

**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO
BACHARELADO EM BIBLIOTECONOMIA**

JOSEILTON LÚCIO DA SILVA

**REPOSITÓRIOS ELETRÔNICOS DE DADOS: INVESTIGANDO A
ADOÇÃO NAS INSTITUIÇÕES FEDERAIS DE ENSINO SUPERIOR
BRASILEIRAS**

**João Pessoa
2017**

JOSEILTON LÚCIO DA SILVA

**REPOSITÓRIOS ELETRÔNICOS DE DADOS: INVESTIGANDO A
ADOÇÃO NAS INSTITUIÇÕES FEDERAIS DE ENSINO SUPERIOR
BRASILEIRAS**

Trabalho de Conclusão de Curso apresentado ao Departamento de Ciência da Informação do Centro de Ciências Sociais Aplicadas da Universidade Federal da Paraíba, para a obtenção do grau de Bacharel em Biblioteconomia.

Orientador: Prof. Dr. GUILHERME ATAÍDE DIAS

**João Pessoa
2017**

Dados Internacionais de Catalogação na Publicação (CIP)

S586r Silva, Joseilton Lúcio da.

Repositórios eletrônicos de dados: investigando a adoção nas instituições federais de ensino superior brasileiras / Joseilton Lúcio da Silva. – João Pessoa, 2017.
59f.: il.

Orientador(a): Pro^{fa} Dr.^a Guilherme de Ataíde Dias.
Trabalho de Conclusão de Curso (Biblioteconomia) – UFPB/CCSA.

1. Dados de pesquisa. 2. Repositórios eletrônicos. 3. e-Science. 4. Instituições Federais de Ensino Superior (IFES). I. Título.

UFPB/CCSA/BS

CDU:02(043.2)

Gerada pelo Catalogar - Sistema de Geração Automática de Ficha Catalográfica do
CCSA/UFPB, com os dados fornecidos pelo autor(a)

JOSEILTON LÚCIO DA SILVA

**REPOSITÓRIOS ELETRÔNICOS DE DADOS: INVESTIGANDO A
ADOÇÃO NAS INSTITUIÇÕES FEDERAIS DE ENSINO SUPERIOR
BRASILEIRAS**

Trabalho de Conclusão de Curso
apresentado ao Departamento de
Ciência da Informação do Centro de
Ciências Sociais Aplicadas da
Universidade Federal da Paraíba, para a
obtenção do grau de Bacharel em
Biblioteconomia.

Aprovada em 14 de junho de 2017.

Guilherme Ataíde Dias

Universidade Federal da Paraíba / Departamento de Ciência da Informação

Genoveva Batista do Nascimento

Universidade Federal da Paraíba / Departamento de Ciência da Informação

Ismaelly Batista dos Santos Silva

Universidade Federal da Paraíba / Departamento de Ciência da Informação

RESUMO

Pesquisa os avanços relacionados ao tratamento dos dados, desde o surgimento da *e-Scienc*. O estudo investiga como se configuram os repositórios eletrônicos nas Instituições Federais de Ensino Superior (IFES) no que diz respeito aos dados de pesquisa. A priori investiga-se quais IFES possuem repositórios eletrônicos, que tipo de repositórios utilizam e quais as características dos mesmos. Trata-se de uma pesquisa exploratória, onde foram realizadas buscas através do Google Acadêmico e do portal da BRAPCI para a coleta dos materiais utilizados e no portal do IBICT para a localização das IFES e seus repositórios, e de abordagem quantitativa e qualitativa para a análise dos dados coletados durante o estudo. Localiza-se quarenta e uma IFES que fazem uso de Repositórios Institucionais (RIs), onde 100% dessas IFES optam por utilizar o *Dspace* em seus repositórios. A implementação desses repositórios ocorre devido aos avanços das tecnologias digitais e ao surgimento da *e-Science*, que revoluciona o cenário da produção científica e acadêmica, através de uma nova maneira de se tratar os dados de pesquisa. Constata-se então que, atualmente, as IFES não possuem repositórios específicos para o tratamento dos dados de pesquisa, e apenas a Universidade Federal de Brasília (UnB) apresenta interesse na implementação de um repositório específico para o tratamento desses dados. Conclui-se que, no que diz respeito a questão dos dados de pesquisa, há uma carência por parte das IFES, que não possuem repositórios adequados para o tratamento desses objetos.

Palavras-chave: Dados de pesquisa. Repositórios eletrônicos. *e-Science*. Instituições Federais de Ensino Superior (IFES).

ABSTRACT

Researches the advances related to data treatment, beginning from the emergence of e-Science. The study investigates how the electronic repositories in the Federal Institutions of Higher Education (IFES) are configured in regard to research data. A priori investigates which IFES have electronic repositories, what type of repositories they use and what their characteristics are. It is an exploratory research. The search for bibliographic resources was made through Google Academic and the BRAPCI portal. The IBICT portal was used to locate the IFES's repositories. A quantitative and qualitative approach was used to analyze the data collected during the study. The research found that forty-one (41) IFES makes use of Institutional Repositories (IRs), where 100% of these IFES choose to use DSpace as the software for their repositories. The implementation of these repositories occurs as a result of the advances in digital technologies and the emergence of e-Science, which revolutionizes the scenario of scientific and academic production through new ways of dealing with research data. It can be verified that at the present moment the IFES do not have specific repositories for the treatment of research data and only the Federal University of Brasília (UnB) is interested in the implementation of a specific repository for data treatment. It is concluded that, regarding the research data issue and the use of specific designed repositories, there is a gap by the IFES, which do not have adequate repositories for the treatment of these objects.

Keywords: Research data. Electronic repository. E-Science. Federal Institutions of Higher Education (IFES).

LISTA DE FIGURAS

FIGURA 1 – Ciclo de vida de curadoria dos dados do modelo <i>Digital Curation Centre</i> – DCC.....	20
FIGURA 2 – Ciclo de vida dos dados de pesquisa <i>DataONE</i>	21
FIGURA 3 – Ciclo de vida dos dados para a Ciência da Informação – CVD- CI.....	23
FIGURA 4 – Página Inicial do <i>DSpace</i>	32
FIGURA 5 – Página Inicial do <i>Fedora</i>	34
FIGURA 6 – Página Inicial do <i>Dataverse</i>	36
FIGURA 7 – Página Inicial do <i>CKAN</i>	38
FIGURA 8 – Gerenciamento de dados no <i>DRYAD</i>	39
FIGURA 9 – Planilha de folha de cálculo do Google.....	39
FIGURA 10 – Página Inicial do Google Acadêmico.....	41
FIGURA 11 – Página Inicial da BRAPCI.....	41

LISTA DE GRÁFICOS E QUADROS

GRÁFICO 1 – IFES com RIs de acordo com cada região	48
GRÁFICO 2 – IFES com os <i>sites</i> de seus RIs personalizados	50
GRÁFICO 3 – IFES que visam a implementação de outro software em seus RIs	52
QUADRO 1 – Repositórios eletrônicos das Instituições Federais de Ensino Superior – IFES	43
QUADRO 2 – IFES que optaram por personalizar os seus RIs.....	49
QUADRO 3 – Características dos repositórios das IFES	51

LISTA DE SIGLAS

BRAPCI	Base de Dados de Periódicos em Ciência da Informação
CI	Ciência da Informação
CVD-CI	Ciclo de Vida dos Dados para a Ciência da Informação
DCC	<i>Digital Curation Centre</i>
FURJ	Universidade Federal do Rio Grande
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
IFES	Instituições Federais de Ensino Superior
IFPB	Instituto Federal da Paraíba
IFRN	Instituto Federal do Rio Grande do Norte
MIT	<i>Massachusetts Institute of Technology</i>
NSB	<i>National Science Board</i>
OAI	<i>Open Archives Initiative</i>
OAI-PMH	<i>Open Archives Initiative Protocol for Metadata Harvesting</i>
OKF	<i>Open Knowledge Foundation</i>
RDF	<i>Resource Description Framework</i>
RDs	Repositórios Digitais
RI	Repositório Institucional
TDICs	Tecnologias Digitais da Informação e da Comunicação
TDs	Tecnologias Digitais
UFAC	Universidade Federal do Acre
UFAL	Universidade Federal de Alagoas
UFBA	Universidade Federal da Bahia
UFC	Universidade Federal do Ceará
UFES	Universidade Federal do Espírito Santo
UFF	Universidade Federal Fluminense
UFG	Universidade Federal de Goiás
UFGD	Universidade Federal de Grande Dourados

UFJF	Universidade Federal de Juiz de Fora
UFLA	Universidade Federal de Lavras
UFMA	Universidade Federal do Maranhão
UFMG	Universidade Federal de Minas Gerais
UFMS	Universidade Federal do Mato Grosso do Sul
UFMT	Universidade Federal do Mato Grosso
UFPA	Universidade Federal do Pará
UFPB	Universidade Federal da Paraíba
UFPE	Universidade Federal de Pernambuco
UFPEL	Universidade Federal de Pelotas
UFPI	Universidade Federal do Piauí
UFPR	Universidade Federal do Paraná
UFPR	Universidade Federal do Paraná
UFRB	Universidade Federal do Recôncavo da Bahia
UFRJ	Universidade Federal do Rio de Janeiro
UFRN	Universidade Federal do Rio Grande do Norte
UFRS	Universidade Federal do Rio Grande do Sul
UFS	Universidade Federal de Sergipe
UFSC	Universidade Federal de Santa Catarina
UFSCar	Universidade Federal de São Carlos
UFT	Universidade Federal de Tocantins
UFU	Universidade Federal de Uberlândia
UFV	Universidade Federal de Viçosa
UFVJM	Universidade Federal dos Vales do Jequitinhonha e Mucuri
UnB	Universidade Federal de Brasília
UNIFEI	Universidade Federal de Itajaíba
UNIFESP	Universidade Federal de São Paulo
UNILA	Universidade Federal da Integração Latino-Americana

UNIPAMPA Universidade Federal do Pampa
UNIR Universidade Federal de Rondônia
UTFPR Universidade Tecnológica Federal do Paraná

SUMÁRIO

1 INTRODUÇÃO	11
2 OBJETIVOS	15
3 A IMPORTÂNCIA DOS DADOS DE PESQUISA NO CONTEXTO DA <i>E-SCIENCE</i>.....	16
3.1 DADOS	16
3.2 CICLO DE VIDA DOS DADOS DE PESQUISA	18
4 REPOSITÓRIOS ELETRÔNICOS.....	25
4.1 REPOSITÓRIOS INSTITUCIONAIS NO BRASIL.....	27
4.2 ACESSO ABERTO	29
4.3 <i>SOFTWARES</i> MAIS UTILIZADOS PARA IMPLEMENTAÇÃO DE REPOSITÓRIOS.....	30
4.3.1 <i>DSPACE</i>	30
4.3.2 <i>FEDORA</i>	32
4.3.3 <i>DATAVERSE</i>	34
4.3.4 <i>CKAN</i>	36
4.3.5 <i>DRYAD</i>	38
5 PERCURSO METODOLÓGICO	40
6 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS.....	43
7 CONSIDERAÇÕES FINAIS	53
REFERÊNCIAS	55

1 INTRODUÇÃO

Desde o final do século XX, e de forma mais marcante a partir do século XXI, as pesquisas científicas passaram a produzir uma grande quantidade de dados, principalmente, em formatos digitais. Isso acontece graças ao grande avanço que as Tecnologias Digitais (TDs) vem tendo, estando presente cada vez mais nos meios de produção científica, trazendo novas formas de compartilhamento de dados e informações. Os computadores estão cada vez mais sofisticados, aliados a Internet, possibilitando novas formas de aplicações para as fontes básicas de pesquisa, dessa forma, dando um novo impulso aos processos de investigação científica.

Em meio a essas mudanças, se discutia o problema ético envolvido no limite de acesso ao conhecimento científico que a própria comunidade científica enfrentava, onde a busca por acesso livre era uma questão presente nesse cenário. Marcondes e Sayão (2009), explicam que a popularização da Internet e o surgimento do *World Wide Web* (WWW), na década de 90, deram início ao movimento de livre acesso. O principal marco desse movimento, aconteceu em 1991, quando o físico Paul Ginsparg criou o primeiro repositório de *pre-prints*, no laboratório de física de Los Alamos, Novo México, EUA. A partir desse marco, os repositórios digitais surgem como uma alternativa factível para os processos associados a comunicação científica, trazendo um novo meio para a comunidade acadêmica de preservar e disseminar o conteúdo dos periódicos científicos.

No Brasil, um grande marco que representa a luta pelo acesso livre aberto à informação científica foi o Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica que destaca que “a informação científica é o insumo básico para o desenvolvimento científico e tecnológico de uma nação [...]” (IBICT, 2005, *online*).

Esse manifesto é resultado de um longo processo de mobilização mundial pelo livre acesso aos periódicos científicos. No Brasil, a iniciativa de Arquivos Abertos, conhecida mundialmente como *Open Archives Initiative* - OAI (1999) e do movimento pelo Livre Acesso a Informação Científica propõe o livre acesso da informação para todo o mundo, sugerindo a construção de repositórios institucionais e a implementação de interoperabilidade entre esses repositórios nas instituições. (MATIAS, 2015).

Nesse contexto, os Repositórios Institucionais (RIs) vem ganhando mais relevância nas instituições, contribuindo com a comunicação científica e disponibilizando o acesso livre à produção científica, trazendo também uma nova forma de preservação dos dados obtidos nas investigações científicas.

A implantação de um repositório institucional nas universidades e instituições de pesquisa é de extrema importância no que tange o armazenamento e preservação dos conteúdos científicos e acadêmicos realizados nas instituições. Dessa forma, é importante que o profissional bibliotecário esteja conectado às novas tecnologias, exercendo um papel fundamental no que diz respeito a preservação e o tratamento dos documentos digitais.

De acordo com Lynch (2003, p. 1-7) apud Fachin et al., (2009, p. 223), os repositórios eletrônicos desenvolvidos pelas universidades são utilizados como “um conjunto de serviços que a universidade oferece aos membros de sua comunidade, visando o gerenciamento e a disseminação dos materiais (informações) digitais criados pela instituição e pelos membros de sua comunidade”.

Segundo Leite (2006, p. 88),

Além de expandir o acesso à pesquisa, reafirmar o controle sobre o saber pela academia, reduzir o monopólio dos periódicos científicos, entre outras mudanças significativas no sistema de comunicação científica, têm o potencial de servir como indicadores tangíveis da qualidade de uma universidade e de demonstrar a relevância científica, social e econômica de suas atividades de pesquisas, aumentando a visibilidade, o *status* e o valor público da instituição.

Devido a um grande aumento no volume da produção de dados científicos, a comunidade científica passa a pensar numa nova maneira de se armazenar, preservar e acessar os dados produzidos pelas pesquisas científicas. Desta forma, surge a necessidade de um repositório específico para os dados obtidos a partir das investigações científicas.

