Sobre os dados pessoais, estes são protegidos pela *Data Protection Act* de 1998 que assegura o respeito ao indivíduo vivo. Nesse caso, recomenda-se que seja realizado por meio de sua identificação pessoal e incluindo qualquer expressão de opinião emitida pelo mesmo. A divulgação só poderá ser realizada após consentimento expresso e procedimentos legais cumpridos. A manipulação dos dados pessoais também faz parte do padrão ético a ser

adotado. Nessa feita, os dados pessoais devem ser tratados com lealdade e de forma lícita e justa, sendo utilizados para fins estritamente específicos, adequados e pertinentes ao domínio da pesquisa.

Quando necessário que os dados pessoais sejam transformados ou alterados, a exemplo do reuso, a autorização deve ser concedida para este fim específico. Sugere-se também que o armazenamento dos dados confidenciais seja preservado em arquivos separados. Se houver exceções na coleta dos dados pessoais no decorrer do ciclo de investigação, esses podem ser mantidos indefinidamente, se necessário, utilizados para outros fins em certas circunstâncias, desde que as pessoas sejam informadas.

Quanto aos dados sensíveis, esses são referentes à raça de um indivíduo ou origem étnica, opiniões políticas, crenças religiosas, filiação sindical, saúde física ou mental, vida sexual, processos ou condenações penais previsto no *Data Protection Act 1998* (REINO UNIDO, LEGISLATION.GOV.UK, 2005). Tais dados só podem ser processados para fins de pesquisa com o consentimento expresso, de preferência por escrito, mediante parecer da investigação médica por um profissional de saúde ou equivalente com o dever de confidencialidade, anonimização, análise das origens raciais/étnicas para fins de igual monitoramento de oportunidades e de grande interesse público e não causar danos substanciais à pessoa.

A abrangência dessas diretrizes éticas não possui efeito jurídico, mas são diretivas normativas aos Estados membros da Comunidade Europeia. As políticas institucionais devem também estabelecer nos seus instrumentos a declaração de consentimento quando se tratar de dados sigilosos, sensíveis ou com impedimentos legais. Todos os envolvidos - colaboradores e sujeitos - no processo devem estar cientes das diretrizes éticas e de normas de conduta. Para o compartilhamento de dados de diferentes naturezas (pessoais, sensíveis ou sigilosos), por exemplo, deve o pesquisador gozar de um comportamento rigoroso baseado em princípios éticos e morais. Assim, a ética requer o consentimento, anonimização e regulação de acesso.

Além dos aspectos político-normativo e ético-legal, outra percepção emerge da LAI que se volta para aspectos culturais no sentido de restaurar um espaço democrático entre Estado e cidadão. Para além do âmbito jurídico, trata-se de um processo associado a uma cultura política que vem sendo dinamizada, nas últimas décadas, em meio às contradições que envolvem tradições autoritárias e uma agenda democratizante nas relações Estado e Sociedade (JARDIM, 2012). No cenário brasileiro, prevalecem ainda aspectos autorais e culturais arraigados aos processos vigentes, representados, em certa medida, pela

desonestidade ética, plágio, cultura do produtivismo intelectual desenfreado e da aplicação de mecanismos eficazes de proteção.

Por último, a análise teve o intuito de refletir sobre o novo fazer científico e as novas formas de entender a autoria no ambiente de práticas de pesquisa colaborativas. A dimensão ético-legal e a interface cultural andam juntas no processo de compreender que as posturas arraigadas em um paradigma proprietário ainda são presentes na ciência e que o redesenho em favor de uma cultura baseada no *commons* e na colaboração científica é a essência da ciência aberta que tem um longo caminho a percorrer. A nova dinâmica não visa à destituição do pesquisador-autor, mas do caráter proprietário e individualista em prol do empoderamento da postura colaborativa em um espaço de múltiplos de autorias.

É mister que nas práticas da *e-Science* o desvendar da autoralidade é promovido pela recompensa autoral. As novas acepções sobre a autoralidade expande o conceito de autoria colaborativa. Porém, a promoção da recompensa autoral ainda está em consolidação e precisa de maturação para um entendimento global. Muito embora, as recomendações internacionais conduzam para o reconhecimento por meio da recompensa autoral. No Brasil, ainda é uma realidade que precisa ser discutida em futuras políticas orientadas à pesquisa colaborativa aberta com dados científicos. As reflexões serviram de direcionamento para a constituição do padrão ético-legal-cultural. O enigma desvendado está norteado pelas reflexões emergidas e constituídas por padrões.

10 ENIGMA DESVENDADO: AUTORALIDADE COLABORATIVA NA *E-SCIENCE*

Durante a pesquisa, foi observado que políticas, diretrizes, práticas e recomendações internacionais não são de caráter mandatório, mas estabelecem o cumprimento de procedimentos obrigatórios para a abertura de projetos com dados científicos, em especial nas pesquisas financiadas com fundos públicos. Geralmente, tais diretrizes não possuem força legal, contudo passam a ser normativas que abrangem os procedimentos necessários.

O desvendar inicial aponta para o redirecionamento de dois paradigmas: o quarto paradigma científico preconizando a mudança na comunicação científica orientada à pesquisa colaborativa com dados científicos e o paradigma proprietário que pressupõe um redirecionamento dos direitos de propriedade intelectual vigentes para uma cultura colaborativa e do *commons* que pressupõe que os bens intelectuais sejam de acesso aberto e público.

Os padrões estabelecidos na propositura do modelo conceitual são recomendados e preveem requisitos mínimos de conformidade baseados em diretrizes que são práticas usuais no modelo emergente da *e-Science*. Assim, o modelo conceitual foi pensado em padrões com interfaces multidimensionais, construído a partir das similaridades temáticas e de características intrínsecas às práticas científicas emergentes. O ponto de partida para o modelo foi o agrupamento de seis dimensões que resultaram em um quadrante de quatro padrões. Esse quadrante representa a fusão de uma interface multidimensional e disciplinar a partir das temáticas estudadas.

O *constructo* do estudo foi assentado na colaboração, compartilhamento, multidisciplinariedade e multidimensionalidade que reflete as dinâmicas contemporâneas preconizadas pelo quarto paradigma científico, além de ampliar o entendimento da autoria na *e-Science*. A multidisciplinariedade entre a Ciência da Informação e as Ciências Jurídicas e suas interfaces com a Tecnologia e Economia estão agrupados nos quatro padrões constitutivos para o modelo conforme é apresentada na Figura 27.

Certamente, a visualização através do modelo conceitual ajudará na aplicação de regras mínimas que auxiliarão nos procedimentos autorais junto as práticas de pesquisadores, colaboradores, universidades, instituições governamentais e agências de fomento.

Figura 27 – Padrões multidimensionais do modelo de autoralidade



Fonte: Elaborado pela autora (2016).

Os quatro padrões constitutivos: epistemológico-teórico, político-normativo, morfológico-técnico e ético-legal-cultural dão sustentabilidade ao modelo conceitual de autoralidade colaborativa.

10.1 PADRÃO EPISTEMOLÓGICO-TEÓRICO

Esse padrão compõe a **primeira parte** do modelo norteador pelas dimensões epistemológica e teórica. Os elementos constitutivos desse padrão substanciam pesquisadores, colaboradores e instituições a observarem, preliminarmente, aspectos teóricos e epistemológicos que conduzirão o projeto de pesquisa com procedimentos mínimos que estabeleçam a coesão, qualidade e conformidade durante todo o processo de investigação. Ademais, o quarto paradigma pressupõe que projetos com dados científicos sejam constituídos em uma abordagem colaborativa e a dimensão epistemológico-teórica se direcione ao *constructo* do estudo.

Gerar conceitos, construir hipóteses e pressupostos, formular objetivos, interpretar fatos e fenômenos com vistas às soluções provisórias fazem parte da dimensão teórica

(SILVA, 2016) e da epistemológica, pois o objeto de estudo, a problemática, o problema de pesquisa e o arcabouço metodológico, apesar de já ser uma prática vigente científica, compõem os procedimentos iniciais da investigação de pesquisa.

A dimensão epistemológica considerou a influência do quarto paradigma científico nas práticas da *e-Science*. O **primeiro axioma** afirma que para uma pesquisa científica colaborativa é necessário que o *constructo* do estudo seja conduzido ao produto primário constituído pelos dados científicos. O **segundo axioma** afirma que as práticas e métodos científicos adotados durante o gerenciamento dos dados são sistematizados pelo ciclo vida e salvaguardam o compartilhamento, a preservação, o acesso para que os mesmos possam ser usados e reusados.

Em particular o padrão epistemológico-teórico estabelece diferenciações conceituais entre a ciência aberta e a *e-Science*, além de apontar para os propósitos e contextos diferentes. A tipologia está relacionada com a originalidade dos dados (primários ou derivados). A natureza se relaciona com a proveniência de contextos (científicos, governamentais, institucionais ou privados). O *status*, por sua vez, está relacionado com o acesso (aberto, restrito e gratuito). A importância de estabelecer os aspectos epistemológicos e teóricos que norteiem a condução da pesquisa primária, deve também dar sustentabilidade para subsidiar pesquisas futuras (derivadas) a partir do compartilhamento, reuso e reprodutibilidade a longo prazo.

10.2 PADRÃO POLÍTICO-NORMATIVO

O padrão político-normativo é a segunda parte do modelo. A dimensão política norteou a constituição desse padrão sob duas perspectivas: o lado formal da política a partir de uma decisão governamental ou institucional; e o lado prático da política a partir da *práxis* e perspectivas de pesquisadores, colaboradores e institucionais.

O lado formal conduz uma política pública constituída por princípios e instrumentos normativos que estabelecem premissas para decisões e direcionamentos pelo Estado e aplicadas nas ações e planejamentos em diversos contextos. Pressupõe motivação política de governantes, gestores e pesquisadores, que direcionem as estratégias políticas em diferentes escopos (nacional, governamental ou institucional). Pode também ser elaborada e concretizada sem um lastro governamental, direcionada a um contexto mais restrito e aplicada em âmbito institucional. Independentemente da amplitude da política, a sua composição deve

pensar minimamente nos aspectos administrativo, orçamentário, legal, científico, tecnológico, pessoal, cultural e social.

A política é composta por elementos constitutivos que estabelecem princípios, diretrizes, instrumentos normativos e procedimentos que conduzem à execução. O sentido de estabelecer políticas na ideia do *soft law* é intrínseca às diretrizes internacionais, a exemplo das recomendações da OECD. Essa ideia visa estabelecer valor normativo às políticas e diretrizes como procedimentos e critérios mínimos, e não mandatórios e, por conseguinte, sem imposição legal jurídica, mas que conduzem à abordagem prática político-normativa em conformidade internacional.

10.2.1 Axiomas para a formalidade da política de dados científicos

A política orientada aos dados científicos requer procedimentos formais mínimos para a conformidade internacional. Os aspectos gerenciais e institucionais conduzem a execução da pesquisa, proporciona a acessibilidade e visibilidade dos dados científicos e o empoderamento das comunidades científicas envolvidas. Os aspectos financeiros e orçamentários envolvem captação de recursos para o financiamento do projeto, pessoas e especialistas, recursos materiais, tecnológicos, reconhecimento autoral (sistemas de recompensa e identificadores persistentes), custos de manutenção (armazenamento e preservação) a longo prazo. As diversas instâncias com instituições governamentais, institutos de pesquisa, universidades, agências de fomento e empresas privadas são necessárias para atender estes aspectos.

A política também contempla também aspectos legais, éticos e educacionais. As questões legais prevê os sistemas de recompensa autoral, atribuição e citação para garantir a recompensa autoral. Ademais, para pesquisas que estejam sob a vigência de direitos autorais vigentes, casos de licenciamento e autorização prévia devem ser definidos. Sob o aspecto ético adota-se na política códigos de conduta e comportamento. E o processo educacional dar suporte ao desenvolvimento de habilidades interdisciplinares aos envolvidos no projeto.

A pesquisa suportada pela ciência aberta proporciona uma mudança no *modus operandi* das práticas científicas e, por conseguinte, o lado formal só é possível a partir de uma decisão política, governamental e institucional, com requisitos políticos, normativos, diretrizes, procedimentos e instrumentos com vistas à harmonização interna dos procedimentos necessários, o empoderamento da comunidade científica e da sociedade local, o crescimento e desenvolvimento econômico, social e cultural, e por fim, a inovação científica com padrões mínimos de qualidade, ética e conformidade internacional.

10.2.2 Axiomas para a *práxis* da política de dados científicos

O lado prático da política envolve a *práxis* da *e-Science* e a execução da política e dos procedimentos pela equipe envolvida. É quando pesquisadores, colaboradores ou grupos de pesquisa assumem o caráter prático político, com a execução e condução de ações, metas e comportamentos. A *práxis* envolve a sistemática do modelo de ciclo de vida dos dados e norteado pela política definida previamente.

Os envolvidos no projeto devem ser motivados a compartilhar e depositar seus dados primários e resultados (preliminares e finais) da pesquisa. Tal motivação parte do princípio que os projetos capitaneados por financiamentos públicos devem ser abertos, muito embora ainda exista a preocupação e a resistência sobre os direitos autorais e patentes. As condições para a execução do processo investigativo são apoiados por recursos materiais, financeiros, tecnológicos e por colaboradores envolvidos no projeto. As questões tecnológicas, o acesso aos dados e a sua exploração requerem uma infraestrutura tecnológica adequada, um amplo acordo internacional sobre a interoperabilidade e controles eficazes de qualidade e proveniência de dados. Os recursos materiais, instrumentos, equipamentos e pessoal são para execução local. Por isso, a importância de políticas, normas, diretrizes e regras que estabeleçam os procedimentos para execução local.

Em suma, o padrão político-normativo deve pensar global e local em um empreendimento científico que se volta para o panorama da ciência aberta e da pesquisa compartilhada. As políticas aspiram a estratégias globais que busquem alavancar a colaboração científica internacional e ao mesmo tempo impulsionem o desenvolvimento científico, econômico, tecnológico, social e cultural em âmbito nacional. Pensar localmente é entender o contexto nacional, institucional, cultural e social como um meio de promover a cidadania, oportunizando as condições de acesso à informação, aos dados primários e métodos de pesquisa subjacentes dos processos investigativos como produtos científicos primários, além de empoderar a comunidade científica local. É tornar a ciência uma provedora de benefícios e melhorias em prol do desenvolvimento científico, econômico, social e sustentável com vistas ao retorno à sociedade em geral.

10.3 PADRÃO MORFOLÓGICO-TÉCNICO

O padrão morfológico-técnico constitui a **terceira parte** do modelo. A dimensão morfológica possui uma versatilidade com base nos aspectos metodológicos, técnicos e

tecnológicos. A morfologia abrange a estruturação formal e representacional (modelos, sistemas, formatos, ferramentas) que envolvem a operacionalização sistemática da *práxis* da *e-Science*. A sistematização do objeto de estudo foi observada sob a ótica do modelo de ciclo de vida dos dados e contemplou os procedimentos metodológicos, técnicos, descritivos e tecnológicos.

No processo de gerenciamento de dados é o ciclo de vida que sistematiza as ações: metodológicas, técnicas e gerenciais. A dimensão técnica é transversal e perpassa pelos procedimentos técnicos e descritivos em cada estágio do ciclo de vida, construindo a relação entre o objeto e a *práxis*. Apoia também as soluções tecnológicas que dão sustentabilidade ao processo de gerenciamento e compõem a ciberinfraestrutura tecnológica. Essa ciberinfraestrutura requer procedimentos e execução de uma arquitetura computacional para cada estágio do ciclo de vida com soluções tecnológicas agregadas à execução (técnica e descritiva) do gerenciamento dos dados.

Assim, as dimensões teórica e morfológica contemplam a execução e sistematização do gerenciamento de dados e as soluções tecnológicas adotadas na pesquisa colaborativa. A fusão dessas dimensões em um padrão representa os procedimentos mínimos que atendam os aspectos morfológicos, técnicos e tecnológicos.

10.3.1 Axiomas morfológicos para a sistematização técnica e tecnológica

Por entender a importância do direcionamento morfológico no que tange a sistematização (técnica e tecnológica) através do modelo de ciclo de vida dos dados levantou-se algumas diretrizes importantes e em conformidade internacional. Diretrizes internacionais não estabelecem um padrão, mas recomendam requisitos mínimos que conduzem as ações de: planejar, coletar, assegurar, descrever, preservar, visualizar e analisar. O padrão tem o intuito de apresentar esses critérios necessários como requisitos para a garantia autoral no reuso de dados científicos a partir de três premissas:

- a) As diretrizes político-normativas constituem os procedimentos de planejamento, suporte e manutenção a longo prazo. A importância de elaborar o planejamento como estágio inicial é requisito primordial para a condução do projeto de pesquisa;
- b) A infraestrutura técnica requer conhecimento técnico com profissionais habilitados, além dos pesquisadores, como bibliotecários, cientistas de dados, tecnólogos e curadores para a execução eficiente dos estágios do ciclo;

- c) A recompensa autoral envolve vários estágios do ciclo de vida. O preenchimento de metadados, informações adicionais do projeto de pesquisa, documentos e procedimentos padronizados e normatizados são necessários para atender os critérios de qualidade, acesso, interoperabilidade, visualização e análise.

Para atender aos critérios mínimos legais, esse padrão estabeleceu alguns critérios com vistas a atender requisitos que auxiliem na recompensa autoral dos dados científicos reusados. Para tanto, alguns procedimentos devem ser seguidos dentro da sistemática do ciclo de vida:

- a) O estágio *planejar* deve estabelecer o plano de pesquisa. Iniciativas internacionais recomendam a adoção de ferramentas abertas de gestão e de fluxos de trabalho, a exemplo dos *DMPTool* e *Taverna*. É uma recomendação internacional mandatória para as pesquisas capitaneadas por fundos públicos. É nesse estágio que informações legais de responsabilidade, reconhecimento, atribuição, citação, licenças e as tecnologias que serão adotadas são pensadas para a execução futura;
- b) O estágio *coletar* deve observar fatores extrínsecos (condições de coleta e acondicionamento), responsáveis pela proveniência (pesquisadores, colaboradores, instituições e universidades) e intrínsecos (natureza, tipologia, integridade e fidedignidade) dos dados como requisitos importantes para que questões legais sejam observadas durante a coleta e curadoria preliminarmente. É aconselhado observar instruções normativas para autorização e consentimento prévios;
- c) A curadoria preliminar é realizada no estágio *assegurar* como um procedimento básico de validação de garantia da integridade e integralidade de dados primários para a descrição futura dos metadados;
- d) É imprescindível o uso de identificadores persistentes para a identificação de dados primários, pessoas e instituições. Os identificadores persistentes chamados de *Persistent IDentifiers (PIDs)* são vinculados às publicações e contribuidores para que dados sejam indexados, atribuíveis e citáveis de forma exclusiva (única) e em longo prazo. Um sistema de identificação além de assegurar a qualidade e identificação, proporciona gerenciar os direitos de propriedade intelectual de dados e objetos digitais;
- e) É recomendado fortemente o uso do *Digital Object Identifier (DOI)* para a identificação dos dados científicos, metadados, documentos subjacentes e publicações com os resultados de pesquisa. E o *Open Researcher and Contributor Identification (ORCID)* como identificador único para cada envolvido

(pesquisadores e colaboradores) no processo de pesquisa assegurando a vinculação dos créditos aos dados e aos responsáveis na sua proporcionalidade;

- f) A utilização de licenças públicas que estabeleçam as condições para uso e reuso, acesso aos diferentes tipos de dados e suas restrições, bem como as exigências sobre os prazos e períodos de embargo são medidas importantes para assegurar a proteção e reconhecimento autorais;
- g) A descrição dos metadados de dados científicos é uma das etapas mais relevantes no processo de gerenciamento executado no estágio *descrever*. Adotar um padrão de metadados é obrigatório. É apregoado que se utilizem padrões de metadados de acordo com o domínio específico da pesquisa, a exemplo, *Darwin Core*, *Genome metadata* e *METSRights*. É na descrição dos metadados que a proveniência, a atribuição e a vinculação dos dados aos pesquisadores, colaboradores, instituições e agências de financiamento são asseguradas;
- h) A conformidade internacional também é prevista no estágio *descrever*. Na descrição de metadados é indicado o uso de instrumentos que auxiliem a padronização internacional. Recomenda-se fortemente o emprego de vocabulários controlados, *thesaurus*, ontologias, especificações de tipologias, formato, coberturas (espaciais e temporais), extensão, idioma e palavras-chave para o enriquecimento, padronização e conformidade dos dados científicos para a visualização, análise, uso e reuso futuros;
- i) A utilização de um serviço de indexação para coletar os metadados com seus identificadores, como, por exemplo, o *CrossRef* e o *DataCite*, faz parte do conjunto de boas práticas recomendáveis para o reconhecimento e citação autorais, visibilidade e análise futuras;
- j) A adoção de ferramentas automáticas que gerem os *codes* automáticos ou repositórios de dados é necessária para que todo o processo de investigação seja preservado, interoperável, acessível, visível, atribuível e reutilizável a longo prazo.

Isto posto, o padrão resume que a infraestrutura técnica é descritiva e constitui-se no planejamento e adoção de recursos de descrição, identificação, indexação e citação. As tecnologias abertas, tais como ferramentas, recursos, padrões de metadados, protocolos de interoperabilidade, *software*, simuladores, conversores e repositórios constituem a infraestrutura tecnológica necessária em todo o ciclo de vida. As recomendações de modelos, tecnologias, códigos abertos, treinamentos, orientações de boas práticas e suportes (técnico e tecnológico) sejam abertos, gratuitos e disponíveis pela *internet*.

A ciberinfraestrutura tecnológica que dá sustentabilidade a todos os estágios é extensa e complexa, portanto as exemplificações acima foram pontuais para atender questões legais que subsidiam a recompensa autoral nesse contexto. Assim, o padrão morfológico-técnico é uma parte do modelo conceitual que corresponde à operacionalização técnica, efetivação tecnológica e procedimentos descritivos que estabelecem a recompensa autoral, tanto pessoal como institucional, visando a atribuição de pesquisadores, colaboradores e parceiros.

10.4 PADRÃO ÉTICO-LEGAL-CULTURAL

O padrão ético-legal-cultural constitui a **quarta parte** do modelo. Para compor este padrão os axiomas elaborados foram mais extensos por apresentarem aspectos subjetivos e intrínsecos às práticas da *e-Science*. O padrão também está constituído por duas dimensões de pesquisa e tem uma interface com os aspectos culturais.

A dimensão ética se volta para o comportamento e conduta dos envolvidos no processo investigativo. O comportamento ético, a honestidade acadêmica e a postura investigativa são importantes e relevantes para o processo de compartilhamento e colaboração científica. Nas práticas da *e-Science*, os critérios de qualidade, fidedignidade, veracidade, autenticidade e sigilo se acentuam em virtude da pesquisa ser aberta e colaborativa, e por conseguinte, fica mais exposta para a validação de pares.

