

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUISTA FILHO”

Faculdade de Filosofia e Ciências

Programa de Pós-graduação em Ciência da Informação

ELIZABETE CRISTINA DE SOUZA DE AGUIAR MONTEIRO

**DIREITOS AUTORAIS NOS REPOSITÓRIOS DE DADOS CIENTÍFICOS: análise  
sobre os Planos de Gerenciamento dos Dados**



Marília  
2017

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUISTA FILHO”

Faculdade de Filosofia e Ciências

Programa de Pós-graduação em Ciência da Informação

ELIZABETE CRISTINA DE SOUZA DE AGUIAR MONTEIRO

**DIREITOS AUTORAIS NOS REPOSITÓRIOS DE DADOS CIENTÍFICOS: análise  
sobre os Planos de Gerenciamento dos Dados**

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

Linha de Pesquisa: Informação e Tecnologia

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

Marília

2017

Monteiro, Elizabete Cristina de Souza de Aguiar.  
M775d Direitos autorais nos repositórios de dados científicos:  
análise sobre os Planos de Gerenciamento dos Dados /  
Elizabete Cristina de Souza de Aguiar Monteiro. – Marília,  
2017.

115 f. : 30 cm.

Orientador: Ricardo César Gonçalves Sant'Ana.  
Dissertação (Mestrado em Ciência da Informação) –  
Universidade Estadual Paulista (Unesp), Faculdade de  
Filosofia e Ciências, 2017.

Bibliografia: f. 103-111

1. Direitos autorais. 2. Tecnologia da informação. 3.  
Repositórios institucionais. 4. Sistemas de informação  
gerencial. I. Título.

CDD\_005.73

ELIZABETE CRISTINA DE SOUZA DE AGUIAR MONTEIRO

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

---

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

---

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

---

Prof. Dr. Rogério Aparecido Sá Ramalho  
Universidade Federal de São Carlos - UFSCar

Marília

2017

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

## AGRADECIMENTOS

Agradeço a Deus por toda obra feita em minha vida.

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

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

Ao meu orientador prof. Dr. Ricardo César Gonçalves Sant'Ana por todo aprendizado e confiança.

Aos profs. José Eduardo Santarém Segundo e Rogério Aparecido Sá Ramalho pelas contribuições que enriqueceram a dissertação.

Ao Grupo de Pesquisa sobre Tecnologias de Acesso a Dados (GPTAD).

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

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

A todos os funcionários da FFC que participam da construção de minha vida acadêmica.

A todos os funcionários da biblioteca da FFC que contribuíram para que concluísse as atividades do mestrado.

A todos os docentes do Programa de Pós-graduação em Ciência da Informação

Aos amigos que conquistei nessa jornada.

A ciência esta cada vez mais dependente de dados.

Fox e Hendler (2011)

## RESUMO

A enchente de dados científicos nos últimos anos apresenta desafios e novas oportunidades. Os desafios incluem a otimização no uso e reuso de dados e as oportunidades, no desenvolvimento de infraestrutura com implementação de repositórios de dados e todas as atividades inerentes a eles como gerenciamento, descrição, disseminação, integração, direitos autorais, privacidade, quando devida, qualidade e preservação de conjunto de dados. O Plano de Gerenciamento de Dados é um documento composto por diretrizes que orientam os pesquisadores na articulação para liberarem seus conjuntos de dados que ficarão depositados nos repositórios. As diretrizes descritas no Plano de Gerenciamento de Dados precisam ser explícitas em relação às questões sobre direito autoral. O licenciamento de dados estabelece claramente os termos de uso evitando problemas jurídicos futuros. As licenças *Creative Commons* e as *Open Data Commons* são usadas para licenciar dados e justificam o objetivo dessa pesquisa que consiste em investigar como os repositórios de dados das cem melhores universidades do mundo explicitam em seus Planos de Gerenciamento de Dados as recomendações relacionadas aos direitos autorais dos dados. A metodologia utilizada foi baseada na pesquisa quantitativa e qualitativa, incluindo estudo exploratório-descritivo. A coleta de dados foi realizada nos Planos de Gerenciamento de Dados dos repositórios implementados nas 100 melhores universidades do mundo. Os resultados demonstram que a maioria das universidades analisadas implementaram repositório de dados, porém não são todos os repositórios que dispõem de PGDs e de orientações relacionadas aos direitos autorais.

Palavras-chave: Repositório de dados. Direito autoral. Gerenciamento de dados. Plano de Gerenciamento de dados. Licenças de uso.



## ABSTRACT

The flood of scientific data in recent years presents challenges and new opportunities. Challenges include optimizing the use and reuse of data and opportunities, developing infrastructure with implementation of data repositories and all activities inherent to them such as management, description, dissemination, integration, copyright, privacy, where appropriate, quality and preservation of data set. The Data Management Plan is a document composed of guidelines that guide the researchers in the articulation to release their datasets that will be deposited in the repositories. The guidelines outlined in the Data Management Plan need to be explicit in relation to copyright issues. The licensing of data clearly establishes the terms of use avoiding future legal problems. Creative Commons and Open Data Commons licenses are used to license data and justify the aims of this research which consists of investigating how the data repositories of the world's top 100 universities spell out in their Data Management Plans the recommendations related to copyright of data. The methodology used was based on quantitative and qualitative research, including an exploratory-descriptive study. The data collection was done in the Data Management Plans of the repositories implemented in the 100 best universities in the world. The results show that most of the universities analyzed have implemented a data repository, but not all repositories have PGDs and copyright guidelines.

Key-words: Data repository. Intellectual property. Data management. Data management plan. Use license.

## LISTA DE GRÁFICOS

Gráfico 1– Total de universidade com Repositório de Dados .....	73
Gráfico 2 – Quantidade de Repositório de Dados por país .....	74
Gráfico 3 – Relação de Repositórios de Dados com PGDs .....	76
Gráfico 4 – PGDs com explicitação de DA .....	77
Gráfico 5 –Explicitação de licenças de uso nos PGDs .....	87
Gráfico 6 – PGDs que mencionam a licença <i>Creative Commons</i> .....	96

## LISTA DE FIGURAS

Figura 1 - Identificação do número de repositórios de dados.....	15
Figura 2 – Interesse sobre “ <i>data repository</i> ” por países do mundo .....	20
Figura 3 – Tendência de busca por “ <i>data repository</i> ” .....	21
Figura 4 - Ciclo de Vida dos Dados na Ciência da Informação (CVD-CI) .....	22
Figura 5 – Fase da coleta com o fator Direitos Autorais.....	23
Figura 6 – Repositórios Institucionais em todo o mundo .....	35
Figura 7 - Modelo de publicação científica com direitos autorais de dados .....	37
Figura 8 – Repositório de Dados no mundo .....	38
Figura 9 - Todos os dados científicos online .....	40
Figura 10 – Países signatários da Convenção de Berna .....	50
Figura 11 - Material licenciado CC remixado.....	59
Figura 12 – Tela inicial do DMPtoll.....	69
Figura 13 – Ciclo de Vida dos Dados no Repositório .....	70
Figura 14 - Ilustração do direcionamento das análises .....	71

## LISTA DE QUADROS

Quadro 1 - Paradigmas da ciência .....	42
Quadro 2 - Aspectos dos direitos autorais .....	52
Quadro 3 - Universidades que explicitam direito autoral em seus PGDs .....	78
Quadro 4 – Relação universidade X licenças de uso .....	94

## LISTA DE ABREVIATURAS

ASPI	Associação Paulista da Propriedade Intelectual
CI	Ciência da Informação
GPL	General Public License
OECD	Organisation for Economic Co-Operation and Development
PGD	Plano de Gerenciamento de Dados
WIPO	World Intellectual Property Organization

## SUMÁRIO

1	INTRODUÇÃO .....	13
1.1	Problema da pesquisa .....	18
1.2	Objetivos .....	19
1.3	Justificativa .....	19
1.4	Delimitação do tema e objeto de pesquisa .....	21
1.4.1	Ciclo de Vida dos Dados .....	22
1.4.1.1	Fases do Ciclo de Vida dos Dados .....	23
1.4.1.2	Fatores das fases do Ciclo de Vida dos Dados.....	26
1.5	Metodologia.....	29
1.6	Estrutura do texto .....	30
2	REPOSITÓRIO DE DADOS CIENTÍFICOS.....	32
2.1	Repositório de dados e a comunicação científica .....	39
2.2	Repositório de dados e e-Science .....	41
3	DIREITOS AUTORAIS DE DADOS CIENTÍFICOS.....	46
3.1	Propriedade intelectual .....	46
3.2	Direitos autorais.....	51
3.2.2	Direito e moral e Direito patrimonial.....	53
3.3	Licenças de uso.....	55
4	PLANO DE GERENCIAMENTO DE DADOS.....	65
5	ANÁLISES E DISCUSSÕES .....	70
5.1	Identificação dos repositórios de dados das universidades .....	72
5.2	Identificação dos Planos de Gerenciamento de Dados.....	75
5.3	Explicitação dos Planos de Gerenciamento de Dados com instruções sobre direitos autorais .....	76
6	CONSIDERAÇÕES FINAIS.....	99
	REFERÊNCIAS.....	103
	APÊNDICE A .....	112
	ANEXO A .....	113

## 1 INTRODUÇÃO

Os dados científicos também são apresentados como dados de pesquisa, dados brutos, dados primários e dados secundários de pesquisa pelos autores Dudziak (2016), Borgmam et al. (2015), Sales e Sayão (2012) e Tolle, Tansley e Hey (2011). Foi adotado nesta pesquisa o termo dados científicos para padronização do termo por estar de acordo com os trabalhos citados.

Dados resultantes da investigação científica de origens computacionais, experimentais, observacionais (NATIONAL SCIENCE BOARD, 2005) e a possibilidade de seu compartilhamento e reutilização têm impulsionado transformações na práxis científica. Estes fatores contribuem para o direcionamento da atenção de pesquisadores na oportunidade de compartilhamento de dados e de instituições para aplicação de recursos na implementação e manutenção de repositórios de dados científicos no apoio à infraestrutura para melhor gerenciamento dos dados (MONTEIRO; SANT'ANA, 2016).

Fox e Hendler (2011) destacam que a ciência está cada vez mais dependente de dados. De acordo com Lynch (2011) o registro científico poderia disponibilizar os dados científicos que embasaram seu desenvolvimento com a finalidade de que outros pesquisadores pudessem reaplicá-los para reproduzir resultados inéditos.

Santos e Sant'Ana (2002) destacam que dado é conceituado como “[...] um elemento básico, formado por signo ou conjunto finito de signos que não contém, intrinsecamente, um componente semântico, mas somente elementos sintáticos.”

De acordo com a *Organisation for Economic Co-Operation and Development* (OECD) (2007, p. 13, tradução nossa) dados científicos

[...] são definidos como registros factuais (números, registros textuais, imagens e sons) utilizados como fontes primárias para a pesquisa científica, e que são geralmente aceitos na comunidade científica como necessários para validar os resultados da pesquisa. Um conjunto de dados de pesquisa constitui uma representação sistemática, parcial do sujeito a ser investigado.<sup>1</sup>

---

<sup>1</sup> [...] are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated.

Dados científicos, segundo Dudziaki (2016) são “[...] os materiais comumente registrados e aceitos na comunidade científica como necessários para validar os resultados de pesquisa [...]” e, nessa ambiência, incluem diversos materiais como documentos, planilhas, cadernos de laboratório ou de campo, questionários, transcrições, áudio, vídeo, texto, imagens, fotografias, slides, sequências de proteínas ou genéticas, amostras, objetos digitais gerados durante a pesquisa, algoritmos, *scripts*, *software* de simulação, fluxo de trabalhos, procedimentos operacionais, padrões e protocolos (DUDZIAKI, 2016).

Agências de financiamento de países como Estados Unidos exigem a disponibilização dos dados científicos para financiamento dos projetos de pesquisas (BORGMAN et al., 2015). Revistas científicas como *Science*, *Nature* e *PloS* orientam os autores que vão submeter seus artigos a liberarem dados científicos no momento da publicação do artigo ou depositem em repositórios de dados.

Os dados coletados pelos pesquisadores nas diversas áreas do conhecimento estão sendo considerados como informação científica e precisam ser tratados de forma a viabilizar a sua organização, recuperação e difusão e, assim, auxiliar a pesquisa colaborativa (MONTEIRO; SANT’ANA; SANTARÉM SEGUNDO, 2016).

Capurro e Hjørland (2007, p.149) apresentam informação no sentido de conhecimento comunicado e destacam que “É lugar comum considerar-se a informação como condição básica para o desenvolvimento econômico juntamente com o capital, o trabalho e a matéria-prima, mas o que torna a informação especialmente significativa na atualidade é sua natureza digital.”

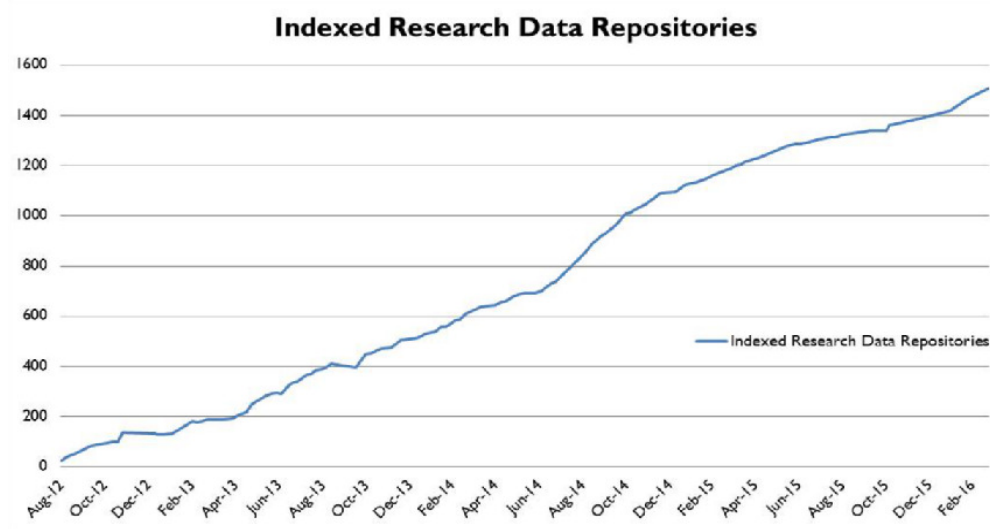
Observa-se a discussão sobre dados na Ciência da Informação (CI) no Brasil com eventos ocorridos na área como o I, II e III Encontro Internacional Dados, Informação e Tecnologia, realizados nos anos de 2014, 2015 e 2016 respectivamente, na Faculdade de Filosofia e Ciências - Unesp Campus de Marília, no ENANCIB 2013, 2014, 2015 e 2016, no GT8 com apresentação e discussão de trabalhos relacionados a dados, no *Workshop em Data Science - I Workshop Desafios no contexto contemporâneo para promover a nova ciência baseada em dados de pesquisa* realizado no Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) em 2015 (MONTEIRO, SANT’ANA, 2016), no VI Seminário em Ciência da Informação ocorrido na Universidade Estadual de Londrina no ano de



2016 e especificamente sobre Repositórios de dados no I Fórum Nacional de Repositórios digitais realizado na Universidade Federal do Rio Grande do Norte em 2016.

Infere-se, assim, que dados científicos estão adquirindo relevância nos estudos e práticas da CI, pois o foco vem sendo ampliado da informação para dados na geração de conhecimento. Concomitante ao aumento de discussões e usos de dados percebe-se, pelo gráfico 1, o aumento de repositórios de dados registrados no re3data.org<sup>2</sup>, um registro global de repositórios de dados de pesquisa que abrange repositórios de diferentes disciplinas acadêmicas. Ele apresenta repositórios para o armazenamento permanente e acesso aos conjuntos de dados para pesquisadores, instituições de financiamento, editores e instituições acadêmicas. O re3data.org promove uma cultura de partilha, maior acesso e melhor visibilidade dos dados da pesquisa.

Figura 1 – Identificação do número de repositórios de dados



Fonte: re3data.org

<sup>2</sup> Disponível em: <<http://service.re3data.org/about>>. Acesso em 25 ago. 2016.

Verifica-se no gráfico 1 que desde 2012 o registro de repositórios de dados no re3data.org aumentou chegando a margem de identificar e listar 1500 repositórios de dados disponíveis na web em fevereiro de 2016.

Dados e informações digitais gerados por atividades da comunidade científica compõem os repositórios de dados científicos e necessitam de gerenciamento (SAYÃO; SALES, 2012). Esta gestão integra o arquivamento, preservação, meios para acrescentar valor aos dados e otimizar a capacidade de reuso (SAYÃO; SALES, 2012).

Os repositórios de dados são mantidos por conjuntos de ações que viabilizam o armazenamento de dados visando à otimização da recuperação, o que amplia as potencialidades de reuso destes dados entre os pesquisadores. Desta forma, agiliza os processos de investigação e, conseqüentemente, o avanço na ciência.

Com uma infraestrutura implementada por repositórios de dados, apoiada por um Plano de Gerenciamento de Dados (PGD) bem fundamentado, os pesquisadores têm aporte propício para depósito de seus conjuntos de dados e busca e recuperação de dados já coletados por outros pesquisadores, que poderão ser reutilizados em suas pesquisas.

Infraestrutura é uma base que suporta a construção de uma estrutura. Sua preparação e implementação dependem do propósito de projetos e de suas respectivas comunidades.

De acordo com Borgman et al. (2015, p. 207, tradução nossa) infraestrutura são ecologias ou sistemas adaptativos complexos “[...] que consistem de muitas partes que interagem através de processos sociais e técnicos, com variados graus de sucesso.”<sup>3</sup> e incluem aparatos tecnológicos, recursos humanos, colaboração, acesso distribuído e informação documentada.

Infraestrutura para gerenciamento de dados depende das características e necessidades de uma comunidade e seu *know-how* para gerir, curar e fazer a análise, favorecendo o acesso a seus dados científicos de modo que outros reutilizem para desenvolvimento de novos métodos de pesquisas (BORGMAN et al.,

---

<sup>3</sup>[...] consist of many parts that interact through social and technical processes, with varying degrees of success.

2014, 2015). A visibilidade que a infraestrutura oportuniza pode tornar os dados mais acessados.

Para os documentos, as informações e instruções referentes ao gerenciamento dos dados científicos disponibilizados nos repositórios foi adotado, nesta pesquisa, o termo Plano de Gerenciamento de Dados para padronização, sendo que os sítios analisados denominaram com mais de uma forma como *Data Management Planning*, *Research Data Management*, *Data Management* e *Policy and Terms on Use*.

A análise deste contexto é permeada pelo estudo dos repositórios de dados implementados pelas universidades. As pesquisas desenvolvidas na Ciência da Informação e áreas afins forneceram embasamento teórico. Foi aplicado o Ciclo de Vida dos Dados, composto por quatro fases e seis fatores (SANT'ANA, 2016) apresentado na subseção 1.4.1. Nesta pesquisa foram adotados a fase da Coleta com o fator Direitos autorais, uma das vertentes da propriedade intelectual.

A propriedade intelectual é a área do direito que protege, por meio de leis, os inventores ou responsáveis por qualquer produção intelectual, seja no domínio industrial, científico, literário ou artístico, facultando aos seus titulares direitos econômicos os quais ditam a forma de comercialização, circulação, utilização e produção dos bens intelectuais ou dos produtos e serviços que incorporam tais criações (BRASIL, 2007; ASSOCIAÇÃO PAULISTA DA PROPRIEDADE INTELECTUAL (ASPI), c2016). A propriedade intelectual possui duas vertentes principais: Direitos autorais e Propriedade industrial.

Direito autoral é um dos ramos da ciência jurídica em que permeia “[...] um conjunto de direitos morais e patrimoniais sobre as criações do espírito, expressas por quaisquer meios ou fixadas em quaisquer suportes, tangíveis ou intangíveis, que se concede aos criadores de obras intelectuais.” (ASPI, c2016, p. 1).

O conjunto de direitos que engloba os direitos autorais inclui, mas não estão limitados, a trabalhos literários, artísticos e cultura imaterial como romances, poemas, peças de teatro, referências, filmes, música, desenhos, desenhos técnicos, fotografias, símbolos, imagens, esculturas, programas de computador, internet, entre outras obras (WORLD INTELLECTUAL PROPERTY ORGANIZATION (WIPO), [20--]).

As universidades, por meio de seus repositórios, podem definir diretrizes que permeiam a propriedade intelectual estabelecendo políticas para melhor apropriação contratual que regem as negociações entre os agentes envolvidos na produção (BASTOS; ODDONE; VIDOTTI, 2010). As universidades propiciam meios de disseminar o conjunto de dados de sua comunidade com o apoio aos repositórios de dados que potencializam a disseminação e acesso aos dados e a visibilidade e reconhecimento de seus pesquisadores.

O desenvolvimento tecnológico dos meios de comunicação está intrinsecamente ligado à evolução da proteção jurídica das obras intelectuais. O avanço científico e a necessidade de arquivamento de dados científicos em repositório de dados mudam consideravelmente a relação de dados e pesquisa, pois a ciência está, atualmente, baseada em dados.

### 1.1 Problema da pesquisa

Os últimos anos têm sido marcados por um aumento demasiado de dados científicos nas diversas áreas do conhecimento, viabilizando o processo de reuso desses dados. Verifica-se que periódicos científicos como o grupo *Science*, *Nature* e *PloS*, entre outros, estão incentivando e, em alguns casos, indicando ser obrigatório a disponibilização dos dados científicos junto ao artigo no momento da submissão ou serem depositados em um repositório de dados científicos. Com esses incentivos, há uma crescente disponibilização de dados nos periódicos científicos e um crescente número de implementação de repositório de dados e de periódicos de dados.

Nessa premissa, muitas discussões emergem na literatura científica sobre o reuso de dados, e a discussão relacionada ao direito autoral nos repositório de dados, se torna propício.

O panorama da reflexão apresentada motiva a problemática deste estudo que está baseada na seguinte questão: Os pesquisadores que vão depositar seus conjuntos de dados nos repositórios de dados têm informações relacionadas ao direito autoral?

## 1.2 Objetivos

O objetivo geral desta pesquisa é investigar como os repositórios de dados das cem melhores universidades do mundo explicitam em seus Planos de Gerenciamento de Dados as recomendações relacionadas aos direitos autorais dos dados.

Os objetivos específicos são:

- localizar as cem melhores universidades do mundo e seus repositórios de dados;
- identificar os Planos de Gerenciamento de Dados de cada repositório de dados encontrado;
- analisar os Planos de Gerenciamento de Dados localizando abordagens referentes aos direitos autorais e licenças de uso;
- evidenciar a explicitação dos direitos autorais identificados nos PGDs.

## 1.3 Justificativa

Os Repositórios de dados científicos são ambientes implementados nas universidades com infraestrutura apropriada para dar suporte aos pesquisadores no gerenciamento e na disponibilização de dados científicos e, dessa forma, contribuir para que outros pesquisadores possam reutilizá-los.

Pesquisadores podem ter uma série de preocupações ao compartilharem seus dados científicos, destacando as premissas de como serão reconhecidos pela autoria dos dados e como esses dados serão liberados para reutilização.

Discutir sobre esses aspectos evidenciando a importância do direito autoral abordado no Plano de Gerenciamento de Dados auxilia os pesquisadores na decisão em disponibilizar seus dados, pois considera as vantagens da liberação planejada e, conseqüentemente, escolha sobre qual licença de uso os dados serão disponibilizados nos repositórios de dados.

Essa pesquisa viabilizou contribuir com a Ciência da Informação ao discutir a explicitação dos direitos autorais apresentados nos Planos de Gerenciamento de Dados dos repositórios de dados.

A Ciência da Informação é um campo interdisciplinar voltado ao uso de informação nos diferentes contextos e tem como objeto de estudo a informação registrada para geração de conhecimento. Segundo Saracevic (1995) a Ciência da Informação tem três características gerais: é interdisciplinar; está ligada à tecnologia da informação e é participante ativa na evolução da sociedade da informação, com uma forte dimensão social e humana.

Observa-se na figura 2 que os países de maior interesse por repositório de dados, conforme *Google Trends*, ferramenta gratuita do *Google* que mede a frequência com que um termo é procurado na Internet, em um determinado período de tempo por determinada localização geográfica, são: Índia, Estados Unidos e Reino Unido. Os valores são calculados em uma escala de 0 a 100, em que 100 é o local com a maior popularidade e 0 indica um local em que o termo tem menos de 1% da popularidade. Foi utilizado o termo “*data repository*” pois com o termo “repositório de dados” não aparecerem resultados. Dessa forma verificou-se a importância da discussão do tema no contexto da Ciência da Informação no Brasil.

Figura 2 – Interesse sobre “*data repository*” por países do mundo

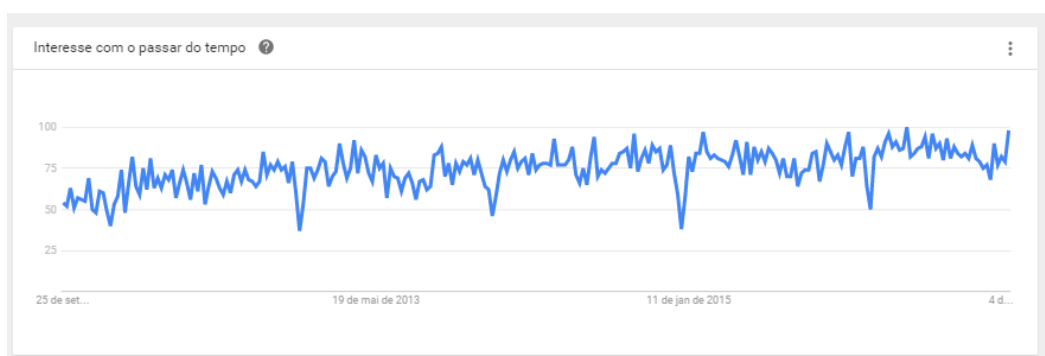


Fonte: <https://www.google.com.br/trends/explore?q=data%20repository>. Acesso em: 25 set. 2016

A geração, organização, gerenciamento, transferência, acesso, utilização e curadoria de dados e de informações nos diversos ambientes científicos, associados aos aparatos tecnológicos, em resumo, constituem alguns dos elementos decisivos para o estudo das relações entre dados, informação, tecnologia e sociedade. Sendo assim, esta pesquisa tem o intuito de contribuir com fornecedores e usuários de dados nos ambientes de repositórios de dados discutindo a relação dos direitos autorais entre as partes envolvidas.

A importância científica na realização dessa pesquisa está em fornecer subsídios com a finalidade de que os profissionais envolvidos no gerenciamento dos dados nos repositórios de dados possam aplicar Planos de Gerenciamento de Dados a fim de que todos os envolvidos participem das discussões dos direitos autorais desde a implementação dos repositórios, contribuindo para diminuir as inquietações e problemas inerentes a cessão dos dados.