Um repositório de dados científicos torna capaz o compartilhamento e reuso de dados já coletados. Medeiros e Caregnato (2012, p. 318), explicam que “[...] não cabe ao pesquisador despender um longo tempo no processo de coleta de dados se este procedimento já foi realizado e os dados estão disponíveis”.

Sayão e Sales (2015), ensinam que o compartilhamento e reuso dos dados científicos implica a necessidade de que os dados de pesquisa gerados e coletados sejam depositados em um ambiente o qual garanta que a preservação ativa dos mesmos seja a longo prazo, mantendo as suas características de autenticidade,

integridade e proveniência, proporcionando que eles estejam sempre disponíveis e prontos para serem usados.

De acordo com o que foi apresentado é necessário que os dados produzidos nas pesquisas científicas sejam armazenados em um ambiente confiável que garanta que esses dados sejam preservados e compartilhados da mesma maneira com que foram depositados.

Segundo Conserva JR (2014, p. 16-17) apud Matias (2010, p. 30),

Um repositório é um espaço – real ou virtual – para armazenamento de grande quantidade de alguma coisa – produtos, *software*, arquivos, dados, informações etc – visando principalmente à segurança e preservação. No ambiente virtual, os repositórios, nesse caso chamados repositórios digitais, tem sido cada vez mais utilizados também para a ampla divulgação do que é armazenado.

Esse desenvolvimento contínuo das tecnologias graças ao surgimento da Internet traz cada vez mais novos questionamentos e novas maneiras de se tratar a comunicação científica. Medeiros e Caregnato (2012), destacam que o desenvolvimento da ciência e da tecnologia proporcionam melhorias no cotidiano e na vida das pessoas envolvendo aspectos sociais e econômicos. Com os avanços tecnológicos acontecendo cada vez mais de maneira rápida, cientistas de diversos ramos discutem questões tratadas há séculos a respeito desses avanços, onde se destaca o tema *e-Science*.

A *e-Science*, como estrutura que visa a colaboração entre cientistas a partir do compartilhamento e gerenciamento de dados científicos primários, parece ganhar corpo, uma vez que é parte essencial de uma descentralização do conhecimento e da aplicação efetiva de recursos públicos em um país com pretensões de avanços significativos em ciência e tecnologia, possibilitando que cientistas de diversos ramos tenham acesso a conteúdo já mapeado. (MEDEIROS; CAREGNATO, 2012, p. 315).

Segundo Cunha (2010), a *e-Science* nunca foi tratada como objeto de preocupação para as bibliotecas universitárias, e que para apoiá-la necessariamente deve ser incorporada uma série de atividades e serviços que exigem o desenvolvimento, coordenação e investimentos em vários setores das bibliotecas universitárias para que seja criado um sistema que forneça a segurança dos dados de pesquisa, a preservação desses dados, o acesso e controle dos metadados.

Neste contexto, a *e-Science* traz para a Ciência da Informação questões relevantes sobre comunicação científica, discutindo sobre a maneira em que os dados são tratados nas bibliotecas universitárias, uma vez que os dados oriundos da

e-Science são de fato dados científicos primários, e que é de grande relevância para as instituições ter livre acesso a esses dados. (CUNHA, 2010).

A *e-Science* revoluciona a Ciência da Informação levantando reflexões a respeito do uso de repositórios nas instituições e da maneira com que se trata a preservação digital, o compartilhamento dos dados científicos e o planejamento das bases de dados nas bibliotecas universitárias.

Com base nas reflexões em torno dos repositórios eletrônicos e do tratamento dos dados de pesquisa, indagamos: **Como se configuram os repositórios eletrônicos nas Instituições Federais de Ensino Superior (IFES) no que tange a questão dos dados de pesquisa?**

A escolha do tema da pesquisa foi motivada pelo interesse em conhecer como se configuram os repositórios eletrônicos nas Instituições Federais de Ensino Superior Brasileiras (IFES) no que diz respeito ao arquivamento dos dados de pesquisa, visando aprofundar o conhecimento a respeito do universo *e-Science* e de como os dados de pesquisa estão sendo tratados no cenário das instituições brasileiras.

O estudo estrutura-se em sete capítulos: o primeiro capítulo diz respeito as considerações iniciais da pesquisa. O segundo capítulo apresentará o objetivo geral e os objetivos específicos em torno do estudo. O terceiro capítulo tratará sobre a importância dos dados de pesquisa no contexto da *e-Science*. O quarto capítulo tratará sobre Repositórios eletrônicos, o que são e quais as características dos mesmos. No quinto capítulo será apresentado o percurso metodológico realizado durante a pesquisa. No sexto capítulo, os resultados da pesquisa serão apresentados e analisados. E por fim, no sétimo capítulo, serão apresentadas as considerações finais.

2 OBJETIVOS

No que se refere ao objetivo geral, esta pesquisa, busca apresentar os repositórios eletrônicos nas Instituições Federais de Ensino Superior (IFES) no que tange a questão dos dados de pesquisa.

Para atender ao objetivo geral proposto neste estudo, foram estabelecidos os seguintes objetivos específicos:

- a) Levantar os repositórios eletrônicos das Instituições Federais de Ensino Superior brasileiras;
- b) Identificar a tipologia dos repositórios que as Instituições Federais de Ensino Superior brasileiras utilizam;
- c) Elaborar um quadro com as características dos repositórios das Instituições Federais de Ensino Superior brasileiras.

3 A IMPORTÂNCIA DOS DADOS DE PESQUISA NO CONTEXTO DA *E-SCIENCE*

A produção e utilização dos dados gerados a partir das atividades científicas tem crescido consideravelmente graças a novas formas de geração e disseminação do conhecimento obtido através de um novo paradigma científico, a *e-Science*. Este novo paradigma está ligado fortemente com as tecnologias da informação e comunicação e ao crescimento exponencial na geração de dados.

3.1 DADOS

O conceito do que são dados pode variar de acordo com os pesquisadores e suas áreas de conhecimento. É constatado que os dados são gerados por diferentes comunidades acadêmicas e científicas, e que são destinados aos mais variados propósitos.

Setzer (2015, *online*), define dado como:

[...] uma seqüência de símbolos quantificados ou quantificáveis. Quantificável significa que algo pode ser quantificado e depois reproduzido sem que se perceba a diferença para com o original. Portanto, um texto é um dado. De fato, as letras são símbolos quantificados, já que o alfabeto, sendo um conjunto finito, pode por si só constituir uma base numérica (a base hexadecimal empregada em geral nos computadores usa, além dos 10 dígitos decimais, as letras de A a E). Também são dados fotos, figuras, sons gravados e animação, pois todos podem ser quantificados ao serem introduzidos em um computador, a ponto de se ter eventualmente dificuldade de distinguir a sua reprodução com o original.

De acordo com Dados (2017, *online*), “Dados são um conjunto de valores ou ocorrências em um estado bruto com o qual são obtidas informações com o objetivo de adquirir benefícios[...]”.

O Relatório da Organização para a Cooperação e Desenvolvimento Econômico (2007, p. 13), conforme mencionado por Sayão e Sales (2015, p. 7), descreve dados de pesquisa como “registros factuais usados como fonte primária para a pesquisa científica e que são comumente aceitos pelos pesquisadores como necessários para validar os resultados do trabalho científico”. No contexto em tela, os dados de pesquisa são utilizados pelos pesquisadores como base primária para novas pesquisas científicas, e seu reuso faz-se necessário para a validação dos resultados de novos trabalhos científicos.

O reuso e a interpretação de dados de pesquisa em diferentes contextos é um desafio importante na área de curadoria digital de dados de pesquisas e para a *e-Science* que tem que lidar com os enigmas colocados pela grande quantidade de dados produzidos pelas disciplinas científicas que se enquadram no quarto paradigma, constituindo-se para ambas as áreas objetos essenciais de pesquisa[...]. (SAYÃO; SALES, 2014, p. 82).

O conjunto de dados de pesquisa estabelece o que é necessário para investigar a veracidade e autenticidade das fontes de conhecimento publicadas nos periódicos, o que incide ser essencial para o crescimento científico. Dessa forma, quanto mais os sistemas de informação oferecerem dados de pesquisas que sejam tratados por metadados, de forma com que possam ser expostos e reutilizados pelo maior número possível de pesquisadores de várias áreas do conhecimento, maior será o estado de transparência, reprodutibilidade e eficiência do processo de construção de conhecimento científico, e maior será o volume dos projetos de pesquisa aplicados para a sociedade. (SAYÃO; SALES, 2014).

Devido a diversidade dos dados de pesquisa, conforme indicado anteriormente por Setzer (2015), é preciso elaborar estratégias de gestão de vasto espectro que incluam os diversos tipos de dados para que se torne possível o reconhecimento dessas diferenças quando se determina o processo de gerenciamento para o arquivamento persistente e para a curadoria digital das coleções de dados de pesquisa. (SAYÃO; SALES, 2015).

Os dados de pesquisa podem ser categorizados de várias formas devido as suas diferenças. De acordo com a *National Science Board* (NSB), os dados se distinguem de acordo com a sua natureza ou origem em: dados observacionais, computacionais e experimentais.

Os dados observacionais são aqueles obtidos através de observações diretas, que se relacionam a um determinado lugar ou tempo e até mesmo o levantamento do comportamento de uma comunidade. Esses dados são qualificados como registros históricos e a sua coleta só pode ser feita apenas uma vez, dessa forma é necessário que passem por processos de curadoria para que os mesmos sejam preservados para sempre. Os dados computacionais são resultados da aplicação dos modelos computacionais e de simulações e podem ser reaplicados ao passar dos tempos, no entanto, não se faz necessária sua preservação a longo tempo. Os dados de pesquisa experimentais são originários de situações controladas mediante experimentos que podem ser reproduzidos sem a necessidade de serem armazenados indefinidamente. (SAYÃO; SALES, 2015).

De acordo com Poliakoff (2013, p. 1) apud Sayão e Sales (2013, p. 5),

Os dados que coletamos hoje podem ser usados no futuro de forma que ainda não conseguimos imaginar. Os exploradores de antigamente que coletavam espécimes de plantas e animais não sabiam nada sobre DNA e

hoje as amostras são submetidas a esse tipo de investigação. Quando você coleta os seus dados, reúne informações que, no futuro, poderão ser analisadas de formas muito diferentes. São coisas que terão um valor enorme para cientistas que nem nasceram.

Podemos afirmar então que independente da tipologia dos dados, a preservação dos mesmos é extremamente necessária para que futuramente esses dados possam ser usados e reutilizados por pesquisadores, professores, alunos, entre outros profissionais que estejam construindo uma nova pesquisa, tornando-se então essencial para a Ciência da Informação (CI).

3.2 CICLO DE VIDA DOS DADOS DE PESQUISA

O conhecimento e as práticas reunidas nas últimas décadas em preservação digital e acesso deram origem a um conjunto de estratégias, abordagens tecnológicas e atividades que juntas agora são conhecidas como “curadoria digital”. Embora seja uma definição em evolução, já está determinado que a curadoria digital abrange a gestão operante e a preservação de recursos digitais durante todo ciclo de vida de interesse do universo acadêmico desses dados, tendo como concepção o desafio temporal de atender as gerações atuais e futuras dos usuários. (SALES; SAYÃO, 2012).

Medeiros e Caregnato (2012), explicam que o gerenciamento dos dados primários, sua preservação e manutenção possibilitam o compartilhamento e o reuso desses dados, facilitando a vida dos pesquisadores na hora da produção do conhecimento científico, já que não será necessário voltar a parte de um processo que já foi realizado. “[...] não cabe ao pesquisador despender um longo tempo no processo de coleta de dados se esse procedimento já foi realizado e os dados estão disponíveis”. (MEDEIROS; CAREGNATO, 2012, p. 318).

Os dados e as coleções de dados de pesquisa possuem um tempo de vida superior aos projetos de pesquisa que os geraram. Isso faz com que pesquisadores, professores, estudantes, entre outros profissionais de diversas áreas possam continuar a trabalhar com esses dados mesmo depois que os projetos tenham sido retidos. Dessa forma, será possível que futuros projetos de pesquisa explorem ou insiram novos elementos a esses dados que poderão ser reusados por outros pesquisadores, recomeçando um novo ciclo de vida. (MEDEIROS; CAREGNATO, 2012).

Dados de pesquisa são recursos informacionais valiosos que geralmente requerem muito tempo e dinheiro para serem produzidos. Se eles forem apropriadamente gerenciados podem ser usados agora e no futuro por outros pesquisadores evitando a duplicação de esforços. Além do mais, os dados de pesquisa podem ser reusados por outras disciplinas, isto por que muitos deles têm um valor extraordinário que extrapola o seu propósito original, podendo ser interpretados em contextos diferentes do que aqueles para os quais foram gerados ou coletados inicialmente. (SAYÃO; SALES, 2015, p. 53).

Existe uma sequência de concepções de modelos de ciclo de vida de dados de pesquisa, cada qual com peculiaridades e objetivos definidos.

Sayão e Sales (2015), destacam a importância desses modelos de ciclo de vida, que apresentam uma estrutura que caracteriza as várias opções que necessitarão ser realizadas, garantindo que os dados possam ter a sua usabilidade potencializada e estendida.

Em meio a esses modelos de ciclo de vida dos dados de pesquisa, destacam-se os seguintes modelos: *Digital Curation Centre (DCC) Curation Lifecycle Model*¹; *DataONE Data Lifecycle*²; *DDI Combined Lifecycle model* e³; *UK Data archive data lifecycle*⁴.

Considerando a necessidade de descrever os processos relacionados a gestão de dados, destaca-se o modelo do *Digital Curation Centre* – DCC, conforme a Figura 1, que tem como eixo a capacitação para a gestão de dados de pesquisa e apresenta a distinção das ações que acontecem durante todo o ciclo de vida desses dados. (SANTANA, 2013).

Com base no modelo do *Digital Curation Centre*, Ball (2012), apresenta os seguintes níveis de ciclo de vida:

Descrição e Representação: a criação, coleta, preservação e manutenção dos metadados é feita para que os mesmos possam ser usados e reutilizados futuramente. (BALL, 2012).

Planejamento de preservação: são elaboradas estratégias, políticas e procedimentos para todas as ações. (BALL, 2012).