A dimensão legal contempla a autoria com requisitos de originalidade relativa, novidade temática e não exclusividade contrapondo aos requisitos de exclusividade, individualidade e propriedade que estão intrínsecos aos direitos autorais vigentes. A dimensão cultural entra como uma interface complementar que influencia diretamente os comportamentos e as condutas dos envolvidos. Essas influências afetam as dinâmicas locais e os comportamentos pessoais e ações institucionais. Os projetos de gestão de dados na maioria das vezes são conduzidos por parcerias colaborativas que envolvem *stakeholders* e diversas instituições em diferentes países. Portanto, pensar nesse padrão ético-legal-cultural é ampliar os horizontes em uma abordagem de autoria colaborativa que conduziram a junção das dimensões éticas e legais com a interface cultural através de cinco axiomas:

10.4.1 Primeiro Axioma abrange a permuta dos requisitos autorais vigentes nos Sistemas de Propriedade Intelectual para o modelo emergente da *e-Science*

Os sistemas de direitos de propriedade intelectual vigentes preconizam requisitos de originalidade, novidade, exclusividade e individualidade dos seus criadores estão calcados em um paradigma proprietário econômico e na exploração comercial por terceiros. Na contramão desses sistemas, emerge o paradigma colaborativo com características que enfatizam a colaboração, o compartilhamento, a coletividade e o bem comum. Na essência da ciência aberta essas características foram empreendidas por movimentos como acesso aberto, ciência do *commons*, economia criativa e agora, nas práticas da comunicação científica contemporânea, com o modelo emergente da *e-Science*. Os requisitos que preconizam a garantia da proteção autoral intrínsecos nos direitos autorais vigentes, foram ressignificados para o contexto a partir de três requisitos norteadores dentro deste padrão: proteção da propriedade intelectual, conformidade legal e responsabilidade formal. Tais requisitos foram fundamentados nas diretrizes internacionais da OECD.

A Comissão Europeia defende que o acesso aberto não interfere na decisão de explorar comercialmente os resultados da investigação, mas devem ser previstas as questões legais dentro da política em observância às leis nacionais e acordos internacionais. Legislações específicas - direito autoral, direito de patentes, direito de privacidade (dados pessoais, sigilosos ou sensíveis) - deverão ser consideradas em uma política nacional. Tais questões quando não acordadas previamente afetam diretamente o compartilhamento, a acessibilidade e o embargo aos dados científicos para o uso e reuso.

10.4.1.1 Proteção da Propriedade Intelectual

A proteção da propriedade intelectual está relacionada ao mínimo de segurança autoral que as criações precisam garantir a originalidade de sua primariedade. Alguns requisitos são considerados nessa primariedade:

- a) O valor estético associado à utilidade prática dos dados científicos como um componente de garantia da originalidade primária e sua derivação a partir do reuso;
- b) O aporte trazido pelo autor está diretamente ligado à criação como resultado de uma atividade intelectual própria acrescida e ressignificada com a geração de novos conteúdos. A *práxis* por meio do reuso de dados científicos pressupõe uma criação derivada. Assim, o aporte intelectual do criador originário é garantido e aos pesquisadores e colaboradores envolvidos na criação derivada, a garantia é

atribuída pela sua responsabilidade proporcional no processo intelectual de pesquisa;

- c) A forma é um componente revestido de subjetividade (*corpus mysticum*), mas dotado de carácter estético intrínseco. Por si só já expressa a subjetividade do autor e, ao mesmo tempo, se materializa exteriormente. No reuso de dados científicos, a forma original já é digital e legível por máquinas. O que difere é o valor subjetivo intrínseco à forma derivada dos dados científicos. A derivação contém elementos de ressignificação, recriação e transformação relacionados ao novo domínio específico da pesquisa a partir do reuso da criação original;
- d) A inserção do conteúdo em um formato é a materialização em um determinado suporte (denominado de *corpus mechanicum*). Os dados científicos são originários ou derivados de pesquisas (observacionais, experimentais, simulacionais) com características intrínsecas e em suporte que assume a forma digital. Abre-se uma exceção para a materialização de ideias na *e-Science*. Em áreas de alta complexidade e de competitividade, já é possível a materialização de ideias preliminares que devem ser exteriorizadas em um formato digital;
- e) Os requisitos de originalidade e novidade são dois requisitos importantes para a garantia autoral. Tais requisitos são elementos de criação única e inconfundível e revestem-se de traços e caracteres próprios. Na *e-Science*, a **originalidade** está revestida pela natureza e contexto do domínio específico em que os dados foram coletados, gerados, reusados, ressignificados e reproduzidos. Portanto, a originalidade assume um carácter relativo, uma vez que são inevitáveis o aproveitamento e o reuso de dados científicos a partir da criação original. A **novidade** consiste na geração de novos sentidos, significados e aplicações, pelos quais, os dados científicos foram reusados e reproduzidos. A partir do momento em que os dados científicos são reusados em um novo contexto temático, diferente daquele que os originaram primariamente, novas abordagens, hipóteses e métodos são geradas, e, por conseguinte, direcionamentos investigativos e novidades temáticas são emergidas.

Dessa forma, o requisito da originalidade está ligado à criação primígena, que é a autoria originária (primária) e com o reuso configurando a autoria derivada (secundária). A novidade temática é associada ao reuso dos dados científicos. Isto posto, os requisitos foram utilizados para a nova acepção direcionada para a autoralidade colaborativa na *e-Science*.

10.4.1.2 Conformidade Legal

A conformidade legal está associada à observância de acordos sobre o acesso aos dados científicos, respeitando os interesses legais e direitos legítimos de todas as partes interessadas. Faz mister a previsão de restrições e limitações para acesso, uso e reuso, quando necessários e prevista nas políticas nacionais e institucionais. Alguns pontos devem ser observados quando o conteúdo dos dados abrangem:

- a) A segurança nacional relaciona-se aos dados provenientes dos sistemas de inteligência, atividades militares ou tomada de decisão política e, portanto, sujeitas a restrição no acesso. O acesso (aberto, parcial, restrito ou embargado) deve ser previsto preliminarmente, considerando aspectos políticos, institucionais, éticos e legislações específicas de cada país. É recomendado que a restrição seja informada em documentos instituídos em âmbito governamental (federal e nacional) ou institucional devido aos conteúdos serem de alta complexidade e sigilosidade. O *status* do acesso deve constar na descrição dos metadados, como os períodos de embargo e tipos de licenças;
- b) A privacidade, confidencialidade e anonimização são atributos relacionados aos dados privativos e pessoais. O direito à privacidade possui legislação específica e pode também ser apoiado por código de conduta ética. Procedimentos de anonimização podem garantir um nível satisfatório de confidencialidade, preservando informações pessoais. Tais procedimentos geralmente são conduzidos no âmbito da instituição e adotados por pesquisadores no momento da intervenção de pesquisa. Os dados relativos à privacidade geralmente são abrigados por leis específicas nacionais, com documentos e autorizações que respaldem previamente os acordos entre os pesquisadores, sujeitos e aprovados em comissões éticas institucionais. Recomenda-se que as restrições e prazos de embargo sejam informados nos metadados e regidos por documentações específicas;
- c) Os segredos comerciais e os direitos de propriedade intelectual relacionam-se aos dados sobre empresas privadas ou instituições governamentais que, no todo ou em partes, possuem informações confidenciais não acessíveis. Os segredos comerciais, na grande maioria das vezes, são advindos de descobertas ou invenções e direcionados ao patenteamento. Neste caso, é recomendado que dados científicos ou governamentais obedeçam aos prazos legais em consonância com a legislação específica em cada país, instrumentos normativos previstos na parceria

(público-privada) e as informações que estabeleçam o período de embargo e o *status* de acesso previstos no início do projeto de pesquisa;

- d) A proteção de espécies raras, ameaçadas ou em perigo são casos específicos que por razões legítimas restringem o acesso aos dados científicos. Tal restrição pode ser parcial ou total. Essa restrição geralmente está direcionada aos dados científicos de recursos naturais (fauna e flora). As orientações legais e normativas da legislação específica e da instituição devem ser consideradas para a abertura desses dados de pesquisa. A previsão das restrições de acesso é feita na descrição dos metadados. A restrição acontece quando parte dos dados ou conjunto de dados, está com acesso (parcial, restrito ou embargado);
- e) O processo legal está direcionado aos dados científicos sob litígio em ações legais (*sub judice*). Assim a restrição está associada as pesquisas que resultam em dados que estão vinculados às investigações que envolvem patentes (indústria farmacêutica, por exemplo) ou descobertas que trazem benefício imediato a uma determinada comunidade ou sociedade. Nesses casos, a legislação específica, as diretrizes que envolvem as parcerias e os códigos de conduta profissional e regulamentações são adotados no âmbito governamental ou institucional.

As recomendações são importantes para que dados científicos que estão restritos, sejam considerados como exceções. O prazo de término do embargo deve prever o término das restrições, para que dados científicos estejam abertos e acessíveis em algum momento.

10.4.1.3 Responsabilidade Formal

A responsabilidade formal trata das atribuições legais e éticas dos envolvidos no projeto. As práticas científicas podem gerar conflitos relacionados entre a autoria e o compartilhamento e o acesso aberto. Nesses casos, as diretrizes político-normativas determinam regras, regulamentos e instrumentos para ações, responsabilidades e comportamentos dos envolvidos. Algumas recomendações devem ser atendidas de acordo com os seguintes procedimentos:

- a) O estabelecimento de acordos institucionais, condições para acesso, uso e reuso antes de iniciar o gerenciamento dos dados científicos evitará conflitos e impedimentos legais. Muitos dos problemas restritivos para o compartilhamento dos dados são ocasionados pela ausência de acordos formais estabelecidos previamente;

- b) A normatização e publicização de acordos estabelecidos para garantir o fluxo de procedimentos padronizados em todos os estágios do ciclo de vida. As iniciativas internacionais disponibilizam documentações, instruções normativas, modelos, padrões e ferramentas abertas que auxiliam na condução das práticas, além de assegurar uma conformidade e normatização na condução das mesmas;
- c) As orientações e as ‘boas práticas’ institucionais e de educação continuada são fortemente recomendadas em diretrizes internacionais;
- d) O incentivo de políticas nacionais que motivem a comunidade científica ao trabalho compartilhado com base em preceitos éticos e colaborativos é um requisito primordial para o sucesso de um projeto de pesquisa colaborativa e aberta.

Este axioma possui características que dão sustentabilidade a dimensão ético-legal sob o prisma de uma pesquisa científica heterogênea e antagônica. Busca atender aos sistemas de PI vigentes, ao mesmo tempo que, incentiva ações e responsabilidades conduzidas por uma ambiência compartilhada e colaborativa na ciência. Reflete um desafio político, mas, sobretudo, uma ruptura no fazer científico vigente das práticas existentes regidos por um paradigma proprietário econômico para um espaço intelectual colaborativo e de múltiplas colaborações.

10.4.2 Segundo Axioma direciona a nova acepção de autoria originária e colaborativa

Esse axioma se volta para a nova acepção da autoria originária baseada na colaboração científica a partir de duas abordagens:

A primeira abordagem dispõe sobre o prestígio da **autoria originária** conferindo o *status* de **pesquisador-autor** ao criador, gerador ou responsável pela criação de dados primários. No bojo do compartilhamento e colaboração é enfatizada a troca de saberes com vários autores, colaboradores e instituições, além de propósitos e contextos distintos, configurando a não exclusividade da propriedade, mas o reconhecimento autoral pela criação. A garantia da paternidade é realizada pela atribuição e citação ao pesquisador-autor dos dados primários originais. Assim, o pesquisador-autor é o **autor originário** (primário).

A segunda abordagem pressupõe que dados científicos, informações e resultados de pesquisa são usados, reusados e ressignificados. Assim, o reuso de dados científicos gera novas criações intelectuais a partir de processos de transformação, incorporação, ressignificação, complementação, redução ou adição de dados primários (originais). As

derivações dos feitos científicos por outro pesquisador, colaborador ou grupo de pesquisa configuram-se em uma **autoria colaborativa**. Na acepção moderna, entendeu-se que a abordagem conceitual se adapta à criação derivada a partir do reuso dos dados científicos.

Esse axioma é norteador para a compreensão de que na *e-Science* as novas dinâmicas de colaboração científica geram mudanças no *status* da autoria tradicional. Preconiza a acepção de uma autoria colaborativa constituída por diferentes colaboradores.

10.4.3 Terceiro Axioma baseado no paradigma subjacente do *commons* e nos dados científicos como bens intelectuais

O axioma é direcionado a modalidade de autoria colaborativa independente do domínio específico. Tal autoria é constituída pela colaboração científica, divisão de tarefas, parcerias entre pares e instituições e espaços interdisciplinares, que:

- a) Admite contribuições, focos, naturezas, graus e contextos científicos diversos;
- b) Abrange diferentes profissionais, tais como: pesquisadores, autores, curadores, bibliotecários, cientistas de dados, indexadores, especialistas tecnológicos e outros profissionais que são colaboradores em diferentes estágios do ciclo;
- c) A figura individualista e proprietária do autor é redirecionada para a sua condição de autoria colaborativa. A idealização, criação e materialização de novos significados e conhecimentos consideram a responsabilidade epistemológica, ética e intelectual no processo investigativo;
- d) A criação colaborativa é representada por *quilts* intelectuais de verdadeira criação e responsabilização, resultando em um trabalho divisível e complementar. O reconhecimento, mérito e recompensa autoral são realizados pela atribuição e citação proporcionais à parte responsável;
- e) A autoria colaborativa é distinguida pela independência, responsabilização, colaboratividade, coletividade e atribuição. Ela exclui os requisitos da exclusividade e individualidade.

Esse axioma reforça o pressuposto que na cultura colaborativa e do compartilhamento científico, o autor é um interlocutor e contribuidor em um espaço coletivo, colaborativo, pluralista e interdisciplinar de criação.

10.4.4 Quarto axioma com ênfase na recompensa autoral

A ênfase do novo paradigma científico direciona a recompensa autoral por meio do reconhecimento, responsabilização, atribuição e citação. Esse axioma traz a reflexão de como a recompensa autoral é conduzida na *e-Science*. As práticas vigentes de recompensa e reconhecimento autoral são reconfiguradas a partir:

- a) Da forma de recompensa autoral compartilhada entre diferentes colaboradores, reconhecida e atribuída imediatamente na devida proporcionalidade. Realidades, contextos e culturas distintas estão integrados no mesmo propósito investigativo;
- b) Do reconhecimento realizado pela atribuição e citação de acordo com a sua responsabilidade. É uma característica de singularidade em um ambiente colaborativo sem a predominância da posse e propriedade individualizada;
- c) Da atribuição que é feita à parte responsável. A abrangência da atribuição se estende a todos os colaboradores responsáveis, não se limita à pessoa física, podendo ser designada a um pesquisador, grupo de pesquisa, instituições científicas, universidades, agências de fomento, repositório de dados, bibliotecas ou a outra parte responsável que contribuiu no processo investigativo. É possível também em domínios específicos, atribuição e citação serem realizadas a partir de ideias preliminares, prevendo o período de embargo;
- d) Da citação que é creditada a cada colaborador a partir de sua participação e responsabilidade pela parte que lhe cabe. A identificação apropriada e completa permite que dados compartilhados e reusados estejam acessíveis, citados, interoperáveis e abertos a longo prazo;
- e) Dos créditos que são atribuídos quando a citação for completa. Utilizam-se identificadores persistentes, aplicação de licenças, descrição de metadados, interoperabilidade e sistemas de recompensa. O crédito deve ser expreso, explícito e direcionado ao responsável pela colaboração como garantia autoral;
- f) Da citação que prevê também a recompensa autoral e os méritos por meio de mecanismos de medição do fator de impacto da qualidade de publicações científicas em nível mundial e local. O crédito pessoal para pesquisadores e colaboradores, como progressão de carreira ou captação de recursos de financiamento, é medido pela citação. Prêmios, méritos e honorarias podem ser concedidos a partir de resultados de pesquisa de grande valia, mensurados por sistemas de recompensa autoral.

Esse axioma estabeleceu a recompensa autoral a partir da responsabilidade, atribuição e citação de forma reconfigurada para a *e-Science*. A atribuição e citação são modalidades

existentes no contexto da comunicação científica, mas foram reconfigurados pela colaboração, responsabilidade e proporcionalidade no bojo da dinâmica contemporânea.

10.4.5 Quinto axioma com ênfase na responsabilidade da autoria colaborativa

A responsabilidade da autoria colaborativa pressupõe comportamento e conduta ética. Mais uma vez leva-se em consideração a honestidade acadêmica e os valores morais necessários ao comprometimento ético e moral dos envolvidos. O desvio de conduta é percebido através da falsa atribuição de autorias, plágio (fraude e violação), apropriação indébita de conteúdos alheios e expropriação econômica por terceiros sem autorização. Nas práticas estejam em consonância com recomendações éticas previstas em códigos de conduta, com:

- a) A adoção de processos que estimulem exatidão, integridade, confidencialidade e clareza nas pesquisas;
- b) A implantação de sistemas de autoria e contribuição que promovam as boas práticas;
- c) O estímulo ao uso dos códigos de conduta que desencorajem a autoria fantasma, convidada ou coagida;
- d) O incentivo à originalidade das submissões e adoção de sistemas de detecção de plágio;
- e) A elaboração de instrumentos normativos para as declarações de consentimento e impedimentos legais;

Quanto ao comportamento ético do pesquisador ou colaborador a rigorosa conduta é baseada em princípios morais, tais quais:

- a) Requer o consentimento, anonimização e regulação de acesso dos dados científicos antes do uso ou reuso;
- b) Questões de privacidade relacionadas aos dados científicos devem ter a autoria atribuída corretamente, observando-se quanto houver, o sigilo, a fragilidade e a fidedignidade dos conteúdos;
- c) Deveres e responsabilidades quanto à confiabilidade dos dados pessoais, dados sensíveis e exceções legais são observados. É necessário o consentimento prévio dos participantes antes do compartilhamento para a preservação da confidencialidade assegurada;

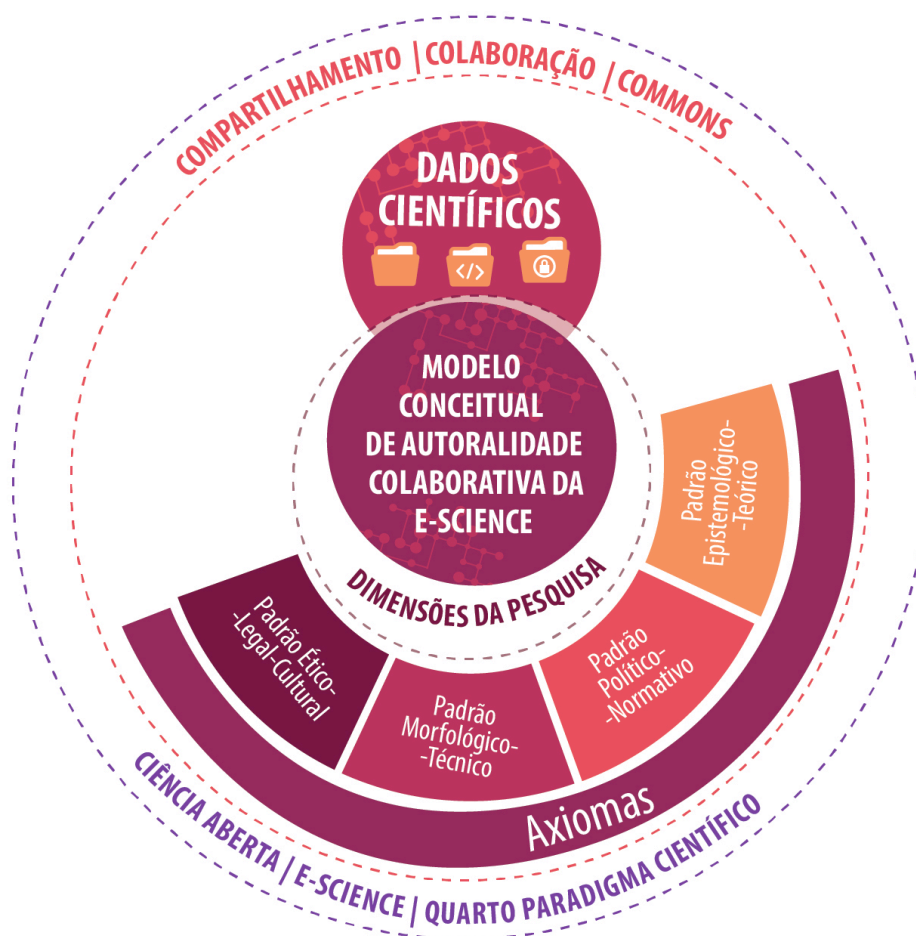
- d) O uso e reuso de dados pessoais só são acessíveis a partir da garantia de sigilo, respeito ao indivíduo vivo, lealdade, licitude e justiça. Recomenda-se proceder com a identificação pessoal e qualquer expressão de opinião emitida mediante autorização prévia do envolvido. A divulgação só poderá ser realizada após consentimento expresso e procedimentos legais cumpridos. A manipulação dos dados pessoais são para fins estritamente específicos, adequados e pertinentes ao domínio da pesquisa. Para o reuso dos dados pessoais requer nova autorização para uma finalidade específica de pesquisa;
- e) Os dados sensíveis possuem conteúdos relacionados a cor, raça, etnia, opiniões políticas, crenças religiosas, filiação sindical, saúde física ou mental, vida sexual, processos ou condenações penais. Recomenda-se que para o uso e reuso em pesquisas haja previamente o consentimento expresso por escrito ou digital. Quando se tratar de conteúdos relacionados à saúde, ao sexo e processo judicial, seja emitido parecer por profissional habilitado em obediência aos critérios de confidencialidade, anonimização e sigilosidade.

Quanto aos aspectos culturais, muitas vezes, arraigados aos processos vigentes, representados, em certa medida, pela desonestidade ética, plágio, cultura do produtivismo intelectual desenfreado são levados em consideração nos códigos de conduta. Os costumes e atitudes comportamentais mudam de acordo com os contextos, sejam institucionais ou nacionais. Entretanto, o código de conduta busca minimizar as discrepâncias culturais e direcionar para a conformidade no comportamento dos envolvidos.

O redesenho em favor de uma cultura colaborativa, presente na essência da ciência aberta requer um pensar diferenciado e apregoa um novo empoderamento de diversos sujeitos, constituídos por pesquisadores-autores, colaboradores, instituições e empresas que configuram diferentes responsabilidades, e por conseguinte, atribuições e citações distintas que representam a recompensa autoral contemporânea. Esse axioma refletiu a dimensão ético-legal-cultural presente em todos os estágios do ciclo de vida da pesquisa aberta e colaborativa.

Ademais, considerando que a ciência aberta preconiza a diversidade de contextos, culturas e propósitos enseja um novo pensar para a autorialidade no cerne das práticas científicas da *e-Science*. A representação visual através do modelo conceitual, é portanto, a materialização da nova acepção de autorialidade colaborativa para os dados científicos da *e-Science*. O Modelo Conceitual de Autorialidade Colaborativa é representado na Figura 28.

Figura 28 – Modelo conceitual de autoria colaborativa da *e-Science*



Fonte: Elaborado pela autora (2016).

O modelo conceitual é apoiado pelos pressupostos do compartilhamento, colaboração e do paradigma do *commons* que dão sustentabilidade a nova compreensão dos direitos de propriedade intelectual vigentes. Entendeu-se que a ciência aberta requer um olhar contemporâneo com diferentes dimensões que convergem entre si, e para tanto não pode ser conduzida sem um olhar multidimensional. Os padrões constitutivos do modelo conceitual teve a finalidade de atender aos objetivos geral e específicos. A questão de pesquisa foi respondida através dos padrões constitutivos do modelo. Os axiomas que constituem cada padrão são direcionamentos como requisitos mínimos que norteiam a *práxis* da *e-Science* com vistas a proporcionar a recompensa autoral de todos os envolvidos. O argumento de tese proposto foi contemplado através dos direcionamentos norteadores em cada padrão que estão em conformidade internacional e reforçam a garantia de uma recompensa autoral adequada.