Figura 3 – Tendência de busca por “*data repository*”



Fonte: <https://www.google.com.br/trends/explore?q=data%20repository>. Acesso em: 25 set. 2016

Conforme a figura 3, o interesse por repositório de dados ampliou no período de 2011 até setembro de 2016, chegando ao pico de popularidade em 2016, confirmando a relevância e importância de se discutir o tema no campo científico.

A proposta de estudar a relação dos direitos autorais no ambiente de repositório de dados advém de interesse particular instigado pelas leituras relacionadas aos temas e as discussões envolvidas com a preocupação dos pesquisadores em disponibilizar seus dados e não serem reconhecidos pelo seu trabalho.

#### 1.4 Delimitação do tema e objeto da pesquisa

O tema do estudo foram os direitos autorais abordados pelos repositórios de dados científicos implementados nas cem melhores universidades do mundo e disponibilizados às suas comunidades. O objeto da pesquisa foi os Planos de Gerenciamento de Dados.

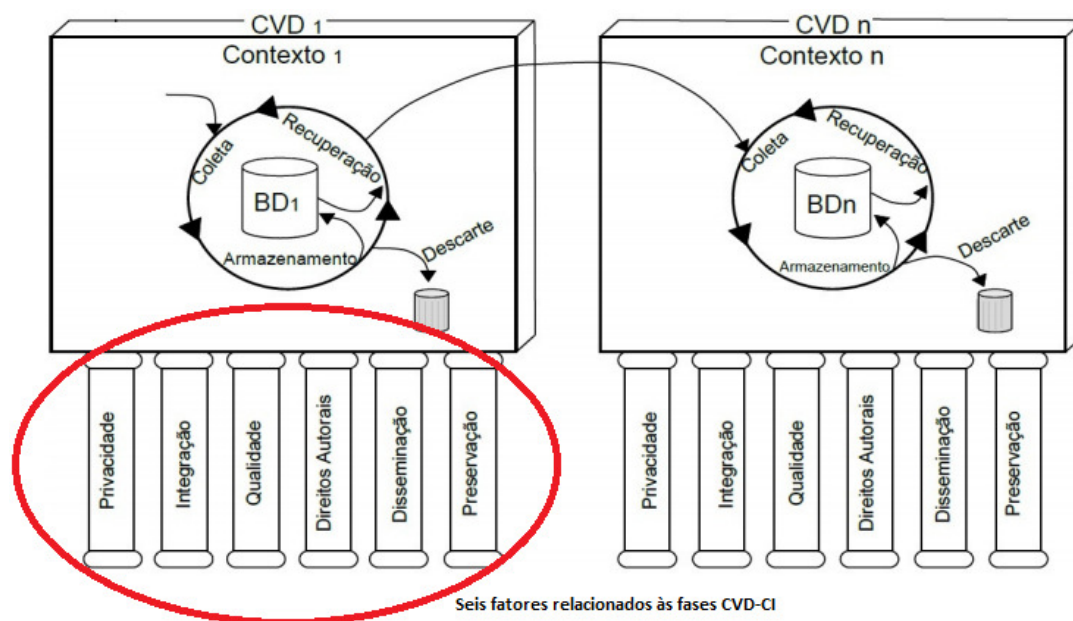
O Ciclo de Vida dos Dados proposto por Sant'Ana (2016) foi utilizado para verificar a explicitação dos direitos autorais nos repositórios de dados.

Os resultados apresentados abrangeram o Plano de Gerenciamento de Dados dos repositórios de dados das cem melhores universidades do mundo. O escopo da pesquisa foi as cem melhores universidades por verificar que a maioria delas implementaram repositórios de dados para sua comunidade.

#### 1.4.1 Ciclo de Vida dos Dados

O Ciclo de Vida dos Dados na Ciência da Informação apresentado por Sant'Ana (2016), é um modelo que leva em conta características e especificidades da Ciência da Informação, partindo-se de pesquisa realizada sobre modelos de ciclo de vida de dados de áreas correlatas e propõe uma estrutura de atividades composta por quatro fases: Coleta, Armazenamento, Recuperação e Descarte passando por seis fatores: Preservação, Disseminação, Direitos Autorais, Qualidade, Integração e Privacidade.

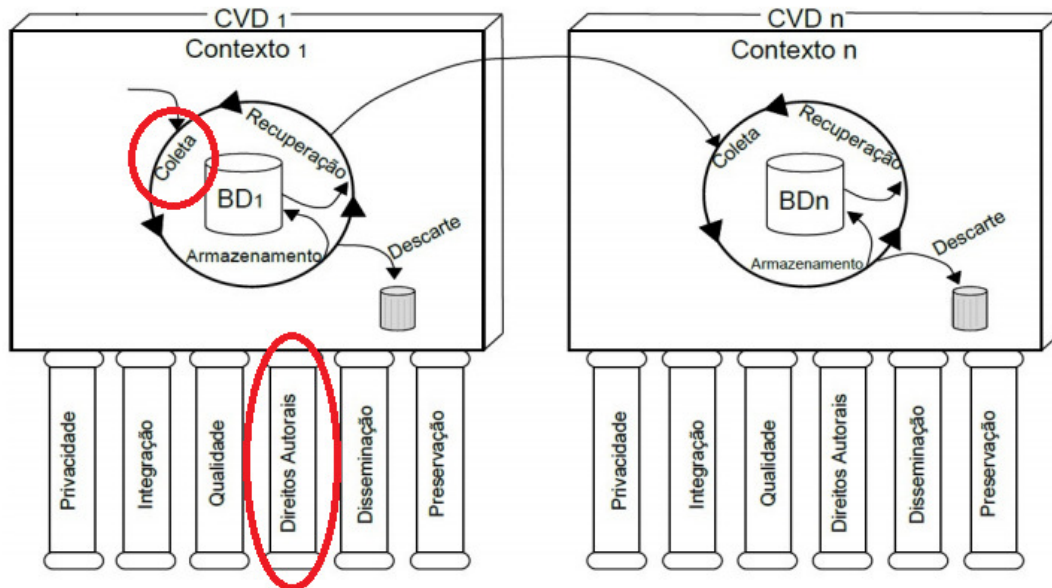
Figura 4 - Ciclo de Vida dos Dados na Ciência da Informação (CVD-CI)





O destaque da figura 4 realça os seis fatores, Preservação, Disseminação, Direitos Autorais, Qualidade, Integração e Privacidade relacionados às 4 fases do CVD-CI.

Figura 5 – Fase da coleta com o fator Direitos Autorais



Fonte: Adaptado de Sant'Ana, 2016, p. 123

Na figura 5 evidencia a Fase da Coleta, perpassando pelos Direitos autorais utilizados para atingir os objetivos propostos na pesquisa. As Fases e os fatores do Ciclo de Vida dos Dados são retratados na subseção 1.4.1.1 para demonstrar como é sua dinâmica.

#### 1.4.1.1 Fases do Ciclo de Vida dos Dados

O CVD-CI possui quatro fases:

- a) Coleta

A fase da coleta envolve atividades relacionadas ao planejamento de como os dados serão obtidos, filtrados e organizados, identificando sua estrutura, formato e meios de descrição (SANT'ANA, 2013).

Nesta fase o repositório poderá estruturar o Plano de Gestão de Dados (PGD) instrumento para a orientação dos pesquisadores. O PGD é um documento formal desenvolvido no início do projeto para que seja propício descrever todos os aspectos da gestão dos dados arquivados, observando as políticas vinculadas ao repositório.

O planejamento de desenvolvimento do PGD garante ao pesquisador quem será responsável pelos seus dados, quem terá os direitos de gestão e divulgação dos dados se é sua agência de fomento, sua instituição, os sujeitos da pesquisa, seus colaboradores e/ou o próprio pesquisador (DATA..., 2016a).

Os direitos de propriedade irão determinar como os dados podem ser processados, descrevendo todas as informações sobre direitos autorais e licenças de uso em seu PGD incluindo a negociação de condições de processamento, divulgação e reutilização (DATA..., 2016a).

#### b) Armazenamento

Essa fase abrange as atividades relacionadas “[...] ao processamento, transformação, inserção, modificação, migração, transmissão e toda e qualquer ação que vise a persistência de dados em um suporte digital.” (SANT'ANA, 2013, p. 18). É a fase com enfoque mais tecnológico com definição de aspectos que garantem o reuso dos dados o que leva a necessidade de estruturação de metodologias e ações relacionadas a persistência dos dados (SANT'ANA, 2016).

Alguns critérios podem ser descritos no PGD e adotados no armazenamento como:

- Armazenar dados em formatos abertos;
- Ter mais de um agente responsável pelo acesso aos dados;
- Fazer *backup* de dados em locais fisicamente distintos;
- Em caso de grandes tamanhos de arquivos, o agente pode decidir *backup* apenas os elementos mais essenciais;

- Regularmente, ou depois de um movimento ou *download*, verificar a integridade dos dados, o tamanho do arquivo e data para averiguar se os dados foram alterados (DATA..., 2016b).

#### c) Recuperação

A fase da recuperação consiste na consulta e visualização dos dados “[...] com as etapas posteriores a obtenção do dado: estruturação, filtro, tratamento, representação, refinamento e interatividade.” (SANT’ANA, 2013, p. 18).

Pesquisadores da Web semântica estão focando nos aspectos formais das linguagens de representação semântica (FOX; HENDLER, 2011). Essas linguagens estão sendo padronizadas e as comunidades estão usando-as para a construção e uso de ontologias (FOX; HENDLER, 2011). As ontologias auxiliam nesse contexto trazendo significados e proporcionando a interação e o compartilhamento de dados e informações o que facilita a recuperação pelos usuários..

Os repositórios que fornecem campos de buscas com filtros, incluindo vocabulários controlados favorecem a busca dos usuários e cooperam para uma melhor recuperação.

#### d) Descarte

A fase de descarte pode ocorrer após os dados cumprirem todos os objetivos planejados ou alcançado o limite de tempo previsto. O descarte pode transcorrer por transferência para outras bases (SANT’ANA, 2013).

Viabilizados todos os objetivos propostos para determinado conjunto de dados eles podem ser descartados. Deve-se sempre atentar ao fato de que exclusão do arquivo não é suficiente para assegurar que os dados armazenados foram completamente removidos. Os arquivos apagados podem ser recuperados usando ferramentas de *software* próprias para esse fim (DATA..., 2016a).

Como verificado na figura 4, todas as fases do Ciclo de Vida dos Dados são permeadas por seis fatores: Preservação, Disseminação, Direitos autorais,

Qualidade, Integração e Privacidade. Os fatores são apresentados na subseção abaixo

#### 1.4.1.2 Fatores das fases do Ciclo de Vida dos Dados

O CVD-CI tem suas fases permeadas por seis fatores evidenciadas como:

##### a) Preservação

O Objetivo da Preservação deve ser pensado e planejado desde a coleta dos dados científicos para assegurar que sejam disponibilizados, pesquisáveis e recuperáveis.

Alguns aspectos contribuem na definição de quais dados serão preservados:

- os custos envolvidos na curadoria e armazenamento dos dados;
- a unicidade dos dados;
- os custos de reprodução dos dados, caso seja necessário;
- o formato de dados e a utilização de metadados;
- os aspectos jurídicos, caso a agência ou financiador exija que os dados sejam armazenados em repositórios de domínio público;
- o valor econômico dos dados no futuro;
- os aspectos éticos, por exemplo, se os dados foram obtidos por experimentação em animais ou seres humanos (DATA..., 2016c).

##### b) Disseminação

Dados produzidos ou gerados por pesquisadores, laboratórios ou instituições são disseminados e disponibilizados para fins como a reutilização. Disseminar “[...] significa, em uma medida, divulgar, difundir, propagar, mediante condições e recursos de que o cerca [...]” (BARROS, 2003, p. 41).

O Fator da Disseminação no contexto dos repositórios envolve aspectos relacionados à disseminação com o intuito de localização e acesso aos dados.

Nesse processo aparecem dois agentes, o emissor, aquele que vai emitir os dados ou a informação e o receptor, aquele que vai receber os dados e a informação

Originalmente referida com base no esquema tradicional de comunicação – emissor, canal, mensagem, receptor –, o debate sobre disseminação da informação associado à transferência sugere, atualmente, abandonar a unidirecionalidade emissor-receptor para contemplar o usuário numa dimensão mais ampla que o inclui como participante ativo do processo informacional. (LARA; CONTI, 2003, p. 27).

A disseminação dos conjuntos de dados auxilia no reconhecimento do trabalho do pesquisador e na localização e reutilização de dados armazenados para evitar retrabalho e gastos já investidos na fase de coleta de dados.

#### c) Direitos Autorais

Direito autoral consiste em dar créditos aos responsáveis pela coleta dos dados e na transferência dos dados a outrem, seja pessoa física ou jurídica, via licenças de uso. Quando os conjuntos de dados estiverem disponibilizados em repositórios de dados para reutilização, os detentores dos direitos autorais devem ser citados. Esses dados estarão disponibilizados através de uma licença de uso.

Depositar os dados em repositórios de dados contribui para torná-los mais visíveis e possivelmente obter os créditos devidos. O pesquisador pode considerar os benefícios de uma liberação planejada de seus dados de pesquisa.

Os direitos autores e licenças de uso de dados serão abordados no capítulo 3.

#### d) Qualidade

A qualidade na organização e gerenciamento dos dados garante uma melhor interpretação por outros pesquisadores e tornam os dados confiáveis para reuso.

Os dados podem ter variadas formas e tamanhos. A documentação e descrição dos dados auxiliam na padronização e confiabilidade dos mesmos. O PGD contribui com a documentação do histórico de processamento dos dados, que envolvem a coleta, os agentes, o histórico, a proveniência, a descrição com metadados, uso de formatos abertos e facilidade na busca e recuperação.

#### e) Integração

No contexto da ciência embasada em grandes quantidades de dados é realidade profissionais que gerenciam dados e pesquisadores que buscam dados que necessitam para o desenvolvimento de suas pesquisas. O processo de busca pode ocorrer em mais de um sistema de informação como repositório de dados, repositórios institucionais que disponibilizam dados e os periódicos científicos e de dados. Com vasta variedade de sistemas espalhados há uma forte necessidade de reunir todos os dados possíveis ou o sistema apontar para onde os dados são encontrados (SALES, 2014).

Lopatemko (2001) destaca a necessidade de encontrar soluções para o problema de integração dos dados, que passa pela implementação de um padrão com três características essenciais:

- fácil de implementar para qualquer participante;
- flexível o suficiente para abraçar a diversidade, a estrutura e o significado dos dados em diferentes estados, organizações, ou áreas da ciência;
- poderoso para fornecer serviços de recuperação de informações sofisticados para os usuários.

#### f) Privacidade

Dados científicos podem abranger leis de privacidade ou confidencialidade e devem ser armazenados em computadores que não estão ligados em rede, tendo o cuidado desses equipamentos não serem roubados ou seus sistemas invadidos para que os dados não sejam adulterados.

Pesquisadores reúnem dados durante a coleta, inclusive, os sensíveis a disponibilização e acesso a todos os usuários. Esses dados abrangem dados

pessoais como nome, endereço, ocupação, raça de uma pessoa, origem, etnia, opiniões políticas, saúde física ou mental, registro criminal, orientação sexual, crenças religiosas, de valor econômico, entre outros.

O pesquisador dispõe no PGD, se oportuno, o consentimento dos participantes de seu projeto para providenciar o compartilhamento, preservação e uso a longo prazo dos dados confidenciais. É oportuno definir e descrever no PGD qual nível de confidencialidade será mantido. A anonimização e criptografia são métodos para garantir a privacidade.

A anonimização de dados é baseada em um protocolo e permite a divulgação legal dos dados e dos metadados sem ferir a integridade dos participantes da pesquisa. Nesse processo os dados e/ou metadados que não serão divulgados recebem pseudônimos ou são criptografados e protegidos. A anonimização consiste no processo de criptografar ou remover informações de identificação pessoal a partir de conjuntos de dados (DATA..., 2016a).

Modelos de privacidade como o K-anonimato podem ser uma alternativa para garantir a privacidade dos dados. K-anonimato é “[...] conceituado como uma forma de definir propriedades que um conjunto de dados deve possuir de forma a garantir a privacidade dos envolvidos em sua eventual divulgação ou acesso.” (AFFONSO; SANT’ANA, 2015, p. 4).

Outro método com a possibilidade de uso é a criptografia que consiste na conversão de dados em uma forma que não pode ser facilmente compreendido por pessoas não autorizadas. Para segurança, os dados não devem ser armazenados ou transmitidos sem criptografia (DATA..., 2016b).

## 1.5 Metodologia

A metodologia utilizada teve como base a pesquisa bibliográfica para a estruturação da fundamentação teórica concomitante à metodologia quantitativa e qualitativa. Foi utilizada a coleta de dados para fazer o levantamento dos repositórios de dados das 100 melhores universidades do mundo.

A coleta de dados se iniciou com a busca das melhores universidades do mundo através do ranking *webometrics.info*. Para tanto, foi necessário definir o escopo com as 100 melhores universidades do mundo. A localização de repositório

de dados nas universidades foram realizadas nos meses de julho à setembro de 2016.

Em seguida foi realizado o método exploratório para o levantamento das páginas oficiais das universidades utilizando os *links* disponibilizados pelo sítio *webometrics.info* e para localização dos repositórios de dados nos sítios das universidades. O processo de recuperação dos dados foi realizado por meio da coleta dos Planos de Gerenciamento de Dados dos repositórios de dados das universidades. Os Planos de Gerenciamento de Dados foram analisados com base no Ciclo de Vida dos Dados no fator dos Direitos autorais na fase da Coleta.

## 1.6 Estrutura do texto

A presente dissertação está estruturada nas seguintes seções:

Seção 1 – INTRODUÇÃO traz o contexto da pesquisa com conceitos iniciais sobre dados científicos, repositórios de dados e propriedade intelectual. Em suas subseções apresenta o problema da pesquisa, objetivo geral e objetos específicos, a delimitação do tema e objeto da pesquisa, a explicitação dos PGDs dentro do Ciclo de Vida dos Dados, a justificativa e a metodologia utilizada.

Na seção 2 – REPOSITÓRIO DE DADOS CIENTÍFICOS, traz conceitos acerca dos repositórios, com breve indicação histórica, abordagem de repositório de dados e a comunicação científica e discussão sobre repositório de dados e *e-Science*.

Na seção 3 – DIREITOS AUTORAIS E LICENÇAS DE USO, há uma contextualização histórica sobre propriedade intelectual, as definições de propriedade intelectual, direitos autorais e licenças de uso, apresentando as licenças de uso que podem ser usadas pelos repositórios de dados.

A seção 4 - PLANO DE GERENCIAMENTO DE DADOS são discutidas as questões relacionadas ao gerenciamento de dados e o papel dos planos de gerenciamento dos dados.

A seção 5 – RESULTADOS E DISCUSSÕES apresenta o resultado das análises dos PGDs dos repositórios de dados localizados, a identificação dos repositórios de dados das universidades, identificação dos Planos de Gerenciamento de Dados de cada repositório e a explicitação dos Planos de Gerenciamento de



Dados apresentando as instruções sobre direitos autorais. Os resultados são apresentados com os gráficos e quadros para melhor visualização dos resultados.

Na seção 6 – CONSIDERAÇÕES FINAIS apresenta as reflexões e considerações sobre os tópicos abordados em toda a pesquisa.

## 2 REPOSITÓRIOS DE DADOS CIENTÍFICOS

A evolução da ciência está altamente relacionada ao aprimoramento do instrumental tecnológico que proporciona à cientistas e pesquisadores coletarem dados nos diversos ramos da ciência.

As comunidades científicas viabilizam ações na ciência que envolvem a disponibilização e o compartilhamento de dados científicos em nível nacional e internacional concomitante a implementação de repositório de dados científicos.

Morin (2007, p. 106) destaca que a ciência “[...] se baseia ao mesmo tempo no consenso e no conflito. Anda ao mesmo tempo sobre quatro patas independentes e interdependentes: a racionalidade, o empirismo, a imaginação, a verificação.”

Os pesquisadores e a evolução de aparatos tecnológicos utilizados nos diversos setores da ciência se relacionam de maneira indissociável para a redefinição de práticas científicas contribuindo para as revoluções que ocorrem no campo científico.

No livro *A estrutura das revoluções científicas* é apontado que as revoluções do pensamento científico “[...] são definidas como um momento de desintegração do tradicional numa disciplina, forçando a comunidade de profissionais a ela ligados a reformular o conjunto de compromissos em que se baseia a prática dessa ciência.” (KUHN, 2006).

A conjuntura apresentada por Kuhn (2006) pode ser observada na ciência, em que as ciências experimentais, teóricas e computacionais estão sendo afetadas pela grande quantidade de dados oriundos de várias áreas do conhecimento auxiliando para conduzir a comunidade científica a repensar o papel que esses dados representam no contexto científico.

Os autores Gray (2009), Mayer-Schonbeeger e Cukier (2013) têm destacado a importância das tecnologias, das mais variadas, na forma de se fazer ciência. Para esses autores, os desafios tecnológicos incluem a necessidade de melhor capturar, analisar, modelar, visualizar e preservar dados e informações científicas, tornando os sistemas de computação vitais para o moderno ambiente de pesquisa (GRAY, 2009; MAYER-SCHONBEERGER; CUKIER, 2013).

O advento do século XXI vem sendo marcado por práticas e descobertas científicas que apontam para campos da ciência no qual o acesso aberto aos dados

científicos gera um avanço na realização dos ideais de compartilhamento, uso e reuso de dados (TOLLE; TANSLEY; HEY, 2011; LYNCH, 2003, 2011). Esses fatores contribuem para um repensar na divulgação e disponibilização de dados nos periódicos científicos e em repositórios de dados.

Gestão de dados está se tornando um aspecto primordial no processo de investigação científica devido às exigências de agências de financiamento e revistas científicas para liberação de dados de pesquisa no momento da publicação do artigo (BORGMAN et al., 2015).

Alinhada a essa premissa, periódicos que publicam dados científicos emergem como mais um canal de comunicação científica para que pesquisadores publiquem seus dados. Periódicos de dados são tendências ímpares nesse contexto e propícios para divulgar conjuntos de dados de forma mais abrangente. Periódicos com essa natureza compreendem mas não estão restritos: *Biodiversity Data Journal*, *Data in Brief da Elsevier*, *Earth system Science Data*, *GeoScience Data Journal*, *Journal of Open Archeology Data*, *Journal of Open Psychology*, *Journal of Physical and chemical Research Data*, *Journal of Open Research Software* e *Open Health Data*.

Agências de fomento de países, como nos EUA, estão apontando no avanço da realização de ideais de depósito de dados em repositórios sugerindo ou mesmo obrigando os pesquisadores que tiveram suas pesquisas financiadas por elas a fazerem o depósito em repositório de dados científicos.

Conforme Pinfield (2009, p. 165, tradução nossa)

Um repositório pode ser definido como um conjunto de sistemas e serviços que facilita o armazenamento, gerenciamento, recuperação, exibição e reutilização de objetos digitais. Os repositórios podem ser estabelecidos por instituições, comunidades, financiadores de pesquisa ou outros grupos. Eles podem fornecer acesso a uma variedade de objetos digitais, incluindo artigos de revistas revisados por pares, capítulos de livros, teses, conjuntos de dados, objetos de aprendizado ou arquivos rich media.<sup>4</sup>

---

<sup>4</sup> A repository may be defined as a set of systems and services which facilitates the ingest, storage, management, retrieval, display, and reuse of digital objects. Repositories may be set up by institutions, subject communities, research funders, or other groups. They may provide access to a variety of digital objects, including peer-reviewed journal articles, book chapters, theses, datasets, learning objects, or rich media files.

Os repositórios de dados são uma evolução dos repositórios digitais no qual o uso de repositórios para a divulgação de trabalhos científicos se iniciou com os repositórios temáticos.

No final de 2002 iniciou-se uma revolução na rede de comunicação científica. Esse processo se instituiu com o desenvolvimento dos repositórios temáticos que “[...] se constitui em um conjunto de trabalhos de pesquisa de uma determinada área do conhecimento, disponibilizados na internet.” (CAFÉ et al. 2003, p. 3). Os diversos departamentos de faculdades podem ter seus Repositórios Temáticos. A união de todos os Repositórios Temáticos sobre a responsabilidade da universidade compõe o Repositório Institucional (CAFÉ et al., 2003).

Os Repositórios Institucionais (RI) surgiram como uma nova estratégia, propiciando às universidades que acelerassem mudanças no desenvolvimento científico e em sua comunicação (LYNCH, 2003).

Segundo Rodrigues (2004, p. 2) os RIs “[...] são coleções digitais que armazenam, preservam, divulgam e dão acesso à produção intelectual de comunidades universitárias.” Lynch (2003, p. 2, tradução nossa) destaca que os RIs são “[...] serviços que uma universidade oferece aos membros de sua comunidade para a administração e disseminação dos materiais digitais criados pela instituição e pelos membros de sua comunidade.”

A grande quantidade de Repositórios Institucionais visualizada na figura 6 demonstra a importância dos RIs no contexto mundial. Destaca-se a grande concentração na Europa e Estados Unidos. No Brasil são várias implementações distribuídas entre Nordeste, Sul e Sudeste.

Figura 6 – Repositórios Institucionais em todo o mundo



Fonte: PINFIELD; SALTER; BATH, 2014

A grande quantidade de instituições que aderiu aos repositórios institucionais, principalmente as universidades, demonstra a importância da produção intelectual de sua comunidade e o papel fundamental da disseminação do conteúdo para dar visibilidade à produção e à comunidade. Os repositórios de dados estão caminhando na mesma direção.

Repositórios de dados científicos têm significativa participação no gerenciamento de dados. A grande quantidade de dados coletados ou gerados precisa de representação e gerenciamento para ser recuperada e reutilizada e, assim, favorecer a colaboração científica internacional.

A colaboração pode ser facilitada evitando repetição de procedimentos de coleta de dados, auxiliando no avanço da ciência. A ambiência de repositórios de dados viabiliza representar, processar, gerenciar, disseminar e dar acesso aos dados nele depositados, incluindo métodos de curadoria. Reunir conjuntos de dados nesses repositórios propicia o compartilhamento, acesso, uso e reuso dos dados.