Vigilância Comunitária e Participação: um grupo predeterminado interessado nos dados é responsável por observar as alterações que os mesmos sofrem, além

¹<<http://www.dcc.ac.uk/resources/curation-lifecycle-model>>

²<<http://escholarship.org/uc/item/7tf5q7n3#page-1>>

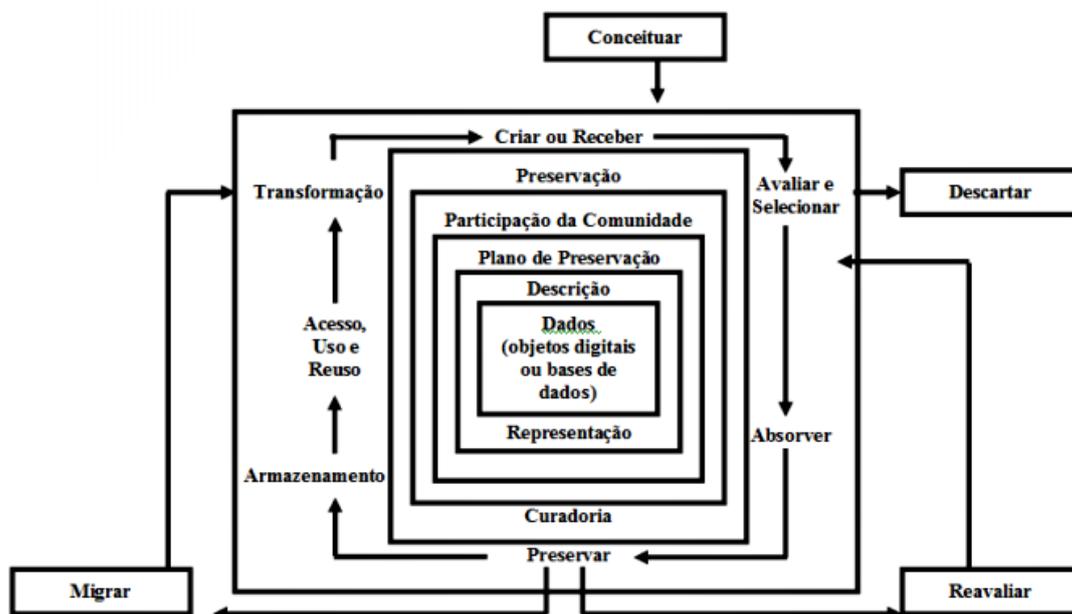
³<<http://www.ddialliance.org/Specification/DDI-Lifecycle/>>

⁴<<http://www.data-archive.ac.uk/create-manage/life-cycle>>

de contribuir no desenvolvimento de padrões, ferramentas e *softwares* relevantes para os dados. (BALL, 2012).

Curadoria e Preservação: representa a execução da gestão planejada e as ações administrativas elaboradas para promover a preservação e curadoria dos dados em seu ciclo de vida. (BALL, 2012).

FIGURA 1 – Ciclo de vida de curadoria dos dados do modelo *Digital Curation Centre* – DCC.



Fonte: DCC (2013) apud Santana (2013).

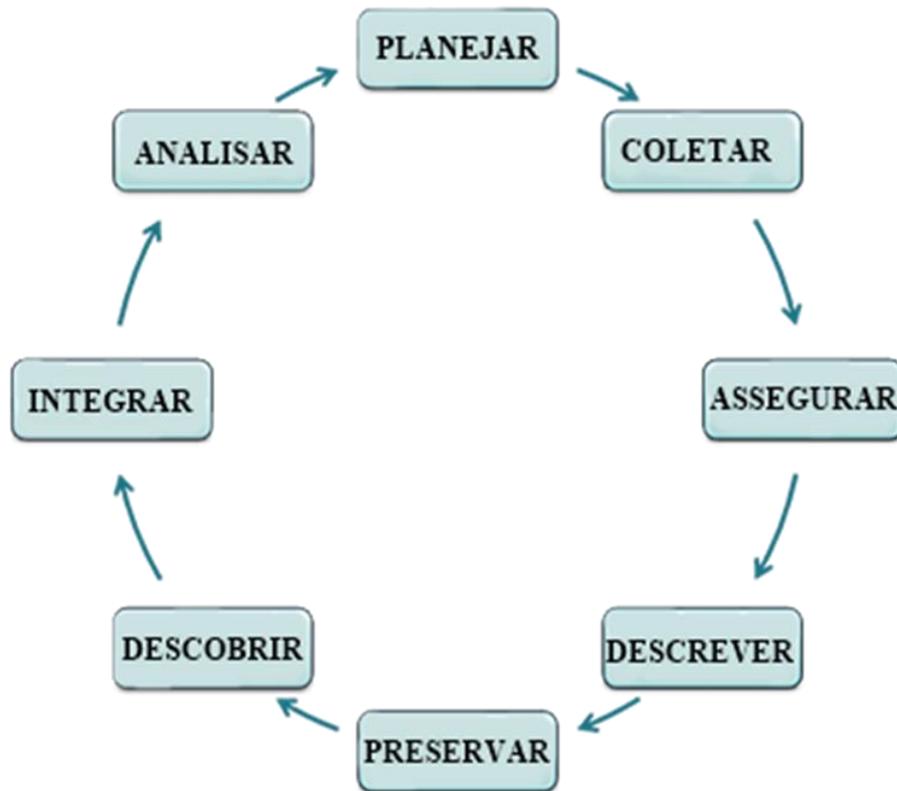
Outro modelo citado por Sayão e Sales (2015), é o *DataONE*, cujo o seu objetivo é:

[...]produzir conjuntos de dados de autodescrição. Se você der seus dados a um cientista ou colega que não esteja envolvido com o seu projeto, eles serão capazes de fazer sentido do mesmo? Eles serão capazes de usá-lo de forma eficaz e adequada? Este primário descreve alguns dados fundamentais para as práticas que lhe permitirão desenvolver um plano de gestão de dados, bem como a forma de criar, organizar, descrever, preservar e compartilhar dados⁵. (DataONE, 2017, online, tradução nossa).

Conforme o ciclo de vida do *DataONE*, existem oito etapas para a gestão de dados, representadas na Figura 2.

⁵ *The goal of data management is to produce self-describing data sets. If you give your data to a scientist or colleague who has not been involved with your project, will they be able to make sense of it? Will they be able to use it effectively and properly? This primer describes a few fundamental data management practices that will enable you to develop a data management plan, as well as how to effectively create, organize, manage, describe, preserve and share data.*

FIGURA 2 – Ciclo de vida dos dados de pesquisa *DataONE*.



Fonte: *DataONE* (2017, tradução nossa).

De acordo com o *DataONE*, Strasser et al. (2015), expõe essas etapas da seguinte forma:

Planejar: descrição dos dados que serão reunidos e como será feito o seu gerenciamento e disponibilização para o acesso durante seu tempo de vida, levando-se em consideração como os dados são gerados, sua metodologia e tempo de vida; qual repositório mais qualificado para o armazenamento desses dados, de acordo com a área do conhecimento que esse repositório é destinado; trata-se também da forma com que esses dados serão gerenciados, descritos, compartilhados e preservados. (STRASSER et al., 2015).

Coletar: são feitas observações manuais ou através de sensores ou outros instrumentos, isso garante que os dados possam ser utilizados posteriormente. É necessário criar um modelo para a coleta de dados para que os mesmos sejam recolhidos, organizar esses dados por meio de planilhas ou bancos de dados, de forma que os dados sejam descritos conforme o seu conteúdo, e fazer a utilização de um *software* e *hardware* não proprietários, possibilitando que os arquivos inseridos possam ser lidos no futuro. (STRASSER et al., 2015).

Assegurar: a qualidade dos dados é assegurada por intermédio de controles e inspeção. Durante a sua coleta, entrada e análise, é necessário descrever quaisquer circunstâncias que possam afetar a qualidade desses dados. Faz-se necessária a utilização de resumos estatísticos e gráficos para levantar se os valores resultantes desse processo são questionáveis ou impossíveis e para apresentar *outliers*. (STRASSER et al., 2015).

Descrever: os dados são descritos utilizando padrões de metadados adequados. A documentação extensiva dos dados (metadados) é a chave para o entendimento futuro dos dados. Recomenda-se descrever em que contexto digital esses dados estão inseridos (nome do conjunto de dados, qual procedimento foi realizado), descrever o pessoal e as partes interessadas (quem realizou o recolhimento dos dados, a quem perguntas devem ser feitas e quais os patrocinadores), em que contexto científico estão inseridos (razão da pesquisa, dados coletados, onde e quais as condições do local onde foram coletados), apresentando todo processo realizado para que esses dados possam ser compreendidos e utilizados. (STRASSER et al., 2015).

Preservar: os dados são inseridos em um arquivo adequado para que os mesmos possam ser utilizados e reutilizados a longo prazo. (STRASSER et al., 2015).

Descobrir: os dados potencialmente utilizáveis são detectados e acessados, junto às informações essenciais sobre metadados. (STRASSER et al., 2015).

Integrar: vários tipos de dados são agregados para formar um conjunto de dados idêntico para serem analisados. (STRASSER et al., 2015).

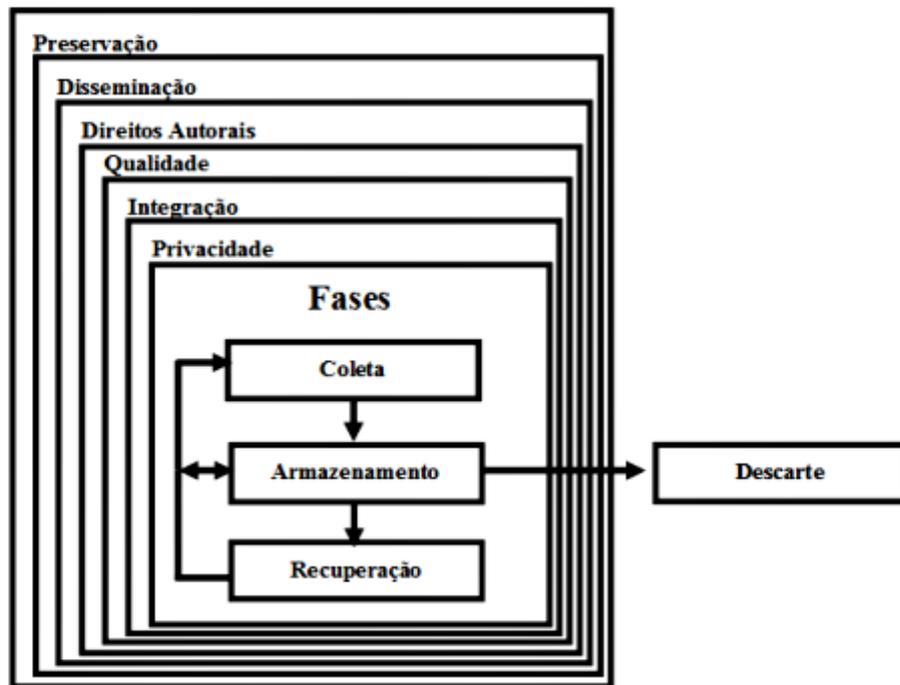
Analisar: é realizada a análise dos dados. (STRASSER et al., 2015).

Um pesquisador ou uma equipe de pesquisadores engajados nos aspectos do ciclo de vida dos dados podem desempenhar tanto o papel de criador quanto o papel de usuário dos dados tratados. Algumas equipes de pesquisadores relacionados a modelagem e sínteses dos dados podem desenvolver novos dados no processo de descobrir, integrar, analisar e sintetizar os presentes dados. (SAYÃO; SALES, 2015).

Santana (2013) propõe um novo modelo de ciclo de vida dos dados para a Ciência da Informação (CVD-CI), conforme a Figura 3, a partir da realização de uma pesquisa bibliográfica que abrange os modelos já existentes. No modelo proposto

por Santana é essencial distinguir fases e objetivos na criação do ciclo de vida dos dados para a Ciência da Informação.

FIGURA 3 – Ciclo de vida dos dados para a Ciência da Informação – (CVD-CI).



Fonte: Santana (2013).

O CVD-CI é dividido em três fases (como descrito na Figura 3): A fase da Coleta inclui as atividades relacionadas a definição inaugural dos dados a serem operados, seja na criação de como esses dados serão acessados, selecionados e estruturados, e os dados devem ser descritos como metadados. A fase do Armazenamento retrata as atividades referentes ao processamento, transformação, inserção, modificação, migração, transmissão e toda e qualquer prática que apontem a persistência de dados por meio de um suporte digital. Essa fase possibilita a execução da fase de recuperação e ainda pode produzir novos dados reavendo ações da fase de coleta. Por sua vez, a Recuperação é a fase em que os dados se tornam concretos, remediando as atividades relacionadas a consulta e visualização dos dados. (SANTANA, 2013).

Nesse contexto, podemos afirmar que os modelos de ciclo de vida dos dados de pesquisa são ferramentas que possibilitam a gestão dos dados, é essencial para a preservação e compartilhamento dos mesmos, e que seguindo esses modelos citados anteriormente, pesquisadores, professores, alunos e demais profissionais

poderão ter acesso a dados confiáveis agora ou futuramente, podendo também reutilizar dados produzidos em pesquisas anteriores.

4 REPOSITÓRIOS ELETRÔNICOS

Em meio às mudanças ocasionadas pelo avanço da tecnologia digital no final do século XX e mais precisamente no início do século XXI, se discutia uma das questões mais importantes para a comunidade científica, o acesso livre aos periódicos científicos.

Marcondes e Sayão (2009), explicam que a popularização da Internet e o surgimento do *World Wide Web* (WWW), ambos na década de 90, determinaram o início do movimento de livre acesso.

O principal marco desse movimento aconteceu em 1991, quando o físico Paul Ginsparg criou o primeiro repositório de *pre-prints*, no laboratório de física de Los Alamos, Novo México, EUA. A partir deste fato a comunidade acadêmica começou a procurar alternativas para ter acesso aos periódicos, já que o custo das assinaturas cobrado pelas editoras comerciais aumentava cada vez mais. Com o surgimento da *internet*, tornou-se possível a publicação de um número maior de artigos a um custo menor, artigos esses que tinham sua submissão e publicação com um alcance mundial ainda maior. A partir dessas mudanças, a questão da cobrança por assinaturas passou a ser confrontada pela comunidade acadêmica, gerando a lógica do livre acesso. (MARCONDES; SAYÃO, 2009).

Sousa Filho et al. (2012), explicam que esse movimento vem crescendo cada vez mais graças às vantagens que o *World Wide Web* traz junto com o Movimento dos Arquivos Abertos, e enfatiza que esse movimento “pressupõe a adoção de uma filosofia aberta que inclui o acesso à informação por todos e para todos, na busca de transformações que contribuam para a melhoria da coletividade”. (SOUSA FILHO et al, 2012, p. 5).

De acordo com Leite (2009), os repositórios de acesso aberto às informações científicas, possibilitam que a comunidade científica faça o uso dessas informações sem barreiras, e que o planejamento e implementação de repositórios com essa política faz com que os resultados das pesquisas realizadas nas instituições ganhem maior visibilidade.

Santos e Miraglia (2009) descrevem repositórios como sistemas de informação cujo sua finalidade é o armazenamento, preservação, disseminação e promoção do acesso à produção intelectual da comunidade científica.

Segundo Conserva Jr. (2014) apud Matias (2010), repositório é um espaço físico ou virtual com a finalidade de armazenar diferentes objetos para que os mesmos estejam em segurança e tenham sua preservação garantida.

Rice e Southall (2016) explicam que os repositórios de dados devem incluir suporte para os seguintes recursos: tipos dos dados de pesquisa; status dos dados de pesquisa; e formato dos arquivos de dados.