Assim, o capítulo que se finda trouxe a visão multidisciplinar e multidimensional pautada na essência da ciência aberta, com um recorte para a ressignificação dos direitos de propriedade intelectual vigentes e as novas acepções da autoria na *e-Science*.

11 CONCLUSÃO

Este capítulo caminha para as reflexões finais da pesquisa. Muito embora a conclusão final é uma utopia, pois a busca por novos *insights* é constante e no momento são realizadas algumas inferências e considerações parciais acerca das temáticas abordadas nesta tese. Visto que o estudo abordou assuntos complexos e emergentes no cenário mundial. Inicialmente, ressalta-se a dificuldade da pesquisadora em convergir em um único estudo, disciplinas tão distintas e específicas, como a Ciência da Informação e as Ciências Jurídicas e suas interfaces com a Tecnologia e a Economia, mas que ao final foi possível confirmar o argumento de tese e atender aos objetivos propostos. A convergência foi necessária para a compreensão das temáticas emergentes com as práticas vigentes da ciência moderna.

Essa compreensão corroborou para a confirmação do argumento de tese e objetivos propostos e levou a **primeira constatação**. Sob o ponto de vista formal, as tendências contemporâneas no escopo científico são idealizadas pelo surgimento da ciência aberta e de novas dinâmicas que direcionam para dados de pesquisa como produto primário de investigação científica. O modelo tradicional de comunicação vigente é predominante, pois ainda utiliza-se dos canais formais de publicação científica, procedimentos de avaliação por pares e publicização dos resultados finais do processo investigativo no contexto científico. Porém, observou-se que os movimentos preconizados pelo *open access*, *commons*, *open science*, *data science* e *open data* estão em crescente evolução.

Concluiu-se que ciência aberta é o escopo formal constituído por políticas, diretrizes, normativas e direcionamentos que dão sustentabilidade para que a investigação no contexto científico orientada para o aproveitamento de dados de pesquisa que dão sustentabilidade as dinâmicas práticas conduzidas pelo modelo emergente da *e-Science*. A ciência aberta propicia as condições para que dados de pesquisa sejam acessíveis, reutilizáveis e reprodutíveis com diferentes propósitos e para diversos contextos. Daí a constatação que a ciência aberta é um conceito guarda-chuva amplo e multidimensional, pois a sua amplitude está além dos ambientes convencionais de comunicação científica. Muito embora não seja uma realidade mundial, pois concentra iniciativas mais consolidadas em alguns países como, Austrália, Estados Unidos, Holanda, Inglaterra e Reino Unido que concentram a maior parte das práticas internacionais voltadas para a ciência aberta, foi constatado que é um movimento em expansão e possui muitas iniciativas para o contexto científico. Dessa maneira, o estudo fez um recorte dimensional apenas do escopo científico e abordou os dados científicos como *constructo* da tese.

Do ponto de vista pragmático o modelo emergente da *e-Science* com base no quarto paradigma científico conduz a *práxis* científica. O quarto paradigma científico traz a premissa básica de que a pesquisa científica tradicional gera uma enorme quantidade de dados que não são aproveitados no decorrer do processo investigativo tradicional. Assim o aproveitamento dos dados de pesquisa como produtos primários na *e-Science* são conduzidos para o aproveitamento e gerenciamento durante todo o processo de investigação científica contemporâneo. O modelo emergente da *e-Science* traz a *práxis* do processo de gerenciamento dos dados científicos para acesso, uso e reuso a longo prazo. Essa mudança altera substancialmente o modo de fazer e publicar a pesquisa científica no modelo tradicional vigente. O uso dos dados científicos como produto primário é o cerne do quarto paradigma que conduz a *práxis* da *e-Science*.

Nesse sentido, a *e-Science* foi compreendida como a *práxis* que envolve aspectos metodológicos, técnicos e tecnológicos conduzidos pelo modelo de ciclo de vida dos dados, com a manutenção dos requisitos formais (epistemológicos, metodológicos e teóricos) da pesquisa científica tradicional. A *práxis* está assentada nos pressupostos da colaboração, compartilhamento e coletividade como fios condutores para o novo fazer científico. Admite uma prática científica de tal forma que outros possam colaborar e contribuir sempre que os dados científicos estejam disponíveis e acessíveis gratuitamente, em termos que permitam o uso, reuso, redistribuição e reprodução da pesquisa.

Apregoa ainda um papel político que conduz as práticas compartilhadas e colaborativas como relevantes para a difusão e acesso aberto à informação, cultura e conhecimento, além dos dados subjacentes da pesquisa. As mudanças no *status quo* da pesquisa tradicional para a ciência aberta exigem mudanças que exigirá conflitos de interesses dos atores envolvidos e reformulações de políticas e legislações específicas em cada país.

A lógica capitalista presente, tanto no contexto científico como nos sistemas de propriedade intelectual vigentes, representa paradoxalmente um impeditivo às atualizações institucionais e legais que dão suporte à *práxis* científica contemporânea, ao mesmo tempo que estimula o compartilhamento e colaboração como estratégia de barateamento e agilização da produção intelectual, científica e tecnológica aberta e pública. Todavia, é um processo complexo e não representa a realidade global, pois envolve aspectos políticos, econômicos, legais e éticos que precisam ser discutidos e reformulados. Essas reflexões conduziram à segunda constatação.

A **segunda constatação** refletiu os direitos de propriedade intelectual vigentes e as novas formas de visualizar o bem intelectual na perspectiva da Teoria do *Commons* e da

Economia Criativa. Essa reflexão foi possível de ser realizada a partir da convergência temática entre o Direito e a interface com a Economia. A ideia subjacente nessa convergência buscou entender que o *commons paradigm* conduz o bem intelectual para uma dinâmica aberta, pública e comum. O *commons paradigm* preconiza que os espaços de criação cultural e intelectual sejam colaborativos, compartilhados e comuns.

Analogamente, a construção foi feita baseada na ideia de que os dados científicos também são bens intelectuais, geradores de informação e conhecimento, e por conseguintes, são bens comuns partindo da noção da Teoria do *Commons* e da Economia Criativa. Na Economia o conhecimento é um bem intelectual e classificado como um bem não rival. Nesse sentido, se refletiu que sendo os dados científicos um bem intelectual é por natureza um bem não rival. Sendo um bem não rival, o acesso simultâneo entre diferentes usuários é possível, e portanto, é de natureza comum e pública. As ressalvas devem ser consideradas quanto ao acesso aos dados científicos devido as peculiaridades e especificidades que cada área possui. Para tanto, diretrizes e procedimentos normativos, legais e técnicos darão sustentabilidade a essa constatação. Ademais, as ressalvas pressupõem que a pesquisa financiada com fundos públicos, deve conduzir para que os dados científicos estejam acessíveis de forma pública e gratuita, considerando-se previamente, legislações específicas, restrições, exceções e embargos.

A reflexão trouxe à tona a compreensão de que os sistemas de direito de propriedade intelectual vigentes estão arraigados em um paradigma proprietário econômico que traz em seu bojo os requisitos de exclusividade, individualidade e propriedade. Estimula a exploração comercial dos bens intelectuais (literários, científicos e culturais) mediante o reconhecimento da propriedade e dos direitos autorais. Dar sustentabilidade a figura do autor e do pesquisador (contexto científico) como proprietários dos direitos autorais ressaltando o requisito de exclusividade e individualidade. Essa reflexão conduziu à terceira constatação da tese.

A **terceira constatação** foi conduzida para a nova acepção de autoria na *e-Science*. A figura do autor passa a ter uma nova acepção e condição de sua função na *práxis* da *e-Science*. A condição da figura do pesquisador-autor assume um papel de autoria colaborativa em um ambiente compartilhado, colaborativo e comum. Tal constatação está diretamente vinculada ao argumento de tese proposto que foi alicerçado por três premissas: a autoralidade na *e-Science* é constituída pelo trabalho colaborativo, compartilhado e de múltiplos colaboradores; a descentralização da figura do pesquisador-autor individual e exclusivo baseada nos direitos de propriedade intelectual vigentes para uma nova acepção de autoria colaborativa; e, a

autoria se alarga para a acepção de autoralidade colaborativa garantida pela recompensa autoral.

As constatações foram conduzidas para confirmar o argumento de tese: na *e-Science* a autoralidade é colaborativa e promovida pela recompensa autoral, responder a questão de pesquisa: a recompensa autoral é promovida na *e-Science* considerando que a autoralidade é colaborativa? e o objetivo geral: elaborar padrões que promovam a recompensa autoral na *e-Science*. A partir das constatações expostas as reflexões seguintes corroboram e reforçam o argumento de tese e afirma a hipótese arguida.

A **primeira reflexão** trouxe as considerações sobre as práticas científicas baseadas em dados científicos gerados em ambientes compartilhados e colaboração científica. Esses ambientes além de promover crescimento, aprimoramento e inovação científica, proporcionam parcerias institucionais e estimulam o desenvolvimento econômico, científico, tecnológico, cultural e social. Constatou-se também que pesquisas geradas em ambientes colaborativos nos moldes da *e-Science* são constituídos por múltiplos colaboradores, pluralidades de contextos e diferentes instituições. A multiplicidade de colaboradores conduziu ao entendimento nesta tese de que a autoria é colaborativa na *e-Science*. A existência dos requisitos de originalidade e exclusividade relacionados à proteção autoral dos sistemas vigentes são reconfigurados para a originalidade relativa e novidade temática.

A **segunda reflexão** focou na descentralização da figura individual, proprietária e exclusivista centrada no pesquisador-autor responsável pela criação primígena para uma criação derivada. A criação derivada foi entendida como um processo de transformação dos dados científicos primários (originários). Essa transformação é entendida como o aprimoramento, acréscimo e ressignificado de dados científicos primários a partir do uso e reuso em um ambiente compartilhado e colaborativo. Dessa forma, a criação derivada foi compreendida como a transformação a partir do reuso dos dados científicos. O reuso dos dados científicos direcionam para novas perspectivas, hipóteses e resultados de uma nova pesquisa. Assim, a criação derivada constituiu a nova acepção de autoria colaborativa. A transformação da criação primígena leva a uma criação derivada criada por múltiplas colaborações e a partir do reuso de dados científicos, pesquisas e resultados preliminares.

No que diz respeito a criação primígena, a autoria originária deve ser preservada e garantida pela recompensa autoral. A recompensa autoral se baseia na responsabilização, atribuição e citação do pesquisador-autor e colaboradores responsáveis e criadores dos dados científicos primários. O redesenho da cultura proprietária assentada nos sistemas de propriedade intelectual vigentes, na perspectiva da ciência aberta, abre espaço para a cultura

comum e colaborativa. É enfatizado o empoderamento do autor colaborativo alicerçado pelas múltiplas colaborações, responsabilidades proporcionais e conduta ética. Essas reflexões encaminham a constatação da terceira premissa.

A **terceira reflexão** constatou que o entendimento da noção de autoria foi alargada para o termo autoralidade. A autoralidade na *e-Science* é colaborativa e garantida pela recompensa autoral. A atribuição e citação são creditadas aos diferentes envolvidos compondo as múltiplas colaborações e a pluralidade de autorias. Compõe essa pluralidade de autorias, pesquisadores, colaboradores, curadores, grupo de pesquisas, instituições, agências de fomento e outros profissionais envolvidos no processo de investigação. Cada envolvido participa com sua colaboração, e por consequência, são responsáveis proporcionais do processo de pesquisa. Os créditos são fixados na proporcionalidade a cada parte responsável. Essa afirmativa é uma característica que configura a autoralidade colaborativa na *e-Science*. Para a atribuição e créditos que constituem parte da recompensa autoral, os aspectos relacionados à infraestrutura são: identificadores persistentes, descrição de metadados, licenças públicas e sistemas de autoria e citação que garantem a autoralidade.

A essa reflexão acrescentam-se os aspectos ético-legais-culturais direcionados para a responsabilização dos envolvidos. A responsabilidade e a conduta ética são realçadas por códigos de conduta que orientam o comportamento ético, honestidade científica, autoria colaborativa, fidedignidade e veracidade na condução da pesquisa. Os aspectos legais estão direcionados à responsabilidade legal na observância preliminar dos aspectos relacionados à privacidade, sigilosidade, confiabilidade, fidedignidade e anonimização, principalmente quando se tratar de dados pessoais, sensíveis e secretos. As condições de uso e restrições devem ser direcionadas por prazos de acesso e embargos. Porém, prevalece o princípio primordial de que dados científicos são acessíveis, públicos e gratuitos é imperioso.

Acerca do aspecto cultural concluiu-se que as dinâmicas culturais interferem nas práticas científicas, principalmente quando condutas éticas sofrem distorções advindas do comportamento pessoal. A ausência de códigos de conduta e instrumentos normativos que norteiem o comportamento moral e as ações no processo científico geram distorções e práticas indevidas. Reforça-se a percepção de que no contexto científico é muito presente as más condutas na ciência, notadamente nos casos de plágios, violação e falsas autorias.

Ademais, as condutas arraigadas pelo paradigma proprietário econômico dos sistemas de direitos de propriedade intelectual, levam ao reforço ou restrição do interesse e a motivação a favor do compartilhamento e da colaboração nos processos investigativos da ciência.

As reflexões expostas estão em consonância com os objetivos específicos elaborados. Os padrões constitutivos do modelo conceitual apresentam direcionamentos mínimos e norteiam através de axiomas a promoção da autoralidade colaborativa nas práticas científicas da *e-Science*. A questão de pesquisa formulada foi respondida por meio da hipótese que se confirmou a partir das três reflexões e representada no modelo conceitual. A ideia precípua do modelo conceitual teve o propósito de representar as acepções contemporâneas da autoralidade em um novo pensar e fazer científico conforme proposto no argumento de tese e conduzidas pelos objetivos geral e específicos.

Por fim, algumas considerações finais foram observadas. No âmbito internacional a ciência aberta e a *e-Science* são conduzidas pelas políticas baseadas em diretrizes e regulamentações normativas e ainda prevalece as regulamentações dos direitos de propriedade intelectual vigentes. As agências de fomento, universidades e instituições de pesquisa, nos últimos anos emitiram declarações e políticas ressaltando a necessidade de arquivamento de dados provenientes de pesquisa. Entendeu-se que as diretrizes internacionais político-normativas estão sendo conduzidas na ideia de *soft law*, muito embora práticas conduzem para tornar os dados científicos acessíveis e públicos. Prevalece a essência das três leis de Eaves presentes nos pressupostos de que os dados precisam ser *encontrados, acessíveis, úteis, interoperáveis, legalizados, abertos, compartilhados e reproduzidos*.

Quanto ao compartilhamento e gerenciamento de dados científicos já é uma realidade consolidada, mas ainda em construção e desafiante, porque depende de outras variáveis referentes às práticas culturais da comunidade científica com peculiaridades de cada área do conhecimento. Cada realidade possui suas infraestruturas organicamente sustentáveis e integradas. Apesar de ser um processo ainda lento, complexo e fragmentado a pesquisa moderna está caminhando para o acesso aberto, gratuito e público baseada nos dados científicos. Entretanto, qualquer mudança desse porte estimula pensar nos impactos políticos, econômicos, legais, sociais e culturais e nos embates que serão reforçados pelos mecanismos jurídicos da propriedade intelectual e mercados editoriais vigentes. O foco da comunicação científica ainda está concentrado na divulgação dos resultados finais da investigação científica. Porém, percebeu-se que o acesso aberto as publicações científicas já é um movimento consolidado e está sendo conduzido nas práticas acadêmico-científicas e apoiadas por instituições de ensino, pesquisa, governamental e agências de fomento.

No plano político-legal há um forte consenso no sentido de que o direito de acesso à informação é considerado como um direito humano fundamental e difuso. As novas práticas de colaboração e compartilhamento têm esse papel relevante para a difusão de informação,

cultura e conhecimento. As reformulações serão sempre permeadas por conflitos de interesses dos atores envolvidos, bem como pela lógica do capitalismo cognitivo que representa, paradoxalmente, um impeditivo às atualizações legais e institucionais que dão suporte à ciência aberta, ao mesmo tempo que serve como um estimulador de práticas de compartilhamento e colaboração como estratégia de acesso, gratuidade e agilização da de resultados preliminares da produção científica.

Os esforços da ciência aberta envolvem instâncias de tomada de decisão e ações, internas e externas à ciência, que vão além do pesquisador, grupos de pesquisa e instituições, mas alargam o espectro para o nível macro que sustentam as políticas públicas e dão suporte às regulações internacionais, passando pelo nível meso das instituições científicas e agências de fomento. Assim, a ciência aberta se constitui um *locus* de poder e negociação, contradições, ideologias, culturas, saberes e conhecimentos.

A dimensão ético-legal é outro desafio que conduz uma agenda global para o entendimento da ciência aberta com o rompimento de paradigmas e práticas de proprietização existentes. A sua promoção vai além do espectro científico-técnico-tecnológico, mas atinge direitos sociais e difusos no sentido de uma sustentabilidade mais ampla, movendo-se para áreas essenciais, como saúde, meio ambiente, água, alimentação, agricultura, segurança nacional, entre outras.

No Brasil, as ações ainda são incipientes e fragmentadas sem consistência política e legal. A pesquisa confirmou a ausência de legislação específica nacional para os dados científicos. As políticas nacionais estão sendo conduzidas no âmbito governamental para os dados abertos e para algumas iniciativas fragmentadas e isoladas que estão trabalhando com parcerias internacionais. No âmbito científico inexistente uma política nacional que conduz à ciência aberta e as práticas científicas com dados científicos.

Os direcionamentos da pesquisa são amplos e longínquos. A ideia preliminar de que a ciência aberta é um conceito guarda-chuva e multidimensional abre um espectro para que a temática seja abordada em diferentes abordagens e diferentes domínios específicos. A *práxis* representada pelo modelo emergente da *e-Science* é um vasto campo de estudo para os aspectos técnicos, descritivos, tecnológicos que sustentam uma infraestrutura da *e-Science* nas disciplinas específicas, como as Ciências Jurídicas, Ciência da Informação, Ciência da Computação, Engenharias, dentre outras, podem fazer uso deste estudo e replicarem novas temáticas com base nas dimensões abordadas pela tese.

As lacunas da pesquisa se deram na amplitude das iniciativas internacionais. O *corpus* de análise foi realizado a partir de um recorte que contemplou as principais iniciativas

internacionais que se voltam para a pesquisa aberta com dados científicos, mas não abrangeu todos os continentes.

A ausência de legislações específicas e diretrizes que norteiem políticas nacionais orientadas a pesquisa colaborativa aberta com dados científicos representa uma lacuna no cenário nacional. Essa lacuna traz uma reflexão latente que o contexto científico brasileiro apresenta um retardo em relação ao cenário internacional e se configura em possibilidades de estudos futuros. O estudo não contemplou jurisprudências e projetos de lei nacionais.

Por fim, a pesquisa trouxe possibilidades de continuidade teórica e prática. Estudos futuros no espectro da ciência aberta traz um leque abrangente com inúmeras possibilidades de pesquisa. As dimensões políticas, legais e éticas da ciência aberta são excelentes focos de estudos, principalmente no cenário brasileiro. A *e-Science* também traz diferentes nuances para estudos com abordagem técnica, tecnológica e experimental. Estudos que envolvam os modelos de ciclo de vida dos dados, metadados, curadoria, preservação, com ênfase nos padrões e repositórios de dados são recomendados.

Para a continuidade dessa pesquisa se vislumbra avançar nas temáticas relacionadas com a Economia Colaborativa e as mudanças no regime de Propriedade Intelectual vigente, Dimensões da Ciência Aberta com a exploração mais aprofundada das mesmas e a Autoralidade Colaborativa com ampliação para as múltiplas colaborações e pluralidade de autorias.

A prática será contínua com a estruturação do protótipo que foi modelado conceitualmente a partir dos capítulos teóricos e que constituiu em uma plataforma preliminar denominada de *e-Quilt Prototype*. Como a tese foi conduzida numa abordagem conceitual, optou-se pela suspensão das questões práticas para ser contemplada em estudos futuros. Dessa forma, o intuito é dar continuidade ao *e-Quilt Prototype* como uma atividade de pesquisa experimental e colaborativa.

REFERÊNCIAS

- AALBERSBERG, I. J.; DUNHAN, J.; KOERS, H. **Connecting scientific articles with research data**: new directions in online scholarly publishing. 2011. Disponível em: <<http://www2.nict.go.jp/isd/ISDS-contents/wds-kyoto-2011.org/pdf/IS704.pdf>>. Acesso em: 19 set. 2016.
- ABDO, A. H. Direções para uma academia contemporânea e aberta. In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. Disponível em: <<https://www.cienciaaberta.net/ciencia-aberta-questoes-abertas-o-livro/>>. Acesso em: 19 set. 2016.
- ALBAGLI, S. Ciência aberta em questão. In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. Disponível em: <<https://www.cienciaaberta.net/ciencia-aberta-questoes-abertas-o-livro/>>. Acesso em: 19 set. 2016.
- ALTMAN, M. Data Citation in the Dataverse Network. In: **For Attribution: Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop**, NAS, 2012. Disponível em: <<https://www.nap.edu/read/13564/chapter/15>>. Acesso em: 29 set. 2016.
- AMBINDER, D. M.; MARCONDES, C. H. Artigos Científicos Digitais na Web: novas Experiências para apresentação, acesso e leitura. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2012, Rio de Janeiro, **Anais eletrônicos...** Rio de Janeiro: ENANCIB, 2012.
- ANDION, C.; SERVA, M. A. A etnografia e os estudos organizacionais. In: GODOI, C. K.; BANDEIRA-DE-MELLO, R.; SILVA, A. **A Pesquisa Qualitativa em Estudos Organizacionais**: paradigmas, estratégias e métodos. São Paulo: Saraiva, 2006.
- ATKINS, D. E. et al. **Revolutionizing science and engineering through cyber infrastructure**: report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyber infrastructure. Washington, DC, Jan. 2003. Disponível em: <<https://www.nsf.gov/cise/sci/reports/atkins.pdf>>. Acesso em: 15 abr.2015.
- AULETE, F. J. C.; VALENTE, A. L. S. **iDicionário Aulete**. Edição brasileira original: Hamílcar de Garcia. Rio de Janeiro: Lexikon Editora Digital, 2016. Disponível em: <http://www.aulete.com.br/site.php?mdl=aulete_digital&op=creditos>. Acesso em: 20 maio 2016.
- AUSTRALIAN NATIONAL DATA SERVICE. **Citation and identifiers**. In: **H2020 Model Grant Agreements**: H2020 General MGA, Australian, Oct. 2015. Disponível em: <<http://www.ands.org.au/working-with-data/citation-and-identifiers>>. Acesso em: 29 set. 2016.
- AVIAN KNOWLEDGE NETWORK. **Data from**: A quantitative survey of local adaptation and fitness trade-offs. Disponível em: <<http://datadryad.org/resource/doi:10.5061/dryad.315>>. Acesso em: 16 jun. 2015.