Rodrigues et al. (2010, p. 22-23, grifo nosso) contextualizam repositório de dados como uma extensão de repositórios

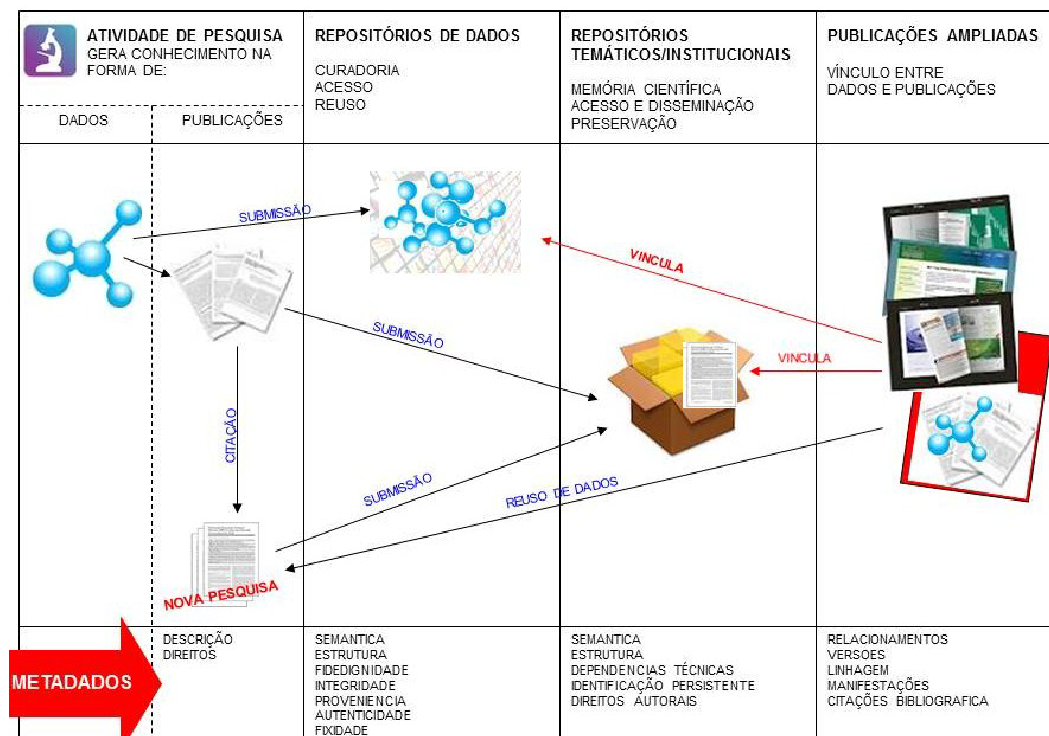
[...] “repositório” designa um sistema informático em que existe uma plataforma de armazenamento de objectos representados em ficheiros, capaz de incorporar novos objectos à medida que são produzidos ou submetidos. O repositório oferece serviços que são dirigidos a quem deposita, a quem pesquisa e aos administradores do sistema. Nos **repositórios de dados** pode ir-se muito além desta visão de repositório de objectos, uma vez que cada conjunto de dados tem características próprias e por isso pode requerer um tratamento diferenciado.

Segundo Sales (2014, p. 63-64) a implementação de repositórios de dados tem vários benefícios:

- Amplia a visibilidade dos resultados de pesquisa, posto que, via de regra, só é formalmente disseminada a fração que está registrada nas publicações acadêmicas.
- Proporciona mecanismos de preservação de longo prazo, em termos de preservação digital, de arquivamento seguro e de curadoria digital.
- Permite que o material depositado esteja disponível on-line continuamente para ser consultado e citado mais frequentemente.
- É um instrumento chave para os processos de reformatação e recriação de dados proporcionados pela curadoria digital.
- Abre a possibilidade de criação de novos serviços de informação para pesquisadores e gestores a partir da análise dos dados arquivados e a integração de dados e publicações acadêmicas.
- Permite a criação de redes de repositórios interoperáveis.
- Aumenta o grau de reuso dos dados minimizando a duplicação de esforços e otimizando os investimentos na geração de dados.

Na figura 7, apresenta-se o modelo de publicação científica com os repositórios de dados para a publicação e divulgação. Os aspectos relacionados à publicação são abordados demonstrando o reuso de dados científicos e a publicação dos periódicos que publicam dados. Destaca-se a questão abordada sobre os aspectos do direito autoral, já implementados nos repositórios que pode ser estendido para os repositórios de dados. A semântica, estrutura, fidedignidade, integridade, proveniência, autenticidade e fixidade são aspectos consideráveis para a recuperação e reuso de dados.

Figura 7 - Modelo de publicação científica com direitos autorais de dados



Fonte: SALES; SAYÃO, 2012.

O trabalho de pesquisa produz dados e textos que são publicados em periódicos científicos ou armazenados em repositórios temáticos, institucionais ou de dados. As publicações de texto ou conjunto de dados são utilizadas para gerar novos conhecimentos que serão registrados para divulgação. Dados e publicações são vinculados e integrados e podem gerar novos resultados e novas publicações. Todos os fatores descritos na figura 7 estão presentes nos meios de comunicação, repositórios e publicações ampliadas, e o fator direito autorai está presente.

Repositórios de dados possibilitam oportunidades sem precedentes em direção ao avanço da pesquisa, instituído em um ambiente com novas ferramentas e serviços para gerenciar, compartilhar e possibilitar a reutilização de dados científicos entre disciplinas e instituições.

Repositórios de dados estão presentes nos diversos países do mundo. São mais de 1.100, conforme apresentado na figura 8. Verifica-se uma maior concentração deles na Europa e Estados Unidos da América, assim como repositórios institucionais.

Figura 8 – Repositório de Dados no mundo



Fonte: KING; FORCE, c2017

Repositórios bem articulados promovem uma colaboração científica transdisciplinar e em grande escala, permitindo que os cientistas construam conhecimentos que proporcionem melhor compreensão e soluções sustentáveis para desafios científicos importantes, ofereçam melhor compreensão pública de problemas tecnológicos através de informações abertamente acessíveis (LEE et al. 2009).

Os dados resultantes da investigação científica podem fornecer evidências que são usadas para fundamentar as discussões apresentadas nas publicações científicas ou auxiliar pesquisadores a testar teorias, dar continuidade aos projetos ou embasar novas pesquisas.

Os conjuntos de dados fundamentam os resultados da investigação e são atualmente reutilizados nas diversas áreas do conhecimento. O avanço exponencial de dados científicos integra as novas formas de publicação científica alinhada a divulgação de conjuntos de dados através da comunicação científica parametrizada



pelos periódicos e repositórios de dados. Sobre esta perspectiva a próxima subseção abordou o papel dos repositórios de dados na comunicação científica.

## 2.1 Repositório de dados e a comunicação científica

A divulgação do conhecimento científico é viabilizada, inclusive, pela publicação. Essa comunicação é uma colaboração em larga escala, com a possibilidade de acesso à pesquisa e possível reprodução de resultados das investigações e experimentos, além de oferecer evidências para a qualidade do trabalho científico (LYNCH, 2011). A comunicação está situada no coração da ciência (MEADOWS, 1999).

A comunicação científica consiste em assegurar o intercâmbio de [dados] e informações sobre os trabalhos em andamento ou já finalizados e viabilizar o contato entre pesquisadores e suas pesquisas sendo possível o trabalho colaborativo (LE COADIC, 2004).

Le Coadic (2004) destaca o paradigma do trabalho coletivo sendo a colaboração, trabalho organizado em rede de pessoas e computadores, novas formas de intercâmbio de informações e de acesso ao conhecimento. O autor destaca os três processos do ciclo da informação: construção, comunicação e uso, que se sucedem e se alimentam reciprocamente (LE COADIC, 2004).

O trabalho colaborativo pode ser potencializado com a disponibilização de dados científicos em periódicos ou repositórios de dados. Desta forma, é possível realizar *inputs* para novas hipóteses, para replicação de estudos, novas descobertas ou refutação das descobertas apresentadas, apresentação de propostas de pesquisas e aprimoramento das atuais.

Dados científicos têm um papel de retroalimentação no desenvolvimento da ciência. Este panorama amplia a possibilidade de pesquisadores de compartilhamento e reutilização de dados resultantes da investigação científica em diversos países e disponibilizados em periódicos científicos, periódicos de dados ou repositórios de dados cooperando com a articulação na colaboração científica internacional.

Na figura 9 verifica-se que os dados científicos estão na base da pirâmide, destacando-se pela grande quantidade com a oportunidade de serem combinados, recombinaados e usados por áreas multidisciplinares, serem unificados, integrados e interoperados em rede. No topo da pirâmide está a literatura publicada resultante das análises, refutações ou aprimoramento dos estudos.

Figura 9 - Todos os dados científicos online



Fonte: adaptado de TOLLE, TANSLEY; HEY, 2011, p. 25.

Os dados coletados e armazenados em repositórios tem a possibilidade de integração entre eles e a unificação de “[...] todos os dados científicos e toda a literatura para criar um mundo em que os dados e a literatura possam interagir.” (TOLLE; TANSLEY; HEY, 2011 p. 24). Os pesquisadores analisam e utilizam dados para embasar o resultado de suas pesquisas, publicam essas junto com os seus dados proporcionando a outros o acesso a todo o conteúdo disponibilizado.

Dessa forma, consegue-se integrar todos os dados e a literatura, aumentando a velocidade e produtividade da comunicação científica. “[...] mas transformar esse volume em valor agregado para os usuários requer um processo de recuperação cada vez mais eficiente.” (SANT’ANA, 2008, p. 146).

Essa integração favorece a produtividade científica por meio do aumento da velocidade com que os dados e as informações são tratados e produzidos (LYNCH, 2011) além de reduzir os custos e investimentos aplicados pelas agências de fomento evitando que esforços na coleta de dados sejam duplicados.

Para Mueller (2000, p. 25) os resultados das pesquisas alcançados por determinados pesquisadores “[...] são freqüentemente retomados por outros

cientistas, teóricos ou aplicadores, que dão continuidade ao estudo, fazendo avançar a ciência ou produzindo tecnologias ou produtos neles baseados.”

Propício, assim, ação para harmonizar os fatores envolvidos nessa dinâmica para a evolução da ciência. A ciência esta sendo impactada por grandes quantidades de dados nas diversas áreas do conhecimento, e emerge um novo contexto baseado no uso intensivo de dados, denominado por pesquisadores de *e-Science*, ou quarto paradigma da ciência abordado na próxima subseção.

## 2.2 Repositório de dados e *e-Science*

Nos paradigmas da ciência, a ciência está sendo afetada pela grande quantidade de dados coletados, gerados ou simulados em grandes laboratórios e observatórios e delinea um novo paradigma, o paradigma do uso intensivo de dados, denominado por alguns pesquisadores como *e-Science*, quarto paradigma da ciência (GRAY, 2007), Ciência orientada a dados (HEY; HEY, 2006), Computação fortemente orientada a dados, Cyberinfraestrutura e Dos dados ao conhecimento (CESAR JUNIOR, 2011).

Thomas Kuhn, em seu livro *A estrutura das revoluções científicas* trata o tema paradigma nas discussões da chamada “ciência normal”, guiada por um paradigma mas que, em seu percurso, surgem descobertas e formulação de novas teorias, com um amadurecimento científico, induzindo a uma revolução científica que, então, ocorre uma mudança de paradigma (KUHN, 2006). As revoluções científicas apresentadas por Kuhn (2006, p. 125) são apontadas como “[...] aqueles episódios de desenvolvimento não-cumulativo, nos quais um paradigma mais antigo é total ou parcialmente substituído por um novo [...]”

Kuhn (2006, p. 20) afirma que “Se a ciência é a reunião de fatos, teorias e métodos reunidos nos textos atuais, então os cientistas são homens que, com ou sem sucesso, empenharam-se em contribuir com um ou outro elemento para essa constelação específica.” Cabe, assim, ação com intuito de harmonizar os fatores envolvidos nessa dinâmica cooperando com a evolução da ciência.

Gray (2007) retrata uma evolução da ciência onde relaciona os quatro paradigmas da ciência:

**Primeiro paradigma:** mil anos atrás, a ciência era empírica, com a descrição dos fenômenos naturais;

**Segundo paradigma:** há poucos séculos, se desenvolveu a ciência teórica, com o uso de modelos e generalizações, mostrando como exemplo as leis de Newton, Kepler e as equações de Maxwell;

**Terceiro paradigma:** nas últimas décadas, apareceu o ramo computacional, com a simulação de fenômenos complexos gerando uma grande quantidade de dados mostrando o caminho para o quarto paradigma destacado por Gray como *e-Science*;

**Quarto paradigma:** a ciência do século XXI é apresentada como exploração de dados, a *e-Science*, que unifica os três paradigmas anteriores (experimento, teoria e simulação) destacando como características:

- grandes quantidades de dados capturados por instrumentos ou gerados por simulações e processados por *softwares*;
- informação e/ou conhecimento armazenados em computadores;
- cientista analisa base de dados e arquivos por meio de gerenciamento de dados e estatísticas.

Quadro 1 – Paradigmas da ciência

PARADIGMA	TEMPO	CIÊNCIA	ASPECTOS
Primeiro	mil anos atrás	empírica	descrição dos fenômenos naturais
Segundo	há poucos séculos	ciência teórica	ciência teórica, com o uso de modelos e generalizações, mostrando como exemplo as leis de Newton, Kepler e as equações de Maxwell;
Terceiro	últimas décadas	computacional	simulação de fenômenos complexos gerando uma grande quantidade de dados
Quarto	século XXI	unifica (teoria, experimento e simulação)	exploração de dados

Fonte: Elaborada pela autora a partir de Gray (2007)

*E-Science* é o ponto onde “[...] TI [tecnologia da informação] encontra cientistas”, onde se usa computador para resolver problemas científicos com uso

intensivo em dados, culminando em uma evolução das fases históricas da ciência dominadas pela experimentação, teoria e simulação (TOLLE; TANSLEY; HEY, 2011, p. 17).

A ambiência da *e-Science* se baseia em uma infraestrutura que visa permitir aos pesquisadores terem acesso a dados científicos distribuídos, com oportunidade de colaboração, com o compartilhamento e o gerenciamento desses dados, possibilitando o reuso de conteúdo já mapeado e, desta forma, favorecendo o avanço da ciência (MEDEIROS; CAREGNATO, 2012). Isso configura um aspecto da ciência em que pesquisadores de países diferentes, falando línguas diferentes, coletando grandes quantidades de dados podem compartilhá-los contribuindo com o aprimoramento científico.

Atuando na mesma direção Ribes e Lee (2010, p. 232) destacam que a *e-Science* é um conceito designado “[...] para as tecnologias de informação em rede de apoio às atividades de investigação científica, como a colaboração de compartilhamento de dados e divulgação dos resultados.” De acordo com os autores, há três aspectos que caracterizam as transformações proporcionadas pela *e-Science*: comunidade ampla e interdisciplinar de colaboração; computação aliada na coleta, representação e análise de dados; integração final (RIBES; LEE, 2010).

Medeiros e Caregnato (2012, p. 315) destacam que “[...] *e-Science* altera consideravelmente a maneira com que os cientistas realizam seu trabalho, as ferramentas que utilizam, os tipos de problemas que abordam e a natureza da documentação e da publicação que resulta da sua pesquisa.”

Todas as capacidades atualmente necessárias à *e-Science*, integração de dados, fusão e mineração; desenvolvimento de fluxos de trabalho, orquestração e execução; captura da proveniência, linhagem e qualidade dos dados; validação, verificação e confiança na autenticidade dos dados; e adequação ao propósito -precisam de representação e mediação para que a *e-Science* possa se tornar realmente intensiva em dados. (FOX; HENDLER, 2011, p. 161).

A “ciência intensiva em dados”, um dos componentes da *e-Science*, deve avançar para permitir o acesso aos dados pelos cientistas e pesquisadores que não fazem parte das equipes dos grandes projetos e “[...] permitir maior integração de fontes e prover interfaces para quem é especialista em ciência, mas não em computação e administração de dados.” (FOX; HENDLER, 2011, p. 159).

Grandes projetos, como Pesquisa Celeste Digital *Sloan*, Grande *Colisor* de Hádrons (GCH) (TOLLE; TANSLEY; HEY, 2011), Projeto Australiano de radiotelescópios ASKAP e o conjunto de telescópios astronômicos Pan-STARRS, estão gerando *petabytes* de dados que são analisados por cientistas do mundo inteiro, em laboratórios diferentes e que falam línguas diferentes.

Na discussão sobre *e-Science*, Fox e Hendler (2011), do Instituto Politécnico Rensselaer, ressaltam três questões:

- Como os dados, que não foram gerados por cientistas e pesquisadores, serão usados por eles?
- Como usar dados, que eles não produziram e nunca viram, junto aos dados que geram e usam todos os dias?
- O que se deve fazer se o cientista ou pesquisador, estudantes e não especialistas precisam dos dados de outra área do conhecimento e não conhecem os termos ou vocabulário da área?

Repositórios de dados científicos contribuem no gerenciamento de grandes quantidades de dados e na resolução das questões levantadas por Fox e Hendler. A vultosa quantidade de dados coletados ou gerados precisam de representação, gerenciamento e curadoria para que possam ser recuperados e reutilizados e, assim, contribuir com a colaboração científica internacional.

Os desafios sustentam-se no fato de que emergiu na ciência um problema destacado por Sayão e Sales (2012, p. 181) o qual aborda a gestão de dados científicos “[...] num mundo digital interligado por redes de computadores, onde há um fluxo intenso de dados, proveniente de diferentes fontes, sendo gerados, processados e compartilhados em ambientes multidisciplinares.”

O Plano de Gerenciamento de Dados (PGD) é um documento que visa a explicitação da gestão de conjunto de dados. O PGD auxilia na descrição de dados indicando qual padrão de metadados ou conjunto de metadados é utilizado, facilitando a busca e localização dos conjuntos de dados incluindo parâmetros para curadoria.

Os pesquisadores têm diretrizes para indicar sobre quais licenças disponibilizará seus conjuntos de dados ao depositá-los, assim como, ao recuperar o conjunto de dados, os pesquisadores têm diretrizes para dar os créditos aos

responsáveis pelos dados. Na seção 4 foram abordadas as questões relacionadas ao PGD, à definição e suas abordagens.

A próxima seção abordou os aspectos relacionados aos direitos autorais, discutindo o arcabouço conceitual, teórico e metodológico de políticas de informação e das questões relacionadas aos direitos dos pesquisadores sobre os dados arquivados nos repositórios de dados científicos e as licenças de uso indicadas na reutilização.

### 3 DIREITOS AUTORAIS DE DADOS CIENTÍFICOS

O direito de autor surgiu de uma lenta evolução dentro do contexto histórico. Sua legitimidade é reconhecida mundialmente. Cada país elaborou sua legislação preservando a integridade intelectual dos autores.

A descoberta da imprensa por Gutenberg em 1455, ano da impressão da Bíblia de 42 linhas, conduziu a adoção de uma política centrada mais nos produtores de obras do que nos autores (UNESCO, 1981).

A legislação sobre o direito do autor é um recurso infraestrutural das comunicações e da educação, chamada a desempenhar um papel gradualmente importante em vista da progressão e da diversificação dos modos de utilização das obras protegidas pelo direito do autor (UNESCO, 1981). Atualmente, as instituições que direta e indiretamente transferem dados requerem modificações do dispositivo jurídico que regula suas transferências.

Os detentores de direitos autorais recorrerem à justiça na ocasião em que o resultado de sua atividade intelectual ou científica é utilizado sem a devida autorização, mostrando-se desejosos de obter formas de proteção legal (UNESCO, 1981).

A natureza do direito autoral fomenta estudos que se aplicam à documentos de caráter pedagógico, cultural, tecnológico e científico, se estendendo para dados, abordagem discutida neste capítulo.

#### 3.1 Propriedade intelectual

Política refere-se a qualquer princípio de governo com plano de ação nos negócios internos da nação (POLÍTICA, c2016). No ambiente institucional, política “[...] são decisões programadas que orientam outras decisões [...]” (MAXIMIANO, 2000). Dessa forma políticas voltadas para as atividades de informação científica e tecnológica pressupõem uso e aplicação de princípios sobre o que é necessário para um país em termos de produção, transferência, acesso e uso de informação (GARCIA, 1980).



A política de informação emergiu no cenário do Pós-Guerra em nível nacional e internacional associada às políticas de ciência e tecnologia através de programas governamentais de políticas públicas e

O nexó da informação com a política seria então estabelecido por sua inclusão na esfera de intervenção do Estado, agora não só como dimensão de racionalidade administrativa, mas como fator estratégico do desenvolvimento científico-tecnológico. (GONZÁLES DE GÓMEZ, 2002, p. 27).

As políticas de informação trabalham com a transferência da informação científica no qual se tem o Estado como agente privilegiado da elaboração e estabelecimento e, a ciência e a tecnologia, como domínio de seu exercício além de definição e implementação de diretrizes para as necessidades que emergem na sociedade da informação.

Na sociedade que necessita de dados e informações, políticas de informação voltadas para a proteção da propriedade intelectual se apresentam como instrumental relevante para o desenvolvimento científico e tecnológico.

#### A propriedade intelectual

[...] abrange os direitos relativos às obras do engenho do espírito humano que garante aos pesquisadores ou responsáveis por qualquer produção do intelecto o direito de receber, por um determinado período de tempo, recompensa pela própria criação. A segurança proporcionada pelas leis de Propriedade intelectual estimula a inovação e o desenvolvimento de novas soluções para os problemas existentes. As criações podem ser protegidas por lei sob as formas de patente, registro de marca, programa de computador, desenho industrial, direitos autorais, indicação geográfica [...]. (GUIA..., [20--]).

A inquietação com a propriedade intelectual veio antes de seu reconhecimento pelas grandes convenções. Burke (2003) destaca que a ideia de comercialização do conhecimento é tão antiga quanto a crítica de Platão aos sofistas por esta prática e que a ideia de conhecimento como propriedade foi formulada por Cícero.

Na antiguidade a concepção de propriedade intelectual não existia (MARTINS, 2001). Na Idade Média compilar foi aceito sugerindo que o sentido da propriedade intelectual se tornasse menos importante (BURKE, 2003). No final da Idade Média deu-se ênfase na exploração do conhecimento para ganhos materiais e na necessidade de proteção do ofício (BURKE, 2003).

A Inglaterra foi o primeiro país a reconhecer a propriedade literária. A corporação dos livreiros reclamou persistentemente uma espécie de proteção do autor (UNESCO, 1981). Um projeto de lei foi submetido à Câmara dos Comuns em 11 de janeiro de 1709 destinado a estabelecer

[...]o *copyright* dos autores ou dos seus cessionários sobre qualquer obra, durante 14 anos a contar da primeira publicação, e por outros 14 anos se, expirado o primeiro prazo, o autor ainda estivesse vivo, sob a condição, todavia, de que o livro fosse registrado na *Stationer's Company* (constituída em 1556 para registrar os títulos dos livros e regular a indústria tipográfica [...]) (MARTINS, 2001, p. 396, grifos do autor).

Este projeto converte-se na Lei de 10 de abril de 1710, conhecida como Lei da Rainha Ana pela qual a Inglaterra começou a reconhecer formalmente o direito autoral. Por meio dessa lei, os direitos do autor passaram a ser protegidos (COPYRIGHT HOUSE, c2009-2010) e, segundo seu preâmbulo “[...] ‘para incentivar os homens de cultura a escreverem livros úteis’, em outras palavras, tendo em mente mais o conhecimento que o romance.” (BURKE, 2003, p. 137).

As primeiras leis sobre o direito de autor nos Estados Unidos da América são anteriores à Revolução Francesa e a primeira lei federal foi promulgada estabelecendo a proteção dos livros e das cartas terrestres ou marítimas (UNESCO, 1981).

Na França, a noção de propriedade literária foi aos poucos substituindo o regime dos privilégios. A Revolução francesa de 1789 com sua exacerbação dos direitos individuais acrescentou ao conceito inglês a primazia do autor sobre a obra (GANDELMAN, 2007).

A importância da propriedade intelectual foi reconhecida pela primeira vez na Convenção de Paris para a Proteção da Propriedade Industrial e da Convenção de Berna para a Proteção das Obras Literárias e Obras artísticas nos anos de 1883 e 1886 respectivamente. Ambos tratados são administrados pela Organização Mundial

da Propriedade Intelectual (OMPI) agência especializada da Organização das Nações Unidas (ONU) (WIPO, [20--]):

Até agosto de 2005, **169** Estados haviam aderido à Convenção de Paris para a Proteção da Propriedade Industrial (adotada em 20 de março de 1883, revisada em Bruxelas em 14/12/1910, em Washington em 2/06/1911, em Haia em 6/11/1925, em Londres em 2/06/1934, em Lisboa em 31/10/1958 e em Estocolmo em 14/07/1967, e emendada em 2/10/1979); **159** à Convenção de Berna para a Proteção de Trabalhos Artísticos e Literários (adotada em 9/09/1886, emendada em Paris em 4/05/1896, revisada em Berlim em 13/11/1908, em Berna em 20/03/1914, em Roma em 2/06/1928, em Bruxelas em 26/06/1948, em Estocolmo em 14/07/1967 e em Paris em 24/07/1971 e emendada em 28/09/1979); e **148** Estados ao Acordo sobre Aspectos de Direitos de Propriedade Intelectual Relacionados ao Comércio (anexo 1C do Acordo de Marrakesh estabelecendo à Organização Mundial do Comércio, adotada em 15/04/1994, vigente desde 1/01/1995) (VASCONCELOS, 2009, p. 54-55, nota de rodapé, grifo nosso).

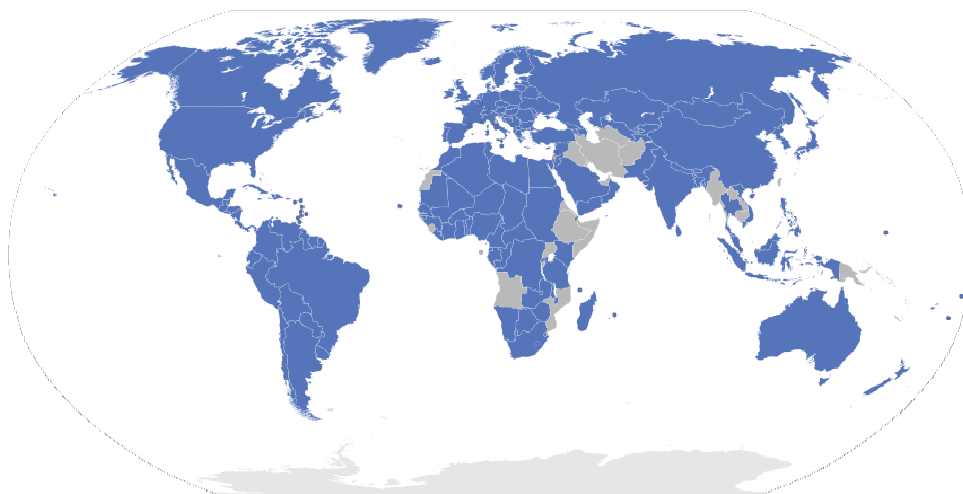
A convenção de Berna foi realizada em 1886 para dar proteção adicional a artistas, músicos, fotógrafos, designers, entre outros e, assim garantir proteção dos trabalhos literários e artísticos. De acordo com o Decreto n. 75.699 de 1975 artigo 2º parágrafo 1º

Os temas "obras literárias e artísticas", abrangem todas as produções do domínio literário, científico e artístico, qualquer que seja o modo ou a forma de expressão, tais como os livros, brochuras e outros escritos; as conferências, alocações, sermões e outras obras da mesma natureza; as obras dramáticas ou dramático-musicais; as obras coreográficas e as pantomimas; as composições musicais, com ou sem palavras; as obras cinematográficas e as expressas por processo análogo ao da cinematografia; as obras de desenho, de pintura, de arquitetura, de escultura, de gravura e de litografia; as obras fotográficas e as expressas por processo análogo ao da fotografia; as obras de arte aplicada; as ilustrações e os mapas geográficos; os projetos, esboços e obras plásticas relativos à geografia, à topografia, à arquitetura ou às ciências. (BRASIL, 1975).