Segundo o IBICT (2017, *online*, tradução nossa),

Repositórios digitais (RDs) são bases de dados online que reúnem de maneira organizada a produção científica de uma instituição ou área temática. Os RDs armazenam arquivos de diversos formatos. Ainda, resultam em uma série de benefícios tanto para os pesquisadores quanto às instituições ou sociedades científicas, proporcionam maior visibilidade aos resultados de pesquisas e possibilitam a preservação da memória científica de sua instituição. Os RDs podem ser institucionais ou temáticos. Os repositórios institucionais lidam com a produção científica de uma determinada instituição. Os repositórios temáticos com a produção científica de uma determinada área, sem limites institucionais.

Os repositórios podem ser classificados como institucionais e temáticos. São repositórios institucionais os que pertencem a universidades, institutos e laboratórios de pesquisa, contendo os mais variados tipos de objetos digitais, e temáticos os que pertencem a uma determinada área do conhecimento. (LEITE, 2009).

Leite et al. (2012) explicam que um repositório institucional de acesso aberto estabelece serviços de informações científicas em um ambiente digital capaz de operar com outro repositório de acesso aberto, sendo este repositório dedicado ao gerenciamento da produção científica e acadêmica de uma universidade ou institutos de pesquisa. O referido autor justifica que um repositório institucional de acesso aberto “contempla a reunião, armazenamento, organização, preservação, recuperação e, sobretudo, a ampla disseminação da informação científica produzida na instituição.” (LEITE et al. 2012, p. 7).

Os repositórios eletrônicos são de grande importância para os profissionais da informação, para os bibliotecários, sendo responsáveis por todo processo que os documentos digitais passam durante o seu ciclo de vida, facilitando não apenas o trabalho dos bibliotecários, mas também de pesquisadores, alunos, e toda a comunidade acadêmica.

Sayão e Marcondes (2009) afirmam que a implementação de um repositório institucional é um modo de reconhecer que as atividades intelectuais e acadêmicas das instituições de pesquisa estão cada vez mais representadas, documentadas e

compartilhadas no formato digital, e que é responsabilidade das instituições exercer a custódia sobre esses conteúdos, a fim de torná-los disponíveis para o acesso e manter sua preservação por longo prazo.

Os repositórios institucionais são, em sua maioria, utilizados nas IFES, conforme Freitas, Silva e Guimarães (2009, p. 334), com “a preocupação de disponibilizar os resultados de pesquisas feitas por seus docentes e discentes, o que dá uma grande projeção acadêmica e permite o acesso de maneira irrestrita, rápida e gratuita por qualquer pessoa com acesso à *internet*”.

Nesse contexto, os repositórios digitais surgem como uma nova alternativa para a comunicação científica, trazendo um novo meio para a comunidade acadêmica de se preservar os periódicos.

4.1 REPOSITÓRIOS INSTITUCIONAIS NO BRASIL

No Brasil, um grande marco que representa a luta pelo acesso livre aberto à informação científica foi o Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica que destaca que “a informação científica é o insumo básico para o desenvolvimento científico e tecnológico de uma nação [...]” (IBICT, 2005, *online*).

O Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) foi criado com o objetivo de registrar e expandir a produção científica brasileira. O surgimento das tecnologias digitais da informação e da comunicação (TDICs), aliadas ao movimento do acesso livre à informação, possibilitaram o surgimento de um novo cenário onde o objetivo inicial do instituto possa ser realizado. Com o lançamento do Manifesto Brasileiro de Apoio ao Movimento de Acesso Livre à Informação, surge uma política nacional de acesso livre à informação científica orientada por esse manifesto. (KURAMOTO, 2006).

Esse manifesto é resultado de um longo processo de mobilização mundial pelo livre acesso aos periódicos científicos. No Brasil, a iniciativa de Arquivos Abertos, conhecida mundialmente como *Open Archives Initiative* - OAI (1999) e do movimento pelo Livre Acesso a Informação Científica propõe o livre acesso da informação para todo o mundo, sugerindo a construção de repositórios institucionais e a implementação de interoperabilidade entre esses repositórios nas instituições. (MATIAS, 2015).

Os repositórios institucionais são hoje um dos desafios mais importantes para os profissionais de informação no Brasil. O livre acesso à informação científica, que é fundamental para o crescimento da pesquisa brasileira, sempre foi o objetivo dos sistemas de Informação, Ciência e Tecnologia (ICT) implantados no Brasil desde a década de 1950. (MARCONDES; SAYÃO, 2009).

No que diz respeito a questão do acesso livre e do acesso aberto a informação, Gauz (2008, p. 279) explica que, “compreendemos a expressão ‘acesso livre à informação’ como uma política, portanto, como algo de abrangência mais ampla e geral. ‘acesso aberto à informação’, expressão igualmente utilizada na literatura especializada é, para nós, o caminho, o mecanismo.”

A implantação de um repositório institucional nas universidades e instituições de pesquisa é de extrema importância no que diz respeito ao armazenamento e a preservação das atividades realizadas na instituição.

Ainda que sejam grandes produtoras de informação e conhecimento, as instituições de ensino, tem grande parte de sua produção indisponível no campo das instituições e na comunidade científica. Isso acontece por causa da grande quantidade de documentos impressos e eletrônicos espalhados pelas coleções e porque não há como acessar muitos desses documentos. Os repositórios institucionais trazem uma nova forma de compartilhamento da informação científica e são meios importantes para que a informação tenha grande visibilidade. A implementação dessa ferramenta visa diminuir a falta de visibilidade que sua produção intelectual possui. (MIRANDA; DELFINO, 2016).

De acordo com Lynch (2003 p. 1-7) apud Fachin et al., (2009, p. 223), os repositórios eletrônicos desenvolvidos pelas universidades são utilizados como “um conjunto de serviços que a universidade oferece aos membros de sua comunidade, visando o gerenciamento e a disseminação dos materiais (informações) digitais criados pela instituição e pelos membros de sua comunidade”.

Segundo Leite (2006, p. 88),

Além de expandir o acesso à pesquisa, reafirmar o controle sobre o saber pela academia, reduzir o monopólio dos periódicos científicos, entre outras mudanças significativas no sistema de comunicação científica, têm o potencial de servir como indicadores tangíveis da qualidade de uma universidade e de demonstrar a relevância científica, social e econômica de suas atividades de pesquisas, aumentando a visibilidade, o *status* e o valor público da instituição.

Dentre os muitos significados que a ideia de ciência aberta traz, o mais persuasivo deles é o que indica o conhecimento científico como patrimônio da humanidade e que desta forma, deve ser disponível de forma livre e sem barreiras para toda a comunidade, científica ou não, e que o uso e reuso desse conhecimento possa ser feito por todos sem que haja constrangimentos por meios tecnológicos, econômicos, sociais ou legais. (SAYÃO; SALES, 2014).

Nesse contexto, podemos afirmar que para a informação e o conhecimento científico serem disponíveis para todas as pessoas, sendo essas cientistas, professores, estudantes ou profissionais de várias áreas é preciso que os repositórios institucionais utilizem meios que possibilitem o acesso livre das informações e de todo o conhecimento científico focalizados nesses repositórios.

4.2 ACESSO ABERTO

A *Open Archives Initiative* (OAI), constituída em 1999, na Convenção de Santa Fé, Novo México, EUA, tem como objetivo principal contribuir para a transformação da comunicação científica por meio de aspectos técnicos e de suporte organizacional definidos por uma estrutura de publicação científica aberta. O modelo *Open Archives*, criado através dessa iniciativa possibilita um alto nível de interoperabilidade entre os repositórios devido a componentes como: mecanismo de submissão; sistema de armazenamento a longo prazo; política de gestão para a submissão e preservação de documentos e uma interface aberta que permite terceiros coletar os metadados dos respectivos arquivos. (OLIVEIRA; CARVALHO, 2011).

O movimento de Acesso Livre trouxe novas alternativas para a Ciência da Informação, as quais surgiram para facilitar a vida de pesquisadores e outros profissionais. No Brasil e no mundo, novas ferramentas para a disseminação da produção científica passaram a ser usadas graças a esse movimento. Os repositórios digitais de acesso livre foram, sem dúvidas, uma das ferramentas mais importante nesse processo de evolução tecnológica. O acesso aberto às publicações científicas facilita a vida dos cientistas, professores, estudantes e quaisquer outros profissionais que buscam obter informações no que diz respeito ao conhecimento científico.

Para atender as exigências do movimento do acesso livre e do arquivamento seguro das publicações, as instituições acadêmicas estão cada vez mais colocando

em vigor o uso de pacotes livres de *softwares* como o *DSpace*, *Fedora*, entre outros em seus repositórios. (SAYÃO, 2010).

4.3 SOFTWARES MAIS UTILIZADOS PARA A IMPLEMENTAÇÃO DE REPOSITÓRIOS

Conforme Leite et al. (2012), os repositórios que seguem uma política de acesso aberto lidam com a produção científica de uma determinada universidade ou instituição de pesquisa, aumentando a visibilidade dos trabalhos realizados por pesquisadores, professores e alunos da instituição.

Nesse contexto, os RIs de acesso aberto possibilitam que as universidades e instituições de pesquisa tratem de suas produções científicas de forma com que estas estejam cada vez mais acessíveis à toda comunidade acadêmica.

4.3.1 DSPACE

Desenvolvido pelo *Massachusetts Institute of Technology* (MIT) e pelos Laboratórios *Hewlett-Packard*, o *DSpace* (Figura 4) é um *software* de acesso livre adequado para a criação e desenvolvimento de bibliotecas digitais e repositórios, construído através das iniciativas de arquivos abertos, *Open Archives Initiative – Protocol for Metadata Harvesting – OAI-PMH*, que fazem uso da tecnologia aberta, podendo ser acessados por vários provedores de serviço disponíveis a nível nacional e internacional, além de seu uso ser indicado tanto para organizações privadas quanto públicas. (BLATTMANN; WEBER, 2008).

O *DSpace* é ideal para criação de repositórios digitais abertos nas organizações acadêmicas, é um *software* gratuito e fácil de instalar, além de ser personalizável, possibilitando que as organizações façam o seu uso de acordo com as suas necessidades. Além de preservar textos, imagens, conjuntos de dados, entre outros tipos de arquivos digitais, o *DSpace* permite o acesso fácil e aberto aos seus conteúdos. (DSPACE, 2017).

Com mais de 1000 organizações utilizando este *software*, o *DSpace* possui a maior comunidade de usuários e desenvolvedores do mundo, podendo ser usado por instituições educacionais, governamentais, privadas e comerciais. Qualquer organização pode usar e modificar essa ferramenta sem ter que pagar algo por isso, já que se trata de um *software* totalmente gratuito. A cada modificação feita visando melhorar o *DSpace* as organizações contribuem não apenas com os desenvolvedores, mas com toda a comunidade. O *software DSpace* é administrado por um pequeno grupo de desenvolvedores voluntários (os *committers*) que

trabalham a favor da comunidade, inserindo novos recursos e fazendo correções. (DSPACE, 2017).

O *DSpace* fornece as seguintes ferramentas para atender as necessidades das organizações:

Personalizar a interface do usuário – Você pode personalizar o visual do seu *site DSpace* de acordo com as necessidades da sua instituição, oferecendo aos seus usuários uma interface mais intuitiva. O *DSpace* oferece duas opções de interface de usuário: uma interface tradicional e outra *Manakin* que disponibiliza vários temas. (DSPACE, 2017).

Personalizar os metadados – O *DSpace* utiliza o *Dublin Core* como formato padrão de metadados e possibilita que você adicione ou altere qualquer campo para personalizá-lo para sua aplicação. (DSPACE, 2017).

Configurar Navegar e pesquisar – O *DSpace* possibilita que você escolha quais campos você gostaria de exibir para navegação de seu *site*, como autor, título, data entre outros. Permite também que você escolha os campos de metadados que serão incluídos na interface de pesquisa. (DSPACE, 2017).

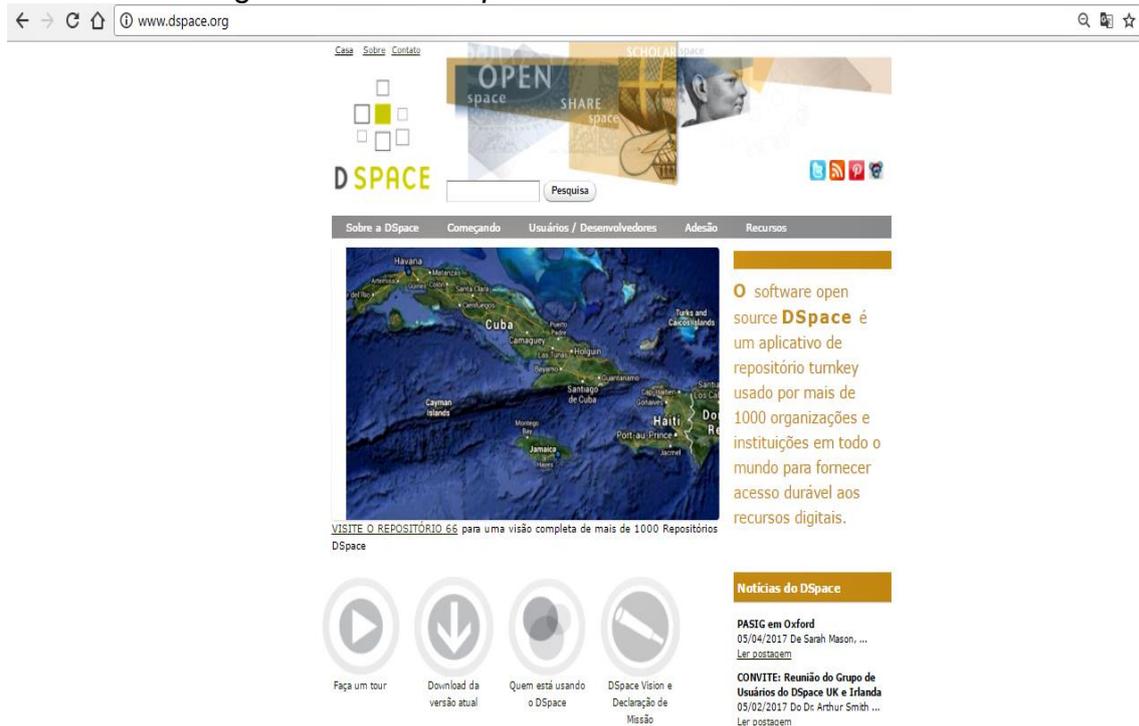
Mecanismos de autenticação local - O *DSpace* fornece uma série de plug-ins para a maioria dos métodos de autenticação. Além disso, possui seu próprio método de autenticação interna ou pode ser configurado para usar outros tipos de métodos de autenticação de uma só vez. O *DSpace* ainda permite que você crie seu próprio *plug-in* de autenticação caso seja feito o uso de um mecanismo de autenticação personalizado. (DSPACE, 2017).

Compatibilidade de padrões - O *DSpace* funciona com vários protocolos padrão: OAI-PMH, *OpenSearch*, *OpenURL*, RSS, entre outros. (DSPACE, 2017).