BADIOU, A. **The Concept of Model**: an Introduction to the Materialist Epistemology of Mathematics. Tradução de Zachary Luke Fraser e Tzuchien Tho. Australia: re.press, 2007.

BANDEIRA-DE-MELLO, R.; CUNHA, C.J.C.A. Grounded Theory. In: GODOI, C. K.; BANDEIRA-DE-MELLO, R.; SILVA, A. **A Pesquisa Qualitativa em Estudos Organizacionais**: paradigmas, estratégias e métodos. São Paulo: Saraiva, 2006.

BARACAT, A. C.; RIGOLIN, C. C. D. Proteção do Conhecimento e Movimento Open Access: Discussões no Âmbito da Organização Mundial da Propriedade Intelectual. **Liinc em Revista**, Rio de Janeiro, v. 8, n. 2, p. 348-364, set. 2012. Disponível em: <<http://liinc.revista.ibict.br/index.php/liinc/article/viewFile/487/376>>. Acesso em: 20 maio 2016.

BARBOSA, C. R. **Propriedade Intelectual enquanto Informação e os aspectos econômicos dos bens intelectuais**. 2007. Tese. (Doutorado em Direito Comercial)-Programa de Pós-graduação em Direito Comercial, Faculdade de Direito da USP, São Paulo, 2007.

BARBOSA, C. R. **Propriedade Intelectual**: introdução à propriedade intelectual como Informação. Rio de Janeiro: Elsevier, 2009.

BARDI, A.; MANGHI, P. Enhanced Publications: Data Models and Information Systems. **Liber Quarterly**, Pisa, v. 23, n. 4, 2014. Disponível em: <<https://www.liberquarterly.eu/articles/10.18352/lq.8445/>>. Acesso em: 20 maio 2016.

BARDIN, L. **Análise de Conteúdo**. São Paulo: Edições 70, 2011.

BARROS, Eduardo Portanova. O autor no imaginário da pós-modernidade: repensando Flusser e Foucault. **Ciências Sociais Unisinos**, São Leopoldo, v. 46, n. 2, p. 152-155, maio/ago. 2010. Disponível em: <<http://www.redalyc.org/pdf/938/93820638004.pdf>>. Acesso em: 27 set. 2016. <http://10.4013/csu.2010.46.2.04>.

BARTHES, R. A morte do autor. In: _____. **O Rumor da Língua**. São Paulo: Martins Fontes, 2004.

BASKERVILLE, R.; PRIES-HEJE, J. Grounded action research: a method for understanding IT in practice. **Accounting, Management & Information Technology**, v. 9, n. 1, p. 1-23, 1999. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0959802298000174>>. Acesso em: 27 set. 2016.

BELL, G. Foreword. In: HEY, A. J. G.; TANSLEY, Stewart; TOLLE, K. M. **The Fourth Paradigm**: Data-intensive Scientific Discovery. Redmond, Wash: Microsoft Research, University of North Texas 2009. Disponível em: <<http://digital.library.unt.edu/ark:/67531/metadc31516/>>. Acesso em: 27 set. 2016.

BIODIVERSITY INFORMATION STANDARDS - TDWG. **Darwin Core**. Disponível em: <<http://rs.tdwg.org/dwc/index.htm>>. Acesso em: 12 mar. 2015.

BIRD, A. T. K. Thomas Kuhn. In: ZALTA, E. N. (Ed.). **The Stanford Encyclopedia of Philosophy**, Fall 2013 Edition, Aug 11, 2011. Disponível em: <<http://plato.stanford.edu/archives/fall2013/entries/thomas-kuhn/>>. Acesso em: 11 set. 2016.

BISHOP, L. **Legal and ethical issues in data sharing**, UK Data Service, oct. 2015. Disponível em: <<https://www.fosteropenscience.eu/sites/default/files/pdf/1894.pdf>>. Acesso em: 20 jul. 2016.

BITTAR, C. A. **Direito de autor**. Rio de Janeiro: Forense Universitária, 2000.

BJÖRK, Bo-Christer. Open access to scientific publications - an analysis of the barriers to change? **Information Research**, v. 9, n. 2, Jan., 2004. Disponível em: <<http://www.informationr.net/ir/9-2/paper170.html>>. Acesso em: 20 jul. 2016.

BOLLEN, J. **Attribution and Credit: beyond print & citations**. California: Indiana University Berkeley, 2011. Disponível em: <https://researchdata.and.s.org.au/ausstage-gateway-to-the-australian-performing-arts/2023/?refer_q=rows=15/sort=list_title_sort%20asc/class=collection/q=BOLLEN/>. Acesso em: 29 set. 2016.

BOMFÁ, C. R. Z. et al. Acesso livre à informação científica digital: dificuldades e tendências. **Transinformação**, Campinas, v. 3, n. 20, p. 309-318, set./dez. 2008.

BORGMAN, C. L. **Big Data, Little Data, No Data: Scholarship in the Networked World**. Cambridge, MA: MIT Press, 2015. Disponível em: <<https://mitpress.mit.edu/big-data-little-data-no-data>>. Acesso em: 16 dez. 2015.

BORGMAN, C. L. Data, Scholarship, and the eHumanities. **eMagazine**, v. 4, 2014. Disponível em: <<http://ehumanities.leapress.com/emagazine-4/openingsscherm4/>>. Acesso em: 17 abr. 2014.

_____. **From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World**. London: MIT Press, 2003. Disponível em: <<https://mitpress.mit.edu/books/gutenberg-global-information-infrastructure>>. Acesso em: 19 abr. 2016.

_____. **Research data: who will share what, with whom, when, and why?** In: CHINA-NORTH AMERICAN LIBRARY CONFERENCE, 5., Beijing, Sep. 2010. Disponível em: <<https://works.bepress.com/borgman/238/>>. Acesso em: 20 maio 2015.

_____. **Scholarship in the digital age: information, infrastructure, and the Internet**. Cambridge: The MIT Press, 2007.

_____. The conundrum of sharing research data. **Journal of the American Society for Information Science and Technology**, v. 63, n. 6, p. 1059-1078, June 2012. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.22634/full>>. Acesso em: 17 abr. 2016.

BORKO, H. Information Science: What It Is? American Documentation (pre – 1986). **ABI/INFORM Global**, v. 19, n. 1, p. 1-3, jan.1968. Disponível em:

<<https://www.marilia.unesp.br/Home/Instituicao/Docentes/EdbertoFerneda/k---artigo-01.pdf>>. Acesso em: 17 abr. 2016.

BOWKER, G. C. et al. Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. **International Handbook of Internet Research**, Dordrecht, Springer Netherlands, p. 97-117, 2010. Disponível em: <<http://pal.lternet.edu/docs/bibliography/Public/351lterc.pdf>>. Acesso em: 02 dez. 2015.

BRADLEY, J. C. et al. Colaboration using Open Notebook Science in Academia. In: **The Future of Collaborations**, parts IV. New York: John Wiley & Sons, 2011. Disponível em: <http://media.wiley.com/product_data/excerpt/36/04706380/0470638036-1.pdf>. Acesso em: 10 set. 2016.

BRADLEY, J. C. et al. Open Notebook Science Challenge Solubilities of Organic Compounds in Organic Solvents. **Nature proceedings**, Mar. 2010. Disponível em: <<http://precedings.nature.com/documents/4243/version/2/files/npre20104243-2.pdf>>. Acesso em: 10 set. 2016.

BRADLEY, J. C. Web 2.0 in Science: success or failure?. **Scientific American**, 9 Jan. 2008. Disponível em: <<https://www.scientificamerican.com/article/science-2-point-0-great-new-tool-or-great-risk/>>. Acesso em: 12 set. 2016.

BRAMAN, S. Defining Information Policy. **Journal of Information Policy**, v. 1, n. 1, p. 1-5, 2011. Disponível em: <<http://pantherfile.uwm.edu/braman/www/html/pub.html>>. Disponível em: 12 set. 2015.

BRASE, J. The DataCite Consortium. In: **For Attribution: Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop**, NAS, 2012. Disponível em: <<https://www.nap.edu/read/13564/chapter/14>>. Acesso em: 07 maio 2016.

BRASIL. Câmara dos Deputados. **Projeto de Lei nº 3396/2015**. Institui a Política Nacional de Incentivo à Economia Criativa, suas definições, princípios norteadores, e objetivos. Disponível em: <<http://www.camara.gov.br/proposicoesWeb/fichadetramitacao?idProposicao=2024504>>. Acesso em: 07 maio 2016.

_____. Congresso Nacional. Lei n. 10.406, 10 de janeiro de 2002. Código Civil. **Diário Oficial da União**, Brasília, 11 jan. 2002. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/2002/L10406.htm>. Acesso em: 29 jul. 2016.

_____. Congresso Nacional. **Decreto-lei nº 2.848, de 07 de dezembro de 1940**. Código Penal, Rio de Janeiro, 7 dez. 1940. Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto-lei/Del2848.htm>. Acesso em: 29 jul. 2016.

_____. Congresso Nacional. Lei nº 9.610, de 19 de fevereiro de 1998. **Diário Oficial da União**, Brasília, 20 fev. 1998. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/L9610.htm>. Acesso em: 29 jul. 2016.

_____. Manual dos dados abertos: governo. Tradução e Adaptação da Comunidade Transparência Hacker, Esfera.mobi e Everton Zanella Alvarenga. In: GRAY, J. et al. **Open**

Data Manual, Open Knowledge Foundation. São Paulo: Laboratório Brasileiro de Cultura Digital e o Núcleo de Informação e Coordenação do Ponto BR (NIC.br), 2011.

_____. Ministério do Meio Ambiente. Instituto Chico Mendes de Conservação da Biodiversidade. **Portal da Biodiversidade.** Disponível em: <http://www.icmbio.gov.br/sisbio/images/stories/Folder_A3_GIZ_06-2a-impressao04092015FINAL.pdf>. Acesso em: 6 set. 2016.

BRUYNE, P.; HERMAN, J.; DE SCHOUTHEETE, M. **Dinâmica da pesquisa em ciências sociais: os pólos da prática metodológica.** Tradução de Ruth Joffily. Rio de Janeiro: Francisco Alves, 1977.

BUFREM, L. S. Configurações da pesquisa em ciência da informação, **DataGramaZero: Revista de Informação**, v. 14, n. 6, p. 1-15, dez. 2013.

BUSH, V. Como podemos pensar?. **Atlantic Monthly**, v. 176, n. 1, p. 101-108, 1945. Tradução Fábio Mascarenhas e Silva. Tradução de: As We May Think. Disponível em: <<http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>>. Acesso em: 21 set. 2016.

CALIFORNIA DIGITAL LIBRARY. Rights Management Group – copyrightMD, Oct. 9, 2016. Disponível em: <<http://www.cdlib.org/groups/>>. Acesso em: 15 ago. 2016.

CANHOS, D. A. L. **Sistemas de Informação em Biodiversidade e a Formulação de Políticas Públicas na Era Digital.** 2013. 174f. Tese (Doutorado) - Instituto de Geociências, Universidade Estadual de Campinas, São Paulo, 2013.

CARBONI, G. C. Aspectos da teoria da função social do direito de autor. In: TIMM, L. B.; MACHADO, R. B. (Org.). **Função Social do Direito.** São Paulo: Quartier Latin, 2009.

_____. **O direito de autor na multimídia.** São Paulo: Quartier Latin, 2003.

_____. Direito autoral e acesso ao conhecimento: em busca do equilíbrio. **Revista Juris**, Faculdade de Direito da FAAP, São Paulo, v. 1, p. 32-98, jan./jun. 2009.

_____. **Direito autoral, diversidade das expressões culturais e pluralidade de autorias.** Disponível em: <https://www.academia.edu/20014613/Direito_Autoral_Diversidade_das_Expressões_Culturais_e_Pluralidade_de_Autorias>. Acesso em: 19 abr. 2016.

_____. Os desafios do direito de autor na tecnologia digital e a busca do equilíbrio entre interesses individuais e sociais. **Revista Juris**, Faculdade de Direito da FAAP, São Paulo, p. 1-56, 2009. Disponível em: <<http://docplayer.com.br/32132-Os-desafios-do-direito-de-autor-na-tecnologia-digital-e-a-busca-do-equilibrio-entre-interesses-individuais-e-sociais.html>>. Acesso em: 21 fev. 2016.

CARVALHO, V. M. B.; LANZILLO, A. S. S.; GUIMARÃES, P. B. V. A Economia Criativa: Conceito e Relação com o Direito. **FIDES**, Natal, v. 6, n. 1, jan./jun. 2015. Disponível em:

<http://www.cidp.pt/publicacoes/revistas/rjlb/2015/3/2015_03_1493_1510.pdf>. Acesso em: 21 fev. 2016.

CARVALHO, J. R.; MORRIS, C. **The Science and Technology**: guidebook for Lawyers. Chicago: American Bar Association, 2014.

COMITÊ GESTOR DA INTERNET NO BRASIL. **Dados Abertos Governamentais**. Tradução de: opendatamanual.org. São Paulo: CGI.br, 2011. Disponível em: <http://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual_Dados_Abertos_WEB.pdf>. Acesso em: 20 maio 2016.

_____. **Manual dos dados abertos: governo**. Tradução de: opendatamanual.org. São Paulo: CGI.br, 2011. Disponível em: <<http://www.w3c.br/pub/Materiais/PublicacoesW3C/dados-abertos-governamentais.pdf>>. Acesso em: 20 maio 2016.

COMMITTEE ON PUBLICATION ETHICS. **Code of conduct and best practice guidelines for journal editors**, 7 Mar., 2011. Disponível em: <<http://publicationethics.org>>. Acesso em: 12 jul. 2016.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. **Recommendation for Space Data System Practices**: Reference Model for an Open Archival Information System (OAIS). Washington, DC, p. 1-10, June, 2012. Disponível em: <<https://public.ccsds.org/pubs/650x0m2.pdf>>. Acesso em: 12 set. 2016.

CORRÊA, P. L. P. et al. Uma arquitetura de sistemas voltada para a integração de bases de dados distribuídas de biodiversidade. **Revista Brasileira de Agroinformática**, São Paulo, v. 6, n. 1, p. 47-63, 2004. Disponível em: <<https://www.sbiagro.org.br/pdf/revista/rbiagro-v6n1-artigo4.pdf>>. Acesso em: 12 set. 2016.

COSTA, M. M.; CUNHA, M. B. da. A literatura internacional sobre e-science nas bases de dados LISA e LISTA. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 20, n. 44, p. 127-144, nov. 2015. ISSN 1518-2924. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2015v20n44p127>>. Acesso em: 06 out. 2016.

DATA ARCHIVING AND NETWORKED SERVICES. **NARCIS**. Disponível em: <<http://www.narcis.nl/search/coll/dataset/Language/en>>. Acesso em: 17 abr. 2016.

DATA DOCUMENTATION INITIATIVE ALLIANCE. **Document, Discover and Interoperate**. Disponível em: <<http://www.ddalliance.org/>>. Acesso em: 15 mar. 2015.

DATA OBSERVATION NETWORK FOR EARTH. **What is DataONE?** Disponível em: <<https://www.dataone.org/>>. Acesso em: 16 fev. 2015.

DATA.GOV (Estados Unidos). **The home of the U.S. Government's open data**. Disponível em: <<https://www.data.gov>>. Acesso em: 16 jun. 2015.

DATAONE. **DataONE Education Module**: Data Entry and Manipulation, 12 Nov. 2012. Disponível em: <<https://www.dataone.org/education-modules>>. Acesso em: 16 fev. 2015.

_____. **Repository**. Disponível em: <<https://www.dataone.org/software-tools/tags/repository>>. Acesso em: 16 jun. 2016.

DATIG, I.; RUSSELL, B. Instructing College Students on the Ethics of Information Use at the Reference Desk: A Guide and Literature Review. **The Reference Librarian: Special Ethics**, v. 55, p. 234-246, 2014. Disponível em: <<https://www.fosteropenscience.eu/sites/default/files/pdf/1894.pdf>>. Acesso em: 29 set. 2016.

DELFANTI, A.; PITRELLI, N. Ciência aberta: revolução ou continuidade? In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015.

DENZIN, N.K; LINCOLN, Y.S. **O planejamento da pesquisa qualitativa: teorias e abordagens**. 2. ed. São Paulo: Artmed, 2006.

DIAS, G. A. Evolução tecnológica dos periódicos científicos eletrônicos: do paradigma gutenberguiano para um cenário hipermidiático. **Revista do Mestrado Profissional Gestão em Organizações Aprendentes**, v. 2, n. 1, p. 1-2, 2013. Disponível em: <<http://periodicos.ufpb.br/ojs/index.php/mpgoa/article/view/16996/9677>>. Acesso em: 04 maio 2016.

DIAS, G. A.; VIDOTTI, S. A. B. G. O Direito da Propriedade Intelectual: relações com os entregáveis da arquitetura da informação. **Perspectivas em Gestão & Conhecimento**, João Pessoa, v. 1, p. 73-85, out. 2011. Disponível em: <<http://periodicos.ufpb.br/ojs2/index.php/pgc/article/view/10825/6100>>. Acesso em: 9 out. 2016.

DIGITAL CURATION CENTRE. **About the DCC**. Disponível em: <<http://www.dcc.ac.uk/about-us>>. Acesso em: 13 set. 2016.

_____. **DataCite Metadata Schema**. Disponível em: <<http://www.dcc.ac.uk/resources/metadata-standards/datacite-metadata-schema>>. Acesso em: 13 set. 2016.

_____. **DCC Curation Lifecycle Model**. Disponível em: <<http://www.dcc.ac.uk/resources/curation-lifecycle-model/>>. Acesso em: 13 set. 2016.

_____. **List of Metadata Standards**. Disponível em: <<http://www.dcc.ac.uk/resources/metadata-standards/list?page>>. Acesso em: 13 set. 2016.

DLYNCH, C. Jim Gray's Fourth Paradigm and the Construction of the Scientific Record. **The Fourth Paradigm**, April, 2009. Disponível em: <<https://www.cni.org/about-cni/staff/clifford-a-lynch/publications>>. Acesso em: 10 set. 2016.

DOWBOR, L. **Capital Cognitivo: a economia da criatividade**. 2011. Disponível em: <<http://dowbor.org/category/artigos-recebidos/>>. Acesso em: 29 nov. 2015.

_____. O sistema financeiro atual trava o desenvolvimento econômico. **Estud.av.**, São Paulo, v. 29, n. 83, p. 263-278, abr. 2015. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-

[40142015000100263&lng=pt&nrm=iso](http://dx.doi.org/10.1590/S0103-40142015000100013)>. Acesso em: 29 nov. 2015.
<http://dx.doi.org/10.1590/S0103-40142015000100013>

DUBLIN CORE METADATA INITIATIVE. **Dublin Core Metadata Element Set**. Disponível em: <<http://dublincore.org/documents/dcmi-terms>>. Acesso em: 15 set. 2016.

DUSOLLIER, S. **Scoping Study on Copyright and Related Rights and the Public Domain**. Belgium: University of Namur, 2010. Disponível em:
 <http://www.wipo.int/edocs/mdocs/mdocs/en/cdip_4/cdip_4_3_rev_study_inf_1.pdf>.
 Acesso em: 20 abr. 2016.

EAVES, David. **The Three Laws of Open Government Data**. Sep. 11, 2009. Disponível em: <<https://eaves.ca/2009/09/30/three-law-of-open-government-data/>>. Acesso em: 25 maio 2015.

EISENSTEIN, E. L. **A Revolução da Cultura Impressa: os Primórdios da Europa Moderna**. Tradução de The Printing Revolution in Early Modern Europe. São Paulo: Ática, 1998. Disponível em: <<http://sapereaude10.blogspot.com.br/2013/05/a-revolucao-da-cultura-impressa-de.html>>. Acesso em: 25 maio 2016.

ERIKSSON, H. The semantic-document approach to combining documents and ontologies. **International Journal of Human - Computer Studies**, v. 65, n. 7, p. 624-639, 2007. Disponível em: <<http://dl.acm.org/citation.cfm?id=1247767>>. Acesso em: 25 set. 2016.

EUROPEAN COMMISSION. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. **Official Journal L 281**, 23 Nov. 1995, p. 0031-0050. Disponível em: <<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML>>. Acesso em: 20 jul. 2016.

_____. **Guidelines on the Implementation of Open Access to Scientific Publications and Research Data in Projects supported by the European Research Council under Horizon 2020**, Version 3.1, 25 August 2016. Disponível em: <https://erc.europa.eu/sites/default/files/document/file/ERC_Guidelines_Implementation_Open_Access.pdf>. Acesso em: 19 set. 2016.

_____. **H2020 Programme: Multi-Beneficiary General Model Grant Agreement**, Version 3.0, 20 Jul. 2016. Disponível em: <https://erc.europa.eu/sites/default/files/document/file/ERC_Guidelines_Implementation_Open_Access.pdf>. Acesso em: 19 set. 2016.

_____. **Law**. Disponível em: <https://ec.europa.eu/info/law_en>. Acesso em: 16 ago. 2016.

_____. **Open Research Data in Horizon 2020**. Version 3.0, 26 Jul. 2016. Disponível em: <http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf>. Acesso em: 19 abr. 2016.

EUROPEAN UNION. **FOSTER: Facilitate Open Science Training for European Research**. Disponível em: <<https://www.fosteropenscience.eu>>. Acesso em: 20 jul. 2016.

FEDERAL GEOGRAPHIC DATA COMMITTEE. **Geospatial Metadata Standards and Guidelines**. Disponível em: <<https://www.fgdc.gov/metadata/geospatial-metadata-standards>>. Acesso em: 11 mar. 2015.

FLICK, U. **Introdução à pesquisa qualitativa**. 3. ed. Porto Alegre: Artmed, 2009.

FOUCAULT, Michel. **Ditos e Escritos**: Estética – literatura e pintura, música e cinema. Rio de Janeiro: Forense Universitária, 2001. v. 3. Disponível: <http://www.foucault.ileel.ufu.br/foucault/bibliografia/ditos-e-escritos-v-7>>. Acesso em: 02 abr. 2016.

FRAGOSO, J. H. R. **Direito Autoral**: da Antiguidade à Internet. São Paulo: Quartier Latin do Brasil, 2009.

FREIRE, G. H. A.; FREIRE, I. M. Hipertexto e comunicação científica na sociedade em rede. **Inf. & Soc.: Est.**, João Pessoa, v. 23, n. 2, p. 09-10, maio/ago. 2013. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/16787/9680>>. Acesso em: 8 out. 2016.

FREIRE, G. H. A.; SILVA, J. L. C. **Enc. Bibli: R. Eletr. Bib. Ci. Inf.**, Florianópolis, v. 17, n. 33, p. 1-29, jan./abr. 2012. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/16787/9680>>. Acesso em: 8 out. 2016.

FRICKÉ, M. Big Data and Its Epistemology. **Journal of the Association for Information Science and Technology**, v. 66, n. 4, p. 651–661, 2015. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.23212/pdf>>. Acesso em: 1 dez. 2016. DOI: 10.1002/asi.23212

FROSIO, G. F. **Open Access Publishing**: A Literature Review. Reino Unido: Research Fellow School of Law University of Nottingham, UK, 2014. Disponível em: <<http://www.create.ac.uk/publications/open-access-publishing-a-literature-review/>>. Acesso em: 8 out. 2016.

FROST, C. **Bounty vs. Overload, and the Eye of the Beholder**. Disponível em: <<http://www.herravendomain.com/bounty-vs-overload-and-the-eye-of-the-beholder/>>. Acesso em: 8 out. 2016.