Todos os países do mundo interessados na proteção recíproca instituída pela Convenção de Berna eram livres a aderir (UNESCO, 1981). Os direitos autorais registrados em um desses países vale automaticamente em todos os outros.

A figura 10 mostra os 164 países, destacados em azul, signatários da Convenção de Berna. Este número tem um aumento de 5 países desde 2005 (VASCONCELOS, 2009).

Figura 10 – Países signatários da Convenção de Berna



Fonte: WIKIPEDIA, 2016.

Direitos de propriedade intelectual são como qualquer outro direito de propriedade e permitem que os criadores ou proprietários das patentes, marcas comerciais ou direitos autorais trabalhem para benefício de sua própria atividade ou investimento em uma criação (WIPO, [20--]). Estes direitos estão descritos no artigo 27 da Declaração Universal dos Direitos do Homem<sup>5</sup>,

1. Todo ser humano tem o direito de participar livremente da vida cultural da comunidade, de fruir das artes e de participar do progresso científico e de seus benefícios.
2. Todo ser humano tem direito à proteção dos interesses morais e materiais decorrentes de qualquer produção científica literária ou artística da qual seja autor (NAÇÕES UNIDAS, 2009).

---

<sup>5</sup> A Declaração Universal dos Direitos Humanos (DUDH) é um documento marco na história dos direitos humanos. Elaborada por representantes de diferentes origens jurídicas e culturais de todas as regiões do mundo, a Declaração foi proclamada pela Assembleia Geral das Nações Unidas em Paris, em 10 de Dezembro de 1948, através da Resolução 217 A (III) da Assembleia Geral como uma norma comum a ser alcançada por todos os povos e nações. Ela estabelece, pela primeira vez, a proteção universal dos direitos humanos (DECLARAÇÃO UNIVERSAL DOS DIREITOS HUMANOS, c2016).

Propriedade intelectual refere-se às criações da mente como invenções, obras literárias e artísticas e aos símbolos, nomes e imagens usadas no comércio (WIPO, [20--]). A propriedade intelectual é dividida em duas categorias:

- Propriedade Industrial: inclui patentes de invenções, marcas, desenhos industriais, indicações geográficas e obtenções vegetais.
- Direitos autorais: abrangem obras literárias como: romances, poemas e peças; filmes, música, trabalhos artísticos como, por exemplo, desenhos, pinturas, fotografias e esculturas e projeto arquitetônico (WORLD INTELLECTUAL PROPERTY ORGANIZATION (WIPO), [20--]).

Os direitos autorais estimulam a inovação e a exploração ordenada das descobertas científicas, protegem o trabalho criativo e o investimento e estão se tornando cruciais na investigação científica (ELLIOT, 2005). À medida que o regime jurídico foi reforçado é importante que os cientistas estejam cientes de seus direitos e convencidos de usá-los no bem público (ELLIOT, 2005).

O direito autoral é dividido em direito moral e direito patrimonial abordados nas próximas subseções.

### 3.2 Direitos autorais

O progresso dos meios de comunicação que difunde obras intelectuais com amplo nível internacional desenvolveu a necessidade de proteção do direito autoral em todos os países. Este fator originou os tratados internacionais nos quais buscam “[...] dar aos autores e titulares dos países aderentes aos convênios a mesma proteção legal que cada país dá a seu autor ou titular nacional. É o princípio de reciprocidade no tratamento jurídico da autoria.” (GANDELMAN, 2007, p. 31).

Direitos autorais são usados como mecanismo legal no sentido da ordenação da vida cultural e social e promovem a maneira mais propícia em conciliar os interesses do público que usam as ideias compartilhadas e os autores que dão expressão a elas (PLOMAN; HAMILTON, 1980). No quadro 2 estão relacionadas as características dos direitos autorais.

Quadro 2 - Aspectos dos direitos autorais

Registro	Declaratório, optativo e gratuito
Titularidade	Desde a primeira utilização ou registro
Requisitos básicos	Originalidade da forma de expressão
Prazo de proteção	Variável, conforme a natureza da obra
Cessão de direitos	Os direitos patrimoniais podem ser total ou parcialmente cedidos; suas utilizações são independentes entre si
Integridade	Controle exercido através dos direitos morais
Divulgação	Essencial após a primeira publicação
Relação empregador X empregado	Titularidade patrimonial estabelecida em contrato, caso a caso
Beneficiário do uso	Autor(es), artistas criadores de obras coletivas ou derivadas; editores, produtores de discos ou filmes cinematográficos e/ou outras mídias, cessionários
Proteção internacional	Automática e independente de registro, segundo reciprocidade dos tratados

Fonte: adaptado de Gandelman, 2007, p. 38

A legislação sobre o direito do autor protege as obras literárias, artísticas e científicas de diversas formas, define os beneficiários da proteção e determina o alcance e a duração dessa proteção (UNESCO, 1981). As leis de direitos autorais concedem aos autores, artistas e outros criadores proteção por suas criações literárias, artísticas, culturais e científicas e, conforme mostrado no quadro 2, os direitos patrimoniais podem ser total ou parcialmente cedidos e os morais controlam a integridade dos titulares.

Praticamente todas as legislações nacionais integram disposições que definem as várias medidas a aplicar em caso de infração do direito de autor. Via de

regra, essas disposições encontram-se dispersas nos inúmeros textos legislativos: leis nacionais sobre o direito do autor e códigos penais. O objeto é impor uma sanção àquele que cometeu infração e indenizar a parte lesada pelo prejuízo sofrido (UNESCO, 1981).

O direito autoral apresenta duas vertentes:

- o direito moral – assegura ao autor o controle à citação de seu nome na divulgação de sua obra, o respeito a sua integridade e o direito de modificá-la ou retirá-la de circulação;
- o direito patrimonial – tem o intuito de regular as relações jurídicas do uso econômico das obras (GANDELMAN, 2007).

### 3.2.2 Direito moral e Direito patrimonial

Direito moral mais fundamental do autor consiste em controlar a publicação da sua obra ou apresentação desta ao público; é a base de todos os outros direitos concedidos pela lei; é um direito ligado a pessoa ou instituição jurídica que criou uma obra e é inalienável e imprescritível (UNESCO, 1981).

O direito de paternidade é o qual o autor tem o reconhecimento de uma obra como sua e de a ela associar ou não o seu nome. O autor pode almejar que seu nome seja mencionado todas as vezes que a sua obra for utilizada (UNESCO, 1981).

Nos países em que a legislação relacionada ao direito de autor reconhece expressamente os direitos morais, os autores conservam esses direitos mesmo que tenham cedido os seus direitos patrimoniais (UNESCO, 1981).

Os direitos patrimoniais do autor são os que lhe permitem viver da sua obra; são os aspectos financeiros inerentes ao direito autoral. O direito patrimonial pode ser transferido, podendo o autor habilitar outras pessoas a exercerem esses direitos no que diz respeito às diversas maneiras de exploração da obra (UNESCO, 1981). O autor, se desejar, cede esses direitos a título oneroso ou gratuito.

As leis reconhecem que os direitos patrimoniais são protegidos ao longo da vida do autor adicionado certo tempo após a sua morte. Os prazos de proteção após a morte do autor são entre os vinte e cinco e cinquenta anos, mas algumas leis

concedem prazos mais curtos ou mais longos dependendo de cada país (UNESCO, 1981).

As leis as quais datam do princípio do século outorga ao autor um direito geral de exploração, de cessão da obra ou de controle desta, sem especificar os modos de utilização. Outras, sem proclamar, embora qualquer direito patrimonial em particular, indicam com mais ou menos precisão as diferentes maneiras de utilização da obra para as quais deve ser obtida a autorização prévia do autor (UNESCO, 1981).

Direito moral e patrimonial podem ser utilizados para usar e reproduzir dados científicos armazenados nos repositórios de dados e citar os autores.

A garantia de referenciar, de uma forma persistente, os dados armazenados de modo que permita a pesquisa, localização e recuperação dos mesmos por longo prazo é um aspecto relevante documentado no PGD.

Pesquisadores, os quais disponibilizam seus conjuntos de dados, demandam de referências estáveis e a garantia de que seus conjuntos de dados armazenados sejam devidamente reutilizados e permaneçam válidos por longo período de tempo. A identificação apropriada permite que o conjunto de dados seja preservado, recuperado, compartilhado e citado (SAYÃO; SALES, 2015).

Repositórios de dados, implementados com o *software Dspace*, tem o aspecto particular da funcionalidade em criação de identificadores persistentes para cada item, coleção e comunidade armazenada no sistema.

A vantagem em agregar um identificador persistente aos conjuntos de dados consiste no fato de que esse identificador aponta sempre para o conjunto de dados, mesmo que tenha sido alterada sua localização. O repositório re3data.org disponibiliza informações sobre repositórios que fornecem identificadores persistentes para os dados. Existem vários sistemas de identificação persistente baseados em padrões abertos como *Digital Object Identifiers (DOI)*, *Handle System* e *PURL* (SAYÃO, 2007).

Um ID Autor é um identificador exclusivo usado para distinguir o autor de outros com o mesmo nome ou nome semelhante. O autor tem a possibilidade de assegurar que todas as suas publicações estejam associadas ao seu perfil em bancos de dados ao utilizar esse tipo de ID. Inscrever-se para ID Autor é uma prática propícia aos pesquisadores para tornar suas publicações mais fáceis de localizar e



identificar. O uso do *Open Researcher & Contributor ID* (ORCID), identificador digital persistente para autor, contribui com esse aspecto (THOMAS; CHEN; CLEMENT, 2015).

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

### 3.3 Licenças de uso

Licença representa a cessão do direito sobre uma obra cuja exploração em outras condições seria ilícita. A licença pode revestir a forma de um direito exclusivo ou não exclusivo. A licença exclusiva habilita seu titular a explorar a obra, com exclusão das outras pessoas (incluindo o próprio autor) da maneira autorizada (UNESCO, 1981).

No caso em que a transferência ou cessão dos direitos for total, o titular adquire todos os direitos patrimoniais que um autor possa ter sobre a obra e está habilitado a exercê-los como se tratasse do próprio autor (UNESCO, 1981). Em caso de transferência ou de cessão parcial, o cessionário só adquire o direito de exercer um ou mais direitos específicos no contrato (UNESCO, 1981). A licença de uso que o titular dos direitos autorais concede deve especificar de forma clara e objetiva o modo de utilização do que está sendo licenciado.

O contrato é o instrumento jurídico que regulamenta o acordo entre o titular dos direitos e o licenciado. No contrato estão determinadas as obrigações e os direitos das partes envolvidas em relação ao uso e exploração dos conjuntos de dados visando sempre o melhor desenvolvimento e aproveitamento dos mesmos.

Licenças de uso são “[...] instrumentos legais que, normalmente, objetivam garantir o reconhecimento da autoria e promover a distribuição livre e aberta do conhecimento construído coletivamente.” (MANTOVANI; DIAS; LIESENBERG, 2006, p. 273).

As licenças de uso asseguram a publicação de dados sem ferir os direitos autorais, transfere o direito patrimonial a outrem, cria clareza quanto a utilização do conteúdo e constitui uma segurança para os usuários no sentido da utilização dos

conjuntos de dados. A licença de maior destaque nos repositórios de dados são as *Creative Commons* (CC).

*Creative Commons* é uma organização global, não governamental, sem fins lucrativos localizada em *Mountain View*, Califórnia, fundada em 2001 por Lawrence Lessig e autoriza a partilha e reutilização da criatividade e conhecimento através do fornecimento de instrumentos legais gratuitos (CREATIVE COMMONS, 2016).

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

O projeto *Creative Commons* estabelece um conjunto de ferramentas jurídicas

[...] que permite a artistas de todo o mundo escolher a forma como querem proteger suas obras. Se quiserem que as obras sejam 100% protegidas, não precisam fazer nada, elas já nascem assim. Mas aqueles que voluntariamente desejam fazer com que sua criação intelectual possa ser compartilhada com o mundo, ou mesmo remixada, alterada, traduzida ou distribuída por outras pessoas, passa a ter o movimento do CC uma ferramenta global para isso (LEMOS, 2005, p. 20).

A licença *Creative Commons* é uma maneira do detentor do direito autoral autorizar previamente a utilização da obra comercialmente ou não, alterando o conteúdo original ou não dependendo de qual é a chancela CC escolhida (LEMOS, 2005).

As licenças *Creative Commons* oferecem uma maneira fácil de gerenciar os termos de direitos autorais que se ligam automaticamente a todo o material criativo sob *copyright* (CREATIVE COMMONS, 2016). Suas licenças permitem que os materiais que serão compartilhados e reutilizados sob os termos sejam flexíveis.

Licenças como a *Creative Commons* (CC) substituem todos os direitos reservados do *copyright* para alguns direitos reservados e permitem o uso mais flexível composto por combinações de cláusulas determinadas pelos autores. Há seis licenças CC padrão. CC-BY é a licença mais utilizada, em que a atribuição é obrigatória quando se utiliza dados (DATA..., 2016d). O pesquisador pode escolher restrições como não-comercial, não derivados, ou partes iguais (DATA..., 2016d).

*Creative Commons* oferece seis licenças de direitos de autor (CREATIVE COMMONS, [20--]):

1 - *Creative Commons* Atribuição (CC-BY) 


Licença CC mais flexível de todas as licenças disponíveis. É recomendada para maximizar a disseminação e uso dos materiais licenciados. Permite a outro distribuir, remixar, adaptar e criar a partir do seu trabalho, mesmo para fins comerciais, desde que lhe atribua o devido crédito pela criação original.

2 - *Creative Commons* Atribuição-Compartilhual (CC BY-SA) 

Esta licença permite que outros remixem, adaptem e criem a partir do trabalho original, mesmo para fins comerciais, desde que lhe atribuam o devido crédito e que licenciem o trabalho derivado sob termos idênticos. CC BY-SA costuma ser comparada com as licenças de *software* livre e de código aberto "*copyleft*". Todos os trabalhos baseados no original terão a mesma licença, portanto quaisquer trabalhos provenientes também permitirão o uso comercial. Esta é a licença usada pela Wikipédia e é recomendada para materiais que seriam beneficiados com a incorporação de conteúdos da Wikipédia e de outros projetos com licenciamento semelhante.

3 - *Creative Commons* Atribuição-SemDerivações (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.

4 - *Creative Commons* Atribuição-NãoComercial (CC BY-NC) 

Esta licença permite à outros remixarem, adaptarem e criarem a partir do trabalho original para fins não comerciais e, embora os novos trabalhos tenham que atribuir o devido crédito ao autor e não possam ser usados para fins comerciais, os usuários não precisam licenciar os trabalhos derivados sob os mesmos termos.

### 5 - *Creative Commons* Atribuição-NãoComercial-Compartilhalgual (CC BY-NC-SA)



Esta licença permite que outros remixem, adaptem e criem a partir do trabalho original para fins não comerciais, desde que atribuam o devido crédito ao autor e que licenciem as novas criações sob termos idênticos.

### 6 - *Creative Commons* Atribuição-SemDerivações-SemDerivados (CC BY-NC-ND)



Esta é a mais restritiva das seis licenças, só permitindo que outros façam *download* dos trabalhos e os compartilhem desde que atribuam crédito ao autor, mas sem que possam alterá-los de nenhuma forma ou utilizá-los para fins comerciais.

A *Creative Commons* fornece também instrumentos que operam no espaço "todos os direitos concedidos", do domínio público. O instrumento CC0 permite aos licenciados renunciarem a todos os direitos e colocarem seus trabalhos no domínio público. A Marca de Domínio Público permite que qualquer usuário da internet "sinalize" um trabalho para indicar que este se encontra no domínio público (CREATIVE COMMONS, [20--])

No rights reserved (CC Zero) ou (CC0)



















CC Zero permite aos criadores e proprietários de conteúdo renunciar os direitos sobre suas obras e, assim, colocá-los tão completamente quanto possível no domínio público, para que outros possam construir livremente sobre a obra original, melhorá-las e reutilizá-las para quaisquer fins sem restrições de direito autoral.

Os usuários que utilizarão as licenças CC podem combinar materiais de diferentes licenças *Creative Commons* observando alguns aspectos: se a combinação não cria uma adaptação, então é permitido combinar qualquer conteúdo licenciado pela CC, desde que não forneça atribuição e cumpra com a restrição NãoComercial que se aplica. Se o usuário quiser combinar material de uma forma que resulte na criação de uma adaptação (isto é, um "remix"), então deve prestar atenção para a licença específica que se aplica ao conteúdo que deseja combinar (CREATIVE COMMONS (2016)).

Os remix podem ser observados na figura 11.

Figura 11 - Material licenciado CC remixado

	 PUBLIC DOMAIN	 PUBLIC DOMAIN	 CC BY	 CC BY SA	 CC BY NC	 CC BY ND	 CC BY NC SA	 CC BY NC ND
 PUBLIC DOMAIN	✓	✓	✓	✓	✓	✗	✓	✗
 PUBLIC DOMAIN	✓	✓	✓	✓	✓	✗	✓	✗
 CC BY	✓	✓	✓	✓	✓	✗	✓	✗
 CC BY SA	✓	✓	✓	✓	✗	✗	✗	✗
 CC BY NC	✓	✓	✓	✗	✓	✗	✓	✗
 CC BY ND	✗	✗	✗	✗	✗	✗	✗	✗
 CC BY NC SA	✓	✓	✓	✗	✓	✗	✓	✗
 CC BY NC ND	✗	✗	✗	✗	✗	✗	✗	✗

Fonte: CREATIVE COMMONS (2016).

O material licenciado-CC pode ser remixado. Para verificação do remix na figura 11, deve-se encontrar uma licença na coluna da esquerda e na linha superior. Se houver uma marca de verificação na caixa de onde essa linha e coluna se cruzam, as obras podem ser remixadas. Se houver um "X" na caixa, as obras não podem ser remixadas (CREATIVE COMMONS ([20--d])).

Lessig afirma que a cultura do remix é a peça central do processo criativo contemporâneo e sempre defendeu que essa cultura “[...] é aquela que vai derrubar as barreiras entre criador e ‘consumidor’ da cultura. [...] Entre Estado e cidadão.” (LEMONS, 2005, p. 18).

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

1 - ODC Domínio Público e Licença de Dedicção (PDDL)

Esta licença é semelhante à licença CC0, na medida em que não impõe restrições sobre o uso de seus dados / banco de dados. Os usuários são livres para copiar, distribuir e usar o trabalho; produzir trabalhos a partir dos dados / bancos de dados armazenados e para modificar, transformar e construir sobre os dados / banco de dados.

## 2 - Licença de Atribuição ODC

Esta licença é semelhante à licença CC BY, na medida em que permite que outras pessoas copiem, distribuam e usem o trabalho e produzam trabalhos a partir dos dados armazenados além de modificar, transformar e construir sobre os dados / banco de dados, desde que eles forneçam a atribuição adequada.

## 3 - Licença Open Database (ODbL)

É semelhante à licença CC BY-SA, uma vez que permite que outras pessoas copiem, distribuam e utilizem o trabalho e produzam trabalhos a partir dos dados / banco de dados e modifiquem, transformem e construam sobre os dados / banco de dados, desde que eles forneçam a atribuição adequada e que todos os trabalhos derivados sejam disponibilizados sob esta mesma licença.

ODC também fornece informações sobre como essas licenças diferem das licenças *Creative Commons* em seu aplicativo para bancos de dados.

Cada área do conhecimento tem suas particularidades na coleta e utilização dos conjuntos de dados. Dependendo desse conjunto é necessário o uso de *software* para a visualização e manipulação dos dados. O conhecimento das particularidades de cada *software* influenciará na escolha. Dessa forma, o *software* não pode ter restrição de uso, o pesquisador que irá utilizar os dados deve ter um *software* livre para ser utilizado. As licenças GNU *General Public Licence* são utilizadas para esse fim.

Richard Stallman é fundador do movimento *software* livre, do projeto GNU, um *software* livre que respeita a liberdade dos usuários (GNU OPERATING

SYSTEM, 2016), da *Free Software Foundation* (FSF) e autor da GNU *General Public Licence*. Stallman é produtor de programas alternativos de *software* e descreve a evolução de direito de autor. As licenças GNU são:



#### 1 - Licença GNU General Public Licence

*General Public License* (GPL) (Licença Pública Geral). A mais difundida dessas licenças é a GNU *General Public License* GNU ou GNU GPL licença designada para *software* (um *software* livre que respeita a liberdade dos usuários (GNU OPERATING SYSTEM, 2016). É utilizada pela maioria dos programas do GNU assim como muitos outros programas de *software* livre que não são parte do Projeto GNU. A versão mais recente é versão 3 (GNU OPERATING SYSTEM, 2016).

#### 2 - GNU General Public License 3 (GPL-3.0)

Permite copiar, distribuir e modificar o *software*, desde que acompanhe as mudanças/datas em arquivos de origem e que mantenha as modificações sob a GPL. Permite distribuir um aplicativo usando uma biblioteca GPL comercialmente, desde que forneça o código-fonte. GPL 3 tenta fechar algumas lacunas na GPL 2 (GNU OPERATING SYSTEM, 2016).

#### 3 - GNU Library or "Lesser" General Public License 3.0 (LGPL-3.0)

Permite copiar, distribuir e modificar o *software*, desde que as modificações sejam descritas dentro dos arquivos modificados e licenciado gratuitamente sob LGPL-2.1 (OPEN SOURCE INITIATIVE, [20--]a).

#### 4 - GNU Affero General Public License 3 (AGPL-3.0)

É uma licença *copyleft* livre para *softwares* e outros tipos de obras, especificamente concebido para assegurar a cooperação com a comunidade no caso de *software* servidor de rede. A licença AGPL difere das outras licenças GNU, em que foi construído para o *software* de rede. Permite distribuir versões modificadas, manter o

controle das mudanças e a data em que as realizou (OPEN SOURCE INITIATIVE, c2007).

Outras licenças de *software* destacadas no trabalho que podem ser usadas para visualização e utilização dos conjuntos de dados:

#### 1 - Artistic License 2.0

É uma licença para os pacotes de *software* com a intenção de dar ao detentor dos direitos autorais originais alguma medida de controle sobre seu *software*, enquanto ainda permaneça *open source*. É flexível e permite distribuir e vender versões modificadas, desde que cumpra uma das várias condições (OPEN SOURCE INITIATIVE, [20--]b).

#### 2 - Mozilla Public License 2.0

Esta é uma licença suave utilizada pela *Mozilla Corporation* que permite uma variedade de liberdades explícitas com o *software*, desde que se mantenham modificações sob esta licença e distribua o código-fonte original (OPEN SOURCE INITIATIVE, [20--c]).

#### 3 - Eclipse Public License 1.0 (EPL-1.0)

Esta licença, feita e utilizada pela Fundação Eclipse, não é muito rigorosa e dá o direito de autor e direitos de patente explícitos (OPEN SOURCE INITIATIVE, [20--d]).

#### 4 - Common Development and Distribution License (CDDL-1.0)

Licença com permissão maior feita pela *Sun Microsystems* que inclui concessões de patentes explícitas. É relativamente simples em suas condições, necessitando apenas de uma pequena quantidade de documentação para redistribuição (OPEN SOURCE INITIATIVE, [20--e]).

.



#### 5 - The MIT License (MIT)

A permissão é concedida, gratuitamente, a qualquer pessoa que obtenha uma cópia deste *software* e dos arquivos de documentação associados, sem restrições, incluindo, sem limitação, os direitos de usar, copiar, modificar, mesclar, publicar, distribuir, sub-licenciar e/ou vender cópias do *software* (OPEN SOURCE INITIATIVE, [20--f]).

#### 6 - The BSD 3-Clause "New" or "Revised" License (BSD)

The BSD 3-clause tem permissão quase ilimitada com o *software*, desde que inclua o aviso de BSD *copyright* nele (OPEN SOURCE INITIATIVE, [20--g]).

#### 7 - The BSD 2-Clause "Simplified" or "FreeBSD" License

Permite uma liberdade quase ilimitada com o *software*, desde que inclua o aviso de BSD *copyright* nele (OPEN SOURCE INITIATIVE, [20--h]).

#### 8 - Apache License 2

Uma licença que permite utilização, reprodução e distribuição do *software*, incluindo o direito explícito de uma patente. "Mudanças de estado" significa que o usuário tem que incluir um aviso em cada arquivo modificado (THE APACHE..., c 2016).

Atribuição de licenças para dados também tem desvantagens. Licenças são estáticas, não mudam com o desenvolvimento rápido no campo de dados científicos. Portanto, alguns repositórios de dados trabalham com uma licença CC0 pelo qual os direitos não são reservados (DATA..., 2016d). Todas as licenças possibilitam a utilização de materiais sem ferir os direitos autorais.

Os dados provêm em vários formatos: consistem em valores numéricos medidos por um equipamento; em textos não estruturados ou uma coleção de respostas estruturadas entre outros. Gerir dados significa coletar, descrever, analisar, organizar, armazenar, preservar e publicar os dados para reutilização. O Plano de Gerenciamento de Dados é uma maneira eficaz de trabalhar através dessas ações e selecionar abordagens consistentes e reproduzíveis para cada um

deles. PGD articulado possibilita o gerenciamento dos dados nos repositórios de dados. A próxima seção apresentou os PGD dos repositórios.

#### 4 PLANO DE GERENCIAMENTO DE DADOS

O PGD é adotado nesta pesquisa como documento ou conjunto de informações e instruções disponibilizadas no próprio sítio dos repositórios o qual documenta, desde o início da pesquisa que aborda dados científicos, como serão os critérios de organização e gerenciamento, incluindo a coleta, disponibilização, acesso, curadoria e questões relacionadas aos direitos autorais na utilização dos dados após sua publicação.

A gestão dos direitos de propriedade intelectual é um componente expressivo de qualquer programa de gestão de dados. Repositórios de dados que disponibilizam esse recurso certificam que todas as obrigações legais e éticas são atendidas em relação aos conjuntos de dados nele armazenados. Pesquisadores podem considerar seguro compartilhar seus conjuntos de dados com outras pessoas visto que este aspecto é persuasivo e documentado.