Banco de dados configurável - Você pode configurar o banco de dados onde o *DSpace* gerencia seus metadados. (DSPACE, 2017).

Idioma padrão – É possível escolher seu idioma padrão de acordo com a sua linguagem. O *DSpace* possui mais de vinte idiomas e permite que você o configure para que o idioma definido em seu navegador Web seja seu idioma preferido. (DSPACE, 2017).

FIGURA 4 – Página inicial do *DSpace*.



Fonte: *DSpace* (2017, tradução nossa).

4.3.2 FEDORA

Desenvolvido pela *University of Virginia* e pela *Cornell University*, o *Flexible Extensible Digital Object and Repository Architecture – Fedora* (Figura 5), é uma plataforma de acesso aberto que possui uma arquitetura projetada para servir como base para a implementação de repositórios digitais para uma diversidade de aplicações, como bibliotecas digitais, repositórios institucionais, entre outros sistemas gerenciadores de informação. (OLIVEIRA; CARVALHO, 2011).

Oliveira e Carvalho (2011) explicam que o núcleo central do *Fedora* é o repositório de serviços, que pode ser acessado utilizando ferramentas por meio de uma integração de serviços da *web* que permitem a criação, gerenciamento, armazenamento, acesso e o reuso dos objetos digitais. Todas as funções do *Fedora* são disponibilizadas por esse repositório de serviços.

Em 2005, o primeiro estágio para a tecnologia da *Web Semântica* foi inserido ao repositório de serviços do *Fedora*, fornecendo suporte para descrições RDF (*Resource Description Framework*) de objetos digitais, relacionamentos RDF e indexação do repositório baseada em um conjunto de objetos relacionados. Com esta implementação, o *Fedora* passa a ser o único sistema de código aberto que

relaciona redes de informações semânticas do conteúdo com o suporte a persistência e gestão do conteúdo digital. (OLIVEIRA; CARVALHO, 2011).

Liderado pelo *Fedora Leadership Group* e administrado pela organização sem fins lucrativos *DuraSpace* que fornece inovação para projetos e soluções de tecnologia aberta que se concentram no acesso a longo prazo dos dados de digitais, o *Fedora* é um repositório robusto, modular e de código aberto que possibilita o gerenciamento e divulgação de todo o conteúdo digital. (FEDORA, 2017).

O seu uso é adequado para bibliotecas e arquivos digitais tanto para o acesso como para preservação dos mesmos, além de fornecer acesso especializado a coleções digitais extensas e complexas de materiais históricos e culturais, bem como dados científicos. O *Fedora* possui uma base de usuários instalada em todo o mundo que conta com organizações de patrimônio acadêmico e cultural, universidades, instituições de pesquisa, bibliotecas universitárias, bibliotecas nacionais e agências governamentais. (FEDORA, 2017).

De acordo com o *Fedora* (2017), suas características são:

Serviços baseados em padrões: conjunto de serviços de repositórios que seguem os padrões modernos da *web*. (FEDORA, 2017).

Qualquer arquivo, qualquer tamanho: armazenamento e preservação de qualquer tipo de arquivo, fornecendo acesso sem restrições aos mesmos. (FEDORA, 2017).

Suporte de dados vinculados nativos: RDF como linguagem padrão e participação como cidadão nativo da *web* de leitura e gravação. (FEDORA, 2017).

Arquitetura Extensível: interação com outros aplicativos e serviços de pesquisa e descoberta. (FEDORA, 2017).

Opções avançadas de armazenamento: várias opções de armazenamento para arquivos e metadados são fornecidas, incluindo sistemas de arquivos, bancos de dados, entre outros. (FEDORA, 2017).

Segurança plugável: proteção aos arquivos digitais com módulos de autorização conectáveis, baseados no controle ao acesso à *Web*. (FEDORA, 2017).

Preservação Pronta: diferentes serviços de restauração de chave são fornecidos. (FEDORA, 2017).

Altamente escalável: milhões de arquivos e registros de metadados são tratados. (FEDORA, 2017).

APIs RESTful: serviços centrais confiáveis são fornecidos aos usuários. (FEDORA, 2017).

Poderosas extensões: conjunto de serviços essenciais e módulos de *plug-in* para disseminação de arquivos abertos. (FEDORA, 2017).

Fluxos de Trabalho Baseados em Mensagens: ferramenta de mensagens para criação de fluxos de trabalho. (FEDORA, 2017).

Implantação fácil: sua implantação como arquivo WAR é altamente fácil e pode ser feita em qualquer ambiente de *Web*. (FEDORA, 2017).

Operações em lote: uma série de ações relacionadas são agrupadas em um repositório evento visando melhorar o seu desempenho. (FEDORA, 2017).

Abrir Dados, Abrir Formatos: dados são facilmente obtidos e exportados em formato RDF ainda que não sejam dependentes do *Fedora*. (FEDORA, 2017).

FIGURA 5 – Página Inicial do *Fedora*.



Fonte: *Fedora* (2017, tradução nossa).

4.3.3 DATAVERSE

Dataverse (Figura 6) é um projeto de *software* de código aberto para o compartilhamento, preservação, citação, investigação e análise de dados de pesquisa. O uso desse *software* facilita a disponibilização de dados de pesquisa para outras pessoas, permitindo que elas reapliquem esses dados em seus trabalhos. Assim, pesquisadores, autores, editores e distribuidores de dados e instituições associadas recebem crédito acadêmico e ganham mais visibilidade na

web. O *software Dataverse* está sendo desenvolvido no Instituto de Ciências Sociais Qualitativas de *Harvard*, além da contribuição de vários colaboradores e contribuintes espalhados por todo o mundo. A construção do *Dataverse* baseou-se no projeto *Virtual Data Center (VDC)*, que se prolongou entre 1999 e 2006 como uma colaboração entre o *Harvard-MIT Data Center* e a *Harvard University Library*. (DATAVERSE, 2017).

De acordo com o IBICT (2015), a plataforma *Dataverse* permite um gerenciamento mais seguro de dados de pesquisa por meio da criação de termos de uso e restrições que limitam o uso e acesso dos dados, além de fornecer cópias de segurança. Dessa forma, devido a facilidade de disseminação de dados e do seu compartilhamento efetivo, as equipes de pesquisa podem trabalhar unidas em um ambiente para acompanhamento dos processos e mudanças nos projetos no decorrer do tempo.

No *Dataverse*, a preservação e acesso de dados a longo prazo é assegurado no reconhecimento persistente que protege os documentos digitais com instrumentos que presumem a limitação da vida útil dos dados. Para sua organização e compatibilidade, o *Dataverse* desenvolve um arquivo de dados pessoal orientado por padrões de metadados que aumentam sua compatibilidade e recuperação. (IBICT, 2015).

Segundo o *Dataverse* (2017), a elaboração de um plano de gerenciamento de dados é essencial para projetos de pesquisa que abrangem a coleta ou disseminação de dados. Os planos de gerenciamento de dados garantem que os dados obtidos por um projeto tenham a integridade, qualidade e procedência necessários para contribuir com a pesquisa proposta, e que os dados necessários para a replicação externa dos resultados da pesquisa estejam a disposição da comunidade de pesquisa.

O gerenciamento de dados é exigido por organizações que patrocinam essas pesquisas, incluindo agências federais e fundações sem fins lucrativos). Muitas revistas, editores e agências de financiamento exigem a utilização de um repositório público para que os pesquisadores possam depositar conjuntos de dados de replicação. O uso do *Dataverse* permite que os pesquisadores cumpram este requisito, além de apoiar o depósito de dados de replicação, tornando estes dados facilmente detectáveis por outros pesquisadores que poderão reutilizar um estudo

sem precisarem gastar tempo entrando em contato com os autores do estudo. (DATAVERSE, 2017).

O *Dataverse* (2017) possui características que auxiliam o gerenciamento de dados tais como: ferramenta para descrição de dados; termos de uso e livros de visitas; notificação de contas e dados; pesquisa facetada; extração de metadados; APIs para interoperabilidade; suporte ao controle de acesso de dados; capacidade de solicitar acesso a arquivos restritos; personalização de *dataverses*; mapeamento de arquivos geoespaciais; entre outras.

Visando atender a necessidade das organizações e melhorar o seu *software*, o projeto *Dataverse* estabelece os seguintes objetivos estratégicos,

- Aumentar a adoção (usuários, *dataverses*, conjuntos de dados, instalações, revistas);
- Concluir as funcionalidades de migração do *Dataverse 4*;
- Desenvolver capacidade para lidar com dados de nível 3 sensíveis, em grande escala e em fluxo contínuo;
- Expandir dados e recursos de metadados para disciplinas existentes e novas;
- Expandir recursos de arquivamento e preservação;
- Aumentar as contribuições da comunidade de desenvolvimento de código aberto;
- Melhorar UX e UI;
- Continuar a aumentar a qualidade do software. (DATAVERSE, 2017, online, tradução nossa).

FIGURA 6 – Página inicial do *Dataverse*.



Fonte: *Dataverse* (2017, tradução nossa).

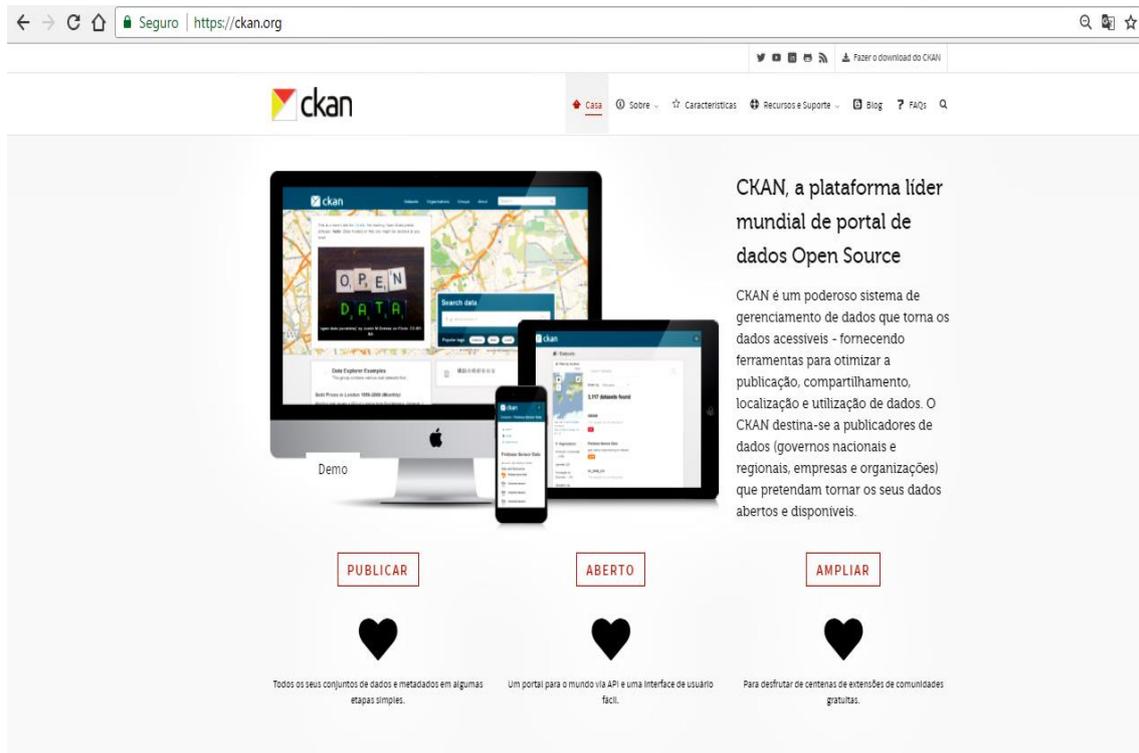
4.3.4 CKAN

O *CKAN* (Figura 7) é um sistema poderoso para criação de *sites* de dados abertos e para o gerenciamento de dados que fornece ferramentas para aprimorar a publicação, compartilhamento, localização e uso de dados. É um *software* utilizado por governos nacionais e locais, instituições de pesquisa e outras organizações que coletam um grande número de dados, e contribui para o gerenciamento e publicação de coleções de dados. Uma vez que os dados são publicados no *CKAN*, desenvolvedores, jornalistas, pesquisadores, ONGs, cidadãos ou até mesmo sua própria equipe podem utilizar seus recursos de pesquisa facetada para buscar e localizar os dados de que necessitam e visualizá-los por meio de mapas, gráficos e tabelas. (CKAN, 2017).

O desenvolvimento do *software Comprehensive Knowledge Network -CKAN*, liderado pela *Open Knowledge Foundation (OKF)*, teve início em Março de 2006, tendo seu primeiro lançamento público em Julho de 2007. A OKF é uma organização sem fins lucrativos criada em 2004 com o objetivo de promover a abertura de todas as formas de conhecimento. O *software* tem como objetivo principal desenvolver, apoiar e promover projetos, comunidades e ferramentas que facilitam a criação, acesso e disseminação de conhecimentos. (WINN, 2013).

O sistema *CKAN* possui uma comunidade ativa de colaboradores que desenvolvem e mantêm sua tecnologia principal, além de ser modificado e estendido por uma comunidade ainda maior de desenvolvedores que contribuem para uma crescente biblioteca de extensões *CKAN*. Sua associação supervisiona e gerencia o sistema por meio do *Steering Group* e *Technical Team*, em nome de toda sua comunidade de usuários e contribuintes. O sistema *CKAN* possui uma arquitetura adaptável que possibilita que extensões sejam desenvolvidas para oferecer recursos adicionais, como coleta ou upload de dados, além de fornecer uma interface de programação de aplicativos poderosa que permite que aplicativos e serviços de terceiros sejam criados em torno dele, além de utilizar seu modelo interno para armazenar metadados sobre diferentes registros e permite que seus usuários naveguem e pesquisem esses metadados através de uma interface *web*. (CKAN, 2017).

FIGURA 7 – Página Inicial do CKAN.



Fonte: CKAN (2017, tradução nossa).

O CKAN facilita a publicação e o compartilhamento de dados. É uma plataforma líder mundial de acesso aberto para catalogar, armazenar e acessar conjuntos de dados.

4.3.5 DRYAD

DryadLab é uma coleção de acesso livre, abertamente licenciado e de alta qualidade que oferece módulos educacionais para que os estudantes participem da investigação científica usando dados reais. Os padrões *DryadLab* podem ser facilmente integrados em cursos avançados de ensino secundário, de graduação e de pós-graduação, e podem ser adaptados para atender às necessidades específicas de seu curso. O *DryadLab* disponibiliza uma vasta variedade de dados de pesquisa implícitos a publicações científicas e médicas que podem ser descobertas, reutilizadas e citadas. (DRYAD, 2017).

FIGURA 8 – Gerenciamento de dados no *DRYAD*.