FURI-PERRY, U. **Social Media Law**: a handbook of cases & use. Chicago: American Bar Association, 2014.

FURNIVAL, A. C.; HUBBARD, B. Acesso aberto às publicações científicas: vantagens, políticas e advocacy. **InCID: R. Ci. Inf. e Doc.**, Ribeirão Preto, v. 2, n. 2, p. 160-177, jul./dez. 2011. Disponível em: <<http://www.revistas.usp.br/incid/article/view/42358>>. Acesso em: 8 out. 2016.

FUTURE OF RESEARCH COMMUNICATIONS AND E-SCHOLARSHIP (FORCE 11). In: MARTONE, M. (Ed.). **Data Citation Synthesis Group**: Joint Declaration of Data Citation Principles: final.. San Diego CA: FORCE11, 2014. Disponível em: <<https://www.force11.org/group/joint-declaration-data-citation-principles-final>>. Acesso em: 19 set. 2016.

GANDELMAN, H. **De Gutenberg à Internet: Direitos Autorais na era digital**. 3. ed. atual. e ampl. São Paulo: Record, 2011.

GARCIA, J. C. R.; TARGINO, M. G.; DANTAS, E. R. F. Conceito de responsabilidade social da informação. **Inf. & Inf.**, Londrina, v. 17, n. 1, p. 1-25, jan./jun. 2012. Disponível em: <<http://periodicos.ufpb.br/ojs/index.php/biblio/article/view/15281>>. Acesso em: 8 out. 2016.

GARVEY, W. D.; GRIFFITH, B. C. Scientific communication as a social system. In: _____. **Communication: the essence of science**. London: Pergamon Press, 1979. Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=&id=3RijBQAAQBAJ&oi=fnd&pg=PA148&ots=8FNSI4bJ6V&sig=0q_ciQ3TUH1AAhjvT8aIwWfoCCQ&redir_esc=y#v=onepage&q&f=false>. Acesso em: 8 out. 2016.

GASQUE, K. C. G. D. **O pensamento reflexivo na busca e no uso da informação na comunicação científica**. 2008. 240 f. Tese (Doutorado) - Departamento de Ciência da Informação e Documentação, Universidade de Brasília, 2008.

GATTI, D.; CORTIZ, D., DAWBOR L. Em busca de uma ciência aberta, colaborativa e conectada. In: DOWBOR, L.; SILVA, H. (Org.). **Propriedade intelectual e direito à informação**. São Paulo: EDUC, 2014. Disponível em: <http://dowbor.org/blog/wp-content/uploads/2013/03/14Livro_Proprietade-intelectual.pdf>. Acesso em: 16 ago. 2016.

GOMES, C. M. **Comunicação Científica: alicerces, transformações e tendências**. São Paulo: Livros LabCom, 2013.

GRAY, J. eScience: a transformed scientific method. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). **The Fourth Paradigm: Data-Intensive Scientific Discovery**. Redmond, Washington. UNT Digital Library. Palestra de Jim Gray em 11 de janeiro de 2007, cedida a NRC-CSTB 1, em Mountain View, CA. Disponível em: <<http://digital.library.unt.edu/ark:/67531/metadc31516/>>. Acesso em: 25 abr. 2016.

HARLEY, D. Issues of Time, Credit, and Peer Review. In: **For Attribution: Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop**. NAS, 2012. Disponível em: <<https://www.nap.edu/read/13564/chapter/13>>. Acesso em: 07 maio 2016.

HARVEY, D. **Condição pós-moderna**. 15. ed. São Paulo: Loyola, 2006.

HENRY A. Murray Research Archive. **Policies**. Disponível em: <<http://murray.harvard.edu/policies>>. Acesso em: 07 maio 2016.

HERSCOVICI, A. Escolha coletiva, governança e direitos de propriedade: uma análise econômica dos commons. **Nova econ.**, Belo Horizonte, v. 23, n. 1, p. 185-208, 2013. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-63512013000100007&lng=pt&nrm=iso&tlng=en>. Acesso em: 19 abr. 2016.

HEY, A.; TANSLEY, S.; TOLLE, K. **The Fourth Paradigm: Data-intensive Scientific Discovery**. Redmond, Washington: UNT Digital Library, 2009. Disponível em: <<http://digital.library.unt.edu/ark:/67531/metadc31516/>>. Acesso em: 29 set. 2016.

HEY, T. The Big Idea: The Next Scientific Revolution. **Harvard Business Review**, nov. 2010. Disponível em: <<https://hbr.org/2010/11/the-big-idea-the-next-scientific-revolution>>. Acesso em: 29 set. 2016.

HEY, T.; TREFETHEN, A. E-Science and its implications. **Philosophical Transactions of the Royal Society (A)**, v. 15, n. 361, p. 1809-1825, June, 2003. Disponível em: <https://www.researchgate.net/publication/10584819_e-Science_and_its_implications>. Acesso em: 29 set. 2016.

HOGENAAR, A. **What is an Enhanced Publication?** OpenAIRE. 2013. Disponível em: <<https://www.openaire.eu/en/component/content/article/76-highlights/344-a-short-introduction-to-enhanced-publications>>. Acesso em: 20 maio 2016.

HOLDREN, John P. **Memorandum for the Heads of Executive Departments and Agencies' John P. Holdren da White House Office of Science and Technology Policy.** February 22, 2013. Disponível em: <<http://spaceref.biz/news/viewstr.html?pid=43429>>. Acesso em: 29 set. 2016.

HOPFER, K. R.; MACIEL-LIMA, S. M. Grounded Theory: avaliação crítica do método nos estudos organizacionais. **Rev. FAE**, Curitiba, v. 11, n. 2, p. 15-24, jul./dez. 2008.

HOWKINS, John. **Economia criativa: como ganhar dinheiro com ideias criativas.** Tradução de Ariovaldo Griesi. São Paulo: M. Books, 2013.

HUFTON, A. **Scientific Data to complement and promote public data repositories.** July 23, 2013. Disponível em: <<http://blogs.nature.com/scientificdata/2013/07/23/scientific-data-to-complement-and-promote-public-data-repositories/>>. Acesso em: 20 jul. 2016.

ILKA, D.; RUSSEL, B. The Fruits of Intellectual Labor: International Student Views of Intellectual Property. **Coll. res. libr.**, London, Sep., 1, 2015. Disponível em: <<http://crl.acrl.org/citmgr?gca=crl%3Bcrl14-677v1>>. Acesso em: 6 nov. 2015.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Diadorim.** Disponível em: <<http://diadorim.ibict.br/>>. Acesso em: 6 nov. 2015.

_____. **Dicionário Eletrônico de Terminologia em Preservação Digital.** Disponível em: <http://wiki.ibict.br/index.php/Preservacao_Digital>. Acesso em: 16 set. 2016.

_____. **DSpace.** Disponível em: <<http://www.ibict.br/pesquisa-desenvolvimento-tecnologico-e-inovacao/Sistema-para-Construcao-de-Repositorios-Institucionais-Digitais>>. Acesso em: 6 nov. 2015.

_____. **Oasisbr.** Disponível em: <<http://oasisbr.ibict.br/vufind/>>. Acesso em: 6 nov. 2015.

_____. **Rede Cariniana.** Disponível em: <<http://cariniana.ibict.br/>>. Acesso em: 6 nov. 2015.

INSTITUTO CHICO MENDES DE CONSERVAÇÃO DA BIODIVERSIDADE. Decreto nº7.515 de 8 de julho de 2011. **Diário Oficial da União**, Brasília, 11 jul. 2011. Disponível em: <<https://portaldbiodiversidade.icmbio.gov.br/portal/>>. Acesso em: 6 nov. 2015.

_____. **Portal da Biodiversidade**. Disponível em:
<<https://portaldabiodiversidade.icmbio.gov.br/portal/>>. Acesso em: 6 nov. 2015.

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH. **Guide to Social Science Data Preparation and Archiving**: introduction. 5th edition. (2012). Ann Arbor, MI. Disponível em:
<<http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/index.html#cycle>>. Acesso em: 21 fev. 2015.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **IFLA Data Protection Policy**. Disponível em: < <http://www.ifla.org/data-protection-policy>>. Acesso em: 27 maio 2016.

JANKOWSKI, N. W. Exploring e-science: an introduction. **Journal of Computer-Mediated Communication**, v. 12, n. 2, p. 549-562, 2007.

JAPIASSÚ, H.; MARCONDES, D. **Dicionário básico de filosofia**. 4.ed. rev. ampl. Rio de Janeiro: Jorge Zahar, 2006.

JARDIM, J. M. A Lei de Acesso à Informação Pública: dimensões político-informacionais. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 5, n. 1, p. 1-21, 2012. Disponível em: <<http://inseer.ibict.br/ancib/index.php/tpbci/article/viewArticle/68>>. Acesso em: 19 abr. 2016.

JOINT INFORMATION SYSTEMS COMMITTEE. **Intellectual property for data**. Disponível em: <<https://www.jisc.ac.uk/search/intellectual%20property%20for%20data>>. Acesso em: 19 set. 2016.

JOINT INFORMATION SYSTEMS COMMITTEE. **Intellectual property law**. Disponível em: <<https://www.jisc.ac.uk/guides/intellectual-property-law>>. Acesso em: 21 jul. 2016.

_____. **Intellectual property**. Disponível em:
<<https://www.jisc.ac.uk/guides/crowdsourcing/intellectual-property>>. Acesso em: 19 set. 2016.

_____. **Research data**. Disponível em: <<https://www.jisc.ac.uk/research-data>>. Acesso em: 19 set. 2016.

KAASE, M. Databases, Core: Political Science and Political Behavior. In: SMELSER, N. J.; BALTES, P. B. (Ed.) **International Encyclopedia of the Social and Behavioral Sciences**, Amsterdam, Elsevier, v. 5, p. 3251-3255, 2001. Disponível em:
<<http://www.sciencedirect.com/science/article/pii/B0080430767000346>>. Acesso em: 17 maio 2015.

KALLINIKOS, J.; AALTONEN, A.; MARTON, A. A theory of digital objects. **First Monday**, v. 15, n. 6, jun. 2010. Disponível em:
<<http://firstmonday.org/ojs/index.php/fm/article/view/3033/2564>>. Acesso em: 17 maio 2015.

KERA, D. Hardware aberto para ciência aberta no sul global: Diplomacia geek?. In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015.

KIRCZ, J. G. New practices for electronic publishing 1: Will the scientific paper keep its form? **Learned Publishing**, v. 14, n. 4, p. 265-272, 2001. Disponível em: <<http://www.kra.nl/Website/Artikelen/learned-publ1.pdf>>. Acesso em: 9 set. 2015.

KNOWLEDGE NETWORK FOR BIOCOMPLEXITY. **Morpho**. Disponível em: <<https://knb.ecoinformatics.org/#tools/morpho>>. Acesso em: 11 mar. 2015.

KOESTLER, A. **Act of creation**: a study of the conscious and unconscious in science and art. New York: Laurel, 1967.

KOWALCZYK, S.; SHANKAR, K. Data Sharing in the Sciences. **Annual Review of Information Science and Technology**, Bloomington, Indiana University, Chapter 6, p. 247-294, 2014. Disponível em: <https://kalpanashankar.files.wordpress.com/2010/06/arist_data_sharing.pdf>. Acesso em: 9 dez. 2015.

KROKOSZ, M. **Outras palavras sobre autoria e plágio**. São Paulo: Atlas, 2014.

KUHN, T. S. The Structure of Scientific Revolutions. Second Edition. **International Encyclopedia of Unified Science**, Chicago, v. 2, n. 2, p. 1-210, 1970. Disponível em: <http://projekintegracija.pravo.hr/download/repository/Kuhn_Structure_of_Scientific_Revolutions.pdf>. Acesso em: 14 set. 2015.

LAFUENTE, A.; ESTALELLA, A. Modos de ciência: pública, abierta y común. In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Orgs.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015.

LEGISLATION.GOV.UK (Reino Unido). **Data Protection Act 1998**: Chapter 29. London: The Stationery Office, 2005. Disponível em: <http://www.legislation.gov.uk/ukpga/1998/29/pdfs/ukpga_19980029_en.pdf>. Acesso em: 19 set. 2016.

LEITE, F. C. L. **Como gerenciar e ampliar a visibilidade da informação científica brasileira**: repositórios institucionais de acesso aberto. Brasília: IBICT, 2009.

LEITE, F. C. L.; COSTA, S. M. S. Gestão do conhecimento científico: proposta de um modelo conceitual com base em processos de comunicação científica. **Ci. Inf.**, Brasília, v. 36, n. 1, p. 92-107, jan./abr. 2007.

LESSARD-HÉBERT, M.; GOYETTE, G.; BOUTIN, G. **Investigação Qualitativa**: fundamentos e práticas. 3. ed. Lisboa: Instituto Piaget, 2008.

LEVITT, C. A.; DAVIS, J. K. **Internet Legal Research on a budget**: free and low-Cost resources for lawyers. Chicago: American Bar Association, 2014.

LIBRARIES AND ARCHIVES COPYRIGHT ALLIANCE. **Copyright**. Disponível em: <<http://www.cilip.org.uk/cilip/copyright>>. Acesso em: 14 set. 2015.

LYNCH, C. A. Jim Gray's Fourth Paradigm and the Construction of the Scientific Record In: HEY, Tony (Ed.). **The Fourth Paradigm**, Microsoft Research, 22 June 2009. Disponível em: <<https://www.cni.org/publications/cliffs-pubs/jim-grays-fourth-paradigm-and-the-construction-of-the-scientific-record>>. Acesso em: 29 set. 2016.

MACHADO, J. Dados abertos e ciência aberta. In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015.

MACHADO, L. A. Economia Criativa. **Economistas**, v. 17, ago. 2015.

MACHADO, R. **Foucault, a ciência e o saber**. 3.ed. rev. ampl. Rio de Janeiro: Zahar, 2007.

MARCONDES, C.H. Um modelo semântico de publicações eletrônicas. **Liinc em Revista**, v. 7, n. 1, 2011. Disponível em: <<http://liinc.revista.ibict.br/index.php/liinc/article/view/404/262>>. Acesso em: 3 mar. 2016.

MARCUM, D. B.; GEORGE, G. (Ed.). **The data deluge: can libraries cope with e-science?** Santa Barbara, California: Libraries Unlimited, 2010.

MARQUES, F. Ciência no palheiro: Programa em eScience busca extrair novos conhecimentos em meio a volumes gigantescos de dados. **Revista Pesquisa FAPESP**, São Paulo, ed. 215, jan. 2014. Disponível em: <<http://revistapesquisa.fapesp.br/2014/01/13/ciencia-palheiro/>>. Acesso em: 19 mar. 2014.

MARQUES, Lilian Emanuelli; PINHEIRO, Marta Macedo Kerr. A cúpula mundial sobre a sociedade da informação: foco nas políticas de informação. **Inf. & Soc.:Est.**, João Pessoa, v.23, n.1, p. 117-131, jan./abr. 2013. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/download/15450/9536>>. Acesso em: 21 set. 2016.

MARTINS FILHO, P. Direitos autorais na Internet. **Ci. Inf.**, Brasília, v. 27, n. 2, 1998. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651998000200011>. Acesso em: 22 maio 2015.

MATTIOLI, M. Disclosing Big Data. **Minnesota Law Review**, v. 99, n. 2/3, p. 535-583, dec. 2014.

MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999.

MEDEIROS, J. S.; CAREGNATO, S. E. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **Liinc em Revista**, Rio de Janeiro, v. 8, n. 2, p. 311-322, set. 2012. Disponível em: <<http://liinc.revista.ibict.br/index.php/liinc>>. Acesso em: 17 jan. 2014.

MENEZES, E. D. **Curso de Direito Autoral**. Belo Horizonte: Del Rey, 2007.

MENEZES, E. T.; SANTOS, T. H. **Verbete multidisciplinaridade**: Dicionário Interativo da Educação Brasileira - Educabrazil. São Paulo: Midiamix, 2015. Disponível em: <<http://www.educabrazil.com.br/multidisciplinaridade/>>. Acesso em: 21 set. 2016.

MENEZES, E. T.; SANTOS, T. H. **Verbete temas transversais**: Dicionário Interativo da Educação Brasileira - Educabrazil. São Paulo: Midiamix, 2001. Disponível em: <<http://www.educabrazil.com.br/temas-transversais/>>. Acesso em: 21 set. 2016.

MENZEL, H. **The flow of information among scientists**: problems, opportunities, and research questions. New York: Columbia University, Bureau of Applied Social Research, 1958.

MERRIAM, S. B. **Qualitative Research**: a guide to design and interpretation. San Francisco: Jossey-Bass, 2009.

NATIONAL ACADEMIC RESEARCH AND COLLABORATIONS INFORMATION SYSTEM. **About NARCIS**. Disponível em: <<http://www.narcis.nl>>. Acesso em: 25 jun. 2014.

NATIONAL ACADEMY OF ENGINEERING. Disponível em: <<https://www.nae.edu>>. Acesso em: 18 abr. 2016.

NATIONAL ACADEMY OF MEDICINE. Disponível em: <<https://nam.edu>>. Acesso em: 18 abr. 2016.

NATIONAL ACADEMY OF SCIENCES. **Copyright in the Digital Era**: Building Evidence for Policy committee on the impact of copyright policy on innovation in the digital era. Washington, DC: National Academies Press, 2013. Disponível em: <<http://www.nasonline.org>>. Acesso em: 18 abr. 2016.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION. **Hubble 2011**: science year in review. Finding Planets: in archived Hubble Data. Disponível em: http://hubblesite.org/hubble_discoveries/science_year_in_review/pdf/2011/about_the_hubble_space_telescope.pdf>. Acesso em: 19 abr. 2016.

NATIONAL FOUNDATION FOR SCIENTIFIC COMPUTING. **Blimunda Project**. Disponível em: <<https://sites.google.com/site/blimundaproject/sobre>>. Acesso em: 19 abr. 2016.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **Standards**. Disponível em: <http://www.niso.org/apps/group_public/projects.php>. Acesso em: 07 mar. 2016.

NATIONAL INSTITUTE OF HEALTH. **NIH Data Sharing Policy and Implementation Guidance**, Washington D.C., Mar., 2003. Disponível em: <http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm>. Acesso em: 5 ago. 2016.

NATIONAL RESEARCH COUNCIL. **A Question of Balance**: Private Rights and the Public Interest in Scientific and Technical Databases. Washington, DC: National Academy Press, 2006. Disponível em: <<https://www.nap.edu>>. Acesso em: 26 maio 2015.

NATIONAL SCIENCE BOARD. **Long-lived digital data collections**: enabling research and education in the 21st century. Washington D.C., Sep. 2005. Disponível em: <<https://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>>. Acesso em: 20 out. 2015.

NATIONAL SCIENCE FOUNDATION. **About**. Disponível em: <<http://www.nsf.gov>>. Acesso em: 18 abr. 2016.

_____. **Today's Data, Tomorrow's Discoveries**: Increasing Access to the Results of Research Funded by the National Science Foundation. March 18, 2015. Disponível em: <<https://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf>>. Acesso em: 18 abr. 2016.

NATURE. **Scientific Data**. Disponível em: <<http://www.nature.com/sdata/about>>. Acesso em: 23 jun. 2016.

NETHERLANDS ORGANISATION FOR SCIENTIFIC RESEARCH. **Open science**. Amsterdam, 2016. Disponível em: <<http://www.nwo.nl/en/policies/open+science>>. Acesso em: 20 abr. 2016.

_____. **Policy**. Amsterdam, 2016. Disponível em: <<http://www.nwo.nl/en/policies/open+science>>. Acesso em: 19 abr. 2016.

NETSCHER, S. Research Ethics and Legal Compliance: Informed Consent and Data Licensing. [s.n.]: Gesis, 2015. 19 slides, color. Disponível em: <<https://www.fosteropenscience.eu/sites/default/files/pdf/2028.pdf>>. Acesso em: 25 jul. 2016.

NIELSEN, H. J.; HJØRLAND, B. Curating research data: The potential roles of libraries and information professionals. **Journal of Documentation**, v. 70, n. 2, p. 221-240, 2014. Disponível em: <<http://www.emeraldinsight.com/doi/pdfplus/10.1108/JD-03-2013-0034>>. Acesso em: 29 set. 2016. <http://dx.doi.org/10.1108/JD-03-2013-0034>

NIELSEN, M. A. **Reinventing discovery**: the new era of networked science. Princeton, N.J: Princeton University Press, 2011.

NOGUEIRA, R. H.; GUIMARÃES, P. B. V.; XAVIER, Y. M. de A. Desenvolvimento e economia criativa: uma análise preliminar das origens e rumos da economia criativa e sua repercussão no direito. In: GUIMARÃES, P. B. V.; XAVIER, Y. M. de A. (Org.). **O Direito da Economia Criativa**: Temas de Desenvolvimento e Propriedade Intelectual. São Paulo: Editora Max Limonad, 2015.

OLIVEIRA, A. C. S. de et al. ; DIAS, G. A.; ANJOS, R. L. dos; SOUZA, Virgínia M. de; CORRÊA, Pedro Luiz P. Quarto Paradigma Científico: adoção do *data life cycle model* para o gerenciamento de dados científicos abertos do experimento *e-Quilt Prototype*. CONFOA, 6., Salvador, 2015. **Anais...** Disponível em: <<https://drive.google.com/folderview?id=0B2CiJWPT1NhwcNiZ1BwZWY5TVU&usp=sharing>>. Acesso em: 20 jul. 2016.

OLIVEIRA, A. C. S. de; SILVA, E. M. Ciência Aberta: dimensões para um novo fazer científico. **Informação & Informação**, Londrina, 2016. No prelo. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/issue/view/1143>>. Acesso em: 20 set. 2016.

OLIVEIRA, A. C. S.; TABOSA, H. R.; FREIRE, G. H. A Natureza da Ciência da Informação: Mudança de Modelos. **Inf. & Soc.: Est.**, João Pessoa, v. 23, n. 1, p. 177-181, jan./abr. 2013. Disponível em: <<http://www.brapci.ufpr.br/brapci/v/a/13330>>. Acesso em: 20 set. 2016.

OPEN ARCHIVES INITIATIVE. **Object Reuse and Exchange**. Disponível em: <<http://www.openarchives.org/ore/>>. Acesso em: 7 mar. 2015.

_____. **The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)**. Disponível em: <<http://www.openarchives.org/OAI/openarchivesprotocol.html>>. Acesso em: 20 jun. 2015.

OPEN DATA COMMONS. **Open Database License (ODbL) v1.0**. Disponível em: <<http://opendatacommons.org/licenses/odbl/>>. Acesso em: 9 dez. 2015.

OPENAIRE. **Open access to research data: the Open Research Data Pilot**. Disponível em: <<https://www.openaire.eu/h2020-oa-data-pilot>>. Acesso em: 19 abr. 2016.

_____. **Open science**. Disponível em: <<https://www.openaire.eu/supporting-open-science>>. Acesso em: 19 abr. 2016.

_____. **What is a Enhanced Publication?** Disponível em: <<https://www.openaire.eu/en/component/content/article/76-highlights/344-a-short-introduction-to-%20enhanced-publications>>. Acesso em: 6 dez. 2013.