A disponibilização de conjuntos de dados viabiliza sua partilha entre grupos específicos ou comunidade científica em geral. Apesar de seus desafios, há crescente reconhecimento de que o compartilhamento de dados proporciona benefícios tais como permitir a verificação dos resultados da investigação de outros pesquisadores, contribuir com dados em amplos campos de atuação e facilitar a reutilização de dados por terceiros (PINFIELD; COX; SMITH, 2014).

Os financiadores de pesquisas, conscientes da necessidade de incentivar boas práticas científicas e alcançar um maior valor para a pesquisa que eles patrocinam, pressionam os pesquisadores a elaborar um Plano de Gerenciamento de Dados (PINFIELD; COX; SMITH, 2014). Tais exigências servem para enfatizar a necessidade de gerenciamento de dados de investigação dentro das instituições (PINFIELD; COX; SMITH, 2014).

As instituições na área científica presenciam desafios no gerenciamento de dados científicos produzidos em grandes quantidades, velocidade e variedades por pesquisadores nas diversas áreas do conhecimento. A gestão de conjuntos de dados é um processo ativo pelo qual os tornam detectáveis e acessíveis, com o potencial de agregação de valor à pesquisa apreciando maior uso entre os pesquisadores. Na ambiência da investigação científica, tal processo de agregação de valor é um contribuinte significativo para a realização tão desejada de impacto

(PRYOR, 2012)<sup>6</sup>. Este contexto gera a necessidade de elaborar políticas para o gerenciamento de dados nas instituições.

Pesquisadores compartilham as mesmas dificuldades associadas ao gerenciamento e compartilhamento de dados com outras pessoas. As dificuldades com a gestão de dados muitas vezes têm origem em falta de treinamento formal. Cientistas não têm um conhecimento prático de conceitos básicos de gerenciamento de dados, assim como de metadados e ontologias (LEE et al., 2009).

Conjuntos de dados de um determinado grupo de pesquisadores podem conter diferentes formatos, tipos e descrições, tornando-os heterogêneos. À medida que o tamanho dos conjuntos de dados aumenta, o gerenciamento de conjuntos de dados heterogêneos tende a se tornar árduo (LEE et al., 2009). Muitos problemas relevantes de informática e da Ciência da Informação, tais como pesquisar e fundir grandes conjuntos de dados, não estão resolvidos bem como pode ser difícil descobrir, compartilhar ou reutilizar dados pois:

- 1) dados valiosos podem ter sido descartados;
- 2) modelos de computador podem ser inutilizáveis ou difíceis de usar devido a problemas anteriores;
- 3) formatos incompatíveis podem tornar os dados difíceis ou impossíveis de integrar;
- 4) o fluxo de dados entre domínios pode ser impedido por metadados incompletos, imprecisos e/ou mal descritos;
- 5) muitos cientistas relutam em compartilhar dados devido à falta de recompensa, às questões de propriedade intelectual e documentação apropriada (LEE et al., 2009).

Um elemento relevante na gestão de dados científicos é assegurar que os mesmos possam ser compreendidos e interpretados por outros pesquisadores ao longo do tempo. Para isso é essencial uma descrição clara e detalhada dos dados, anotações adicionais e informações que contextualizam os dados e possibilitem que transmitam informação e conhecimento no tempo e no espaço (SAYÃO; SALES,

---

<sup>6</sup> conceito de Pryor retirado de informações referente ao livro PRYOR, G. (Ed.) **Managing research data**. United Kingdom: Facet Publishing, 2012. Disponível em: <<http://www.dcc.ac.uk/news/book-managing-research-data>>. Acesso em: 27 set. 2016.

2015). A documentação dos dados é o instrumento que apoia e efetiva essas ações (SAYÃO; SALES, 2015).

O Plano de Gerenciamentos de Dados é um plano

[...] constituído por um número de diferentes atividades e processos associados ao ciclo de vida de dados, envolvendo a concepção e criação de dados, armazenamento, segurança, preservação, recuperação, partilha e reutilização, todos tendo em conta as capacidades técnicas, considerações éticas, questões legais e estruturas de governança.”<sup>7</sup> (COX; PINFIELD, 2014, tradução nossa).

A preparação de um PGD envolve atividades em diferentes graus de formalidade, extensão, periodicidade, metas e objetivos. O desenvolvimento de seu conteúdo envolve as particularidades de cada área abrangida pelo repositório. É imprescindível analisar e planejar o conteúdo observando os aspectos legais como a identificação da titularidade, licenças e contratos, o que evitará conflitos e violações no futuro (GANDELMAN, 2007).

O plano das atividades tem o planejamento como fase inicial e minimiza riscos, aproveita as oportunidades, define objetivos, sequências de ações decisórias e fases, verifica variáveis e prevê mudanças, pois viabiliza e controla as ações, traz segurança para os colaboradores, aumenta a colaboração e facilita o controle (ALMEIDA, 2000).

O planejamento na gestão de dados se inicia antes da implementação do repositório e, para o pesquisador, no momento do delineamento da pesquisa e deve-se considerar como os dados serão gerenciados durante o desenvolvimento do projeto e como eles serão compartilhados no final com a disponibilização (SAYÃO; SALES, 2015).

Os procedimentos adotados na execução de um PGD são instrumentos que definem e estabelecem métodos de execução das atividades e detalham a forma exata pela qual os procedimentos serão realizados. O planejamento é um processo cíclico, dinâmico e interativo, em que as fases não precisam ser lineares, pois há uma dinâmica no processo (ALMEIDA, 2000). É constituído por um complexo de

---

<sup>7</sup> “[...] consists of a number of different activities and processes associated with the data lifecycle, involving the design and creation of data, storage, security, preservation, retrieval, sharing, and reuse, all taking into account technical capabilities, ethical considerations, legal issues and governance frameworks.”

instrumentos, cada um com uma função específica que garantem a eficácia do processo quando devidamente articulado (ALMEIDA, 2000).

O planejamento é, portanto, um conjunto altamente complexo de atividades que envolvem uma série de desafios técnicos, bem como um grande número de questões culturais, administrativas, jurídicas e políticas (PINFIELD; COX; SMITH, 2014).

Geralmente a instituição que implementa repositório de dados tem em seu sítio um documento ou informações relacionadas ao plano de gestão de dados para orientar os autores que depositarão seus conjuntos de dados.

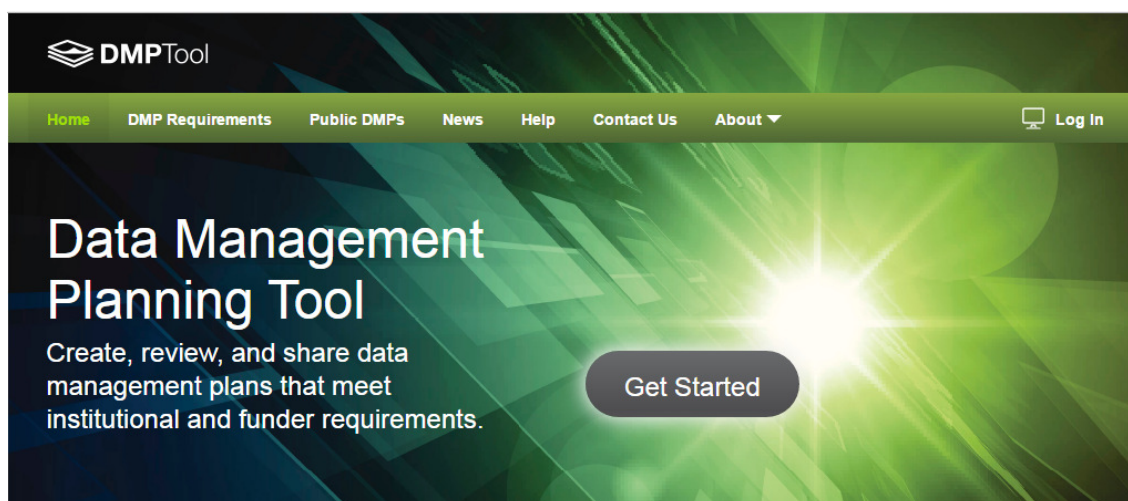
O uso de PGD pelos repositórios de dados proporciona benefícios aos pesquisadores, pois orienta sobre vários aspectos como conforme descrito nos vários repositórios pesquisados:

- Fornece opções de acesso flexíveis: os dados ficam acessíveis a todos, ou com acesso moderado mediante solicitação, dependendo das opções do pesquisador;
- Acesso a longo prazo: identificadores persistentes e DOIs tornam mais fácil para os pesquisadores poderem localizar e citar os dados;
- Os dados recebem um URL permanente sob a forma de um identificador de objeto digital (DOI) para que o pesquisador possa conectar seus dados para suas publicações;
- Análise quantitativa: mostra com que frequência os seus dados são vistos e feito *download*;
- Maximiza a reutilização: pode ter consulta com o pesquisador para garantir que os dados estejam em um formato e estrutura que melhor facilita o acesso em longo prazo, descoberta e reutilização;
- O repositório e os dados podem ser indexados pelo Google, o que aumenta a possibilidade de outros pesquisadores encontrarem os dados;
- O servidor mantém com *backups* e manutenção regular para evitar a perda de dados.

Nos sítios também contém indicação de ferramentas que auxiliam os autores a montarem seus PGDs.

O DMPtool<sup>8</sup> é uma ferramenta da Universidade da Califórnia que auxilia o pesquisador criar Plano de Gerenciamento de Dados incluindo os requisitos necessários para tal. Fornece orientações sobre instituições financiadoras específicas que exigem PGD e um guia para a elaboração do documento de PGD. Os PGDs são propriedade intelectual de quem os criou. O pesquisador que elabora o PGD no DMPtool pode optar em compartilhar seu PGD publicamente contribuindo com outros pesquisadores.

Figura 12 - Tela inicial do DMPtool



Fonte: <https://dmptool.org/>.

Os usuários do DMPTool podem visualizar amostra de planos, requisitos das agências financiadoras e exibir as alterações mais recentes feitas em seus planos uma vez que permite ao usuário criar um documento editável para apresentar a uma agência de financiamento e pode acomodar versões diferentes à medida que os requisitos de financiamento mudam.

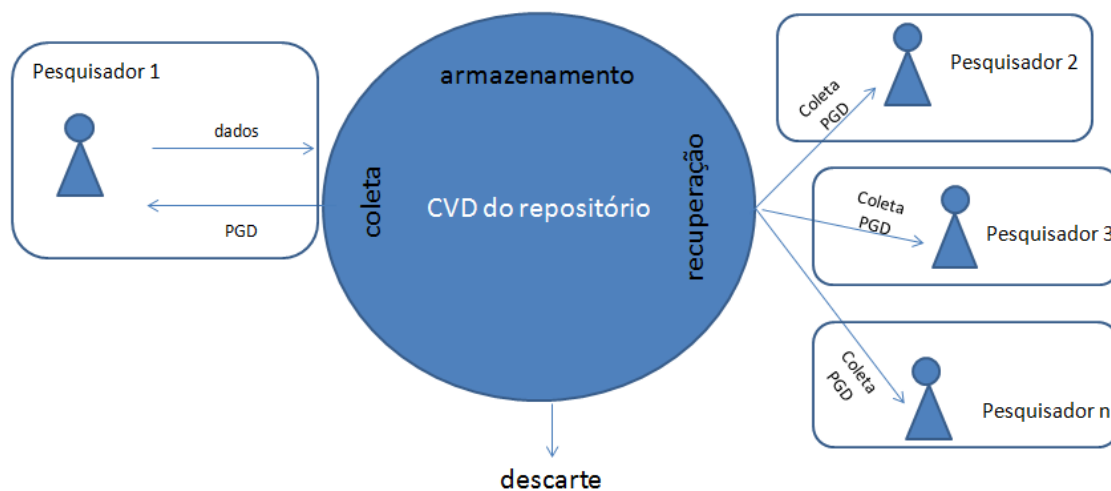
---

<sup>8</sup> <https://dmptool.org/>

## 5 ANÁLISES E DISCUSSÕES

Nessa seção são apresentadas as discussões abordadas com base na análise dos Planos de Gerenciamento de Dados dos Repositórios de dados de cada universidade relacionadas às questões de Direitos autorais na fase da Coleta do CVD-CI. A fase da Coleta está relacionada à coleta dos repositórios quando os pesquisadores depositam seus conjuntos de dados, conforme a figura 13.

Figura 13 – Ciclo de Vida dos Dados no Repositório



Fonte: Elaborada pela autora

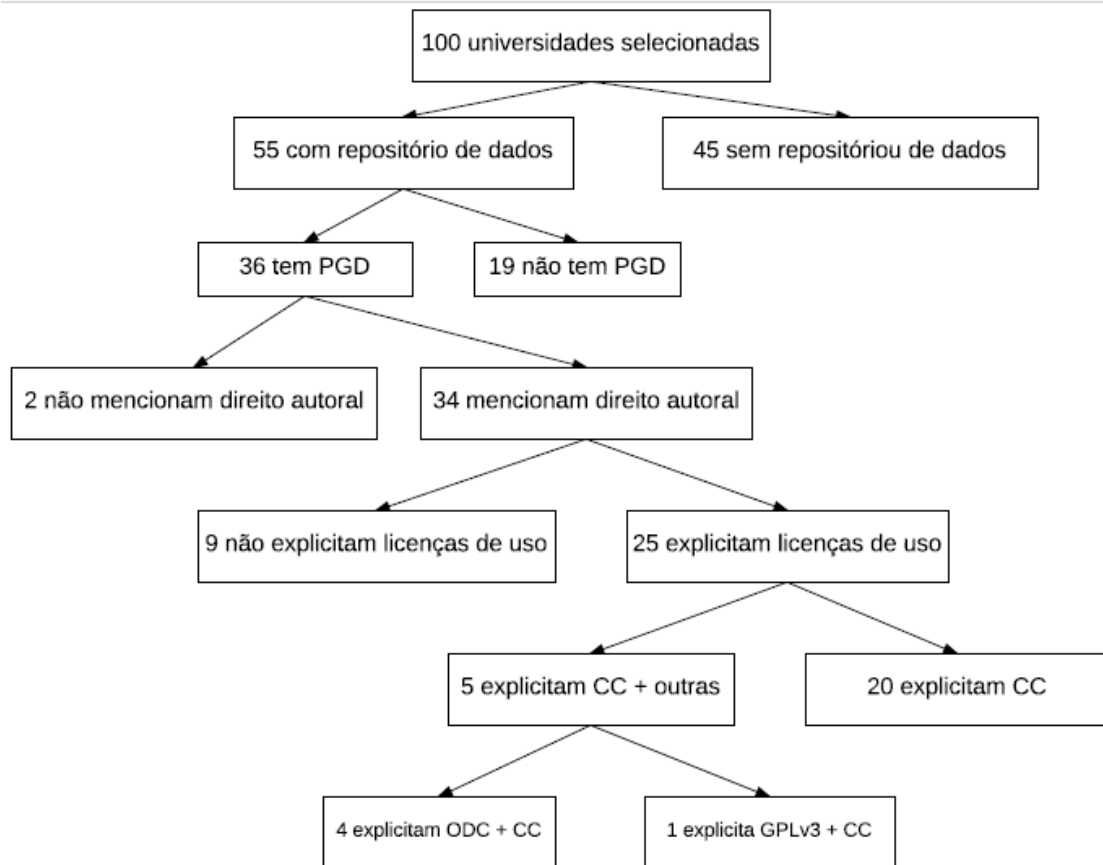
Verifica-se na figura 13 que o pesquisador 1 envia seu conjunto de dados para o repositório. Esse processo é a fase da coleta do CVD do repositório, pois coleta os dados do pesquisador 1 que ficarão arquivados e disponíveis para outros pesquisadores como o 2, 3, n pesquisarem e coletarem no repositório. Os pesquisadores 2, 3, n recuperarão os dados no repositório iniciando uma nova fase de coleta, no caso coleta dos pesquisadores, iniciando um novo ciclo de vida dos dados e não mais o ciclo do repositório. O Plano de Gerenciamento de Dados (PGD) é o documento disponível pelo repositório o qual orientará os pesquisadores sobre as políticas do repositório, entre elas, os aspectos relacionados ao direito autorais.

Para investigar se os repositórios de dados das cem melhores universidades do mundo têm PGD com critérios de direitos autorais utilizou-se os sítios oficiais dos



repositórios de dados das universidades. As análises incluíram a identificação dos repositórios de dados das universidades, identificação dos Planos de Gerenciamento de Dados e explicitação das instruções sobre direitos autorais dos Planos de Gerenciamento de Dados.

Figura 14: Ilustração do direcionamento das análises



Fonte: Elaborada pela autora

A figura 14 ilustra os caminhos que foram seguidos para explicitar o direito autoral nos PGDs dos repositórios de dados abordados nas próximas subseções.

As análises foram sobre os 34 PGDs que mencionaram aspectos relacionados ao direito autoral, explicitando, inclusive as licenças abordadas.

## 5.1 Identificação dos repositórios de dados das universidades

Para localização das 100 melhores universidades do mundo foi utilizado o ranking *webometrics.info*. Entre as listas dos rankings, Exame da editora Abril, do Estudarfora.org, Shanghai ranking, e do sítio *webometrics.info*., utilizou-se este último por ser a maior classificação acadêmica do mundo, ser um ranking do *Cybermetrics Lab*, grupo de pesquisa que pertence ao *Consejo Superior de Investigaciones Científicas* (CSIC), o maior organismo público da Espanha, que está entre as primeiras organizações de pesquisa básica na Europa e é vinculado ao Ministério da Educação (RANKING..., 2016). O *webometrics.info* ranqueia também os repositórios, com o objetivo de apoiar iniciativas de Acesso Aberto e o livre acesso às publicações científicas em formato eletrônico e outros materiais acadêmicos.

No ranking do *webometrics.info* são registradas 11.996 universidades. Para o escopo da pesquisa utilizou-se as 100 melhores, conforme apresentado no Anexo A.

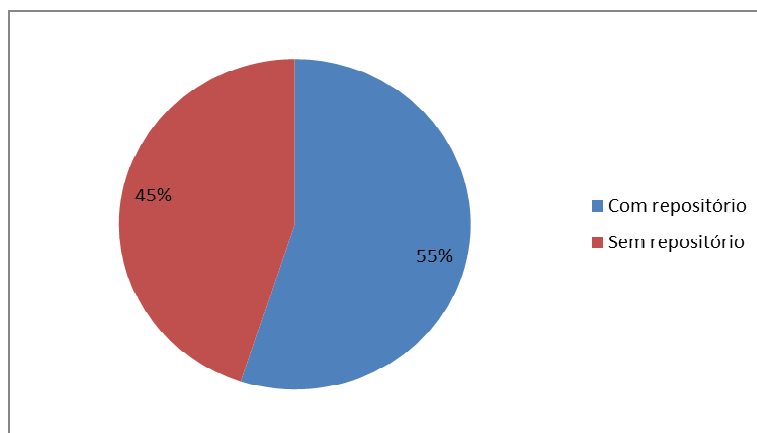
Para a localização dos repositórios de dados das universidades foram utilizados os sítios oficiais de cada universidade disponibilizados pelo *webometrics.info*. Após, adotou-se o método exploratório, iniciando-se a coleta de dados dentro dos sítios de cada universidade para localização de repositório de dados na seguinte ordem:

- 1) Busca, dentro do sítio oficial, de *link* direto para o repositório de dados;
- 2) Quando não localizado o *link* direto para o repositório de dados na página inicial, foram acessados todos os *links* da página para a localização;
- 3) Não localizado, foi feita a busca na caixa de busca dentro do sítio inicial das universidades pelo descritor “*data repository*” nas universidades de idioma inglês, para as universidades que tem opção de verificação do site no idioma inglês e com o descritor, “repositorio de datos” para universidades com idioma espanhol e o descritor “repositório de dados” para idioma português;
- 4) Não localizado, foi feita a busca no *Google* com o nome da universidade adicionada aos descritores apontados;
- 5) Com a localização de repositório de dados por um dos caminhos acima mencionados, foi verificado se o repositório possuía PGD.

O uso do descritor de pesquisa dentro dos sítios oficiais das universidades foi baseado no conhecimento prévio da pesquisadora conforme leituras e pesquisas feitas considerando que são descritores utilizados na literatura para nomear as páginas dos repositórios de dados.

Com os resultados alcançados das 100 melhores universidades do mundo foram localizados repositórios de dados em 55 delas correspondendo a 55% e que em 45 universidades não foram localizadas repositórios de dados pelos caminhos indicados na metodologia, prevalecendo repositório de dados na maioria delas.

Gráfico 1– Total de universidade com Repositório de Dados



Fonte: Elaborado pela autora

Na verificação das 55 universidades que disponibilizam repositórios de dados verificou-se que:

- a *Duke University* e a *Rutgers The State University of New Jersey* - não tem informações na página inicial, solicitam *login* e senha para entrar no ambiente do repositório. Não foi possível fazer a análise para a localização dos PGDs;

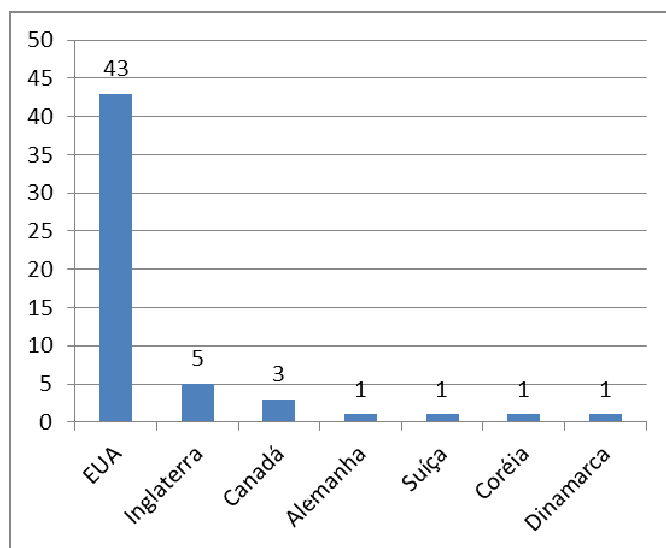
Com a análise das 45 universidades que não foram localizados repositório de dados, destaca-se que:

- na *Carnegie Mellon University* as bibliotecas recebem e gerenciam os conjuntos de dados de suas comunidades e tem PGD;
- *University of Utah* - sua biblioteca aceita e gerencia os conjuntos de dados de sua comunidade mas não tem PGD;

- na *University of Rochester*, a biblioteca aceita os conjuntos de dados e enviam para o repositórios de dados *Dryad* indicados pelas mesmas no ato do recebimento.
- na *University of New South Wales*, a biblioteca recebe os conjuntos de dados e enviam para um repositório de dados sem mencionar no PGD para qual enviarão.

O gráfico 2 apresenta a quantidade de repositórios de dados nas universidades de cada país com base nas cem melhores universidades .

Gráfico 2 – Quantidade de Repositório de Dados por país



Fonte: Elaborado pela autora

Nota: Baseado nas cem melhores universidades do mundo

Observa-se no gráfico 2 que a maioria dos repositórios de dados que foram localizados estão nas universidades dos Estados Unidos no total de 43, seguida da Inglaterra com 5 e Canadá com 3. Os países que menos tem universidades com repositórios de dados são a Alemanha, Suíça, Coreia e Dinamarca com apenas um em cada universidade. Evidencia-se que prevalecem universidades de idioma inglês.

Com a localização do repositório de dados das universidades foi feita análise do sítio para verificar se disponibilizavam PGDs. Para a localização dos PGDs foram averiguados todos os links do sítio oficial dos repositórios de dados observando

informações que abordassem sobre plano de gerenciamento de dados, políticas de uso, gerenciamento de dados e dados e informações no item Sobre o repositório de dados tópico da próxima subseção.

## 5.2 Identificação dos Planos de Gerenciamento de Dados

O processo de recuperação dos dados foi realizado por meio da coleta dos Planos de Gerenciamento de Dados (PGDs) dos repositórios de dados das universidades. Os PGDs foram analisados com base no Ciclo de Vida dos Dados que perpassa o fator dos Direitos autorais na fase da Coleta, conforme figura 5. No entanto, para entendimento do Ciclo de Vida dos Dados foram apresentadas as quatro fases dentro dos seis fatores propostos por Sant'Ana (2016) na seção 1.4.1.

O PGD descreve como o repositório armazena, gerencia, descreve, fornece acesso, dissemina, integra e preserva os dados durante o período que os dados ficam armazenados, incluindo aspectos relacionados aos direitos autorais.

Na explicitação do PGD dos repositórios estão envolvidos procedimentos que englobam todo o gerenciamento dos dados até seu reuso por outros pesquisadores.

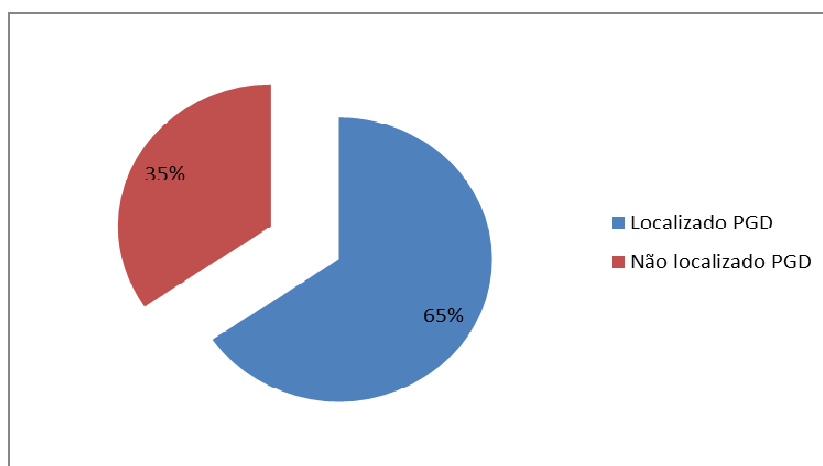
Os repositórios também indicam como o pesquisador pode elaborar seu próprio PGD. A ferramenta *DMPTool*, abordado na seção 4, auxilia nesse processo. A organização do PGD pelo pesquisador desde o início da pesquisa economiza tempo, garante que os dados sejam devidamente documentados e reutilizáveis e favorece oportunidades para colaboração. A disponibilização de PGD pelos repositórios orientam os pesquisadores no compartilhamento de seus conjuntos de dados.

A localização dos PGDs foi realizada através da exploração dos sítios dos repositórios de dados de cada universidade sendo verificadas as informações referentes ao gerenciamento de dados com abordagens sobre direitos autorais.

A metodologia quantitativa foi utilizada na quantificação dos repositórios de dados que disponibilizam PGDs e explicitam aspectos sobre direitos autorais. A quantificação foi feita também com as licenças apropriadas para que os autores liberem seus conjuntos de dados propiciando o reuso, por ser verificado variações da indicação das licenças.