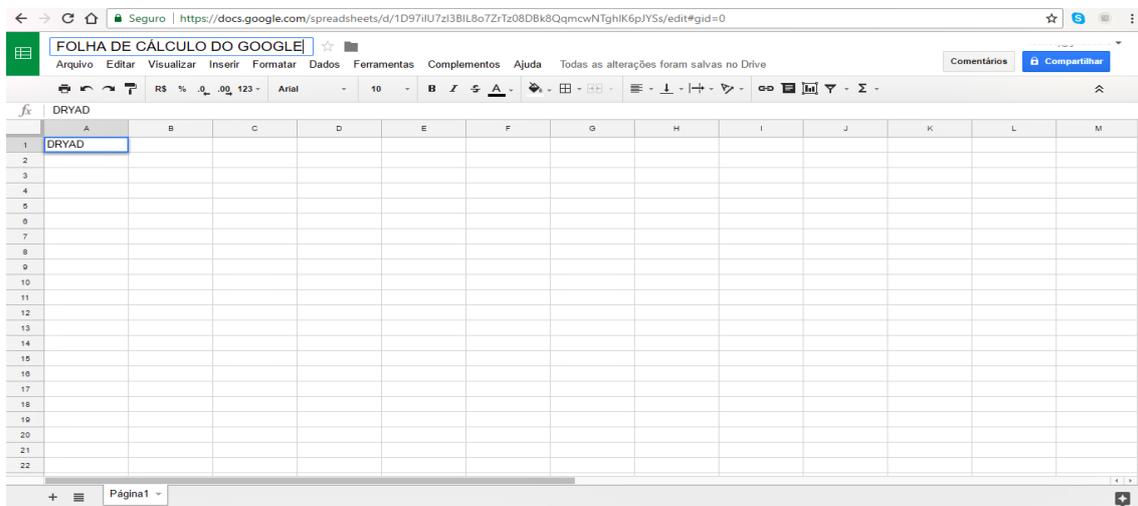


Fonte: *DRYAD* (2017).

Conforme a Figura 8, os dados passam por quatro etapas: na primeira, os dados são depositados; na segunda, adquire-se um identificador permanente; na terceira etapa, as citações tornam-se maiores; e por fim, os dados são descobertos e seguros.

Os dados incluídos em seu pacote estão extremamente curados e disponíveis para uso. Os módulos foram criados para serem usados com a Folha de Cálculo do Google (Figura 9), já que está disponível para qualquer pessoa usar. É recomendado que os estudantes tenham acesso aos dados antes de de se iniciar um módulo, isso permitirá que possíveis problemas possam ser resolvidos antes de uma atividade ser iniciada. (*DRYAD*, 2017).

FIGURA 9 – Planilha de Folha de Cálculo do Google.



Fonte: Folha de Cálculo do Google, 2017.

5 PERCURSO METODOLÓGICO

A pesquisa caracteriza-se como: exploratória, bibliográfica e de abordagem quantitativa e qualitativa. Quanto aos objetivos, trata-se de uma pesquisa exploratória que tem como objetivo se familiarizar com a temática estudada, coletando informações a respeito do problema, levantando hipóteses, aprimorando ideias, visando a descoberta de informações que antes não eram evidentes sobre o tema de estudo. (GIL, 2002).

Quanto aos procedimentos técnicos, a pesquisa caracteriza-se como bibliográfica, constituindo-se como uma pesquisa construída com auxílio de materiais elaborados por diversos autores, tais como livros, artigos científicos e materiais disponibilizados na internet. (GIL, 2002).

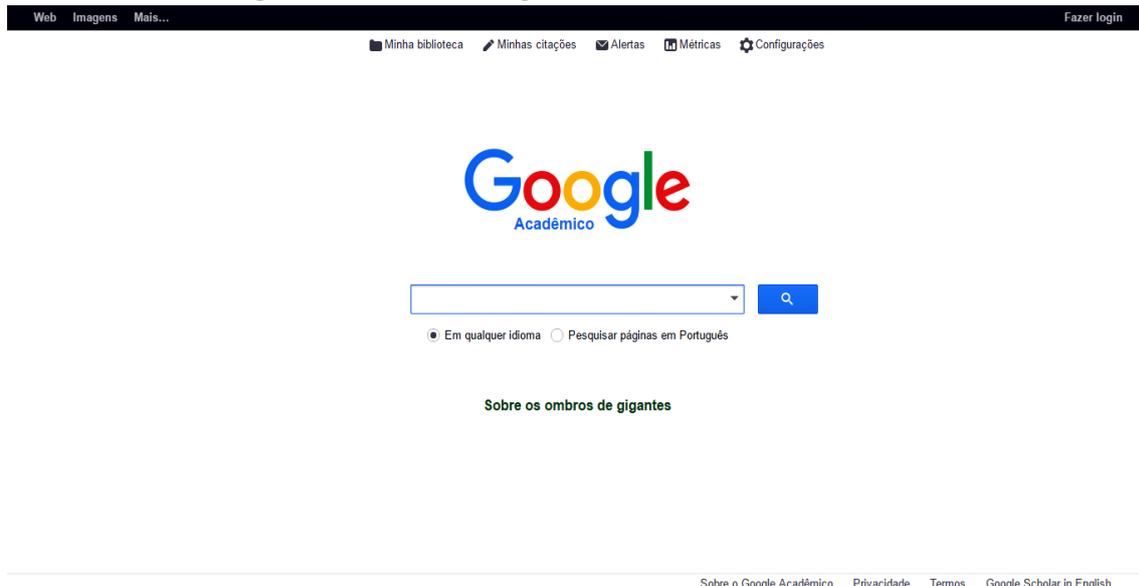
Quanto a sua abordagem, a pesquisa assume caráter quantitativo e qualitativo no que tange a análise das informações coletadas. Silva e Menezes (2005, p. 20) explicam que a pesquisa quantitativa “considera que tudo pode ser quantificável, o que significa traduzir em números opiniões e informações para classificá-las e analisá-las,” fazendo-se necessário o uso de recursos e técnicas estatísticas. Já na pesquisa qualitativa, o uso de recursos e técnicas estatísticas não se faz necessário, uma vez que “o ambiente natural é a fonte direta para a coleta de dados e o pesquisador é o instrumento-chave. [...] os pesquisadores tendem a analisar seus dados indutivamente. O processo e seu significado são os focos principais da abordagem.” (SILVA; MENEZES, 2005, p. 20).

A pesquisa foi realizada durante o período de 5 de abril a 2 de junho de 2017, através de buscas realizadas na internet utilizando o Google Acadêmico (Figura 10) e a Base de dados de periódicos em Ciência da Informação – BRAPCI (Figura 11) para obter os materiais utilizados na pesquisa.

Escolheu-se a base de dados do Google Acadêmico por conter diversas fontes acadêmicas, e a BRAPCI por ser uma base de dados bastante relevante para a Ciência da Informação.

Para obter os materiais utilizados na pesquisa, foram utilizados como descritores os termos “dados de pesquisa”, “repositórios eletrônicos”, “repositórios institucionais”, “curadoria digital”, “preservação digital”, “gestão de dados”, “e-Science” e “acesso aberto”.

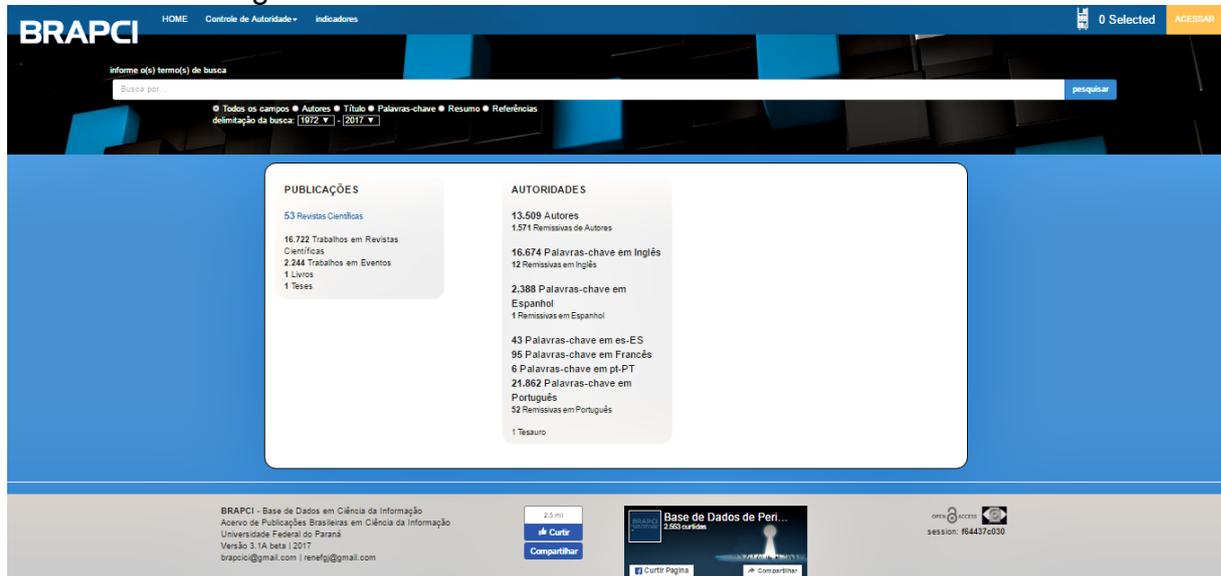
FIGURA 10 – Página inicial do Google Acadêmico.



Fonte: Google Acadêmico, 2017.

A página inicial do Google Acadêmico é simples e fácil de se navegar, além de oferecer a busca por materiais apenas no idioma Português ou em qualquer tipo de idioma, conforme a Figura 10.

FIGURA 11 – Página inicial da BRAPCI.



Fonte: BRAPCI, 2017.

O campo de busca para os textos da BRAPCI é oferecido logo em sua página inicial. A busca pode ser realizada através do nome do autor, título, palavras-chave, resumo, referência, ou por todos os campos anteriores.

Para localizar os repositórios eletrônicos das IFES, realizou-se uma pesquisa através do portal do Instituto Brasileiro de Informação em Ciência e Tecnologia

(IBICT), o referido portal disponibiliza uma lista com repositórios de instituições públicas e privadas brasileiras. A partir da análise desta lista, delimitamos os repositórios eletrônicos das Instituições Federais de Ensino Superior (IFES), apresentando as instituições e seus repositórios de acordo com cada região.

Para identificar o tipo de repositório utilizado pelas IFES, foram realizadas buscas nos *sites* de cada um desses repositórios, visando identificar suas características e como estes repositórios estão configurados.

Para apresentar estatisticamente os dados, fez-se o uso da planilha do Excel, que possui uma interface intuitiva e ferramentas que auxiliam nos cálculos e na construção de gráficos.

No seguinte capítulo, os resultados obtidos juntamente com sua análise serão apresentados.

6 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Segundo o portal do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), existem quarenta e uma (41) Instituições Federais de Ensino Superior (IFES) que possuem repositórios institucionais, todas utilizando o *DSpace* como *software* de acesso aberto em seus repositórios (Quadro 1).

QUADRO 1 – Repositórios eletrônicos das Instituições Federais de Ensino Superior (IFES).

REGIÃO	INSTITUIÇÃO	NOME DO REPOSITÓRIO	SOFTWARE UTILIZADO
CENTRO – OESTE	Universidade Federal de Brasília – UnB	Repositório Institucional da Universidade de Brasília – RIUnB	<i>DSPACE</i>
	Universidade Federal de Goiás – UFG	Repositório Institucional da Universidade Federal de Goiás – RIUFG	<i>DSPACE</i>
	Universidade Federal de Grande Dourados - UFGD	Repositório Institucional da Universidade Federal de Grande Dourados – DSpace UFGD	<i>DSPACE</i>
	Universidade Federal de Mato Grosso – UFMT	Repositório Institucional da Universidade Federal de Mato Grosso – RIUFMT	<i>DSPACE</i>
	Universidade Federal de Mato Grosso do Sul – UFMS	Repositório Institucional da Universidade Federal de Mato Grosso do Sul – RIUFMS	<i>DSPACE</i>
	Instituto Federal de Educação, Ciência e Tecnologia da Paraíba – IFPB	Repositório Digital do Instituto Federal da Paraíba	<i>DSPACE</i>
	Instituto Federal de Educação, Ciência e	MEMORIA – Repositório Institucional	<i>DSPACE</i>

NORDESTE	Tecnologia do Rio Grande do Norte – IFRN		
	Universidade Federal de Alagoas – UFAL	Repositório Institucional da Universidade Federal de Alagoas – RIUFAL	<i>DSPACE</i>
	Universidade Federal da Bahia – UFBA	Repositório Institucional da Universidade Federal da Bahia – RIUFBA	<i>DSPACE</i>
	Universidade Federal do Ceará – UFC	Repositório Institucional da Universidade Federal do Ceará – RIUFC	<i>DSPACE</i>
	Universidade Federal do Maranhão – UFMA	Repositório Institucional da Universidade Federal do Maranhão - RIUFMA	<i>DSPACE</i>
	Universidade Federal da Paraíba – UFPB	Repositório Eletrônico Institucional – REI	<i>DSPACE</i>
	Universidade Federal de Pernambuco – UFPE	Repositório Institucional da Universidade Federal de Pernambuco – RIUFPE	<i>DSPACE</i>
	Universidade Federal do Piauí – UFPI	Repositório Institucional da Universidade Federal do Piauí – RIUFPI	<i>DSPACE</i>
	Universidade Federal do Recôncavo da Bahia – UFRB	Repositório Institucional da Universidade Federal do Recôncavo da Bahia – RIUFRB	<i>DSPACE</i>
	Universidade Federal do Rio Grande do Norte – UFRN	Repositório Institucional da Universidade Federal do Rio Grande do Norte – RIUFRN	<i>DSPACE</i>
	Universidade Federal de Sergipe – UFS	Repositório Institucional da Universidade Federal	<i>DSPACE</i>

		de Sergipe – RIUFS	
NORTE	Universidade Federal do Acre – UFAC	Repositório Institucional da Universidade Federal do Acre – RIUFAC	<i>DSPACE</i>
	Universidade Federal do Pará – UFPA	Repositório Institucional da Universidade Federal do Pará – RIUFPA	<i>DSPACE</i>
	Universidade Federal de Rondônia – UNIR	Repositório Institucional da Universidade Federal de Rondônia – RIUNIR	<i>DSPACE</i>
	Universidade Federal de Tocantins – UFT	Repositório Institucional da Universidade Federal de Tocantins – RIUFT	<i>DSPACE</i>
SUL	Universidade Federal da Integração Latino-Americana – UNILA	Dspace@UNILA - Repositório Institucional	<i>DSPACE</i>
	Universidade Federal do Pampa – UNIPAMPA	Repositório Institucional da Universidade Federal do Pampa – RI DSPACE UNIPAMPA	<i>DSPACE</i>
	Universidade Federal do Paraná – UFPR	Acervo Digital UFPR	<i>DSPACE</i>
	Universidade Federal de Pelotas – UFPEL	GUAIACA – Repositório Institucional	<i>DSPACE</i>
	Universidade Federal do Rio Grande – FURG	Repositório Institucional da Universidade Federal do Rio Grande – RIFURG	<i>DSPACE</i>
	Universidade Federal do Rio Grande do Sul – UFRS	LUME – Repositório Digital	<i>DSPACE</i>
	Universidade Federal de Santa Catarina – UFSC	Repositório Institucional da Universidade Federal de Santa Catarina –	<i>DSPACE</i>