ORCID. Connecting Research and Researchers. **About**. Disponível em: <<http://orcid.org/about>>. Acesso em: 27 jan. 2016.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **OECD Principles and Guidelines for Access to Research Data from Public Funding**. Paris: OECD, 2007. Disponível em: <<https://www.oecd.org/sti/sci-tech/38500813.pdf>>. Acesso em: 15 ago. 2016.

OSTROM, E. **Governing the Commons: the evolution of institutions for collective action**. Cambridge: Cambridge University Press, 1990.

_____. **Governing the Commons: the evolution of institutions for collective action**. Cambridge: Cambridge University Press, 1990. Disponível em: <http://wtf.tw/ref/ostrom_1990.pdf>. Acesso em: 20 set. 2016.

OSTROM, E.; HESS, C. (Ed.). **A framework for analyzing the knowledge commons: Understanding Knowledge as a Commons**. Cambridge, Massachusetts, London, England: The Mit Press, 2007. Disponível em: <<https://mitpress.mit.edu/books/understanding-knowledge-commons>>. Acesso em: 20 ago. 2016.

OTLET, P. The science of bibliography and documentation. In: RAYWARD, W. B. **The international organization and dissemination of knowledge: selected essays of Paul Otlet**. Amsterdam: Elsevier, 1990.

OXFORD ENGLISH DICTIONARY. **Data**. Disponível em: <<http://www.oxforddictionaries.com/definition/english/data?q=data>>. Acesso em: 21 set. 2016.

PAESANI, Liliana Minardi. **Direito e Internet**: Liberdade de informação, privacidade e responsabilidade civil. 5.ed. Rio de Janeiro: Atlas, 2012.

PATHOSYSTEMS RESOURCE INTEGRATION CENTER (PATRIC). **Genome Metadata**. Disponível em: <<http://enews.patricbrc.org/faqs/2-genome-data-and-tools/genome-metadata-faqs/>>. Acesso em: 5 nov. 2015.

PERALTA, P. P.; SILVA, E. F. da; TERUYA, D. Y. Busca de consenso entre o direito do autor e o acesso à informação pelo público na rede de computadores: uma ótica dos tratados relativos ao direito autoral. **Perspect. ciênc. inf.**, Belo Horizonte, v. 16, n. 3, p. 116-130, set. 2011. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362011000300007&lng=en&nrm=iso>. Acesso em: 11 nov. 2015.
<http://dx.doi.org/10.1590/S1413-99362011000300007>.

PERUCCHI, V.; GARCIA, J. C. R. Indicadores de produção dos grupos de pesquisa do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 8, n. 1, p. 50-64, jan./jul. 2012.

PEZZI, Rafael Peretti. Ciência aberta: dos hipertextos aos hiperobjetos. In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015.

PIWOWAR, H. A. A new task for NSF reviewers: Recognizing the value of data reuse. **Research Remix vers**, May. 2011. Disponível em: <<https://researchremix.wordpress.com/2011/05/28/dear-nsf-reviewers/>>. Acesso em: 15 set. 2016.

PIWOWAR, H. A. et al. Big Data Sharing and Intellectual Capital Workspace. Towards a Data Sharing Culture: Recommendations for Leadership from Academic Health Centers. **PLoS Med.**, v. 5, n. 9, p. 183, 2008. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2528049/>>. Acesso em: 19 abr. 2016.

POPPER, K. R. **Conjecturas e refutações**. Brasília: Editora da UnB, 1980.

PRESERVATION METADATA MAINTENANCE ACTIVITY. Disponível em: <<http://www.loc.gov/standards/premis/>>. Acesso em: 20 jun. 2016.

PRÍNCIPE, P. **OpenAIRE integra dados dos projetos financiados pela FCT na sua infraestrutura de Acesso Aberto**. 2015. Disponível em: <<http://openaccess.sdm.uminho.pt/?p=3947>>. Acesso em: 19 abr. 2016.

REBELO, I. B. **Modelos conceituais**. Brasília: UNIEURO, 2009. Disponível em: <<https://irlabr.wordpress.com/apostila-de-ihc/parte-1-ihc-na-pratica/modelos-conceituais/>>. Acesso em: 20 jul. 2016.

RICHARDS, N. **Intellectual privacy: rethinking civil liberties in the digital age**. New York: Oxford University Press, 2015.

RODRIGUES, N. A. Introdução ao METS: Preservação e Intercâmbio de Objetos Digitais. **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, n. 26, p. 1-16, 2008.

SALES, L. F.; SAYÃO, L. F. Repositório do Instituto de Engenharia Nuclear: curadoria digital, publicações ampliadas e gestão de pesquisa. In: Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação, 25., 2013, Florianópolis. **Anais eletrônicos...** Florianópolis FEBAB, 2013. p. 982-986. Disponível em: <<https://portal.febab.org.br/anais/article/view/1315>>. Acesso em: 28 jan. 2016.

SALES, L. F.; SAYÃO, L. F.; SOUZA, R. F. Publicações ampliadas: um novo modelo de publicação acadêmica para o ambiente de *e-science*. In: Encontro Nacional de Pesquisa em Ciência da Informação, 14., 2013, Florianópolis. **Anais eletrônicos...** Disponível em: <<http://enancib.ibict.br/index.php/enancib/xivenancib/paper/viewFile/4393/3516>>. Acesso em: 20 jan. 2016.

SANTOS, P. L. V. A. C.; SIMIONATO, A. C.; ARAKAKI, F. A. Definição de metadados para recursos informacionais: apresentação da metodologia BEAM. **Inf. Inf.**, Londrina, v. 19, n. 1, p. 146-163, abr. 2014. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/15251>>. Acesso em: 11 set. 2016.

SÃO PAULO (Estado). **Guia de incentivo ao (re)uso de dados abertos**. São Paulo: Núcleo de Informação e Coordenação do Ponto BR (NIC.br), (CGI.br), Governo do estado de São Paulo, Brasília: Embaixada Britânica, 2016. Disponível em: <<http://www.cgi.br/publicacao/guia-de-incentivo-ao-reuso-de-dados-abertos/>>. Acesso em: 11 set. 2016.

_____. **Decreto nº 58.052, de 16 de maio de 2012**. Regulamenta a Lei federal nº 12.527, de 18 de novembro de 2011, que regula o acesso a informações, e dá providências correlatas. São Paulo, 16 de maio de 2012. Disponível em: <<http://www.legislacao.sp.gov.br/legislacao/dg280202.nsf/5fb5269ed17b47ab83256cfb00501469/0d8cf8dcbd4ef45f83257a010046ef75?OpenDocument>>. Acesso em: 21 set. 2016.

SAYÃO, L. F.; SALES, L. F. Curadoria Digital: um novo patamar para preservação de dados digitais de pesquisa. **Inf. & Soc.:Est.**, João Pessoa, v. 22, n. 3, p. 179-191, set./dez. 2012. Disponível em: <<https://infobci.wordpress.com/category/informacao-sociedade-estudos/>>. Acesso em: 21 set. 2016.

_____. Dados de Pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Tendências da Pesquisa Brasileira em Ciência da Informação**, Belo Horizonte, v. 6, n. 1, 2013. Disponível em: <<http://inseer.ibict.br/ancib/index.php/tpbci/article/viewArticle/102>>. Acesso em: 21 set. 2015.

SERINGHAUS, M. R.; GERSTEIN, M. B. Publishing perishing? Towards tomorrow's information architecture. **BMC Bioinformatics**, v. 8, n. 17, Jan. 2007. Disponível em:

<<http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-8-1>>. Acesso em: 15 abr. 2016.

SHERPA-JULIET. **Research funders' open access policies**. Disponível em: <<http://www.sherpa.ac.uk/juliet/index.php>>. Acesso em: 19 abr. 2016.

SHERPA-ROMEO. **Publisher copyright policies & self-archiving**. Disponível em: <<http://www.sherpa.ac.uk/romeo/index.php>>. Acesso em: 19 abr. 2016.

SHUTTLEWORTH, M. What Is a Paradigm?. **Philosophy of Science, Explorabe.com**, Oct. 24, 2008. Disponível em: <<https://explorable.com/what-is-a-paradigm>>. Acesso em: 11 set. 2015.

SILVA, E. M. **O Pacto pela Vida no estado de Pernambuco**: informação, política e poder João Pessoa. 2016. 230 f. Tese (Doutorado) - Curso de Ciência da Informação, Universidade Federal da Paraíba, Paraíba, 2016.

SILVA, E. M.; GARCIA, J. C. R. Política de Informação Científica e Tecnológica no Brasil. **DataGramaZero**, Rio de Janeiro, v. 15, n. 4, ago. 2014. Disponível em: <<https://infobci.wordpress.com/2014/10/09/politica-de-informacao-cientifica-e-tecnologica-no-brasil/>>. Acesso em: 11 set. 2016.

SIMON, I.; VIEIRA, M. S. O rossio não rival. **Revista USP**, São Paulo, n. 86, p. 66-77, jun./ago. 2010. Disponível em: <<http://www.revistas.usp.br/revusp/article/viewFile/13814/15632>>. Acesso em: 25 jun. 2014.

SMITH, M. Institutional Perspective on Credit Systems for Research Data. In: **For Attribution: Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop**. Washington, DC: National Academy of Sciences, 2012. p. 77-80. Disponível em: <<https://www.nap.edu/catalog/13564/for-attribution-developing-data-attribution-and-citation-practices-and-standards>>. Acesso em: 29 set. 2016.

STERN, S. From Author's Right to Property Right. **University Of Toronto Law Journal**, v. 62, 2012. Disponível em: <http://www.law.utoronto.ca/utfl_file/count/documents/stern/Stern%20-%20Authors%20-%20UTLJ.pdf>. Acesso em: 25 set. 2016.

STRAUSS, A.; CORBIN, J. **Pesquisa Qualitativa**: técnicas e procedimentos para o desenvolvimento de teoria fundamentada. 2. ed. Porto Alegre: Artmed, 2008.

SURF FOUADATION. **Digital rights**. Disponível em: <<https://www.surf.nl/en/themes/digital-rights>>. Acesso em: 19 abr. 2016.

TEETERS, J.L. et al. Data Sharing for Computational Neuroscience. **Neuroinform**, jan. 2008. Disponível em: <<http://redwood.berkeley.edu/fsommer/papers/teetersetal08.pdf>>. Acesso em: 15 maio 2015.

TENOPIR, C. et al. Data Sharing by Scientists: Practices and Perceptions. **Plos One**, v. 6, n. 6, p. 1-21, 29 jun. 2011. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>>. Acesso em: 25 ago. 2016.

TENOPIR, C.; KING, D.W. Communication patterns of engineers. [S. l.]: **Wiley-IEEE Computer Society Pr**, 2004. Disponível em: <<http://dl.acm.org/citation.cfm?id=966251>>. Acesso em: 12 set. 2015. ISBN:047148492X (Book).

THIBODEAU, K. The State of Digital Preservation: an international perspective. **Conference Proceedings**, p. 1-31, June 2005. Disponível em: <https://chnm.gmu.edu/digitalhistory/links/pdf/preserving/8_37e.pdf>. Acesso em: 12 set. 2016.

TROTTA, Rosyane. Autoralidade, grupo e encenação. **Sala Preta**, São Paulo, v. 6, p. 155-164, nov. 2006. ISSN 2238-3867. Disponível em: <<http://www.revistas.usp.br/salapreta/article/view/57305>>. Acesso em: 06 oct. 2016. doi: <http://dx.doi.org/10.11606/issn.2238-3867.v6i0p155-164>.

UNITED STATES GEOLOGICAL SURVEY. **Community for Data Integration**. Disponível em: <<https://my.usgs.gov/confluence/display/cdi/Home>>. Acesso em: 21 fev. 2016.

_____. **Supporting and Enabling USGS Data Management**. Disponível em: <<https://www2.usgs.gov/datamanagement/>>. Acesso em: 21 fev. 2016.

URQUHART, C.; LEHMANN, H.; MYERS, M. D. Putting the “theory” back into grounded theory: guidelines for grounded theory studies in information systems. **Information Systems Journal**, v. 20, n. 4, p. 357-381, maio 2010.

VALÉRIO, P. M.; PINHEIRO, L. V. R. Da comunicação científica à divulgação. **TransInformação**, Campinas, v. 20, n. 2, p. 159-169, maio/ago. 2008.

VAN DE SOMPEL, H. et al. An interoperable fabric for scholarly value chains. **D-Lib Magazine**, v. 12, n. 10, p. 1082-9873, 2006.

VAN DEN HEUVEL, C.; RAYWARD, W. B. Facing interfaces: Paul Otlet's visualizations of data integration. **Journal of the American Society for Information Science and Technology**, v. 62, n. 12, p. 2313-2326, 2011. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.21607/abstract?deniedAccessCustomisedMessage=&userIsAuthenticated=fa+lse>>. Acesso em: 11 set. 2016

VERNOOY-GERRITSEN, M. (Ed.). **Enhanced publications: linking publications and research data in digital repositories**. Amsterdam: Amsterdam University Press, 2009. Disponível em: <<http://dare.uva.nl/cgi/arno/show.cgi?fid=150723>>. Acesso em: 02 dez. 2013.

WACHOWICZ, M.; SILVA, R. O. C. Direito autoral e economia criativa: a construção de uma economia preocupada com a criatividade. **Liinc em Revista**, Rio de Janeiro, v. 7, n. 2, p. 556-572, out. 2011. Disponível em: <<http://liinc.revista.ibict.br/index.php/liinc/article/view/442>>. Acesso em: 13 abr. 2014.

WALL, M. Big Data: Are you ready for blast-off?. **BBC News**, 4 Mar. 2014. Disponível em: <<http://www.bbc.com/news/business-26383058>>. Acesso em: 25 maio 2015.

WALLIS, J. C. et al. Moving archival practices upstream: An exploration of the life cycle of ecological sensing data in collaborative field research. **International Journal of Digital Curation**, v. 3, n. 1, p. 114 – 126, nov. 2008. Disponível em: <http://www.ijdc.net/index.php/ijdc/article/view/67/46>>. Acesso em: 26 maio 2016.

WIGGINS, A. et al. Data Management Guide for Public Participation in Scientific Research. **DataONE Public participation in Scientific Research Working Group**, p. 3, Feb. 2013. Disponível em: <https://www.dataone.org/sites/all/documents/DataONE-PPSR-DataManagementGuide.pdf>>. Acesso em: 15 jun. 2015.

WORLD INTELLECTUAL PROPERTY ORGANIZATION. **What is Intellectual Property?** Disponível em: <http://www.wipo.int/about-ip/en/>>. Acesso em: 15 abr. 2016.

WORLD TRADE ORGANIZATION. Disponível em: <https://www.wto.org>>. Acesso em: 15 abr. 2016.

YAMAOKA, E.; GAUTHIER, J. F. O. Objetos digitais: em busca da precisão conceitual. **Inf. Inf.**, Londrina, v. 18, n. 2, p. 77-97, maio/ago. 2013. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/viewFile/16162/13077>>. Acesso em: 06 out. 2016.

ZUCOLOTO, G. F.; FREITAS, R. E. **Propriedade Intelectual e aspectos regulatórios em Biotecnologia**. Rio de Janeiro: IPEA, 2013.

APÊNDICES

APÊNDICE A - LICENÇAS LIVRES, ABERTAS E PÚBLICAS

No arcabouço jurídico, o caráter público do *commons*, *semion*, *open access* e *open science* está impulsionando as mudanças nos instrumentos de propriedade intelectual consolidados. O conceito *open* é da ciência 2.0, inspirado na *Web 2.0* que se baseou numa nova abordagem da ciência baseada no compartilhamento e colaboração em rede. A palavra *open* tem sido cada vez mais utilizada para os “modelos proprietários – com restrições para cópia, distribuição e reusos da informação. As expressões *open science*, *open access*, *open research*, *open education*, *open contents*, *open source*, *open notebook*, *open licenses*, *open courses* etc.”(MACHADO, 2015, p.210) e são utilizadas atualmente para definir as diferentes iniciativas, novos negócios e modelos contemporâneos científicos.

A iniciativa do Acesso Aberto foi lançada, em 2002, na Declaração de Budapeste. Do ponto de vista político, a publicação de Budapeste e a Declaração de Berlim estabeleceram em nível mundial as bases ao movimento de acesso aberto. O *Open Data Government Working Group* estabeleceu um conjunto de princípios que foi desenvolvido por defensores públicos da Califórnia em 2007. A finalidade foi definir um entendimento mais robusto sobre a importância dos dados governamentais abertos. O *Open Government Data Principles* estabeleceu este entendimento em oito princípios para nortear o processo de abertura de dados e a ampliação dos usos e reusos da informação com impactos também na ciência.

Quadro 1 – Oito princípios para dados abertos

PRINCÍPIOS	DEFINIÇÃO
1. Completo	Todos os dados públicos devem ser disponibilizados. Dados públicos são aqueles que não estão sujeitos às restrições de privacidade, segurança ou privilégios de acesso.
2. Primários	Os dados devem ser coletados na fonte, com o maior nível de granularidade (detalhamento) possível, não de forma agregada ou modificada.
3. Oportunidade	Os dados devem ser disponibilizados tão rapidamente quanto necessários para preservar o valor dos dados.
4. Acessibilidade	Os dados devem estar disponíveis para a mais ampla gama de usuários e os mais diversos propósitos.
5. Processável por máquinas	Os dados devem ser razoavelmente estruturados de modo a permitir o processamento automatizado.
6. Não discriminatório	Os dados devem estar disponíveis para qualquer pessoa sem necessidade de registro.
7. Não proprietário	Os dados devem estar disponíveis em um formato sobre o qual nenhuma entidade tenha o controle exclusivo.
8. Licença livre	Os dados não estão sujeitos a quaisquer direitos de autor, patentes, marcas comerciais ou regulamento secreto. Podem ser permitidas uma razoável privacidade e restrições de privilégio e segurança. A conformidade deve ser reapreciada.

Fonte: *Open Government Data Principles* (2007).

O entendimento destes princípios é voltado para os dados abertos, de forma a torná-los transparentes, acessíveis e gratuitos no âmbito governamental. Tais princípios são norteadores e podem ser aplicáveis ao contexto da ciência aberta e a diferentes naturezas de dados. Nesta direção, deve-se levar em consideração que os princípios que especificam as condições de dados públicos devem ser cumpridos para serem considerados ‘abertos e compartilhados’. Dos princípios elencados, os que merecem destaque pela sua relação com os direitos de propriedade intelectual são os princípios primeiro, segundo, sétimo e oitavo, que estão explicados brevemente a seguir.

O **primeiro princípio** refere-se aos dados completos abertos e públicos. Estabelece que todos os dados públicos devem ser disponibilizados por meio de um *dataset*, sem restrições de privacidade, segurança ou privilégios de acesso. No modelo tradicional da ciência, a disponibilidade de dados abertos gera um embate com os direitos de propriedade intelectual estabelecidos como vantagem competitiva para autores ou grupos de pesquisa, muito embora, *a priori*, a disponibilidade dos dados mantenha a garantia de que os autores “obtenham o primeiro crédito pelo uso dos dados” (MACHADO, 2015, p.215).

O **segundo princípio** refere-se aos dados primários. Os dados devem ser coletados na fonte, com o maior nível de detalhamento possível, sem alterações e agregações com outros dados ou documentos. Por suas características, “dados em formatos abertos oferecem uma grande vantagem para a reutilização, enquanto dados agregados restringem as possibilidades de uso da base” (MACHADO, 2015, p.215). No âmbito governamental, a manipulação errada pode mascarar a veracidade e fidedignidade dos dados. No aspecto que tange a cientificidade, a coleta e a manipulação dos dados devem obedecer a critérios metodológicos adequados, mantendo o rigor exigido pela ciência.

O **sétimo princípio** está relacionado com os dados não proprietários. Os dados devem estar disponíveis em um formato sobre o qual nenhuma entidade tem o controle exclusivo, considerando-se, assim, de código aberto. Este princípio é especialmente importante em áreas que estabelecem um padrão de *software*. Para evitar restrições, é sempre recomendável o uso de formatos livres/não proprietários, como as plataformas gratuitas de gerenciamento de periódicos e repositórios eletrônicos.

O **oitavo princípio** está relacionado à licença livre. Os dados não estão sujeitos a quaisquer direitos de autor, patentes, marcas comerciais ou regulamento secreto, muito embora possam ser permitidas uma razoável privacidade e restrições de privilégio e segurança. Explica Machado (2015, p.215) que na “prática, dados precisam estar protegidos com algum tipo de licença para que não sejam apropriados por outra parte que venha

restringir a cadeia de inovação com base nessa informação”. A atribuição de uma licença é necessária para estabelecer a proteção autoral para os conteúdos que estão sob domínio público. Na academia “a adoção de licenças continua sendo algo bastante alternativo, em que pesem bons exemplos, que usam Creative Commons como o do PLoS (Public Library of Science) (CC by 4.0) e o SciELO (CC by NC-SA 3.0)” (MACHADO, 2015, p.215).

Em 2010, foram lançados novos princípios específicos para dados abertos na ciência (*Open Data in Science*) estabelecidos pelo *Panton Principles*. Estes princípios focam no licenciamento de dados abertos da ciência, os quais garantem, por meio da *Internet* pública, a disponibilidade do compartilhamento, da distribuição, do reuso e da produção de obras derivadas para qualquer usuário que possa baixar, copiar, analisar, reprocessar, reutilizar, transferir para *software* ou usar para qualquer outra finalidade sem outras barreiras financeiras, jurídicas ou técnicas, ao invés de seu uso para aqueles inseparáveis do acesso à própria internet. (PANTON PRINCIPLES, 2010). Os princípios estabelecidos pela *Panton Principles* estabelecem recomendações para a aplicação dos critérios legais aos dados abertos oriundos da ciência.

Quadro 2 - Princípios para os dados abertos em ciência

PRINCÍPIOS		RECOMENDAÇÕES
1.	Onde as coleções de dados ou dados são publicadas, é fundamental que eles serão publicados com uma declaração clara e explícita dos desejos e expectativas dos editores no que diz respeito à reutilização e a uma reafirmação de propósito de elementos de dados individuais, que constituem-se em toda a coleta de dados e subgrupos da coleção. Esta declaração deve ser precisa, irrevogável e baseada em uma instrução jurídica adequada e reconhecida sob a forma de uma renúncia ou licença. Ao publicar dados, fazer uma declaração explícita e contundente de seus desejos.	Ao publicar dados fazer uma declaração explícita e robusta de seus desejos.
2.	Muitas licenças amplamente reconhecidas não se destinam e não são apropriadas para os dados ou coleções de dados. Uma variedade de isenções e licenças que são projetadas para e adequadas para o tratamento de dados é descrita aqui. Licenças <i>Creative Commons</i> (além de CCZero), GFDL, GPL, BSD, etc., não são apropriadas para os dados e a sua utilização é fortemente desencorajada. Use uma isenção ou licença reconhecida que é apropriada para os dados.	Use uma renúncia ou licença que é apropriada para dados.
3.	O uso de licenças que limitam a reutilização comercial ou limitam a produção de trabalhos derivados, excluindo o uso para fins particulares ou por pessoas ou organizações específicas é fortemente desencorajado. Se os dados forem usados de forma eficaz e adicionados por outros, eles deveriam ser abertos como definido pelo <i>Open Knowledge/Data Definition</i> - especificamente as cláusulas não comerciais e outras restritivas não deveriam ser usadas.	Se você quiser que seus dados sejam utilizados de forma eficaz e adicionados por outros, ele deve ser aberto, tal como definido pela <i>Open Knowledge / Definição de Dados</i> - em particular as cláusulas restritivas não comerciais e outras não devem ser usadas.
4.	Além disso, na ciência é altamente recomendável que os dados, especialmente quando com financiamento público, sejam explicitamente colocados no domínio público através da utilização da Dedicção de Domínio Público e Licença <i>Creative Commons</i> .	Dedicção explícita de dados subjacentes da ciência publicada para o domínio público através <i>PDDL</i> ou <i>CCZero</i> é fortemente

	<i>Commons Zero Waiver</i> . Isto está de acordo com o financiamento público da maioria da pesquisa científica e do <i>ethos</i> geral do compartilhamento e reutilização dentro da comunidade científica. Dedicção explícita de dados subjacentes à ciência publicada no domínio público via PDDL ou CCZero é fortemente recomendada e garante conformidade com tanto o protocolo <i>Science Commons</i> para implementar os dados de acesso aberto quanto o <i>Open Knowledge/Data Definition</i> .	recomendado e garante a conformidade com o Protocolo de ambos <i>Science Commons</i> para Implementação <i>Open Data Access</i> e do Conhecimento Aberto / Definição de dados.
--	---	--

Fonte: *Panton Principles* (2010).