O gráfico 3 ilustra a quantidade de repositórios de dados que disponibilizam PGDs em seus sítios. Nota-se que, dos 55 repositórios de dados das universidades apresentadas, 36 deles possuem PGDs, correspondendo 65% do total e 19 deles não foram localizados PGDs correspondendo 35% da totalidade.

Gráfico 3 – Relação de Repositórios de Dados com PGDs



Fonte: Elaborado pela autora

Os PGDs nos repositórios de dados foram analisados para verificar a orientação sobre direitos autorais para os pesquisadores que arquivam seus conjuntos de dados.

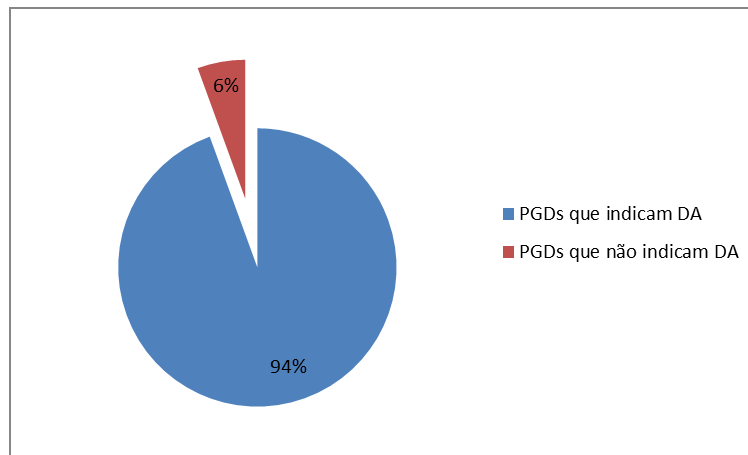
### 5.3 Explicitação dos Planos de Gerenciamento de Dados com instruções sobre direitos autorais

Nessa seção foram analisados como os repositórios de dados evidenciam a questão do direito autoral, por meio de seus PGDs na fase da Coleta do Ciclo de Vida dos Dados dos repositórios, indicado na figura 13. Os nomes das universidades que disponibilizam em seus sítios PGDs, por meio dos repositórios de dados, e o texto referenciado naquele sobre o direito autoral foram apresentados.

O gráfico 4 apresenta a quantidade de PGDs dos repositórios de dados que explicitam questões relacionadas ao direito autoral. Dos 36 PGDs localizados, 34

explicitam questões relacionadas ao direito autoral correspondendo a 94% e dois não explicitam correspondendo a 6%.

Gráfico 4 – PGDs com explicitação de DA



Fonte: Elaborado pela autora

No quadro 3 são apresentadas os 34 repositórios que explicitam questões relacionadas ao direito autoral em seus PGDs. Foi indicado na primeira coluna os nomes das universidades representando os repositórios, na segunda coluna o trecho do texto indicativo nos PGDs, na terceira coluna a indicação se os repositórios abordam o direito autoral, na quarta coluna se mencionam licenças, na quinta coluna se indica uma das licenças *Creative Commons* e na sexta coluna, a indicação de quais outras licenças mencionam.

Quadro 3 - Universidades que explicitam direito autoral em seus PGDs

Universidade	Techo do PGD	Direito autoral	Licenças	Creative Commons	Outras
Harvard University	<p>Os dados não serão dificultados com direitos de propriedade intelectual (incluindo direitos autorais, direitos de base de dados, restrições de licença, segredo comercial, patente ou marca comercial) por qualquer parte (incluindo os investigadores, instituições de investigadores e fornecedores de dados); Nem estão sujeitos a quaisquer requisitos legais adicionais. Depositando no Harvard Dataverse não requer uma transferência de direitos autorais, mas, em vez disso, concede permissão para o Harvard Dataverse para re-disseminar os dados e transformar os dados necessários para a preservação e acesso.</p> <p>Todos os dados sob este conjunto de dados também serão disponibilizados para replicação por qualquer parte sob a licença de atribuição CC, utilizando os protocolos LOCKSS - que é totalmente suportado pelo aplicativo Dataverse.</p> <p>Dataverse padroniza a citação de conjuntos de dados para tornar mais fácil para os pesquisadores publicarem seus dados e obter crédito, bem como reconhecimento para o seu trabalho. Quando se cria um conjunto de dados no Dataverse, a citação é gerada e apresentada automaticamente.</p>	Sim	Sim	Sim	Não
Massachusetts Institute of Technology	<p>Promover o compartilhamento e o uso ilimitado de seus dados, tornando-o disponível sob uma licença <i>Open Data Commons</i> ou <i>Creative Commons</i>.</p> <p>Cite dados no seu trabalho / apresentação para que você possa:</p> <p>Dar ao produtor de dados o crédito apropriado.</p>	Sim	Sim	Sim	ODC



Stanford University	<p>Uma licença definirá o que outros podem ou não fazer com seus dados. Você pode optar por atribuir uma licença ampla que permita a qualquer pessoa fazer o que quiser com seus dados, ou você pode atribuir uma licença mais restrita que restringe seu uso a atividades estritamente não comerciais e exige a atribuição do criador de dados sempre que ele for usado.</p> <p>As duas principais fontes atuais de licenças são <i>Creative Commons</i> e <i>Open Data Commons</i>. Se você optar por depositar dados no Repositório Digital Stanford, você terá a opção de escolher qualquer uma das licenças abaixo.</p> <p>CC BY CC BY-SA CC BY-ND CC BY-NC CC BY-NC-SA CC BY-NC-ND</p> <p><i>Open Data Commons</i> <i>Open Data Commons</i> (ODC) oferece três opções de licença criadas especificamente para dados / bancos de dados:</p> <p>ODC Domínio Público e Licença de Dedicção (PDDL):</p> <p>ODC Atribuição Licença:</p> <p>Licença de banco de dados aberto (ODbL):</p>	Sim	Sim	Sim	ODC
University of California Berkeley	<p>Selecionando uma licença para seus dados</p> <p>Uma licença <i>Creative Commons</i> (CC) é uma das várias licenças públicas de direitos autorais que permitem a distribuição gratuita de uma obra protegida por direitos autorais. Uma licença de CC é usada quando um autor quer dar às pessoas o direito de compartilhar, usar e construir sobre um trabalho que eles criaram.</p>	Sim	Sim	Sim	Não

University of Michigan	Os depositantes devem optar por aplicar uma das várias licenças <i>Creative Commons</i> aos dados. Os depositantes autorizam a Biblioteca a distribuir os dados sob os termos da licença que eles selecionaram.	Sim	Sim	Sim	Não
University of Pennsylvania	Antes de começar a carregar materiais para o <i>Scholarly Commons</i> , você precisará avaliar as permissões de direitos autorais para seu (s) trabalho (s).	Sim	Sim	Não	Não
University of Oxford	A Universidade reivindica a posse de toda a propriedade intelectual especificada na seção 6 deste estatuto que foi concebido, feito ou criado:	Sim	Não	Não	Não
Yale University	Conteúdo digital no Arquivo de Dados do ISPS está licenciado sob uma <i>Licença Creative Commons Atribuição-Não Comercial-Sem Derivados 3.0</i>	Sim	Sim	Sim	Não
University of Cambridge	As seguintes licenças são oferecidas para você escolher quando enviar seus dados de pesquisa para nós: CC BY CC BY-SA CC BY-ND CC BY-NC CC BY-NC-SA CC BY-NC-ND CC0 GPLv3  Nossa licença recomendada é CC BY.	Sim	Sim	Sim	GPL
University of Wisconsin Madison	Formalmente, recomendamos a adoção dos seguintes princípios: **  1. Quando forem publicados dados ou conjuntos de dados, é essencial que sejam publicados com uma declaração clara e explícita dos desejos e expectativas dos editores no que se refere à reutilização e re-definição de elementos de dados individuais, a totalidade da recolha de dados, e subconjuntos da coleção. Esta declaração deve ser precisa, irrevogável e baseada em uma declaração legal apropriada e reconhecida na forma de uma renúncia ou licença.  * Use uma renúncia reconhecida ou	Sim	Sim	Sim	Sim

	<p>licença que seja apropriada para dados. *</p> <p>3. A utilização de licenças que limitem a reutilização comercial ou limitem a produção de obras derivadas, excluindo a utilização para fins específicos ou por pessoas ou organizações específicas, é FORTEMENTE desencorajada.</p> <p>4. Além disso, na ciência, É FORTEMENTE recomendado que os dados, especialmente quando financiados por fundos públicos, sejam explicitamente colocados no domínio público através do uso da Dedicção e Licença do Domínio Público ou <i>Creative Commons Zero Waiver</i>.</p> <p>* A dedicação explícita dos dados subjacentes à ciência publicada no domínio público via PDDL ou CCZero é fortemente recomendada</p> <p>Padrões de citações de dados não existem em muitas disciplinas, embora a iniciativa <i>DataCite</i> está trabalhando neles. [...] Cite <i>datasets</i> pelas mesmas razões que você cita livros e artigos de revistas: para que os criadores de conjuntos de dados recebam o crédito apropriado pelo seu trabalho [...]</p>				
Michigan State University	<p>Aqui estão algumas opções para controle de acesso: [...] Liberar dados sem restrições de acesso usando a licença CCZero</p>	Sim	Sim	Sim	Não
University of Texas Austin	<p>Você, ou seus cessionários de direitos autorais mantém todos os seus direitos de propriedade intelectual. Os direitos limitados que você concede ao Repositório não são exclusivos, e sua capacidade de conceder, ceder ou reter todos os direitos que você tinha antes de sua submissão não mudam com o resultado de sua submissão.</p>	Sim	Não	Não	Não
University of California San Diego	<p>Todos os dados carregados podem ter licenças associadas a eles. O MIDAS suporta várias opções de licenças e a seleção da licença apropriada é deixada ao usuário que envia os dados</p>	Sim	Sim	Não	Não

	para o repositório.				
Pennsylvania State University	Abordar quaisquer questões de direitos de propriedade intelectual (por exemplo, quem detém esses direitos sobre os dados).	Sim	Não	Não	Não
University of Illinois Urbana Champaign	11. Meus dados incluem informações de potencial valor comercial. Estou obrigado a disponibilizar essa informação?  Não necessariamente. É a forte expectativa da NSF de que os pesquisadores compartilhem com outros pesquisadores [...] os dados primários [...]. No entanto, também é necessário proteger os direitos de propriedade intelectual e o potencial valor comercial. O Plano de Gestão de Dados deve descrever a abordagem proposta, que será então sujeita a revisão por pares e gestão de programas.	Sim	Não	Não	Não
University of North Carolina Chapel Hill	Se o depositante detém os direitos autorais, o depositante retém os direitos autorais dos materiais depositados no CDR. O CDR exige que o detentor dos direitos autorais autorize a distribuição não exclusiva e a preservação do trabalho ao depositar o conteúdo. [...] Os detentores de direitos autorais podem preferir usar o Modelo de Licenciamento da <i>Creative Commons</i> .	Sim	Sim	Sim	Não
Princeton University	Ao clicar nesta licença, o (s) autor (es) ou criador (es) ou detentor (es) dos direitos autorais concedem à Universidade de Princeton o direito não-exclusivo para reproduzir, traduzir e / ou distribuir o seu envio (incluindo o Resumo) em formato eletrônico. [...]  Você declara que a submissão é sua obra original e / ou que você tem o direito de conceder os direitos contidos nesta licença. Você representa que sua submissão não infringe, na medida do seu conhecimento, os Direitos autorais.	Sim	Não	Não	Não
University College	A menos que sejam cobertos por acordos contratuais de terceiros, obrigações legislativas ou disposições	Sim	Sim	Sim	Não

London	<p>relativas à propriedade, os dados de pesquisa da UCL serão fornecidos usando uma licença <i>Creative Commons</i> CC0; [...]</p> <p>UCL suporta o uso de licenças <i>Creative Commons</i> para todos os resultados da pesquisa, incluindo dados de pesquisa.</p>				
University of British Columbia	<p>Eu assino sobre meus direitos reservados quando eu deposito meus trabalhos no cIRcle?</p> <p>Não, o status de direitos autorais do seu trabalho permanece o mesmo.</p> <p>Você meramente concede uma licença (ou seja, sua permissão) para:</p> <p>Disponibilizar livremente o seu trabalho para os usuários do cIRcle's; Preservar o seu trabalho migrando-o para novos formatos [...]</p>	Sim	Sim	Não	Não
University of Maryland Baltimore	<p>Você pode enviar somente materiais para os quais você detém os direitos autorais. Envie-nos um e-mail para obter assistência com questões de direitos autorais e consulte o nosso guia de direitos de autor para obter maiores informações.</p> <p>Se você transferiu direitos para um editor ou outra entidade, talvez seja necessária a permissão para enviar material.</p>	Sim	Não	Não	Não
Purdue University	<p>Você concede à <i>Purdue University</i> uma licença mundial, livre de royalties, não exclusiva, perpétua (durante a vigência do direito autoral aplicável) para exercer os direitos da Obra, no todo ou em parte, a serem depositados em PURR conforme indicado abaixo:</p> <p>a. Para usar, duplicar e distribuir o trabalho [...]</p> <p>Você também representa e garante que tem autoridade e / ou obteve todos os direitos necessários para permitir que a <i>Purdue University</i> use, duplique e distribua o trabalho e que qualquer conteúdo de terceiros seja claramente identificado e reconhecido no trabalho.</p>	Sim	Sim	Não	Não

	<p>Você mantém o direito de publicar e / ou compartilhar seu trabalho que é depositado no site do projeto em PURR. [...]</p> <p>A <i>Purdue University</i> se reserva o direito de liberar o trabalho sob diferentes termos de licença ou de parar de distribuir o trabalho a qualquer momento. Apesar desta falta de um padrão formal neste ponto, as citações de dados devem incluir os seguintes elementos [...]</p>				
California Institute of Technology Caltech	Menciona que deve ser verificada a questão legal como propriedade intelectual	Sim	Sim	Não	Não
University of Virginia	<p>Ao clicar nesta licença, você concede à Universidade da Virgínia o direito não exclusivo de reproduzir, exibir e / ou distribuir sua submissão [...]</p> <p>Você reconhece que o contrato de licença de uso de dados padrão do <i>Libra Dataverse</i> para todos os materiais enviados é uma renúncia de dedicação de Domínio Público <i>Creative Commons Zero</i> ("CC0").</p> <p>Os usuários também têm a opção de elaborar um contrato de licença de uso de dados personalizado.</p> <p>Você pode encontrar a citação para o conjunto de dados na parte superior da página do conjunto de dados em uma caixa azul. Além disso, existe um botão "Cite Data" que oferece a opção de fazer o download da citação como <i>EndNote XML</i>, <i>RIS Format</i> ou <i>BibTeX Format</i>.</p>	Sim	Sim	Sim	Não
University of California Irvine	A melhor maneira de doar conjuntos de dados é preencher nosso formulário da web que permitirá que você faça <i>upload</i> de seu arquivo de dados em nosso repositório. [...]	Sim	Não	Não	Não
University of Arizona	Se você quiser promover o compartilhamento e o uso ilimitado de seus dados, você pode torná-los disponíveis sob uma Declaração CC0 para torná-los explícitos. Existem outras licenças <i>Creative Commons</i> com proteção adicional.	Sim	Sim	Sim	Não

University of Edinburgh	<p>Você tem duas opções na fase de licenciamento. A licença CC BY 4.0 é recomendada pelo <i>Edinburgh DataShare</i>,</p> <p>CC-BY 4.0 é escrito especificamente para bases de dados / <i>datasets</i> em vez de expressões criativas</p> <p>Como alternativa, selecione "Sem licença" e você será solicitado a preencher sua própria instrução de direitos..</p>	Sim	Sim	Sim	Não
Washington University Saint Louis	<p>Se você quiser aplicar uma licença para seus próprios dados, você é encorajado a torná-lo tão aberto quanto apropriado para que outros possam usar e construir em seu trabalho. Para promover o compartilhamento e o uso ilimitado de seus dados, você pode usar uma licença CC0 da <i>Creative Commons</i> [...]</p>	Sim	Sim	Sim	Não
Simon Fraser University	<p>Ao fazer upload de conteúdo para o Radar, você deve selecionar uma das seguintes opções de licenciamento:</p> <p><i>Creative Commons Zero</i>  <i>Creative Commons Atribuição 4.0 (CC BY)</i>  <i>Creative Commons Atribuição-Uso Não-comercial 4.0 (CC BY-NC)</i>  <i>Creative Commons Atribuição-NoDerivatives 4.0 (CC BY-ND)</i></p>	Sim	Sim	Sim	Não
Virginia Polytechnic Institute and State University	<p>Ao aceitar esta licença, você não desiste dos direitos autorais de seu trabalho. [...]</p> <p>Você pode optar por aplicar uma de várias licenças <i>Creative Commons</i> ao seu envio.</p> <p>Se você não escolher uma licença, você concede à Virginia Tech o direito não exclusivo de distribuir sua submissão (incluindo os metadados e resumo) em todo o mundo, em qualquer formato ou meio para fins não comerciais, de pesquisa ou educacionais. Você também autoriza a Virginia Tech a permitir que os usuários do repositório distribuam sua apresentação (incluindo os metadados e resumo) em todo o mundo, em qualquer formato ou meio para fins não</p>	Sim	Sim	Sim	Não

	<p>comerciais, de pesquisa ou educacionais ou relacionados. A Virginia Tech identificará claramente seu (s) nome (s) como autor (es) ou proprietário (s) da submissão, e não fará qualquer alteração, além da permitida por esta licença, à sua submissão.</p> <p>Para fins de publicação, o campo "Criador" para seu conjunto de dados deve ser preenchido e deve incluir criadores de todos os itens incluídos no seu conjunto de dados.</p>				
Tufts University	<p>Você pode especificar quais restrições (se houver) que deseja aplicar ao seu conjunto de dados. [...] Selecione uma licença padrão do <i>Creative Commons</i> e aplique-a aos dados de pesquisa: CC BY CC BY-SA CC BY-ND CC BY-NC CC BY-NC-SA CC BY-NC-ND</p>	Sim	Sim	Sim	Não
Ruprecht Karls Universität Heidelberg	<p>heiDATA permite uma variedade de diferentes modelos de licenciamento para a publicação de dados de pesquisa. Recomendamos a escolha da <i>Creative Commons</i> ou licenças <i>Open Data Commons</i>. A variante licença CC-BY garante nosso ponto de vista atualmente a usabilidade livre dos dados de pesquisa publicados em conformidade com o acesso aberto e Open Science.</p>	Sim	Sim	Sim	ODC
University of Copenhagen	<p>Considere como os dados serão licenciados para reutilização. Existem restrições à partilha devido a Direitos de Propriedade Intelectual, falta de acordos de consentimento, acordos de não divulgação, etc.</p>	Sim	Não	Não	Não
University of Amsterdam /	<p>Se você mesmo publicar dados, pode especificar em uma licença como os outros podem usar seu conjunto de dados e o que é proibido. [...]. Licenças <i>Creative Commons</i> CC0: domínio público Open Data Commons</p> <p>Assim como um artigo ou livro, um</p>	Sim	Sim	Sim	Sim



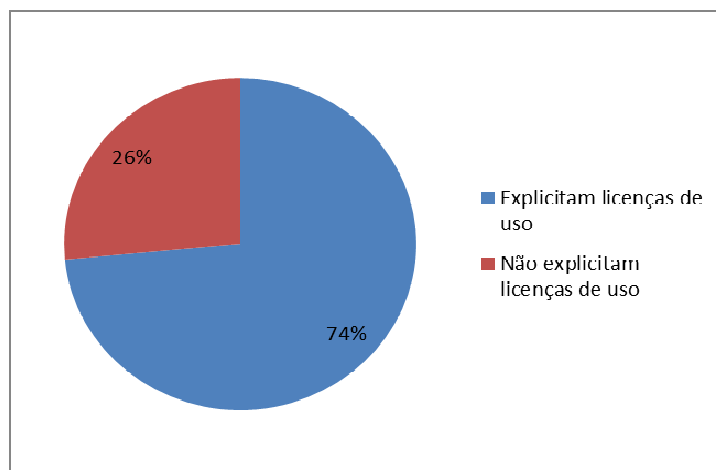
	conjunto de dados pode funcionar como uma fonte de informação para uma publicação acadêmica. O conjunto de dados deve ser explicitamente citado como fonte, também se você mesmo for o criador do conjunto de dados.				
University of California Los Angeles UCLA	A citação de dados é um componente importante da partilha e reutilização de dados. Citar dados dá crédito aos criadores de dados para criar e compartilhar seu trabalho, e cria uma trilha de progresso de pesquisa semelhante à citação de artigos e livros.	Sim	Não	Não	Não

Fonte: Elaborado pela autora baseado nas indicações e recomendações dos PGDs dos Repositórios de dados

Com base na análise do quadro 3, verifica-se os 34 PGDs que explicitam questões relacionadas ao direito autoral. Destaca-se que em 25 PGDs são evidenciadas sobre quais licenças os dados depositados serão cedidos e em nove não são evidenciadas as licenças.

Observa-se no gráfico 5 que em 25 PGDs estão explicitados licenças de uso, correspondendo a 74% e em 9 PGDs não explicitam, correspondendo a 26% .

Gráfico 5 –Explicitação de licenças de uso nos PGDs



Fonte: Elaborado pela autora

As universidades que explicitam direitos autorais em seus PGDs são:

- A *Harvard University* deixa claro em seu PGD quais licenças de uso o pesquisador que vai publicar seus dados poderá adicionar em seus conjuntos de dados. Esse pesquisador não transfere os direitos autorais, mas concede uma permissão para disseminação dos dados. Destaca-se que, o repositório foi implementado com o *software Dataverse* que possui PGD com todas as instruções para os pesquisadores e disponibiliza o padrão de citação de dados do *Dataverse Project*. Esse padrão expõe como citar os dados para deixar mais claro aos usuários a forma de citação e dar os créditos aos autores. Ao criar um conjunto de dados a citação é gerada e apresentada automaticamente. Na página tem exemplos de citação com explicação de cada elemento que a compõe. Nas instruções do *Dataverse Project* declara que, por padrão, todos os conjuntos de dados adicionados ao *Dataverse* recebem a licença CC0 e que a usa por padrão para todos os conjuntos de dados por causa do reconhecimento do nome na comunidade científica. Se o pesquisador que vai depositar preferir, pode não usar a CC0.
- *Massachusetts Institute of Technology* - deixa claro que o compartilhamento dos dados deve estar sob a licença *Open Data Commons* ou uma das licenças *Creative Commons*. O detentor dos direitos autorais deve escolher umas das licenças para seu conjunto de dados definindo o que outros podem ou não fazer com seus dados. É orientado citar os dados utilizados e mostra as formas de citação com exemplos. Tem o intuito de fornecer fácil acesso aos dados nele depositado, fator que influencia na escolha das licenças.
- *Stanford University* - esclarece que o detentor dos direitos autorais pode optar por uma licença mais ampla que permita que outrem faça uso dos dados de forma ilimitada ou uma licença mais restrita limitando o uso de acordo com seus critérios. Explicita que ao depositar os dados no repositório o detentor terá a opção de escolher umas das licenças *Creative Commons* ou *Open Data Commons*. Indica todas as CC, pois incentiva o licenciamento para que pesquisadores *on-line* compreendam como eles podem usar o conteúdo.

- *University of California Berkeley* – indica ao detentor dos direitos autorais para selecionar uma licença para seus dados, indicando a licença *Creative Commons* (CC) e explica para que o uso da licença permita a distribuição gratuita de uma obra protegida por direitos autorais. Destaca a importância de citar os conjuntos de dados.
- *University of Michigan* – adverte aos pesquisadores que depositarão os dados que devem optar por aplicar uma das várias licenças *Creative Commons* aos dados e que os dados serão distribuídos sob a licença escolhida.
- *University of Pennsylvania* – Orienta que antes da submissão deve-se avaliar as permissões de direitos autorais e disponibiliza a política de *copyright* para orientação dos pesquisadores.
- *A University of Oxford* - não determina em seu PGD sobre quais licenças de uso o autor concederá seus conjuntos de dados a outrem, mas deixa claro que a Universidade reivindica a posse de toda a propriedade intelectual especificada no estatuto do seu PGD. O autor deve consultar o estatuto para verificar como os dados serão cedidos.
- *Yale University* – expressa que o conteúdo do repositório está licenciado sob a licença *Creative Commons Atribuição-Não Comercial-Sem Derivados 3.0*. Não direciona para outras opções de licenças.
- *University of Cambridge* – indica todas as possibilidades das licenças *Creative Commons* sob as quais o pesquisador que vai depositar licenciará seus dados, entretanto deixa explícito que a licença recomendada pelo repositório é a *GPLv3*. Conjuntos de dados arquivados no repositório precisam de *software* específico para sua visualização. Na descrição de cada conjunto de dados é indicado qual *software* é recomendado para visualização dos mesmos. Cada arquivo tem a informação sobre qual é a licença atribuída e informações de como cita-lo.

- *University of Wisconsin Madison* – recomenda que os dados que foram financiados pelo poder público sejam licenciados sob a licença *Creative Commons Zero* – CCZero ficando no domínio público. Destaca que é essencial que se tenha uma declaração clara e explícita do intuito e expectativas dos editores no que se refere à reutilização e re-definição dos dados. A declaração deve ser precisa, irrevogável e baseada em uma declaração legal apropriada e reconhecida na forma de uma renúncia ou licença. Indica a importância da citação dos dados utilizados, porém não dá exemplos de como citar.
- *Michigan State University* – designa que libere os dados sem restrições de acesso usando a licença CCZero. Dá exemplo de como citar o conjunto de dados.
- *A University of Texas Austin* - não menciona sobre a cessão dos conjuntos de dados através de licenças, porém deixa claro que o publicador manterá todos os seus direitos de propriedade intelectual. Os direitos limitados que concederá ao Repositório não serão exclusivos.
- *University of California San Diego* – Deixa a critério do pesquisador que vai depositar a licença que utilizará em seus dados. Informa que o repositório suporta vários tipos de licenças.
- *A Pennsylvania State University* - menciona sobre o tema de propriedade intelectual, contudo não deixa explícito sobre a cessão dos conjuntos de dados por meio de licenças de uso.
- *University of Illinois Urbana Champaign* - menciona ser necessário proteger os direitos de propriedade intelectual principalmente dados de valor comercial. As propostas deverão ser escritas no PGD e deixa clara a importância de ter os aspectos relacionados ao direito autoral documentado, pois a expectativa é compartilhar os dados com outros pesquisadores.