		RIUFSC	
	Universidade Tecnológica Federal do Paraná – UTFPR	Repositório Institucional da Universidade Tecnológica Federal do Paraná – RIUT	<i>DSPACE</i>
SUDESTE	Universidade Federal do Espírito Santo – UFES	Repositório Institucional da Universidade Federal do Espírito Santo – riUfes	<i>DSPACE</i>
	Universidade Federal Fluminense – UFF	Repositório Institucional da Universidade Federal Fluminense – RIUFF	<i>DSPACE</i>
	Universidade Federal de Itajaúba – UNIFEI	Repositório Institucional da Universidade Federal de Itajaúba – RIUNIFEI	<i>DSPACE</i>
	Universidade Federal de Juiz de Fora – UFJF	Repositório Institucional da Universidade Federal de Juiz de Fora – RIUFJF	<i>DSPACE</i>
	Universidade Federal de Lavras – UFLA	Repositório Institucional da Universidade Federal de Lavras – RIUFLA	<i>DSPACE</i>
	Universidade Federal de Minas Gerais – UFMG	Repositório Institucional da Universidade Federal de Minas Gerais – RIUFMG	<i>DSPACE</i>
	Universidade Federal do Rio de Janeiro – UFRJ	Pantheon – Repositório Institucional	<i>DSPACE</i>
	Universidade Federal de São Carlos – UFSCar	Repositório Digital Livre Saber – LISA	<i>DSPACE</i>
	Universidade Federal de São Paulo – UNIFESP	Repositório Institucional UNIFESP	<i>DSPACE</i>
	Universidade Federal de	Repositório Institucional	<i>DSPACE</i>

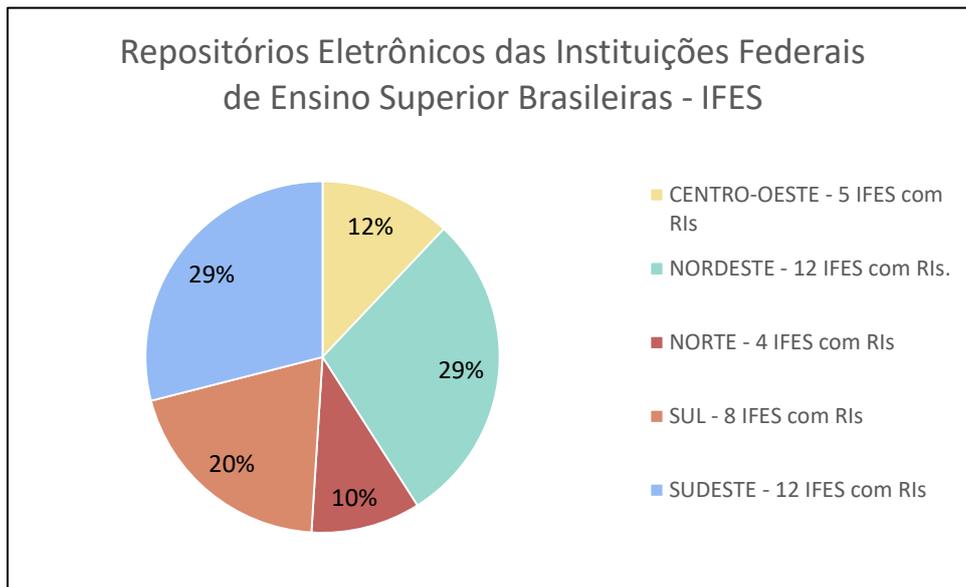
	Uberlândia – UFU	da Universidade Federal de Uberlândia – RIUFU	
	Universidade Federal dos Vales do Jequitinhonha e Mucuri – UFVJM	Repositório Institucional da Universidade Federal dos Vales do Jequitinhonha e Mucuri – RIUFVJM	<i>DSPACE</i>
	Universidade Federal de Viçosa – UFV	LOCUS - Repositório Institucional	<i>DSPACE</i>

Fonte: Dados de pesquisa, 2017.

Do total de quarenta e uma IFES que possuem RIs apresentadas, cinco IFES pertencem a região Centro-Oeste, composta pelos estados do Distrito Federal, Goiás, Mato Grosso e Mato Grosso do Sul. Doze IFES pertencem a região Nordeste, composta pelos estados de Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte e Sergipe. Quatro IFES pertencem a região Norte, composta pelos estados do Acre, Pará, Rondônia e Tocantins. Oito IFES pertencem a região Sul, composta pelos estados do Paraná, Rio Grande do Sul e Santa Catarina. E doze IFES pertencem a região Sudeste, composta pelos estados do Espírito Santo, Minas Gerais, Rio de Janeiro e São Paulo.

Tanto a região Nordeste quanto a Sudeste, são responsáveis por 29% das IFES que utilizam RIs. Em seguida, vem a região Sul, representando 20% das IFES que utilizam RIs. Logo depois, a região Centro-Oeste, responsável por 12% das IFES com RIs. E em último lugar, a região Norte, responsável por 10% das IFES que utilizam RIs.

No Gráfico 1, a seguir, a porcentagem de cada região de acordo com o número de IFES e seus respectivos RIs será apresentada.

GRÁFICO 1 – IFES com RIs de acordo com cada região.

Fonte: Dados da pesquisa, 2017.

Constatou-se também que alguns estados possuem um maior número de IFES, sendo estes responsáveis por uma porcentagem maior dentro de suas regiões.

Na região Centro-Oeste, por exemplo, o estado do Mato Grosso do Sul possui duas das cinco IFES que utilizam RIs, sendo responsável por 4,8% dos 12% que correspondem a sua região, enquanto os estados do Distrito Federal, Goiás e Mato Grosso possuem apenas uma IFES com RI, cada um representando 2,4% do valor total de sua região.

Na região Nordeste, os estados da Bahia, Paraíba e Rio Grande do Norte representam a mesma porcentagem de 4,9%, cada um desses estados possui duas IFES com RIs. Os estados de Alagoas, Ceará, Maranhão, Pernambuco, Piauí e Sergipe apresentam apenas uma IFES com RI, cada um representando 2,4% do total de 29% de sua região.

Na região Sul, o estado do Rio Grande do Sul possui quatro IFES com RIs, sendo responsável por 10% das IFES com RIs, metade do valor total de sua região. Em seguida, vem o estado do Paraná, com 3 IFES que possuem RIs, representando 7,5% do valor total de sua região. E por fim, o estado de Santa Catarina, possuindo apenas uma IFES com RI, representando 2,5% dos 20% que sua região representa.

Na região Sudeste, o estado que se destaca por possuir o maior número de IFES com RIs é Minas Gerais, composto por sete IFES que utilizam RIs, com uma porcentagem de 17%, mais da metade do valor total de IFES com RIs em sua região. Os estados do Rio de Janeiro e São Paulo representam a mesma

porcentagem de 4,8%, cada um desses estados possui duas IFES com RIs. O estado do Espírito Santo possui apenas uma IFES com RIs e representa 2,4% do valor total de 29% de sua região.

A região Norte é a única região em que todos os seus estados representam o mesmo número de IFES. Os estados do Acre, Pará, Rondônia e Tocantins possuem apenas uma IFES com RIs e representam 2,5% cada, totalizando os 10% de sua região.

Com base nesta análise, constatou-se que, Minas Gerais, além de possuir o maior número de IFES com RIs de sua região, é também o estado com mais IFES utilizando RIs, levando em consideração todos os outros estados de todas as regiões.

A respeito de como os repositórios são configurados, constatou-se através das pesquisas feitas nos *sites* de cada um desses repositórios, que 100% das IFES utilizam como ferramenta o *software DSpace* em suas configurações.

O *DSpace* como *software* de acesso aberto possibilita que as instituições possam configurá-lo de acordo com as suas necessidades, além de possibilitar o acesso livre e sem custos a toda comunidade acadêmica.

Constatou-se que a maioria das IFES optaram por denominar seus repositórios de acordo com a sigla RI seguida pela sigla da universidade e seguem um mesmo padrão em seus RIs, como exemplo o Repositório Institucional (RI) da Universidade Federal de Brasília (UnB) que denomina seu repositório como RIUnB e o Repositório Institucional (RI) da Universidade Federal de Minas Gerais (UFMG), cujo repositório denomina-se RIUFMG.

No Quadro 2, a seguir, serão expostas as IFES que optaram por outras denominações para seus RIs.

QUADRO 2 – IFES que optaram por personalizar os seus RIs.

IFES	DENOMINAÇÕES DOS RIs
IFRN	MEMORIA
UFTPR	RiUT
UFPEL	GUAIIACA
UFRS	LUME
UFRJ	Pantheon
UFSCar	LISA

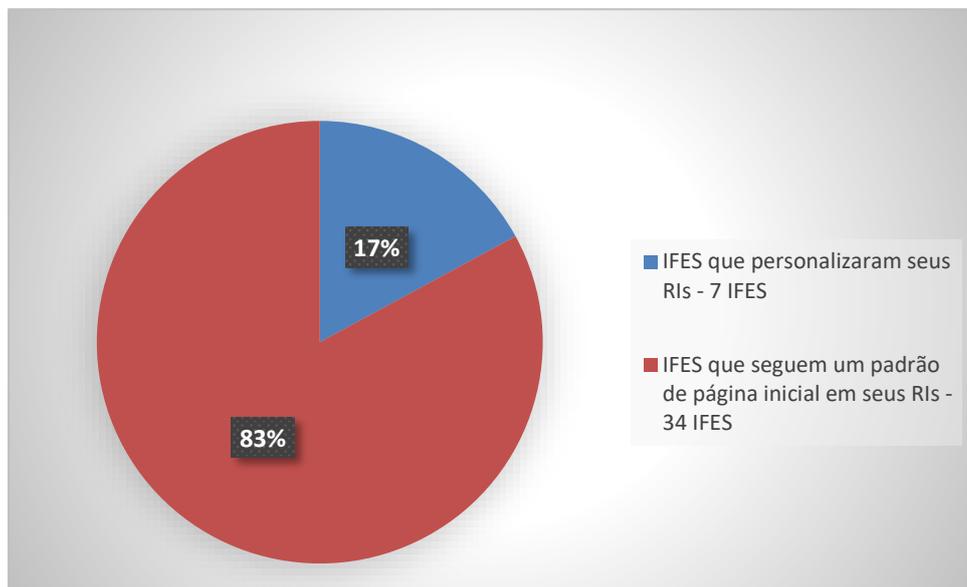
UFV	LOCUS
-----	-------

Fonte: Dados da pesquisa, 2017.

Conforme exposto no Quadro 2, sete IFES optaram por outras definições para seus RIs. Dessa forma, essas IFES se destacam das demais por possuírem uma página inicial diferenciada. Isso é possível porque o *DSpace* permite que as IFES personalizem seus repositórios de acordo com suas necessidades.

No Gráfico 2, a seguir, a porcentagem das IFES que se destacam quanto a sua denominação e página inicial será apresentada:

GRÁFICO 2 – IFES com os *sites* de seus RIs personalizados.



Fonte: Dados de pesquisa, 2017.

Através das pesquisas nos mecanismos de buscas dos repositórios e da constatação de que todas as IFES utilizam o *DSpace* como ferramenta de acesso aberto para arquivar e disponibilizar seus respectivos conteúdos em seus RIs, identificou-se quais as características desses repositórios, expostas no Quadro 3.

QUADRO 3 – Características dos repositórios utilizados nas IFES.

SOFTWARE UTILIZADO PELAS IFES	CARACTERÍSTICAS DOS REPOSITÓRIOS DAS IFES
<i>DSPACE</i>	Fontes de informação: teses e dissertações, artigos de periódicos, trabalhos de conclusão de curso, entre outros.
	Políticas instituídas e disponibilizadas aos usuários
	Foco na comunidade
	Mecanismo de busca para pesquisas simples e avançadas
	<i>Download</i> dos arquivos
	Cumulativo e contínuo
	Conteúdo produzido na instituição
	<i>Software</i> de acesso aberto

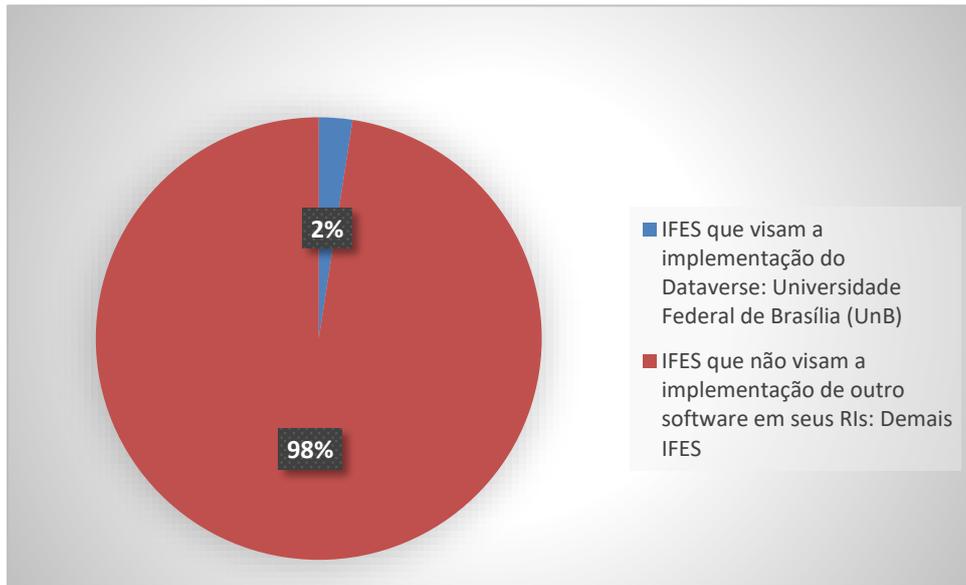
Fonte: Dados da pesquisa, 2017.

Apesar de 100% das IFES optarem por utilizar o *DSpace*, um dos *softwares* abertos mais utilizados para implementação de repositórios institucionais no mundo, as instituições não trabalham com um repositório específico para dados de pesquisa, como é o caso da Universidade de Harvard, que faz uso do *Dataverse*, um sofisticado repositório para todos os tipos de dados.

Os repositórios utilizados nas IFES não apresentam características que os definam repositórios de dados, como: quais os tipos de dados de pesquisa, qual o status desses dados e em que formato os arquivos de dados estão, que de acordo com Rice e Southall são essenciais em um repositório de dados. Os repositórios utilizados pelas IFES tratam quase que exclusivamente apenas de materiais como: Teses e Dissertações e trabalhos de conclusão de curso.

Verificou-se também se as IFES visam a implementação de um *software* de acesso aberto que trabalhe com dados brutos de pesquisa, constatando que apenas a Universidade Federal de Brasília (UnB) tem planos de implementar o *Dataverse* em seus repositórios, conforme Gráfico 3.