A *Panton* disponibiliza também o *Protocol for Implementing Open Access Data*, um protocolo definido aqui que se destina a estar em conformidade com a Definição *Open Knowledge* e estender as ideias da Declaração de Budapeste para dados e bases de dados. Disponibiliza um padrão para a *ciência Commons Open Access Mark* e metadados de dados, usando as ferramentas legais e normas que estejam em conformidade com o protocolo especificado. As recomendações baseadas nos princípios do *Panton Principles* serão analisadas para a adoção no modelo proposto.

Outra iniciativa foi definida pela *Open Knowledge Foundation* (2014) com definições que abrangem aspectos técnicos, legais e procedimentais no uso e distribuição da informação. Machado (2015, p.210) apresenta no Quadro 3 uma síntese composta por 11 (onze) itens.

Quadro 3 – Aspectos técnicos que dever ser considerados para os dados abertos

1. Acesso	A obra deve ser disponibilizada na íntegra, em uma forma utilizável e modificável, por um preço que não exceda o custo razoável de reprodução; preferencialmente gratuita na <i>Internet</i> .
2. Redistribuição	A licença não deve restringir a possibilidade de venda ou distribuição da obra em si ou como parte de um pacote com obras de fontes diversas. A licença não deve exigir pagamento de direitos ou outra taxa para venda ou distribuição.
3. Reutilização	A licença deve permitir modificações e obras derivadas; sejam distribuídas sob as mesmas condições da obra original.
4. Ausências de restrições tecnológicas	Não deve haver restrições tecnológicas. A disponibilização da obra deve estar em formato cuja especificação seja disponível pública e gratuitamente, e cujo uso não esteja sujeito a restrições monetárias ou outras.
5. Atribuição	Como condição para a redistribuição e a reutilização, a licença pode exigir atribuição da autoria – desde que de forma não onerosa – aos autores da obra.
6. Integridade	É aceitável que, como condição para a distribuição da obra, se exija, no caso de sua modificação, nome ou número de versão diferente daquele da obra original.
7. Não discriminação de pessoas ou grupos	A licença não discriminará indivíduos ou grupos de indivíduos.
8. Não discriminação de domínios de atividade	A licença não pode restringir o uso da obra a uma área de atividade específica.
9. Distribuição da Licença	Os direitos da obra devem ser aplicados na redistribuição da mesma, sem necessidade de uma licença adicional.
10. A licença não deve ser específica de um pacote	Os direitos à obra não devem depender de sua inserção em um determinado pacote. Cada obra deve ter os mesmos direitos que o conjunto do pacote.
11. A licença não deve restringir a distribuição de outras obras	A licença não deve restringir outras obras que sejam distribuídas conjuntamente com a obra licenciada. Por exemplo, a licença não deve impor que todas as obras distribuídas pelo mesmo meio sejam abertas.

Fonte: Adaptado da *Open Knowledge Foundation* (OKF, 2014) e Machado (2015, p. 210).

Estes itens foram desenvolvidos pela *Open Knowledge Foundation* para orientar o uso e a aplicação das licenças *Creative Commons* no contexto aberto e gratuito. Dos itens apresentados, os que serão analisados com o presente estudo estão relacionados com o acesso e restrições autorais para dados de pesquisa abertos. Machado (2015, p. 212) afirma que para o contexto contemporâneo “o conceito passou a ter seu uso ampliado, passando a incluir dados científicos ou mesmo de organizações privadas. Seus princípios afirmam que qualquer dado, para ser aberto, deve ser utilizado por qualquer um para qualquer propósito”. Tal definição visa orientar o processo de abertura de dados de modo que possa ser considerado aberto.

As licenças *Creative Commons* são os instrumentos jurídicos adotados às publicações eletrônicas e se estenderam aos domínios dos dados de pesquisa, *software* livre e marcas, além das políticas e diretrizes gerais internacionais que são estabelecidas pelos acordos e convenções que servem de balizadores jurídicos para os direitos de propriedade intelectual. No que diz respeito ao licenciamento, Sayão e Sales (2015, p. 58) afirmam que “dados de pesquisa podem ser licenciados, portanto a forma como você licencia seus dados pode determinar suas possibilidades de uso por parte de outro pesquisador. A **Creative Zero License** proporciona um acesso bastante amplo.

Dowbor (2014, p.6) afirma que “*Massachusetts Institute of Technology (MIT)* coloca a produção de todos os seus cientistas *on-line*, em *Creative Commons*, e teve, em poucos anos, mais de 50 milhões de textos científicos acessados. No Brasil, estamos impressionantemente atrasados [...]”. A China acompanhou e criou o *China Open Resources for Education (Core)*. Harvard se lançou no EDX, também de acesso aberto *on-line*. O *Mass Open Online Courses (Mooc)*, com cursos gratuitos *on-line* e certificados prestigiosos como de Princeton, hoje, envolve mais de três milhões de participantes só no segmento Coursera.

Como é possível perceber, muitas iniciativas já utilizam diferentes instrumentos legais para a disponibilidade dos seus recursos digitais. Entretanto, são as licenças *Creative Commons* as mais usuais pela sua flexibilidade e popularidade. Tais licenças atendem, primordialmente, ao princípio do acesso aberto, livre e gratuito.

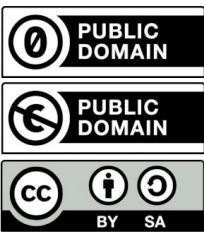
As licenças *Creative Commons* surgiram, a partir do movimento mundial denominado de acesso aberto (*open access*), com o propósito de estabelecer um conceito de aberto a *Open Knowledge Foundation (OKF)*. Essa fundação desenvolveu uma definição bem abrangente, passando a nortear os movimentos em prol do *open access*. *Creative Commons (CC)* é uma organização sem fins lucrativos que permite o compartilhamento e uso da criatividade e do conhecimento através de instrumentos jurídicos gratuitos. Ao mesmo tempo em que é a

nomenclatura utilizada para denominar as licenças públicas de direitos autorais, as “licenças CC permitem você alterar facilmente os seus termos de direitos autorais do padrão de ‘todos os direitos reservados’ para ‘alguns direitos reservados’” (CREATIVE COMMONS BR, 2015, *on-line*).

A finalidade da organização *Creative Commons* é desenvolver, apoiar e administrar uma infraestrutura jurídica e técnica que maximiza a criatividade digital, o compartilhamento e a inovação. O *Creative Commons* é um projeto mundial de gestão de direitos autorais e que “tem como missão oferecer um sistema de licenciamento público. [...] por meio do qual os autores podem licenciar alguns de seus direitos ao público em geral” (KROKOSZ, 2015, p. 31). As licenças *Creative Commons* não são contrárias aos direitos de autor. São de direitos autorais livres e fáceis de usar, fornecendo uma maneira simples e padronizada para dar ao público permissão para compartilhar e utilizar o seu trabalho criativo sob condições de escolhas. Sua identificação e atribuição estão representadas por símbolos e abreviações que recomendam a sua aplicação para diferentes suportes e usos demonstrado no Quadro 4.

Quadro 4 - Representação simbólica e descritiva das licenças *Creative Commons*

SÍMBOLO	ATRIBUIÇÃO	APLICAÇÃO
 BY	Atribuição CC BY	Esta licença permite que outros distribuam, remixem, adaptem e criem a partir do seu trabalho, mesmo para fins comerciais, desde que lhe atribuam o devido crédito pela criação original. É a licença mais flexível de todas as licenças disponíveis. É recomendada para maximizar a disseminação e uso dos materiais licenciados.
 BY SA	Atribuição – Compartilha Igual CC BY-SA	Esta licença permite que outros remixem, adaptem e criem a partir do seu trabalho, mesmo para fins comerciais, desde que lhe atribuam o devido crédito e que licenciem as novas criações sob termos idênticos. Esta licença costuma ser comparada com as licenças de software livre e de código aberto “copyleft”. Todos os trabalhos novos baseados no seu terão a mesma licença, portanto quaisquer trabalhos derivados também permitirão o uso comercial.
 BY NC	Atribuição - Sem Derivações CC BY-ND	Esta licença permite a redistribuição, comercial e não comercial, desde que o trabalho seja distribuído inalterado e no seu todo, com crédito atribuído ao autor.
 BY NC SA	Atribuição – Não Comercial CC BY-NC	Esta licença permite que outros remixem, adaptem e criem a partir do seu trabalho para fins não comerciais e, embora os novos trabalhos tenham de lhe atribuir o devido crédito e não possam ser usados para fins comerciais, os usuários não têm de licenciar esses trabalhos derivados sob os mesmos termos.
 BY NC SA	Atribuição – Não Comercial – Compartilha Igual CC BY-NC-SA	Esta licença permite que outros remixem, adaptem e criem a partir do seu trabalho para fins não comerciais, desde que atribuam a você o devido crédito e que licenciem as novas criações sob termos idênticos.

 BY NC ND	Atribuição – Sem Derivações – Sem Derivados CC BY-NC-ND	Esta é a mais restritiva das nossas seis licenças principais, só permitindo que outros façam download dos seus trabalhos e os compartilhem desde que atribuam crédito a você, mas sem que possam alterá-los de nenhuma forma ou utilizá-los para fins comerciais.
CC BY 4.0  (for content AND data; compatible with UK-OLi2)	Atribuição - 4.0 International CC BY 4.0	Compartilhar - copiar e distribuir o material em qualquer meio ou formato; Adaptar - Remixar, transformar e construir em cima do material para qualquer finalidade, inclusive comercial. O licenciante não pode revogar essas liberdades contanto que você siga os termos da licença.
	Atribuição - CCZero CC0 BY 1.0 Atribuição - PD PD BY 1.0 Atribuição-CC CC-BY-SA	Estas licenças CC0 e Public Domain permitem aos cientistas, educadores, artistas e outros criadores e proprietários de conteúdo por direitos autorais ou protegidos contra banco de dados a renunciar a esses interesses em suas obras e dados, assim, colocá-los tão completamente quanto possível no domínio público, para que outros possam construir livremente em cima, melhorar e reutilizar as obras para quaisquer fins sem restrições sob a legislação autoral ou banco de dados. O CC0 é o 'No Rights Reserve' que efetivamente significa abrir mão de todos os direitos autorais e outros direitos similares que você tem em um trabalho e dedicar esses direitos para o domínio público.

Fonte: Baseado nas licenças *Creative Commons* (2015).

A *Creative Commons* acredita que os dados científicos devem estar disponíveis gratuitamente para todos. Esta ideia é denominada de *Open Data*. O *CC Open Data* disponibiliza ferramentas legais para serem usadas para tornar os dados e bancos de dados disponíveis gratuitamente. As implementações bem-sucedidas no ambiente de dados taxonômicos, de energia, genômicos, a pesquisa de doenças, geoespaciais, polares, e disciplinas bibliométricas que fornecem orientação aos financiadores, instituições, fundações privadas, governos, empresas e outras partes interessadas (CREATIVE COMMONS, 2015).

A licença com a atribuição CC-PD-Domínio Público (Atribuição - PD BY 1.0) pode ser usada em bancos de dados para maximizar a reutilização de dados. Quando aplicada esta licença, o efeito é o de renunciar a todos os direitos de autor, bem como os direitos conexos no banco de dados, colocando o seu conteúdo mais para o domínio público em todo o mundo. Em certos domínios, tais como ciência e governo, há razões importantes a considerar usando CC-PD. A dispensa dos direitos do autor e direitos conexos elimina toda a incerteza para os potenciais utilizadores e encoraja a reutilização máxima e compartilhamento de dados e informações de pesquisa. A CC não recomenda o uso de sua *NonCommercial* (NC) ou *NoDerivatives* (ND) licenças em bancos de dados destinados para uso acadêmico ou científico.

Outra versão mais recente é a CC BY 4.0 que pode ser usada para bancos de dados de licença sujeita aos direitos de autor e, quando aplicável, pelos direitos de banco de dados *sui*

generis. Os direitos *sui generis* de banco de dados impedem a cópia e reutilização de partes substanciais de um banco de dados, incluindo a extração frequente de partes não substanciais. No entanto, ao contrário dos direitos autorais, os direitos de banco de dados protegem o investimento, isto é, a não originalidade do fabricante.

As licenças também garantem “a segurança dos dados pode também ser necessária para **proteger os direitos de propriedade intelectual, interesses comerciais** – por exemplo, dados que serão usados para patenteamento - ou para **manter sigilo sobre dados pessoais ou para proteger informações sensíveis**” (SAYÃO; SALES, 2015, p.44, grifo nosso). Para a atribuição da licença deve ser dado o crédito adequado ao autor, fornecer um *link* para a licença, indicando, assim, se as mudanças foram feitas. Sayão e Sales (2015) argumentam que a licença *Creative Commons* não é adequada para dados, sendo a mais apropriada a licença *Open Data Commons*. Essa licença é um conjunto de ferramentas legais que provem o uso dos dados abertos. Essas licenças são blocos de construção de conhecimento aberto e os dados tornem-se úteis e utilizáveis (OPEN KNOWLEDGE, 2015, *on-line*).

A *Open Knowledge* disponibiliza licenças aplicáveis a diferentes ambientes e conjunto de dados. A *Open Database License* (ODbL) v1.0, é a licença disponibilizada para *databases*. A ODbL é um contrato de licença destinado a permitir que os usuários compartilhem livremente, modificar e utilizar este banco de dados, mantendo essa mesma liberdade para os outros. Muitos bancos de dados estão abrangidos pelos direitos de autor e, portanto, este documento licencia esses direitos. Algumas jurisdições, principalmente na União Europeia, têm direitos específicos que cobrem os bancos de dados, e assim o ODbL aborda esses direitos, também. Finalmente, o ODbL é também um acordo no contrato para os usuários deste banco de dados para agir de determinadas maneiras em troca de acesso a este banco de dados. Os *datasets* podem conter uma grande variedade de tipos de conteúdo (imagens, material audiovisual, e soa todos no mesmo banco de dados, por exemplo), e assim a ODbL regula somente os direitos sobre a base de dados, e não o conteúdo do banco de dados individualmente.

A outra licença disponível é a *2.0 What this License covers*, que possui:

- a) A licença de direitos de autor e direitos conexos aplicáveis;
- b) A licença do direito de banco de dados; e,
- c) um acordo no contrato entre o pesquisador ou depositário e o licenciante.

Disponibiliza ainda os direitos legais que abrange o banco de dados, incluindo:

a) Direitos Autorais. Quaisquer direitos de autor ou conexos no banco de dados. O copyright licenciado inclui quaisquer elementos individuais do banco de dados, mas não abrange os direitos de autor sobre os conteúdos independentes deste banco de dados.

b) Direitos de Banco de Dados que dizem respeito à extração e reutilização da totalidade ou de uma parte substancial do conteúdo. Direitos de banco de dados pode aplicar-se mesmo quando não há nenhum *copyright* sobre o banco de dados.

c) Contrato. Este é um acordo entre usuário e o licenciante para o acesso ao banco de dados. Em troca, você concorda com certas condições de utilização deste acesso.

Prevê os direitos concedidos, por exemplo para dados que já foram compartilhados:

- a) Extração e reutilização da totalidade ou de uma parte substancial do conteúdo;
- b) Criação de Bases de Dados Derivados;
- c) Criação de Bases de Dados Coletivos;

d) Criação de reproduções temporárias ou permanentes, por quaisquer meios e sob qualquer forma, no todo ou em parte, inclusive de quaisquer bancos de dados derivativos ou como parte de bancos de dados Coletivos; e

e) Distribuição, comunicação, exposição, empréstimos, colocação à disposição, ou o desempenho ao público por qualquer meio e em qualquer forma, no todo ou em parte, incluindo banco de dados de qualquer derivado ou como parte de bancos de dados coletivos. (*OPEN DATA COMMONS*, 2015)¹⁰. O teor destas licenças será analisado com a finalidade de nortear o modelo autoral proposto.

¹⁰ Open Data Commons Licenses. Disponível em: <<http://opendatacommons.org/licenses/odbl/1-0/>>. Acesso em: 9 dez. 2015.

APÊNDICE B – PADRÕES DE METADADOS COM ÊNFASE NO DOMÍNIO ESPECÍFICO

Com base no entendimento de que os padrões de metadados são essenciais para este estágio do ciclo, o presente levantamento descreve a ênfase de alguns padrões de metadados adotados para a descrição de recursos informacionais, dados de pesquisa e esquemas. Alguns dos padrões apresentados merecem destaque por possuírem a aplicação em domínios específicos da ciência.

Quadro 1 – Padrões de metadados com ênfase em domínios específicos

ORGANIZAÇÃO	PADRÕES	ÊNFASES
U.S. GEOLOGICAL SURVEY (USGS)	<i>Content Standard for Digital Geospatial Metadata (CSDGM)</i>	Ênfase em dados geoespaciais.
	Website: http://www.fgdc.gov/metadata/geospatial-metadata-standards	
	<i>Geographic Information: Metadata ISO 19115/19139.</i>	Ênfase em dados e serviços geoespaciais.
	Website: http://www.fgdc.gov/metadata/geospatial-metadata-standards#fgdcendorsesdisostandards	
	<i>Geography Markup Language (GML)</i>	Ênfase em dados geográficos (rotas, rodovias e pontes) em linguagem XML.
	Website: http://www.opengeospatial.org/standards/gml	
	<i>The Federal Geographic Data Committee (FGDC)</i>	Ênfase em padrão de metadados para conteúdos digitais geoespaciais.
KNOWLEDGE NETWORK FOR BIOCOMPLEXITY (KNB)	<i>Ecological Metadata Language (EML)</i>	Ênfase em dados ecológicos.
	Website: http://www.dcc.ac.uk/resources/metadata-standards/eml-ecological-metadata-language	
NATURAL HISTORY MUSEUM AND BIODIVERSITY RESEARCH CENTER	<i>Darwin Core - Biodiversity Information Standards (TDWG)</i>	Ênfase em espécies de museus. Um padrão para descrição de objetos que contêm coleções de espécimes de história natural e conjunto de dados de observação de espécies.
	Website: http://rs.tdwg.org/dwc/	
DataCite CONSORTIUM	<i>DataCite Metadata Schema</i>	Um esquema de padrão de metadados com propriedades para a segurança e identificação consistente de dados para citação e recuperação.
	Website: http://www.dcc.ac.uk/resources/metadata-standards/datacite-metadata-schema	
DDI ALLIANCE	<i>Data Documentation Initiative (DDI)</i>	Um padrão internacional de fácil uso para descrever dados comportamentais, sociais e ciências econômicas.
	Website: http://www.dcc.ac.uk/resources/metadata-standards/list?page=1	
GENOME METADATA OVERVIEW	<i>Genome Metadata</i>	Ênfase em metadados que são dados descritivos sobre genomas individuais. Metadados genoma em PATRIC consistem de 61 diferentes campos de metadados, chamados atributos, organizados em sete grandes categorias.

	Website: http://enews.patricbrc.org/faqs/2-genome-data-and-tools/genome-metadata-faqs/	
UNIVERSITY OF OXFORD ISA-TAB	ISA-Tab	Ênfase na coleção e comunicação de metadados complexos para experimentos “omics-based” e a combinação de tecnologias.
	Website: http://isatab.sourceforge.net	
MIBBI INITIATIVE	Minimum Information for Biological and Biomedical Investigations (MIBBI)	Ênfase em informações mínimas para várias disciplinas biológicas.
	Website: http://www.dcc.ac.uk/resources/metadata-standards/list?page=2	
CALIFORNIA DIGITAL LIBRARY (CDL)	CopyrightMD	Ênfase em um esquema de XML para metadados de direitos autorais.
	Website: http://www.cdlib.org/groups/rmg/	
ANSI/NISO	Data Dictionary - Technical Metadata for Digital Still Images (Z39.87-2006)	Ênfase em elementos de dados técnicos relevantes ao gerenciamento de imagens fixas digitais.
	Website: http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key=b897b0cf3e2ee526252d9f830207b3cc9f3b6c2c	
LIBRARY OF CONGRESS	Metadata for Images in XML Standard (MIX)	Ênfase em metadados para imagens fixas em esquema XML.
	Website: http://www.loc.gov/standards/mix/	
DUBLIN CORE INITIATIVE	Dublin Core Metadata	Ênfase em publicações e recursos web.
	Website: http://dublincore.org/documents/dces/	

Fonte: Elaborado pela autora (2015).

O *Federal Geographic Data Committee (FGDC)* é o comitê pertencente à organização governamental *U.S. Geological Survey (USGS)*, que desenvolveu o *Content Standard for Digital Geospatial Metadata (CSDGM)* estruturado nos padrões da *ISO* e com ênfase em dados geoespaciais.

A *Knowledge Network For Biocomplexity (KNB)* é uma rede de conhecimento para biocomplexidade que provê um repositório internacional destinado a facilitar a pesquisa ambiental e ecológica. A *KNB* desenvolveu o padrão *Ecological Metadata Language (EML)* com ênfase em dados ecológicos.

O padrão *Darwin Core Biodiversity Information Standards (TWDG)* possui sua ênfase em espécimes naturais. Um padrão para descrever objetos contidos na história natural de coleções de espécimes e bancos de dados de observação da espécie, desenvolvido pelo *Natural History Museum and Biodiversity Research Center*.

A *California Digital Library (CDL)* desenvolveu o padrão *Copyright MD*, com ênfase em um esquema XML para metadados de direitos autorais.

O *Genome Metadata* é um padrão com ênfase em metadados de dados descritivos sobre genomas individuais. O *Pathosystems Resource Integration Center (PATRIC)* consiste

de 61 (sessenta e um) diferentes campos de metadados, chamados atributos, organizados em sete grandes categorias.

O padrão *Dublin Core* foi desenvolvido pela *Dublin Core Metadata Initiative* (DCMI) e possui ênfase nos recursos e publicações *web*.

A *Data Documentation Initiative* (DDI) está sendo desenvolvido por um grupo internacional chamado DDI Alliance. A Aliança é uma organização autossustentável cujos membros têm uma voz no desenvolvimento da especificação DDI. É um formato importante para estruturar conjuntos de dados. O DDI é uma especificação XML para metadados em ciências sociais e outras disciplinas específicas. Esse padrão é adotado para a estruturação e agregação de publicações ampliadas, que também precisam de algum outro formato de metadados XML para estruturarem seus conjuntos de dados associados às publicações.