- *University of North Carolina Chapel Hill* - exige que o detentor dos direitos autorais autorize a distribuição não exclusiva dos dados ao depositar. Informa que se preferir, podem usar uma das licenças *Creative Commons*.
- *Princeton University* - indica que o autor concede à Universidade de Princeton o direito não-exclusivo para reproduzir, traduzir e/ou distribuir o seu envio, porém não indica sobre quais licenças de uso será a cessão dos dados.
- *University College London* – esclarece que suporta o uso de *licença Creative Commons zero* para os dados, ao menos que sejam cobertos por acordos contratuais de terceiros, obrigações legislativas ou disposições relativas à propriedade. Dá dois exemplos de como citar o conjunto de dados distinguindo entre os estilos de citação da *Harvard* e *Vancouver*.
- *University of British Columbia* – informa que será cedida uma licença para o repositório sem mencionar qual. Destaca-se que, o repositório foi implementado com o *software Dataverse*, assim como a *Harvard University* e *University of Virginia* e possui PGD com todas as instruções para os pesquisadores e disponibiliza o padrão de citação de dados do *Dataverse Project*.
- *University of Maryland Baltimore* – determina que somente os pesquisadores que detém o direito autoral podem submeter os materiais. Não menciona o uso de licenças. O repositório menciona que se o autor tiver dúvidas relacionadas ao direito autoral pode enviar um e-mail.
- *Purdue University* – indica que é concedido à *Purdue University* uma licença mundial, livre de *royalties*, não exclusiva, perpétua (durante a vigência do direito autoral aplicável) sendo reservado à *Purdue* o direito de liberar o trabalho sob diferentes termos de licença. Menciona os elementos que devem compor a citação e dá exemplos evidenciando a importância do reconhecimento do autor.

- *California Institute of Technology Caltech* - menciona que devem ser verificadas as questões legais como propriedade intelectual. Não indica no PGD sobre quais licenças os autores cederão seus dados.
- *University of Virginia* – aponta que os dados enviados estão sob a licença *Creative Commons Zero - CC0*. Deixa a opção de elaborar um contrato de licença de uso de dados personalizado. O pesquisador que depositará poderá escolher como licenciará os dados. Mesmo que o repositório tenha sido implementado com o *software Dataverse*, assim como a *Harvard University* e *University of British Columbia* e possui PGD com todas as instruções para os pesquisadores e disponibiliza o padrão de citação de dados do *Dataverse Project*, indica que a citação para o conjunto de dados está localizada na parte superior da página do conjunto de dados em uma caixa azul e que existe um botão "Cite Data" que oferece a opção de fazer o *download* da citação como *EndNote XML*, *RIS Format* ou *BibTeX Format*.
- *University of California Irvine* – aponta que os dados serão doados e que a melhor forma de fazê-lo é preenchendo o formulário disponível pelo repositório.
- *University of Arizona* – recomenda que seja utilizada a licença *Creative Commons Zero - CC0*. Entretanto, o detentor dos direitos autorais pode utilizar as outras licenças *Creative Commons*. No PGD tem modelos de citação de conjuntos de dados.
- *University of Edinburgh* – destaca que se tem duas opções na fase de licenciamento. A licença *CC BY 4.0* é recomendada pelo *Edinburgh DataShare*, e que como alternativa, tem a opção "Sem licença" em que será solicitado a preencher a própria instrução de direitos.
- *Washington University Saint Louis* – encoraja usar a *Creative Commons Zero - CC0*. Não menciona outras opções de licenças. Tem modelo de citação dos

dados com as informações e explicação de cada elemento que compõe a referência.

- *Simon Fraser University* – deixa claro que ao fazer *upload* de conteúdo para o repositório o autor deve selecionar uma das licenças indicadas: CC0, CC BY, CC BY-NC, CC BY-ND.
- *Virginia Polytechnic Institute and State University* – indica o uso das licenças *Creative Commons*. Se o detentor do direito autoral não escolher uma licença, destaca que a universidade tem o direito não exclusivo de distribuir os dados em todo o mundo, em qualquer formato ou meio para fins não comerciais, de pesquisa ou educacionais, além de os usuários poderem fazer o mesmo. Explicita que identificará claramente o(s) nome (s) como autor (es) ou proprietário (s) dos dados.
- *Tufts University* – indica o uso de uma das Licenças *Creative Commons*.
- *Ruprecht Karls Universität Heidelberg* – Esclarece que permite uma variedade de diferentes modelos de licenciamento para a publicação de dados de pesquisa, e recomenda o uso da *Creative Commons* CC-BY ou licenças *Open Data Commons*. Destaca que a CC-BY é indicada porque garante o ponto de vista do repositório que é a usabilidade livre dos dados de pesquisa.
- *University of Copenhagen* - aponta que os dados serão licenciados, mas não discute nem aponta sob quais licenças serão a cessão. Implementou o repositório com o *software Dataverse*, assim como *University of Virginia*, *Harvard University* e *University of British Columbia*, porém não menciona sobre quais licenças os dados serão licenciados e não menciona para utilizar as orientações do *Dataverse* que possui políticas claras de direitos autorais.
- *University of Amsterdam* – menciona o uso das licenças *Creative Commons* e da *Open Data Commons* e destaca o uso da CC0. Indica que o conjunto de

dados deve ser explicitamente citado, pois pode ser uma fonte de informação para a publicação acadêmica.

- *Universiteit Utrecht* – menciona o uso das Licenças *Creative Commons*. Disponibiliza um guia para o detentor dos direitos autorais determinar qual licença irá utilizar.
- *University of California Los Angeles UCLA* - destaca que a citação de dados é um componente importante da partilha e reutilização de dados além de criar uma trilha de progresso de pesquisa semelhante à citação de artigos e livros. Não menciona sobre quais licenças o pesquisador deve licenciar seus dados.

As universidades destacadas com explicitação de licenças de uso estão elencadas no quadro 4 para melhor visualização. As universidades que não estão tidas explicitam a questão do direito autoral, porém não deixam claro sobre quais licenças de uso os dados serão licenciados.

Quadro 4 – Relação universidade X licenças de uso

UNIVERSIDADE	CC BY	CC BY-AS	CC BY-ND	CC BY-NC	CC BY-NC-SA	CC BY-NC-ND	CC 0	ODC	GPL v3
Harvard University	X	X	X	X	X	X	X		
Massachusetts Institute of Technology	X	X	X	X	X	X	X	X	
Stanford University	X	X	X	X	X	X	X	X	
University of California Berkeley	X	X	X	X	X	X	X		
University of Michigan	X	X	X	X	X	X	X		
University of Pennsylvania									
University of Oxford									
Yale University					X				



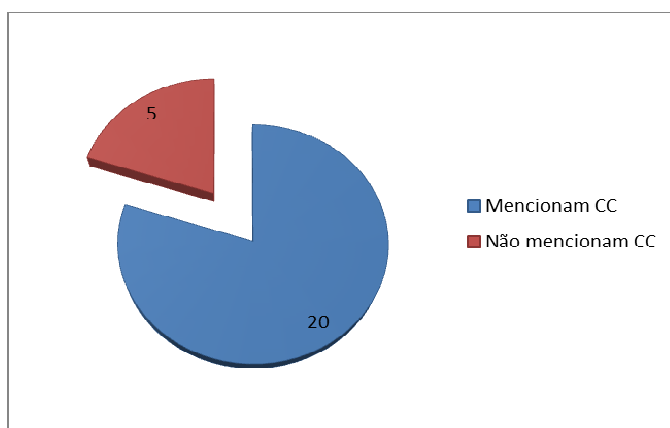
University of Cambridge	X	X	X	X	X	X	X		X
University of Wisconsin Madison							X	X	
Michigan State University							X		
University of Texas Austin									
University of California San Diego									
Pennsylvania State University									
University of Illinois Urbana Champaign									
University of North Carolina Chapel Hill	X	X	X	X	X	X	X		
Princeton University									
University College London	X	X	X	X	X	X	X		
University of British Columbia									
University of Maryland Baltimore									
Purdue University									
California Institute of Technology Caltech									
University of Virginia	X	X	X	X	X	X	X		
University of California Irvine									
University of Arizona	X	X	X	X	X	X	X		
University of Edinburgh	X								
Washington University Saint Louis	X	X	X	X	X	X	X		
Simon Fraser University	X		X	X			X		
Virginia Polytechnic Institute and State University	X	X	X	X	X	X	X		
Tufts University	X	X	X	X	X	X	X		
Ruprecht Karls Universität	X							X	

Heidelberg									
University of Copenhagen									
University of Amsterdam	X	X	X	X	X	X	X	X	
University of California Los Angeles UCLA									

Fonte: elaborado pela autora

Observa-se no gráfico 6 que dos 25 PGDs, em 20 são indicados aos pesquisadores que vão depositar as licenças *Creative Commons* para licenciarem seus dados, em cinco não são indicadas as licenças *Creative Commons*.

Gráfico 6 – PGDs que mencionam a licença *Creative Commons*



Fonte: Elaborado pela autora

Em seis PGDs indicam aos seus pesquisadores outras licenças de uso sendo cinco a *Open Data Commons* e uma a *GPLv3*.

As universidades que possuem PGDs que indicam outras licenças são:

- Massachusetts Institute of Technology – ODC
- *Stanford University* – ODC
- *Ruprecht Karls Universität Heidelberg* – ODC
- *University of Amsterdam* – ODC
- *University Wisconsin Madison* - ODC
- *University of Cambridge* – GPLv3

Observa-se que a maioria dos repositórios indicam as licenças *Creative Commons*. *Stanford University* e *University of Amsterdam* indicam todas as licenças da CC e a ODC. *University of Cambridge* indicam a as CC e ODC além a GPLv3 pois foi a única que deixou claro uso dessa última pela necessidade de manipulação dos dados.

Os repositórios de dados documentam as instruções e normativas nos PGDs no qual mencionam, também, instruções sobre direito autoral dos dados depositados orientando os pesquisadores sobre como e quais licenças poderão disponibilizar seus conjuntos de dados para reutilização por outros pesquisadores e sobre o *software* necessário para visualização ou manipulação dos dados. A atribuição pode ser ampla na qual permite ao usuário fazer maior uso de seus dados sem restrições ou uma licença mais restrita que limita o uso. A escolha de licenciamento depende do tipo de dados disponibilizados. A preocupação de mostrar que os autores serão citados, com demonstração de formas padronizadas de referenciar os dados também está presente.

As instruções inclusas no PGD variam dependendo das características dos repositórios e dos conjuntos de dados neles depositados. As agências de financiamento podem ter requisitos específicos de licenciamento dos dados das pesquisas financiadas por elas que existem princípios de gerenciamento que se aplicam às diversas áreas do conhecimento. O PGD do repositório auxilia o pesquisador que vai depositar os dados atender aos requisitos que se aplicam aos seus conjuntos de dados.

A missão, os objetivos e áreas de conhecimento que compõem os repositórios são indícios de escolhas sobre a liberação dos conjuntos de dados e a indicação de quais licenças os dados serão liberados para reutilização. Repositórios multidisciplinares tendem a deixar mais opções para os autores liberarem seus dados devido ao fato da diversidade e características de área e de cada conjunto de dados. Os pesquisadores distinguem as particularidades dos dados e terão cautela e propriedade no momento da atribuição da licença. Repositórios temáticos variam na indicação da licença dependendo das características da área e dos tipos de dados disponibilizados. A comunidade que o repositório contempla tem influência na elaboração das diretrizes do PGD.

A precaução dos repositórios relacionada ao direito autoral é evidente na maioria deles. As peculiaridades de cada repositório são evidenciadas nos PGDs sendo uma forma de assessorar os autores na liberação de seus dados ponderando os diversos aspectos envolvidos e assegurando questões éticas e jurídicas.

As diretrizes do PGDs precisam ser claras aos autores para não gerar dúvidas. Na análise das orientações sobre questões de direitos autorais observou-se que cada repositório elaborou seu PGD de acordo com as necessidades e particularidades de sua comunidade e do tipo de conteúdo abordado nos conjuntos de dados. Algumas diretrizes são mais flexíveis para que os pesquisadores aplicassem determinada licença sobre seus dados assegurando suas particularidades, pois diferentes tipos de conteúdo exigem diferentes tipos de licenciamento.

## 6 CONSIDERAÇÕES FINAIS

A evolução nos métodos de desenvolvimento da ciência, nas tecnologias, incluindo a Tecnologia de Informação e Comunicação (TIC) e no processo de comunicação científica está impactando o âmbito científico, chamando a atenção dos pesquisadores para o desenvolvimento e a manutenção de repositório de dados científicos, auxiliando no avanço na ciência. Profissionais da informática, bibliotecários, arquivistas, gestores, pesquisadores e professores trabalham no desenvolvimento e gestão de repositórios de dados.

A colaboração multidisciplinar e internacional, a integração e compartilhamento de dados e novos aparatos tecnológicos utilizados nos diversos setores estão presentes nos campos da ciência. Esses fatores se relacionam de maneira indissociável para a redefinição de práticas científicas e da evolução dos aspectos relacionados à utilização de dados para a construção de conhecimento no indivíduo e na sociedade favorecendo a participação colaborativa entre pesquisadores.

Esses episódios são observados na evolução da relação entre dados, informação e conhecimento. Esses elementos formam um ciclo em que dados científicos têm um papel de retroalimentação no desenvolvimento das várias áreas da ciência. Nesse ciclo, os cientistas têm a possibilidade de compartilhamento e reutilização de dados coletados por satélites e/ou tratados em laboratórios, centro de pesquisa, entre outros em diversos países cooperando com a articulação na colaboração científica internacional. A soma dos dados científicos pode gerar um maior ou mais amplo resultado nas pesquisas.

Os dados científicos podem ser gerados por diversos meios como a observação, computação, experiência, simulação, capturados por satélites ou em campo, entre outros. Esses conjuntos de dados fornecem evidências que são usadas para justificar as conclusões que são apresentadas nas publicações científicas, propor novas pesquisas ou refutar as apresentadas.

A estrutura e o fluxo da comunicação e publicação científica permeados pelo aprimoramento das Tecnologias de Informação e Comunicação, computação e redes eletrônicas e interoperabilidade dos sistemas de informação estão propícios a novas

práxis que os amplia, diversifica e os torna mais rápidos e abrangentes, viabilizando novos parâmetros no âmbito acadêmico/científico.

Com o desenvolvimento de repositório de dados científicos, a estrutura e o fluxo da comunicação e publicação científica disponíveis vêm se diversificando, proporcionando a criação de uma nova realidade com a publicação de dados científicos. Os repositórios de dados colaboram para o gerenciamento, disponibilização e curadoria dos dados.

Os repositórios contribuem para expandir o impacto e a visibilidade das pesquisas, dos pesquisadores e das instituições das quais pertencem criando oportunidades adicionais de colaboração e conexão dentro e além de seu domínio.

As possibilidades são muitas e repositórios de dados estão instituídos nas universidades, com grande número na Europa e Estados Unidos. As implementações podem atingir seu potencial pleno à medida que os profissionais envolvidos estiverem engajados nesses aspectos trabalhando para estabelecer um sistema em que as descobertas científicas e os dados coletados pelos pesquisadores ou pelas instituições possam ser disponibilizados, compartilhados e reusados de maneira mais rápida.

Nesse contexto, implementar uma infraestrutura para apoiar os pesquisadores no ciclo de vida dos dados científicos, torna-se essencial. Os repositórios de dados científicos auxiliam no gerenciamento de dados e é fundamental que seja planejado e estruturado desde o início de sua implementação.

Plano de Gerenciamento de Dados é um documento que contribui para o desenvolvimento e gerenciamento dos dados nos repositórios com instruções aos pesquisadores no gerenciamento dos dados desde a coleta e aos profissionais que neles trabalham orientando no gerenciamento, descrição, armazenamento, qualidade, integração, disseminação e preservação dos conjuntos de dados.

A instrução sobre direitos autorais para os pesquisadores que arquivam seus dados nos repositórios fornece segurança e apoio para que todos os agentes envolvidos tenham crédito pelo seu trabalho e as licenças de uso orientam aqueles que reutilizarão os dados evitando o uso indevido e problemas jurídicos posteriores.

Esses agentes podem ser os pesquisadores, agências de fomento, financiadores do projeto, instituições, laboratórios, centros de pesquisa e outros envolvidos no processo de trabalho.

O principal trunfo como pesquisador é o seu trabalho intelectual. É primordial compreender como os repositórios resguardarão seus direitos sobre seus conjuntos de dados depositados nos repositórios.

No direcionamento das análises foi verificado que dentre as 100 melhores universidades do mundo, em 45 delas não foram localizados repositórios de dados. Esse fator destaca que, mesmo com várias plataformas para implementação de repositório de dados disponíveis e a importância da gestão de dados, ainda há um longo caminho na direção de planejamento e implementação de repositório por parte das universidades. Entre essas 45 universidades foram destacadas que em quatro delas suas bibliotecas intermediam o recebimento e encaminhamento dos conjuntos de dados para repositórios de dados. Essa mediação é um caminho para que seja implementado repositórios de dados, pois demonstra a preocupação das universidades no gerenciamento dos dados de sua comunidade. Destaca-se também o papel relevante das bibliotecas no gerenciamento de conjuntos de dados.

A gestão de dados é imprescindível para o bom andamento da pesquisa, porém, dentre as 55 universidades com repositório de dados, em 19 delas não foram localizados PGDs. As universidades implementaram o repositório de dados, no entanto, a gestão de dados ainda não está evidenciada. Esse aspecto deixa o pesquisador sem orientação sobre os caminhos a seguir no arquivamento de seus dados e sobre como o repositório vai gerir seus conjuntos de dados.

A questão de grande relevância dentro do PGD está relacionada ao direito autoral, pois orienta tanto o pesquisador que vai depositar seus conjuntos de dados sobre como ele poderá libera-los além do fato de resguardar o pesquisador sobre o reconhecimento de seu trabalho com a menção da citação dos autores, quanto o pesquisador que vai reutilizar os dados que será orientado no tocante de como cada conjunto de dados pode ser reutilizado seguindo as restrições de uso, quando propício. Esse fator não foi considerado em dois dos 36 PGDs localizados.

Os resultados das análises demonstram que das 100 melhores universidades do mundo, apenas 34 delas explicitam a questão dos direitos autorais. Quando se considera que foram analisadas as melhores universidades do mundo, esse fator comprova que ainda se tem um longo caminho para a conscientização por parte dos envolvidos na gestão dos dados, da importância do direito autoral para evitar

problemas éticos e jurídicos no futuro e resguardar a autoria dos conjuntos de dados dos pesquisadores.

As vertentes apresentadas corroboram que a área da Ciência da Informação dispõe de arcabouço teórico e prático para contribuir substancialmente com a implementação de repositórios de dados, na gestão de conjunto de dados e direito autoral.



## REFERÊNCIAS

AFFONSO, E. P.; SANT'ANA, R. C. G. Anonimização de metadados de imagem digital por meio do modelo k-anonimato. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 16., João Pessoa. **Anais eletrônicos...** João Pessoa: ANCIB, 2015. Disponível em: <<http://www.ufpb.br/evento/lti/ocs/index.php/enancib2015/enancib2015/paper/viewFile/2972/1182>>. Acesso em 25 ago. 2016.

ANDRADE, M. M. **Introdução à metodologia científica**. 10. ed. São Paulo: Atlas, 2010.

ASSOCIAÇÃO PAULISTA DA PROPRIEDADE INTELECTUAL. **Propriedade intelectual**. c2016. Disponível em: <<http://www.aspi.org.br/pt-br/propriedadeintelectual.aspx>>. Acesso em: 20 set. 2016.

BARROS, M. H. T. C. **Disseminação da informação**. Marília: [s.n.], 2003.

BASTOS, F. M.; VIDOTTI, S. ODONNE, N. O direito de cópia da produção científica para os repositórios institucionais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 11., 2010. Rio de Janeiro. **Anais eletrônicos...** Rio de Janeiro: Ibict, 2010. Disponível em: <<http://enancib.ibict.br/index.php/enancib/xienancib/paper/viewFile/3502/2627>>. Acesso em: 29 set. 2016.

BRASIL. **Decreto nº 75.699, de 6 de Maio de 1975**. Promulga a convenção de Berna para a Proteção das Obras Literárias e Artísticas, de 9 de setembro de 1886, revista em Paris, a 24 de julho de 1971. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/decreto/1970-1979/D75699.htm](http://www.planalto.gov.br/ccivil_03/decreto/1970-1979/D75699.htm)>. Acesso em 14 set. 2016.

BRASIL. Ministério da Cultura. **Direitos autorais**. Brasília, DF, 2007. Disponível em: <<http://cultura.gov.br/documents/18021/130362/direito-autoral-27-11-2007.pdf/d3967b82-3f51-4fe5-9021-4825af66f668>>. Acesso em: 20 set. 2016.

BORGMAM, C. L. et al. Knowledge infrastructures in science: data, diversity, and digital libraries. **International Journal on Digital Libraries**, New York, v. 16, n. 3, p. 207-227, 2015. Disponível em: <<http://link.springer.com/article/10.1007%2Fs00799-015-0157-z#page-1>>. Acesso em: 6 nov. 2015.

BURKE, P. A comercialização do conhecimento: o mercado e a impressão gráfica. In: \_\_\_\_\_. **Uma história social do conhecimento**: de Gutenberg a diderot. Rio de Janeiro: Jorge Zahar, 2003. p. 136-158.

CAFÉ, L. et al. Repositórios institucionais digitais: nova estratégia para publicação científica na Rede. In: Encontro Nacional de Informação em Ciências da Comunicação, 12, 2003, Belo Horizonte. **Anais...** Belo Horizonte: [PUC MG], 2003. p. 1-12.

CAPURRO, R.; HJORLAND, B. O conceito de informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 12, n. 1, p. 148-207, jan./abr. 2007.

CESAR JUNIOR, R. M. Apresentação à edição brasileira. In: HEY, T.; STEWARD, T.; TOLLE, K. (Org.). **O quarto paradigma: descobertas científicas na era da eScience**. Tradução Leda Beck. São Paulo: Oficina de textos, 2011. p. 7-8.

COPYRIGHT HOUSE. **Lei de direitos autorais**. c2009-2010a. Disponível em: <<http://br.copyright-house.com/copyright/lei-de-direitos-autorais.htm>>. Acesso em 14 set. 2016.

COX, A. M. PINFIELD, S. Research data management and libraries: current activities and future priorities. **Journal of Librarianship and Information Science**, London, v. 46, n. 4, p. 299-316, 2014. Disponível em: <<http://lis.sagepub.com/content/46/4/299.full.pdf+html>>. Acesso em: 27 set. 2016.

CREATIVE COMMONS (CC). **Frequently asked questions**. California, [2016]. disponível em: <<https://creativecommons.org/faq/#what-is-creative-commons-and-what-do-you-do>>. Acesso em: 20 set. 2016.

CREATIVE COMMONS (CC). **Sobre as licenças**. California, [20--]. Disponível em: <<https://creativecommons.org/licenses/>>. Acesso em: 20 set. 2016.

DATA Management Plan 4: data access and security. Utrecht: Utrecht University, 2016a. Disponível em: <<http://libguides.library.uu.nl/c.php?g=202272&p=1330220&preview=1776d74a25e3418c4d63aafdbf99598f#s-lg-box-wrapper-4739017>>. Acesso em: 18 ago. 2016.

DATA Management Plan 3: data storage and backup. Utrecht: Utrecht University, 2016b. Disponível em: <<http://libguides.library.uu.nl/c.php?g=202272&p=1330140&preview=1776d74a25e3418c4d63aafdbf99598f>>. Acesso em: 18 ago. 2016.

DATA Management Plan 6: data preservation and archiving. Utrecht: Utrecht University, 2016c. Disponível em: <<http://libguides.library.uu.nl/c.php?g=202272&p=1330269&preview=1776d74a25e3418c4d63aafdbf99598f>>. Acesso em: 18 ago. 2016.

DATA Management Plan 5: data sharing and reuse. Utrecht: Utrecht University, 2016d. Disponível em: <<http://libguides.library.uu.nl/c.php?g=202272&p=1330225&preview=1776d74a25e3418c4d63aafdbf99598f>>. Acesso em: 18 ago. 2016.

DECLARAÇÃO UNIVERSAL DOS DIREITOS HUMANOS. **A Declaração Universal dos Direitos Humanos**. c2016. Disponível em: <<http://www.dudh.org.br/declaracao/>>. Acesso em: 22 set. 2016.

DUDZIAK, E. **Dados de Pesquisa agora devem ser armazenados e citados**. 2016. Disponível em: <<http://www.sibi.usp.br/?p=6189>>. Acesso em: 10 ago. 2016.

ELLIOT, R. Who owns scientific data? The impact of intellectual property rights on the scientific publication chain. **Learned Publishing**, v. 18, n. 2, p. 91-94, Apr. 2005. Disponível em: <[file:///C:/Users/bete.MARILIA/Downloads/Elliott-2005-Learned\\_Publishing.pdf](file:///C:/Users/bete.MARILIA/Downloads/Elliott-2005-Learned_Publishing.pdf)>. Acesso em: 05 jan. 2017.

FOX, P.; HENDLER, J. eScience semântica: o significado codificado na próxima geração de ciência digitalmente apropriada. In: HEY, T.; STEWARD, T.; TOLLE, K. (Org.). **O quarto paradigma: descobertas científicas na era da eScience**. Tradução Leda Beck. São Paulo: Oficina de textos, 2011. p. 159-163.

GANDELMAN, H. **De Gutenberg à internet: direitos autorais das origens à era digital**. 5. ed. Rev. e atual. Rio de Janeiro: Record, 2007.

GARCIA, M. L. A. Políticas e programas nacionais de informação científica e tecnológica. **Ciência da Informação**, Brasília, v. 9, n. 1/2, p. 5-39, 1980. Disponível em: <<http://ridi.ibict.br/bitstream/123456789/374/1/MARIALUCIACi1980.pdf>>. Acesso em: 26 set. 2016.

GNU OPERATING SYSTEM. Licenças. 2016. Disponível em: <<http://www.gnu.org/licenses/gpl-faq.html#GPLTranslations>>. Acesso em: 21 set. 2016.

GUIA prático de inovação: propriedade intelectual e transferência de tecnologia. São Paulo: Agência Unesp de inovação, [20--].

GONZÁLEZ DE GÓMEZ, M. N. Novos cenários políticos para a informação. **Ciência da informação**, Brasília, v. 31, n. 1, p. 27-40, 2002. Disponível em: <<http://revista.ibict.br/ciinf/article/view/975/1013>>. Acesso em: 26 set. 2016.

GRAY, J. **eScience: a transformed scientific method**. 2007. Disponível em: <[http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB\\_eScience.ppt](http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB_eScience.ppt)>. Acesso em: 18 ago. 2015.

GRAY, J. Jim Gray on escience: a transformed scientific method. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). **The fourth paradigm: data-intensive scientific discovery**. Washington: Microsoft Research, 2009. Disponível em: <<http://languagelog ldc.upenn.edu/myl/JimGrayOnE-Science.pdf>>. Acesso em: 05 maio 2015.

GRAY, J. et al. Scientific data management in the coming decade. **SIGMOD Record**, v. 34, n. 4, p. 34-41, 2005. Disponível em: <<https://www.microsoft.com/en-us/research/publication/scientific-data-management-in-the-coming-decade/>>. Acesso em 22 set. 2016.