GRÁFICO 3 – IFES que visam a implementação de outro *software* em seus RIs.



Fonte: Dados da pesquisa, 2017.

De acordo com o IBICT, a diretoria da Universidade de Brasília (UnB), localizada na região Centro-Oeste, juntamente com diretores de centros de tecnologia, reuniram-se no dia 10 de abril de 2015 para discutir a ideia de implementação do *Dataverse* em seu repositório, mas até então, nada foi decidido. Se isso ocorrer, a UnB vai ser a primeira IFES a fazer uso de um *software* aberto que trabalha com dados brutos de pesquisa em seu RI.

CONSIDERAÇÕES FINAIS

Durante o percurso desta pesquisa foram analisados vários artigos de periódicos obtidos através do Google Acadêmico e no portal da BRAPCI, que tratam da *e-Science*, curadoria dos dados, implementação de repositórios com políticas de acesso aberto nas organizações, da importância de se ter um repositório de dados nas universidades, além da análise dos repositórios utilizados pelas Instituições Federais de Ensino Superior (IFES).

De fato, todos os objetivos específicos estabelecidos foram realizados para que o objetivo geral desta pesquisa fosse alcançado.

O primeiro objetivo específico propôs o levantamento dos repositórios eletrônicos das IFES, que resultou na localização de quarenta e uma instituições que utilizam repositórios. Dessas quarenta e uma IFES, trinta e nove são universidades e duas são institutos federais.

O segundo objetivo específico estabelecido visou identificar que tipo de repositório as IFES utilizam, sendo constatado que 100% das instituições localizadas utilizam Repositórios Institucionais (RIs) e optaram pelo uso do *DSpace* como ferramenta.

Verificou-se ainda, que do total de IFES que possuem RIs, 29% pertencem a região Nordeste, outros 29% pertencem a região Sudeste, 20% pertencem a região Sul, 12% pertencem a região Centro-Oeste e 10% pertencem a região Norte.

O terceiro objetivo específico requisitou as características dos repositórios utilizados pelas IFES. Foram apontadas características como: acesso aberto e fácil às fontes de informação produzidas nas próprias IFES ou em instituições externas, foco na comunidade, mecanismos de busca simples e avançada, conteúdos cumulativos e contínuos, download dos arquivos, personalização do repositório de acordo com as necessidades de cada IFES, e políticas institucionais.

No que diz respeito aos dados de pesquisa, não foram constatadas características nos repositórios utilizados pelas IFES que os definissem como um repositório de dados. Os repositórios utilizados pelas IFES tratam quase que exclusivamente apenas de materiais como: Teses e Dissertações e trabalhos de conclusão de curso.

Considerando todas essas informações, afirmamos que, atualmente, as IFES brasileiras não trabalham com repositórios que tratem todos os tipos de dados de pesquisa.

Segundo o portal do IBICT, a Universidade de Brasília (UnB), localizada na região Centro-Oeste, cujo repositório denomina-se RIUnB, é a única IFES que visa a implementação de um *software* de acesso aberto que trate todos os tipos de dados, como é o caso do *Dataverse*, utilizado pela Universidade de *Harvard*.

Com a realização dessa pesquisa, nota-se que há uma carência no que se trata a questão dos dados de pesquisa, uma vez que as IFES não possuem repositórios capacitados para o tratamento desses dados, e apenas uma das quarenta e uma IFES localizadas visam a implementação de um repositório que possibilite o tratamento de todos os tipos de dados de pesquisa.

Desde o surgimento da *e-Science*, a maneira como os dados são tratados estão em constante mudança graças as tecnologias digitais, e é preciso que as IFES estejam a par dessas mudanças para que avancem ainda mais no desenvolvimento de seus repositórios.

Sugere-se que, futuramente, novas pesquisas sejam realizadas em torno dos assuntos “*e-Science*,” “dados de pesquisa,” “curadoria dos dados,” “acesso aberto” e da implementação de repositórios que tratem todos os tipos de dados de pesquisa, a fim de constatar se as IFES tiveram algum avanço no que diz respeito ao tratamento destes dados em seus RIs.

REFERÊNCIAS

- BALL, A. **Review of Data Management Lifecycle Models**. University of Bath: REDm-MED, jan., 2012. Disponível em: <<http://opus.bath.ac.uk/28587/1/redm1rep120110ab10.pdf>>. Acesso em: 26 de abril de 2017.
- BLATTMANN, U; WEBER, C. DSpace como repositório digital na organização. **Revista ACB: Biblioteconomia em Santa Catarina**, Florianópolis, v.13, n.2, p.467-485, jul./dez., 2008. Disponível em: <<https://revista.acbsc.org.br/racb/article/view/593>>. Acesso em: 6 de maio de 2017.
- CUNHA, Murilo Bastos da. A biblioteca universitária na encruzilhada. **DataGramZero: Revista de Ciência da Informação**, v. 11, n. 6, dez. 2010. Disponível em: <<http://repositorio.unb.br/handle/10482/14869>>. Acesso em: 6 de abril de 2017.
- DADOS. In: **WIKIPÉDIA**, a enciclopédia livre. Flórida: Wikimedia Foundation, 2017. Disponível em: <<https://pt.wikipedia.org/w/index.php?title=Dados&oldid=48017724>>. Acesso em: 20 de abril de 2017
- DATADRYAD. Disponível em: <<http://datadryad.org/>>. Acesso em 9 de maio de 2017.
- DATAVERSE. Disponível em: <<http://dataverse.org/>>. Acesso em: 8 de maio de 2017.
- DSPACE. Disponível em: <<http://www.dspace.org/>>. Acesso em: 7 de maio de 2017.
- FACHIN, G. R. B. et al. Gestão do conhecimento e a visão cognitiva dos repositórios institucionais. **Perspectivas em Ciência da Informação**. Belo Horizonte, v.14, n.2, p. 220-236, 2009. Disponível em <<http://www.brapci.ufpr.br/brapci/index.php/article/view/0000006565/f8182342b56619efb73f91a24656a77d>>. Acesso em 5 de abril de 2017.
- FEDORA. Disponível em: <<http://fedorarepository.org/>>. Acesso em: 7 de maio de 2017.
- FREITAS, M. A. de; SILVA, P. N; GUIMARÃES, J. de F. Repositórios institucionais: a experiência da Universidade de Brasília. In: SAYÃO, L. et. al (Orgs). **Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação**. Salvador: EDUFBA, 2009. p. 333-344. Disponível em: <<https://repositorio.ufba.br/>>. Acesso em: 10 de maio de 2017.
- GAUZ, V. A alma da Internet e o acesso livre à informação científica. **Liinc em Revista**. v.4, n.2, setembro 2008, Rio de Janeiro, p. 274-285. Disponível em: <<http://revista.ibict.br/liinc/article/view/3142>>. Acesso em: 3 de maio de 2017.
- GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002. Disponível em: <https://professores.faccat.br/moodle/pluginfile.php/13410/mod_resource/content/1/Como_elaborar_projeto_de_pesquisa_-_antonio_carlos_gil.pdf>. Acesso em: 27 de abril de 2017.
- INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Manifesto brasileiro de apoio ao acesso livre à informação científica**. Disponível em: <<http://livroaberto.ibict.br/Manifesto.pdf>>. Acesso em: 5 de abril de 2017.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA.

Promovendo o uso do repositório de dados de pesquisa DATAVERSE: O Ibict e a UnB dão início à primeira rede de repositórios de pesquisa DATAVERSE no Brasil. Disponível em: <<http://cariniana.ibict.br/index.php/noticias/193-a-cariniana-promove-o-uso-de-repositorio-de-dados-de-pesquisa-dataverse>>. Acesso em: 8 de maio de 2017.

KURAMOTO, H. Informação científica: proposta de um novo modelo para o Brasil. **Ciência da Informação**. Brasília, v. 35, n. 2, p. 91-102, maio/ago. 2006. Disponível em: <http://www.scielo.br/scielo.php?pid=S0100-19652006000200010&script=sci_abstract&tlng=pt>. Acesso em: 3 de maio de 2017.

LEITE, F. C. L. **Como gerenciar e ampliar a visibilidade da informação científica brasileira:** repositórios institucionais de acesso aberto. Brasília: IBICT, 2009. 124 p. Disponível em: <<http://repositorio.unb.br/handle/10482/4841>>. Acesso em: 10 de maio.

LEITE, F. C. L. **Gestão do conhecimento científico no contexto acadêmico: proposta de um modelo conceitual.** Universidade de Brasília, 2006. 159 f. Dissertação (Mestrado em Ciência da Informação). Universidade de Brasília, 2006. Disponível em <<http://repositorio.unb.br/>>. Acesso em: 5 de abril de 2017.

LEITE, F. C. L. et al. **Boas práticas para a construção de repositórios institucionais da produção científica.** Brasília: IBICT, 2012. Disponível em: <<http://livroaberto.ibict.br/handle/1/703>>. Acesso em: 10 de maio.

MARCONDES, C. H.; SAYÃO, L. F. À guisa de introdução: repositórios institucionais e livre acesso. In: SAYÃO, L. F. et al (Orgs). **Implantação e gestão de repositórios institucionais:** políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p. 9-21. Disponível em <<https://repositorio.ufba.br/>>. Acesso em: 5 de abril de 2017.

MATIAS, M. S. O. **Base referencial para o povoamento de repositórios institucionais:** coleta automatizada de metadados da Plataforma Lattes. 2015. 94 f. Dissertação (Mestrado em Gestão de Organizações e Sistemas Públicos) – Universidade Federal de São Carlos, São Paulo, 2015. Disponível em: <<https://repositorio.ufscar.br/handle/ufscar/6932>>. Acesso em: 6 de abril de 2017.

MEDEIROS, J. S. CAREGNATO, S. E. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **Liinc em revista**. V. 8, n. 2, set. Rio de Janeiro. 2012, p. 311 – 322. Disponível em: <<http://www.brapci.ufpr.br/brapci/index.php/article/view/0000012500/96eedfc73fe3ce2c59ae795bf77c63e>>. Acesso em: 5 de abril de 2017.

MIRANDA, I. A. A; DELFINO, J. G. M. Repositórios Institucionais: novos desafios para as Bibliotecas e para os Bibliotecários. **Bibliotecas Universitárias:** pesquisas, experiências e perspectivas. v.3, n.1. Belo Horizonte, 2016. Disponível em: <<http://www.brapci.ufpr.br/brapci/index.php/article/view/0000021561/e291112a955a122a510e4550bdc2d73d>>. Acesso em: 3 de maio de 2017.

OLIVEIRA, R. R; CARVALHO, C. L. **Bibliotecas Digitais e o Repositório Fedora.** Goiás: INF-UFG, 2011. Disponível em: <http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_002-11.pdf>. Acesso em: 7 de maio de 2017.

RICE, R; SOUTHALL, J. **The data librarians: Handbook**. Londres: Facet Publishing, 2016. p. 113.

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**. v. 17, n. esp. 2. Rio de Janeiro, 2012. Disponível em: <<http://www.brapci.ufpr.br/brapci/index.php/article/view/0000012634/5444889e257059c68882ad8449e46598>>. Acesso em: 23 de abril de 2017.

SANTANA, R. C. G. Ciclo de vida dos dados e o papel da Ciência da Informação, 2013. In: XIV Encontro Nacional de Pesquisa em Ciência da Informação, 2013, Florianópolis. **Anais...** Florianópolis: XIV ENANCIB, 2013. Disponível em: <<http://enancib2013.ufsc.br/index.php/enancib2013/XIVenancib/paper/viewFile/284/319>>. Acesso em: 26 de abril de 2017.

SANTOS, E. A; MIRAGLIA. S. G. E. K. Arquivos abertos e instrumentos de gestão da qualidade como recursos para a disseminação da informação científica em segurança e saúde no trabalho. **Ciência da Informação**, Brasília, v. 38, n. 3, set./dez. 2009. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652009000300006>. Acesso em: 16 de maio de 2017.

SAYÃO, L. F. REPOSITÓRIOS DIGITAIS CONFIÁVEIS PARA A PRESERVAÇÃO DE PERIÓDICOS ELETRÔNICOS CIENTÍFICOS. **Ponto de Acesso**, Salvador, v.4, n.3, p. 68-94, dez 2010. Disponível em: <<http://www.brapci.ufpr.br/brapci/index.php/article/view/0000009600/26e6521a4c4f78d8a2db0b4fd844273f>>. Acesso em: 5 de maio de 2017.

SAYÃO, L. F.; MARCONDES, C. H. Software livres para repositórios institucionais: alguns subsídios para a seleção. In: SAYÃO, L. et. al (Orgs). **Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação**. Salvador: EDUFBA, 2009. p. 23 – 54. Disponível em: < <https://repositorio.ufba.br/>>. Acesso em: 15 de maio de 2017.

SAYÃO, L. F; SALES, F. S. **Dados abertos de pesquisa**: ampliando o conceito de acesso livre. Rio de Janeiro: CNEN/IEN, 2014. Disponível em: < <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/viewFile/611/1252>>. Acesso em: 20 de abril de 2017.

SAYÃO, L. F; SALES, L. F. DADOS DE PESQUISA: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Tendências da Pesquisa Brasileira em Ciência da Informação**. v.6, n.1. Rio de Janeiro, 2013. Disponível em: <<http://www.brapci.ufpr.br/brapci/index.php/article/view/0000014157/0f3127bfd6475f45e6950e1e0da20371>>. Acesso em: 23 de abril de 2017.

SAYÃO, L. F; SALES, F. S. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. Disponível em: <http://carpedien.ien.gov.br/bitstream/ien/1624/1/GUIA_DE_DADOS_DE_PESQUISAS.pdf>. Acesso em: 6 de abril de 2017.

SETZER, V.W. **Dado, informação, conhecimento e competência**. 2015. Disponível em: <<https://www.ime.usp.br/~vwsetzer/dado-info.html>>. Acesso em: 20 de abril de 2017.

SOUZA FILHO, A. L. et al. **Importância dos repositórios institucionais na preservação intelectual: em foco a gestão do conhecimento**. Múltiplos Olhares em Ciência da Informação. v.2, n.2. Belo Horizonte, 2012. Disponível em: <<http://periodicos.ufpb.br/ojs/index.php/pbcib/article/view/26900>>. Acesso em: 3 de maio de 2017.

STRASSER, C. et al. **Primer on Data Management: What you always wanted to know**. California: CDL, 2012. Disponível em: <<http://escholarship.org/uc/item/7tf5q7n3#page-1>>. Acesso em: 25 de abril de 2017.

VECHIATO, F. L. **Repositório digital como ambiente de inclusão digital e social para usuários idosos**. Marília, 2010. Disponível em: <https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/vechiato_fl_me_mar.pdf>. Acesso em: 5 de abril de 2017.

WINN, J. **Open data and the academy: an evaluation of CKAN for research data management**. Cologne: IASSIST, 2013. Disponível em: <<http://eprints.lincoln.ac.uk/9778/>>. Acesso em: 9 de maio de 2017.