O *Isa Tools* é a suíte do *ISA-Tab Metadata* que tem ênfase na coleta e comunicação de metadados complexos desenvolvidos pela *Nature*.

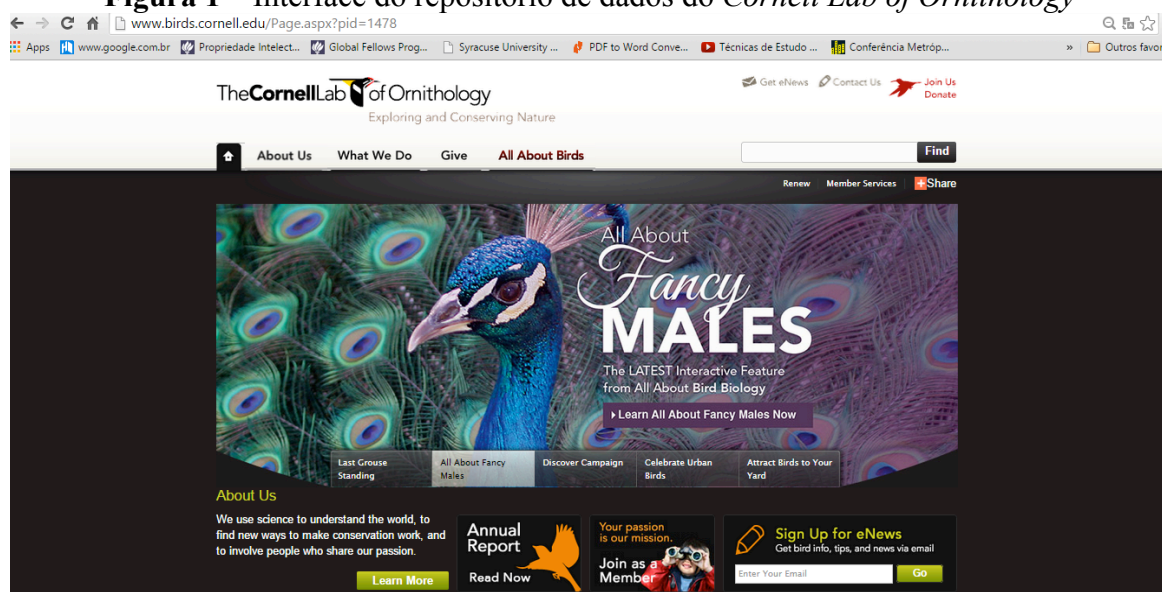
Adiciona-se, ainda, o *Preservation Metadata Maintenance Activity* (PREMIS) é um dicionário de dados para a preservação de metadados e é o padrão internacional para metadados que apoia a preservação de objetos digitais e assegura a sua usabilidade em longo prazo.

O *Statistical Data and Metadata Exchange* (SDMX) é uma iniciativa para promover normas para o intercâmbio, agregação de dados e ferramentas de repositórios de informações estatísticas oficiais.

APÊNDICE C – REPOSITÓRIOS DE DADOS DIGITAIS

O *Cornell Lab of Ornithology* é líder mundial no estudo, valorização e conservação de aves. Possui uma comunidade atuante que *inclui 200.000 (duzentos milhões) de participantes da ciência-cidadã de todas as esferas da vida e 12 (doze) milhões de entusiastas de pássaros de todas as idades que se conectam com o laboratório on-line acerca de tudo dos pássaros.* (CORNELL LAB, 2015, on-line). A Figura 36 apresenta a interface do repositório *Cornell Lab*.

Figura 1 – Interface do repositório de dados do *Cornell Lab of Ornithology*



Fonte: Cornell Lab of Ornithology (2015).

Os dados disponíveis apresentam diferentes formatos e mídias (textos, planilhas, áudio, vídeo, mapas) integrados em uma única interface aberta e *on-line*.

No Brasil, a incidência maior de repositórios institucionais está associada às universidades como um instrumento para disponibilizar a sua produção intelectual. Além das universidades, alguns institutos de pesquisa possuem iniciativas de promover o acesso aberto às publicações científicas e oferecem serviços de apoio. Ainda não há uma realidade consolidada para repositórios especializados em dados de pesquisa abertos, mas existem iniciativas de serviços de preservação digital, diretórios e repositórios de acesso aberto fomentados por institutos de pesquisa e governo federal.

A Rede Brasileira de Serviços de Preservação Digital – CARINIANA - é promovida pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). A rede Cariniana é responsável pela preservação digital compartilhada da informação científica, tecnológica e

cultural. Busca disponibilizar serviços de preservação digital para instituições com publicações de acesso livre, além de mecanismos que facilitem a automatização dos processos de identificação, armazenamento, validação e conversão para novos formatos digitais. (IBICT, CARINIANA, 2015).

O IBICT também promove o Sistema para Construção de Repositórios Institucionais Digitais (*DSpace*). O *DSpace* foi customizado pelo IBICT para possibilitar a criação de repositórios digitais com funções de armazenamento, gerenciamento, preservação e visibilidade da produção intelectual, permitindo sua adoção por outras instituições em forma consorciada federada. Os repositórios *DSpace* permitem o gerenciamento da produção científica em qualquer tipo de material digital, dando-lhe maior visibilidade e garantindo a sua acessibilidade ao longo do tempo (IBICT, *DSPACE*, 2015). A interface do *DSpace* está demonstrada na Figura 37.

Figura 2 – Interface do repositório *DSpace*



Fonte: IBICT (2015).

Apesar de não ser um repositório específico para dados de pesquisa abertos, o escopo já apresenta uma gama variável de materiais digitais, tais como: documentos (artigos, relatórios, projetos, apresentações em eventos, etc.), livros, teses, programas de computador; publicações multimídia, notícias de jornais, bases de dados bibliográficas, imagens, arquivos de áudio e vídeo, coleções de bibliotecas digitais, páginas *Web*, entre outros.

O Diretório de Políticas de Acesso Aberto das Revistas Científicas Brasileiras (Diadorim) é outra iniciativa do IBICT. É um diretório com serviços de informações

da política editorial das revistas científicas brasileiras sobre o armazenamento dos seus artigos em Repositórios Institucionais de Acesso Aberto (RI). Seu objetivo é auxiliar na localização de informações sobre a política editorial das revistas sobre a permissão para o armazenamento de artigos em RI. É um serviço direcionado para autores, editores de revistas e gestores de repositórios institucionais (DIADORIM, 2015).

O Portal Brasileiro de Acesso Aberto à Informação Científica (oasisbr) é outra iniciativa promovida pelo IBICT. O portal oasisbr é um mecanismo de busca multidisciplinar que permite o acesso gratuito à produção científica de autores vinculados a universidades e institutos de pesquisa brasileiros. Por meio do oasisbr, é possível também realizar buscas em fontes de informação portuguesas. A partir de uma única interface, é possível realizar buscas simultâneas em revistas científicas, repositórios institucionais, repositórios temáticos, bibliotecas digitais de teses e dissertações e outras fontes de informação de natureza científica e tecnológica ou academicamente orientada (OASISBr, 2015).

Outra iniciativa brasileira é o Portal da Biodiversidade que disponibiliza os dados de biodiversidade; esse é mantido pelo Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio) e voltado à sociedade brasileira. Possibilita o acesso público via *internet* e a utilização de dados de biodiversidade atualmente disponíveis nos bancos de dados do ICMBio, a partir de buscas textuais e geoespaciais, visualização e *download* de registros de ocorrência de espécies.

O Portal já conta com mais de um milhão de registros de ocorrências (coordenadas geográficas) de espécies, resultantes da integração de nove das bases de dados mantidas pelo ICMBio. O portal é uma iniciativa conjunta lançada em 2015 a partir de bases de dados de Biodiversidade do ICMBio integradas no portal (ICMBIO, 2015).

A seguir, serão expostos alguns repositórios em evidência no cenário internacional desenvolvidos para a ambiência dos dados científicos. No Quadro 9 se apresenta uma sinopse de alguns repositórios especializados para a preservação de dados e a sua ênfase.

Quadro 1 – Repositórios especializados em dados científicos abertos

ORGANIZAÇÃO	REPOSITÓRIOS ESPECIALIZADOS
University of New Mexico (UNM)	<i>LoboVault</i> é um repositório institucional que abriga publicações acadêmicas, teses e dissertações de estudantes de pós-graduação, e registros administrativos e acadêmicos da UNM. Possui as seguintes funcionalidades: autoarquivamento que ajudar a negociar e manter seus direitos autorais; utiliza a plataforma do <i>software</i> livre <i>DSpace</i> e possui uma interface específica por coleções e de dados. Website: https://www.dataone.org/software-tools/tags/repository
Oak Ridge National Laboratory e Distributed Active Archive Center	O repositório de ORNL DAAC é para a dinâmica biogeoquímica vinculado aos centros de dados da NASA Earth Observing System Data and Information System (EOSDIS) geridos pelo Earth Science Data and Information System (ESDIS) Project. Oferece sistema de busca de metadados Mercury que fornece opções para a palavra-chave,

(ORNL DAAC) NASA	dados espaciais e pesquisas temporais de dados. Possui dados de validação de terreno, Regional e Global, e dados de pesquisas ecológicas de longo prazo oriundos das redes de organização de estações de campo biológico. Possui as seguintes funcionalidades: arquivamento de dados; desenvolvimento de produto; distribuição e suporte ao usuário para dados e modelos biogeoquímicos e ecológicos; gerenciamento de dados de missões <i>NASA's Earth Science Missions</i>. Website: https://daac.ornl.gov/get_data.shtml
University of California Curation Center of the California Digital Library	<i>Merritt</i> é um serviço geral de repositório que permite gerenciar, arquivar e compartilhar conteúdo digital valioso com o propósito de apoiar a curadoria com as funções de preservação e acesso para toda a comunidade UC. Fornece recursos significativos para objetos digitais: armazenamento permanente; acesso via URLs persistentes; ferramentas para a gestão em longo prazo; metadados descritivos para acompanhar cada objeto digital, coleções digitais e apresentação de relatórios significativos. Website: https://merritt.cdlib.org/help/about_merritt
DataONE Dash	É uma nova ferramenta de <i>self-service</i> para os pesquisadores que descrevem, fazem <i>upload</i> e compartilham seus dados de pesquisa via <i>ONEShare</i>, membro do repositório da rede <i>DataONE</i>. Ele é projetado para ser uma ferramenta simples curadoria de <i>self-service</i> para os investigadores. Possui as seguintes finalidades: prepara dados para curadoria; seleciona dados para curadoria através de uma operação de procura de arquivos ou <i>drag-and-drop</i> local; descreve os dados em termos do esquema de metadados <i>DataCite</i>; identifica os dados com um identificador persistente digital do objeto (DOI) para citação permanente e descoberta; preserva, gerencia e compartilha dados por meio do <i>upload</i> para o repositório <i>ONEShare</i>; descoberta e recuperação de dados através de pesquisa e ver facetado. Website: https://search.dataone.org/#data/query=dash/page/0
ONEShare	<i>ONEShare</i> é um repositório de dados abertos para os investigadores que não têm acesso a repositórios institucionais ou disciplinares locais, ou cujos dados não cumprem os requisitos de elegibilidade prescritivos. <i>ONEShare</i> é uma coleção pública no repositório <i>Merritt</i> integrado com seu próprio nó DataONE-Membro que é copatrocinado pela e hospedado na Universidade do Novo México. Os dados podem ser enviados e acessados a partir de <i>ONEShare</i> através da interface nativa do <i>Merritt</i>. O principal modo de interação com <i>ONEShare</i> é através do portal de dados de pesquisa de curadoria <i>DataONE</i>. Website: https://oneshare.cdlib.org/xtf/search
Dryad Digital Repository	O repositório <i>Dryad</i> é um repositório de uso e curadoria geral que faz publicações científicas com os dados subjacentes detectáveis, livremente reutilizáveis e citáveis. <i>DataDryad.org</i> oferece uma plataforma de propósito geral para uma grande diversidade de tipos de dados. Quanto à funcionalidade: integra envio de dados para uma lista crescente de revistas e a apresentação de dados de outras publicações também é bem-vinda; armazena dados usados em artigos de biociência revistos por pares sobre 150 jornais; disponibiliza a opção do autoarquivamento de dados e partes relacionadas com a publicação principal. Website: http://datadryad.org/pages/organization
Avian Knowledge Network (AKN)	O <i>AKN</i> é um repositório de apoio à conservação das aves e seus <i>habitats</i> baseados em dados e no paradigma de gestão adaptativa e a melhor ciência disponível. Sua funcionalidade: permite o compartilhamento por meio do intercâmbio de dados de monitorização de aves (BMDE); gestão de dados e ferramentas de visualização para pessoas e máquinas descobrirem, avaliarem, integrarem, analisarem, visualizarem e descarregarem dados, o acesso aos dados através de um ambiente integrado de dados federado. Website: http://www.avianknowledge.net
Cornell Lab	O <i>Cornell Lab</i> é um repositório para a conservação e uso sustentável dos recursos biológicos do planeta. A comunidade inclui 200.000 (duzentos milhões) de participantes de ciência-cidadã. A excelência científica e inovação tecnológica para avançar a compreensão da natureza e de envolver as pessoas de todas as idades na aprendizagem sobre as aves e proteger o planeta. Website: http://www.birds.cornell.edu/Page.aspx?pid=1478
Global Biodiversity Information Facility	O <i>GBIF</i> é uma infraestrutura internacional de repositório gratuito, aberto e para o futuro sustentável da ciência e da sociedade, financiada pelos governos para acesso a

(GBIF)	dados de Biodiversidade. Possui as funcionalidades: permite que qualquer pessoa, em qualquer lugar, acesse os dados sobre todos os tipos de vida na terra, compartilhados além das fronteiras nacionais por meio da <i>Internet</i>; opera através de uma rede de nós, coordenando os serviços de informação da biodiversidade dos países participantes e organizações, colaborando entre si e compartilhando habilidades, experiências e capacidade técnica. Website: http://www.gbif.org/
Data.Gov	O repositório disponibiliza dados, ferramentas e recursos para realizar pesquisas, desenvolvimento <i>web</i>, aplicações móveis e visualizações de desenhos de dados. Uma abordagem que está sendo construída com dados abertos e a colaboração cidadã. ferramentas do governo, da comunidade e de negócios através de dados governamentais abertos dos Estados Unidos acessíveis, controlados e disponíveis. Oferece uma interface com um repositório de dados abertos de coleta e análise de dados de várias fontes e <i>datasets</i> de instituições federais dos Estados Unidos; cria perfis informativos detalhados de todas as cidades nos Estados Unidos; trabalha com o padrão de metadados <i>Data.json Metadata</i> e a licença pública Creative Commons Attribution Share-Alike; permite uma interatividade com o conteúdo e diversos formatos de dados por meio de mapas, textos e <i>links</i>. Website: http://www.data.gov/
ONEMercury (DataONE)	O <i>ONEMercury</i> é o repositório de dados do <i>Data Observation Network for Earth (DataONE)</i>, com uma rede de repositórios robusta que oferece inúmeros recursos, ferramentas e boas práticas para serem utilizadas por outras instituições, parceiros e pesquisadores. O <i>DataONE</i> é uma rede de repositórios que integra diferentes instituições, ferramentas e recursos desenvolvidos inicialmente para o gerenciamento de dados observacionais da terra para a <i>NASA</i> e atualmente um largo espectro de repositórios integrados de dados. O repositório <i>OneMercury</i> disponibiliza os dados em diferentes formatos (textos, imagens, planilhas, tabelas, mapas, entre outros) permitindo o acesso imediato às fontes primárias. O <i>ONEMercury</i> possui interface de busca baseada na plataforma <i>Web</i> permite que o usuário acesse os dados, ferramentas de pesquisas adicionais e uma interface de programação do aplicativo (API). Website: https://www.dataone.org/software-tools/tags/repository
NARCIS (DANS)	O repositório <i>NARCIS</i> apresenta uma experiência de <i>Enhanced Publications</i> com dados científicos agregados ao artigo principal e a outros objetos de pesquisa em <i>datasets</i>. O Portal <i>NARCIS</i> foi constituído em 2007 e tem sido parte da <i>DANS</i> desde 2011. É o portal nacional de publicações científicas de pesquisadores da Holanda, que incluem o acesso aberto a publicações dos repositórios de todas as universidades holandesas e institutos de pesquisa, conjuntos de dados de alguns arquivos de dados, bem como descrições de projetos de pesquisa, pesquisadores e institutos de pesquisa. Disponibiliza e armazena dados de pesquisas, permitindo seu acesso livre e permanente, proporcionando compartilhamento, enriquecimento, agregação, mineração e curadoria de dados de pesquisa em larga escala por meio de <i>datasets</i>. Website: http://www.narcis.nl/?Language=en
Nature Publishing Group	A <i>Nature</i> é uma plataforma que hospeda no sítio <i>nature.com</i>, aproximadamente 80 revistas, publicadas e agregando milhões de cientistas mundialmente, com acesso aberto, publicação <i>on-line</i> contínua e uma experiência de agregação de dados científicos com o artigo principal, fornecendo possibilidades de acesso ao conjunto dos dados de pesquisas. O acesso a esses dados ocorre por meio de descritores de dados, a partir da combinação da publicação científica e as informações estruturadas em <i>datasets</i>, projetados para maximizar, reutilizar e permitir o acesso aos dados de pesquisa. Apresenta ainda funcionalidades de acesso a redes e mídias sociais. No acesso ao artigo, é possível disponibilizar o texto em PDF, métricas, citações, direito de permissões. Informações sobre os autores, contribuições e afiliações também estão disponíveis por meio de <i>hiperlinks</i>. Website: http://isacommons.org/

Fonte: Elaborado pela autora (2015).

Algumas instituições internacionais merecem destaque no desenvolvimento de ferramentas tecnológicas e nos serviços para a gestão e curadoria de dados e conteúdos

digitais, a exemplo da Universidade da Califórnia. A *University of California Curation Center* (UC3) é uma parceria criativa que reúne as competências e recursos da *California Digital Library* (CDL), os dez campi da Universidade da Califórnia (UC), e a comunidade em geral de curadoria internacional. O grupo promove a análise de colaboração, projetos e soluções para assegurar a viabilidade da curadoria em longo prazo e usabilidade de conteúdos curados digitais. (UC3, 2015). Exemplos de ferramentas e serviços desenvolvidos pela UC, incluem o *DMPTool*, o Repositório Serviço de *Merritt*, o serviço de arquivamento da *Web* (*WAS*), e a Ferramenta de Planejamento de Gerenciamento de Dados.

Outra iniciativa que merece destaque é o Repositório *ONEMercury*. O repositório *DataONE* é uma rede de repositórios que integra diferentes instituições, ferramentas e recursos. Desenvolvido inicialmente para o gerenciamento de dados observacionais da terra para a *NASA*, apresenta atualmente um largo espectro de repositórios integrados de dados.

O *ONEMercury* é o repositório de dados do *DataONE*. A interface de busca baseada em web *ONEMercury* permite que o usuário acesse os dados, ferramentas de pesquisas adicionais e uma interface de programação do aplicativo (API). A interface de busca avançada permite o acesso às partes específicas dos metadados, como o título, resumo e datas. Oferece recursos geográficos que permitem definir uma caixa delimitadora em torno de uma região geográfica de interesse e restringir a pesquisa aos conjuntos, possibilidades de buscas, relacionando questões geográficas e históricas. Permite a busca refinada com base no autor dos dados, palavras-chave utilizadas nos metadados, organizações listadas com contatos nos metadados, e o nó membro que hospeda os dados (provedor de dados). Em alguns casos, os metadados contêm *links* para baixa de dados primários de conjuntos de dados que são totalmente integrados no *DataONE*. A busca por dados permite o acesso a diferentes repositórios e *datasets* que estão integrados com o *DataONE*.

A rede de repositórios *DataONE* é robusta e oferece inúmeros recursos, ferramentas e boas práticas para serem utilizadas por outras instituições, parceiros e pesquisadores. O repositório *OneMercury* disponibiliza os dados em diferentes formatos (textos, imagens, planilhas, tabelas, mapas, entre outros) permitindo o acesso imediato às fontes primárias. A busca por dados permite o acesso a diferentes repositórios e *datasets* que estão integrados com o *DataONE*. A interface de busca no *ONEMercury* está demonstrada na Figura 38 a seguir.

Figura 3 – Interface do repositório de dados *ONEMercury*

The screenshot shows the ONEMercury DataONE Search Tool interface. The top navigation bar includes links for About, News, Participate, Resources, Education, and Data. The main content area displays search results for the query "biodiversity and author:david tilman".

Data Package Files

Identifier	Type	Size	Download
doi:10.6073/pasta/7bb33bcf03fa86d37ec6c2321b88a6	Package	4.0 kB	Download
https://pasta.knet.net/package/data/eml/knb-lter-cdr/386/1/34db2945b271c1a142d73fec2e298917	text/csv	239.3 kB	Data
https://pasta.knet.net/package/report/eml/knb-lter-cdr/386/1	text/xml	22.6 kB	Data
https://pasta.knet.net/package/metadata/eml/knb-lter-cdr/386/1	eml/econformatics.org/eml-2.1.0	12.8 kB	Metadata

Document Information:

- Document Identifier: <https://pasta.knet.net/package/metadata/eml/knb-lter-cdr/386/1>
- Data Set Citation: Tilman D. 2013. Plant aboveground biomass data BAC: Biodiversity and Climate. <https://pasta.knet.net/package/metadata/eml/knb-lter-cdr/386/1>
- Alternate Identifier: knb-lter-cdr/386.1
- Short Name: ple249
- Data Set Owner(s): David Tilman
- Individual: David Tilman
- Metadata Provider(s): Cedar Creek LTER
- Organization: Cedar Creek Ecosystem Science
- Address: 100 Ecology Building, 1987 Upper Buford Circle, Saint Paul, MN 55108 USA
- Phone: 612-625-5740 (voice)
- Phone: 612-624-6777 (fax)
- Email Address: webmaster@cedar-creek.umn.edu
- Web Address: <http://www.cedar-creek.umn.edu>

Abstract:

Climate changes forecast for or interactive effects are unknown community composition, temper in CDR grassland ecosystems. The experiment uses a subset of 42 plots from the Biodiversity experiment: 14 monoculture plots (one randomly chosen replicate for each of the 14 non-woody species that became well established in monoculture), 14 plots planted to 4 species (randomly chosen from 4 species plots planted with combination of these 14 species), and 14 plots planted to 16 species (randomly chosen from all 16 species plots containing these 14 species). Plant species are from 4 different functional groups (4 species, each, of C3 grasses, C4 grasses, legumes and non-legume forbs) that are likely to respond differently to changes in temperature and water availability. Treatments to be applied to 4 subplots (3 m x 4 m each) within each of these 42 plots, will create conditions that are (a) warmer than ambient, (b) warmer and watered, (c) ambient temperature but watered, and (d) unwarmed and un-watered, i.e., a full factorial of ambient or elevated temperature and of ambient or elevated precipitation. Temperature will be increased using infrared heat lamps (DeVogele and Harte 2001, Kimball 2005), and increased precipitation using automatic sprinklers, both methods were tested and refined in 2005.

Code of Ethics and Rules for Use of Cedar Creek LTER and Related Data

Fonte: DataONE (2015).