JORENTE, M. J. V.; SANTOS, P. L. V. A. C. Mídias de pós-vanguarda, direito de autor, cultura livre e produtos de criação contemporânea. In: GUIMARÃES, J. A. C.; MOLINA, J. C. F. (Org.). Aspectos jurídicos e éticos da informação digital. São Paulo: Cultura acadêmica, 2008, p. 113-132.

KING, T.; FORCE, M. **Data Citation Index**. C2017. Disponível em: <<http://slideplayer.com/slide/5272632/>>. Acesso em 05 jan. 2017.

KUHN, T. **A estrutura das revoluções científicas**. São Paulo: Perspectiva, 2006.

LARA, M. L. G.; CONTI, V. L. Disseminação da informação e usuários. **São Paulo em Perspectiva**, São Paulo, v. 17, n. 3/4, p. 26-34, 2003. Disponível em: <<http://www.scielo.br/pdf/spp/v17n3-4/a04v1734.pdf>>. Acesso em: 20 set. 2016.

LE COADIC, Y. F. **A ciência da informação**. 2. ed. Brasília, DF: Brinquet de Lemos livros, 2004.

LEE, J. W. DataNet: an emerging cyberinfrastructure for sharing, reusing and preserving digital data for scientific discovery and learning. **AIChE Journal**, New York, v. 55, n. 11, p. 2757-2764, Nov. 2009. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/aic.12085/epdf>>. Acesso em: 05 jan. 2017.

LEMOS, R. Prefácio à edição brasileira. In: LESSIG, L. **Cultura livre**: como a grande mídia usa a tecnologia e a lei para bloquear a cultura e controlar a criatividade. São Paulo, Trama, 2005. p. 15-24.

LYNCH, C. A. Institutional Repositories: essential infrastructure for scholarship in the digital age. **ARL Bimonthly Report** 26. 2003. Disponível em: <<http://www.arl.org/newsltr/226/ir.html>>. Acesso em: 25 out. 2004.

LYNCH, C. O quarto paradigma de Jim Gray e a construção do registro científico. In: HEY, T.; STEWARD, T.; TOLLE, K. (Org.). **O quarto paradigma: descobertas científicas na era da eScience**. Tradução Leda Beck. São Paulo: Oficina de textos, 2011. p. 187-193.

LOPATENKO, A. **Information retrieval in current research information systems**. (arXiv preprint cs/0110026). 2001. Disponível em: <<http://arxiv.org/ftp/cs/papers/0110/0110026.pdf>>. Acesso em: 22 set. 2016.

MANTOVANI, O.; DIAS, M. H. P.; LIESENBERG, H. Conteúdos abertos e compartilhados: novas perspectivas para a educação. **Educação & Sociedade**, Campinas, v. 27, n. 94, p. 257-276, jan./abr. 2006.

MARTINS, W. Direitos autorais. In: \_\_\_\_\_. **A palavra escrita**: história do livro, da imprensa e da biblioteca. Com um capítulo referente a propriedade literária. 3. ed. São Paulo: Ática, 2001. p. 391-413.

MAXIMIANO, A. C. A. Administração estratégica. In: \_\_\_\_\_. **Teoria geral da administração**: da escola científica à competitividade e economia globalizada. São Paulo: Atlas, 2000. p. 391-426.

MAYER- SCHONBERGER, V.; CUKIER, K. **Big data**: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana. Rio de Janeiro: Elsevier, 2013.

MEADOWS, A. J. **A comunicação científica**. Brasília, DF: Brinquet de Lemos livros, 1999.

MEDEIROS, J. C.; CARAGNATO, S. E. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. Rio de Janeiro, **Liinc em Revista**, Rio de Janeiro, v. 8, n. 2, p. 311-322, set. 2012. Disponível em: <<http://www.ibict.br/liinc>>. Acesso em: 8 abr. 2015.

MONTEIRO, E. C. S. A.; SANT'ANA, R. C. G. Repositório de dados científicos nas universidades brasileiras e portuguesas. In: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO, 6.; 2016, Londrina. **Anais eletrônicos...** Londrina: UEL, 2016. p. 652-664. Disponível em: <<http://www.uel.br/eventos/cinf/index.php/secin2016/secin2016/paper/viewFile/338/166>>. Acesso em: 18 ago. 2016.

MONTEIRO, E. C. S. A.; SANT'ANA, R. C. G. SANTARÉM SEGUNDO, J. E. e-SCIENCE SEMÂNTICA: integração dos dados na comunicação científica. **Informação em Pauta**, Fortaleza, v. 1, n. 1, p. 9-29, 2016. Disponível em: <<http://www.periodicos.ufc.br/index.php/informacaoempauta/article/view/2942/2688>>. Acesso em: 18 ago. 2016.

MORIN, E. **Introdução ao pensamento complexo**. 3. ed. Porto Alegre: Sulina, 2007.

MUELLER, S. P. M. A ciência, o sistema de comunicação científica e a literatura científica. In: CAMPELLO, B. C.; CENDÓN, B. V.; KRÉMER, J.M. (Org) **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte : UFMG, 2000. p. 21-34.

NAÇÕES UNIDAS. **Declaração Universal dos Direitos Humanos**. Rio de Janeiro: UNIC, 2009. Disponível em: <<http://www.dudh.org.br/wp-content/uploads/2014/12/dudh.pdf>>. Acesso em: 22 set. 2016.

NATIONAL SCIENCE BOARD. **Long-lived digital data collections**: enabling research and education in the 21st century. 2005. Disponível em: <<http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>>. Acesso em: 30 maio 2016.

OPEN SOURCE INITIATIVE. **The GNU Lesser General Public License, version 3.0 (LGPL-3.0)**. [20--a]. Disponível em: <<http://opensource.org/licenses/LGPL-3.0>>. Acesso em: 21 set 2016.

OPEN SOURCE INITIATIVE. **Artistic License (Perl) 1.0**. [20--b]. Disponível em: <<http://opensource.org/licenses/Artistic-Perl-1.0>>. Acesso em: 21 set. 2016.

OPEN SOURCE INITIATIVE. **Mozilla Public License 2.0 (MPL-2.0)**. [20--c]. Disponível em: <<https://opensource.org/licenses/MPL-2.0>>. Acesso em: 21 set. 2016.

OPEN SOURCE INITIATIVE. **Eclipse Public License 1.0 (EPL-1.0)**. [20--d]. Disponível em: <<https://opensource.org/licenses/EPL-1.0>>. Acesso em: 21 set. 2016.

OPEN SOURCE INITIATIVE. **Common Development and Distribution License (CDDL-1.0)**. [20--e]. Disponível em: <<https://opensource.org/licenses/CDDL-1.0>>. Acesso em: 21 set. 2016.

OPEN SOURCE INITIATIVE. **The MIT License (MIT)**. [20--f]. Disponível em: <<https://opensource.org/licenses/mit-license.php>>. Acesso em: 21 set. 2016.

OPEN SOURCE INITIATIVE. **GNU AFFERO GENERAL PUBLIC LICENSE, Version 3 (AGPL-3.0)**. c2007. Disponível em: <<https://opensource.org/licenses/AGPL-3.0>>. Acesso em: 21 set. 2016.

OPEN SOURCE INITIATIVE. **The BSD 3-Clause License**. [20--g]. Disponível em: <<http://opensource.org/licenses/BSD-3-Clause>>. Acesso em: 21 set. 2016.

OPEN SOURCE INITIATIVE, [20--g]. **The BSD 2-Clause License**. [20--h]. Disponível em: <<http://opensource.org/licenses/BSD-2-Clause>>. Acesso em: 21 set. 2016.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OCDE). **OECD Principles and Guidelines for Access to Research Data from Public Funding**. Paris: OCDE, 2007. Disponível em: <<http://www.oecd.org/sti/scitech/38500813.pdf>>. Acesso em: 20 set. 2016.

PINFIEL, S. Journals and repositories: an evolving relationship? **Learned Publishing**, v. 22, n. 3, p. 165-175, Jul. 2009. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1087/2009302/epdf>>. Acesso em: 05 jan. 2017.

PINFIELD, S; COX, A. M.; SMITH, J. Research data management and libraries: relationships, activities, drivers and Influences. **PLoS ONE**, v. 9, n. 12, p. 1-28, 2014. Disponível em: <<http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=1&sid=fbf9aa61-6c97-445e-a489-9d76b4383641%40sessionmgr107&hid=128>>. Acesso em: 27 set. 2016.

PINFIELD, S. SALTER, J.; BATH, P. A. Open-access repositories worldwide, 2005–2012: past growth, current characteristics, and future possibilities. **Journal of the Association for Information Science and Technology**, New York, v. 65, n. 12, p. 2404–2421, 2014. <<http://onlinelibrary.wiley.com/doi/10.1002/asi.23131/epdf>>. Acesso em: 05 jan. 2017.

PINFIELD, S. et al. Open-Access Repositories Worldwide, 2005–2012: Past

Growth, Current Characteristics, and future possibilities. **Journal of the Association for Information Science and Technology**, New York, v. 65, n. 12, p. 2404–2421, 2014. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.23131/epdf>>. Acesso em: 05 jan. 2017.

PLOMAN, E. W.; HAMILTON, L. C. **Copyright**: intellectual property in the information age. London: Routledge & Kegan Paul, 1980.

POLÍTICA. Dicionário Michaelis. c2016. Disponível em: <<http://michaelis.uol.com.br/busca?r=0&f=0&t=0&palavra=pol%C3%ADtica>>. Acesso em: 26 set. 2016.

RANKING WEB OF UNIVERSITIES. **About us**. 2016. Disponível em: <[http://webometrics.info/en/About\\_Us](http://webometrics.info/en/About_Us)>. Acesso em: 25 set. 2016.

RIBES, D.; LEE, C. P. Sociotechnical studies of cyberinfrastructure and e-research: current themes and future trajectories. **Computer Supported Cooperative Work**, Dordrecht, v. 19, n. 3-4, p. 231-244, 2010. Disponível em: <<http://www.davidribes.com/storage/Ribes%20Lee%20-%20Cyberinfrastructure%20Studies.pdf>> Acesso em: 6 maio 2012.

RODRIGUES, E. et al. **RepositóriUM**: criação e desenvolvimento do Repositório da Universidade do Minho. 2004. Disponível em: <<http://www.apbad.pt/pdf8cong/com/com14.pdf>>. Acesso em: 13 set. 2004.

RODRIGUES, E. et al. **Os repositórios de dados científicos**: estado da arte. 2010. Disponível em: <[http://projeto.rcaap.pt/index.php?option=com\\_remository&Itemid=2&func=startdown&id=271&lang=pt](http://projeto.rcaap.pt/index.php?option=com_remository&Itemid=2&func=startdown&id=271&lang=pt)>. Acesso em: 5 jun. 2016.

SALES, L. F. **Integração semântica de publicações científicas e dados de pesquisa**: proposta de modelo de publicação ampliada para a área de ciências nucleares. 2014. 264 f. Tese (Doutorado em Ciência da Informação) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014. Disponível em: <<http://carpedien.ien.gov.br/bitstream/ien/853/1/LUANA%20SALES%20D.pdf>>. Acesso em 19 set. 2016.

SALES, L. F.; SAYÃO, L. F. O impacto da curadoria digital dos dados de pesquisa na comunicação científica. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 17, p.118-135, 2012. Especial 2. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17nesp2p118/23573>>. Acesso em: 21 set. 2016.

SAYÃO, L. F. Interoperabilidade das bibliotecas digitais: o papel dos sistemas de identificadores persistentes - URN, PURL, DOI, Handle System, CrossRef e OpenURL. *TransInformação*, Campinas, v. 19, n. 1, p. 65-82, jan./abr., 2007. Disponível em: <<http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/621/601>>. Acesso em: 14 jul. 2016.

SANT'ANA, R. C. G. A importância do papel do profissional da ciência da informação nos processos de recuperação de conteúdos digitais estruturados. In: GUIMARÃES, J. A. C.; FUJITA, M. S. L. (Org.). **Ensino e pesquisa em biblioteconomia no Brasil: a emergência de um novo olhar**. Marília : Fundepe ; São Paulo : Cultura Acadêmica, 2008. p. 145-154.

SANT'ANA, R. C. G. Ciclo de vida dos dados e o papel da Ciência da Informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 14., Florianópolis. **Anais eletrônicos...** Rio de Janeiro: ANCIB, 2013. Disponível em: <<http://enancib2013.ufsc.br/index.php/enancib2013/XIVenancib/paper/viewFile/284/319>>. Acesso em: 14 jul. 2016.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação e informação**, Londrina, v. 21, n. 2, p. 116-142, maio/ago. 2016. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940/20124> >. Acesso em: 20 out. 2016.

SANTOS, P. L. V. A. C.; SANT'ANA, R. C. G. Transferência da informação: análise para valoração de unidades de conhecimento. **DataGramZero: Revista de Ciência da Informação**, v. 3, n. 2, abr. 2002. Disponível em: <[http://www.dgz.org.br/abr02/Art\\_02.htm](http://www.dgz.org.br/abr02/Art_02.htm)>. Acesso em: 01 maio de 2015.

SARACEVIV, T. A natureza interdisciplinar da ciência da informação. **Ciência da Informação**, Brasília, DF, v. 29, n. 1, p. 1-9, 1995.

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para a preservação de dados digitais de pesquisa. **Informação & sociedade: estudos**, João Pessoa, v. 22, n. 3, p. 179-191, set. dez. 2012. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/12224/8586>>. Acesso em 20 nov. 2015.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários de pesquisadores**. Rio de Janeiro: CNEN, 2015. Disponível em: <[http://carpedien.ien.gov.br:8080/bitstream/ien/1624/1/GUIA\\_DE\\_DADOS\\_DE\\_PESQUISA.pdf](http://carpedien.ien.gov.br:8080/bitstream/ien/1624/1/GUIA_DE_DADOS_DE_PESQUISA.pdf)>. Acesso em: 5 out. 2016.

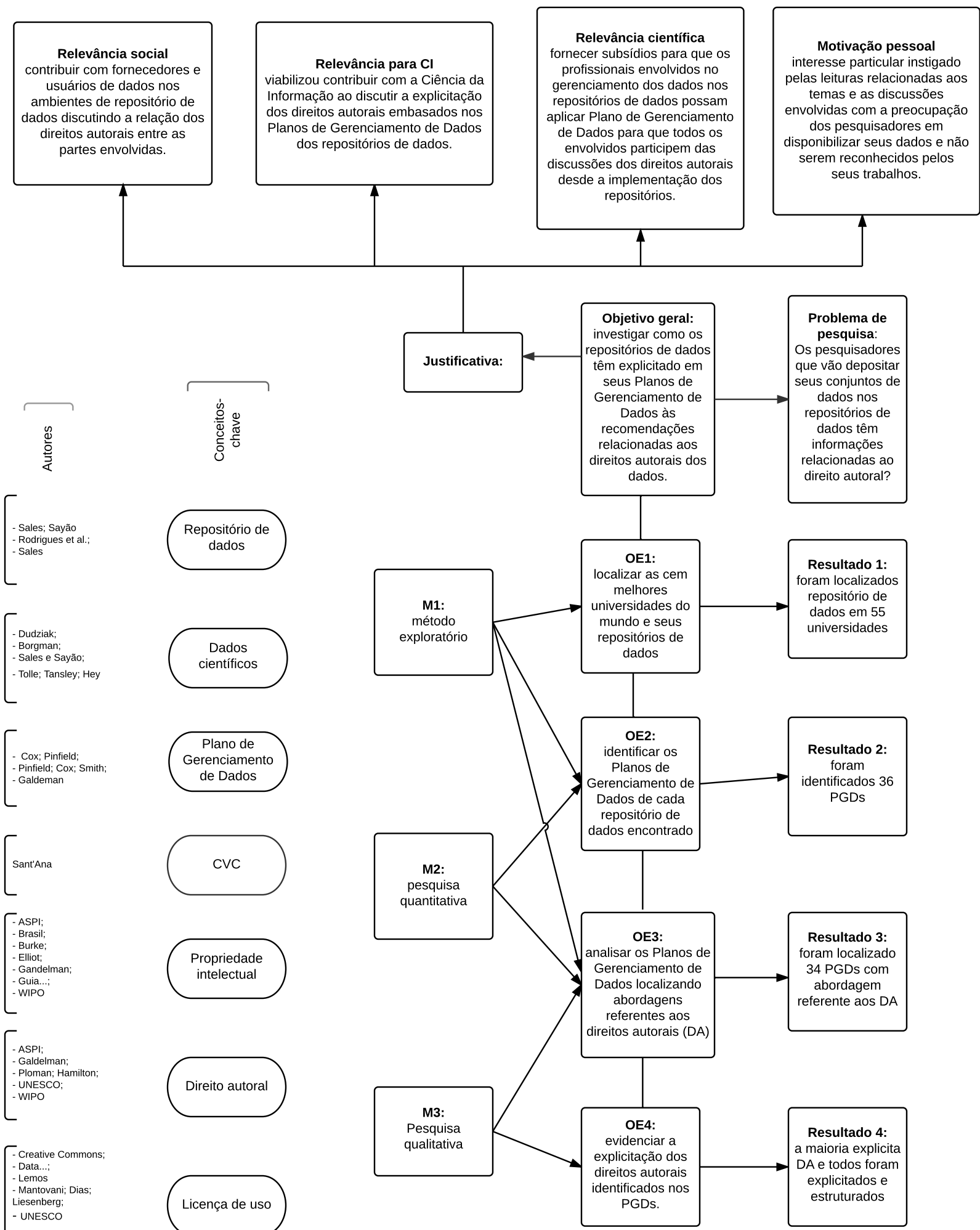
THE APACHE SOFTWARE FOUNDATION. **Apache license**.c2016. Disponível em: <<http://www.apache.org/licenses/LICENSE-2.0>>. Acesso em: 21 set. 2016.

TOLLE, K.; TANSLEY, S.; HEY, T. Jim Gray e a eScience: um método científico transformado. In: HEY, T.; STEWARD, T.; TOLLE, K. (Org.). **O quarto paradigma: descobertas científicas na era da eScience**. Tradução Leda Beck. São Paulo: Oficina de textos, 2011. p. 17-29.

THOMAS, J.; CHEN, B.; CLEMENT, G. ORCID identifiers: planned and potential uses by associations, publishers, and librarians. **The Serials Librarian**, v. 68, p. 332-

















## ANEXO A - Ranking das 100 melhores universidades do mundo

ranking	University	Det.	Country	Presence Rank*	Impact Rank*	Openness Rank*	Excellence Rank*
1	<a href="#">Harvard University</a>	19		1	1	1	1
2	<a href="#">Massachusetts Institute of Technology</a>	19		2	1	6	11
3	<a href="#">Stanford University</a>	19		3	3	2	3
4	<a href="#">University of California Berkeley</a>	19		15	4	4	13
5	<a href="#">University of Michigan</a>	19		7	7	12	4
6	<a href="#">Cornell University</a>	19		4	5	31	21
7	<a href="#">University of Washington</a>	19		23	8	23	6
8	<a href="#">Columbia University New York</a>	19		29	9	10	14
9	<a href="#">University of Pennsylvania</a>	19		14	10	43	12
10	<a href="#">University of Oxford</a>	19		20	18	9	5
11	<a href="#">University of California Los Angeles UCLA</a>	19		60	15	17	7
12	<a href="#">Johns Hopkins University</a>	19		35	42	3	2
13	<a href="#">Yale University</a>	19		18	13	14	20
14	<a href="#">University of Cambridge</a>	19		33	21	7	10
15	<a href="#">University of Wisconsin Madison</a>	19		11	14	52	25
16	<a href="#">Michigan State University</a>	19		63	6	8	90
17	<a href="#">University of Texas Austin</a>	19		85	12	30	46
18	<a href="#">University of Chicago</a>	19		26	19	5	56
19	<a href="#">University of California San Diego</a>	19		79	33	13	15
20	<a href="#">University of Toronto</a>	19		42	43	24	8
21	<a href="#">Pennsylvania State University</a>	19		71	17	61	38
22	<a href="#">University of Illinois Urbana Champaign</a>	19		5	25	32	42
23	<a href="#">New York University</a>	19		111	22	25	40
24	<a href="#">Duke University</a>	19		138	32	16	18
25	<a href="#">University of North Carolina Chapel Hill</a>	19		129	26	33	31
26	<a href="#">Princeton University</a>	19		58	16	19	76
27	<a href="#">University College London</a>	19		136	52	11	9
28	<a href="#">University of Utah</a>	19		129	11	96	95
29	<a href="#">University of British Columbia</a>	19		84	46	36	24
30	<a href="#">(1) University of Maryland Baltimore</a>	19		55	24	50	69
31	<a href="#">University of Florida</a>	19		93	34	41	50
32	<a href="#">University of Southern California</a>	19		118	31	49	52
33	<a href="#">Purdue University</a>	19		183	20	72	83
34	<a href="#">Zhejiang University (National Che Kiang University) / 浙江大学</a>	19		129	36	97	41
35	<a href="#">California Institute of Technology Caltech</a>	19		6	57	54	36
36	<a href="#">Eidgenössische Technische Hochschule ETH Zürich / Swiss Federal Institute of Technology Zurich</a>	19		44	48	109	32
37	<a href="#">University of California Davis</a>	19		173	44	57	34

<u>ranking</u> ▲	<u>University</u>	<u>Det.</u>	<u>Country</u>	<u>Presence Rank*</u>	<u>Impact Rank*</u>	<u>Openness Rank*</u>	<u>Excellence Rank*</u>
38	<a href="#">Peking University / 北京大学</a>	👉		54	65	111	23
39	<a href="#">Northwestern University</a>	👉		222	53	44	28
40	<a href="#">Ohio State University</a>	👉		155	59	27	37
41	<a href="#">University of Virginia</a>	👉		45	27	87	108
42	<a href="#">Tsinghua University / 清华大学</a>	👉		350	60	91	19
43	<a href="#">Carnegie Mellon University</a>	👉		63	29	20	128
43	<a href="#">University of Pittsburgh</a>	👉		142	55	65	35
45	<a href="#">University of California Irvine</a>	👉		339	35	39	84
46	<a href="#">Texas A&amp;M University</a>	👉		56	41	71	94
47	<a href="#">Rutgers The State University of New Jersey</a>	👉		92	40	66	96
48	<a href="#">University of Arizona</a>	👉		160	37	127	87
49	<a href="#">University of Edinburgh</a>	👉		138	62	55	54
50	<a href="#">Universidade de São Paulo USP</a>	👉		16	54	110	78
51	<a href="#">Boston University</a>	👉		69	71	34	64
52	<a href="#">Washington University Saint Louis</a>	👉		207	78	22	49
53	<a href="#">Arizona State University</a>	👉		147	50	26	110
54	<a href="#">University of Colorado Boulder</a>	👉		197	47	106	89
54	<a href="#">Georgia Institute of Technology</a>	👉		80	75	29	72
56	<a href="#">University of Tokyo / 東京大学</a>	👉		90	104	160	27
57	<a href="#">University of California San Francisco</a>	👉		238	126	53	17
58	<a href="#">Vanderbilt University</a>	👉		132	83	58	65
59	<a href="#">University at Buffalo</a>	👉		83	28	186	208
60	<a href="#">McGill University</a>	👉		147	91	51	60
60	<a href="#">Simon Fraser University</a>	👉		145	23	102	277
60	<a href="#">Indiana University Bloomington</a>	👉		138	38	70	183
63	<a href="#">Virginia Polytechnic Institute and State University</a>	👉		111	39	178	171
64	<a href="#">National Taiwan University</a>	👉		9	76	170	106
65	<a href="#">University of Nottingham</a>	👉		254	49	197	118
66	<a href="#">University of California Santa Barbara</a>	👉		188	73	40	100
67	<a href="#">Utrecht University / Universiteit Utrecht</a>	👉		37	123	150	43
68	<a href="#">National University of Singapore</a>	👉		334	133	35	29
69	<a href="#">University of Melbourne</a>	👉		170	131	114	30
70	<a href="#">Shanghai Jiao Tong University / 上海交通大学</a>	👉		90	103	202	61
71	<a href="#">University of Alberta</a>	👉		201	90	108	79
72	<a href="#">Nanyang Technological University</a>	👉		631	100	128	47
73	<a href="#">University of Queensland</a>	👉		96	158	84	39
73	<a href="#">Tufts University</a>	👉		45	58	79	194
75	<a href="#">(1) Seoul National University / 서울대학교</a>	👉		51	143	37	68
76	<a href="#">University of Science &amp; Technology of China / 中国科学技术大学</a>	👉		473	63	213	109

<u>ranking</u> ▲	<u>University</u>	<u>Det.</u>	<u>Country</u>	<u>Presence Rank*</u>	<u>Impact Rank*</u>	<u>Openness Rank*</u>	<u>Excellence Rank*</u>
77	<a href="#">North Carolina State University</a>	👉		75	66	107	170
78	<a href="#">Ruprecht Karls Universität Heidelberg</a>	👉		40	117	280	63
79	<a href="#">Huazhong University of Science &amp; Technology / 华中科技大学</a>	👉		282	77	184	111
80	<a href="#">Universidad Nacional Autónoma de México</a>	👉		17	30	367	331
81	<a href="#">University of Iowa</a>	👉		205	82	125	114
82	<a href="#">Imperial College London</a>	👉		269	269	21	16
83	<a href="#">University of Manchester</a>	👉		279	148	68	45
84	<a href="#">Brown University</a>	👉		119	95	63	116
85	<a href="#">University of Rochester</a>	👉		12	115	28	130
86	<a href="#">University of Oslo / Universitetet i Oslo</a>	👉		13	92	251	126
87	<a href="#">Shandong University / 山东大学</a>	👉		298	51	232	161
87	<a href="#">University of Helsinki / Helsingin yliopisto</a>	👉		39	110	104	101
89	<a href="#">Catholic University of Leuven</a>	👉		190	171	74	44
89	<a href="#">Australian National University</a>	👉		287	84	64	132
91	<a href="#">University of New South Wales</a>	👉		282	132	130	57
92	<a href="#">Emory University</a>	👉		508	113	77	67
93	<a href="#">University of Copenhagen / Københavns Universitet</a>	👉		48	190	98	55
94	<a href="#">University of Hong Kong</a>	👉		53	107	99	119
95	<a href="#">University of Amsterdam / Universiteit van Amsterdam</a>	👉		287	145	172	51
96	<a href="#">Ludwig Maximilians Universität München</a>	👉		375	130	167	58
97	<a href="#">Università degli Studi di Roma La Sapienza</a>	👉		214	127	198	70
98	<a href="#">École Polytechnique Fédérale de Lausanne</a>	👉		242	159	38	80
99	<a href="#">University of Georgia</a>	👉		201	69	133	224
100	<a href="#">King's College London</a>	👉		364	196	47	53