



**UNIVERSIDADE DO ESTADO DA BAHIA
PROGRAMA DE PÓS-GRADUAÇÃO GESTÃO E TECNOLOGIAS
APLICADAS À EDUCAÇÃO
ÁREA DE CONCENTRAÇÃO 1 – GESTÃO DA EDUCAÇÃO E REDES SOCIAIS**

NEUBLER NILO RIBEIRO DA CUNHA

**A GESTÃO DE DADOS CIENTÍFICOS:
Uma proposta para o repositório institucional Acesso Aberto da
Universidade do Estado da Bahia (UNEB)**

Salvador
2021

**UNIVERSIDADE DO ESTADO DA BAHIA
PROGRAMA DE PÓS-GRADUAÇÃO GESTÃO E TECNOLOGIAS
APLICADAS À EDUCAÇÃO**

NEUBLER NILO RIBEIRO DA CUNHA

**A GESTÃO DE DADOS CIENTÍFICOS:
Uma proposta para o repositório institucional Acesso Aberto da
Universidade do Estado da Bahia (UNEB)**

Dissertação apresentada ao Programa de Pós-Graduação Gestão e Tecnologias Aplicadas à Educação (GESTEC), modalidade Pós-Graduação Profissional/Capes, pela Universidade do Estado da Bahia (UNEB), como requisito parcial para obtenção do título de Mestre em Educação.
Área de concentração: GESTÃO DA EDUCAÇÃO E REDES SOCIAIS

Orientadora: Prof^ª. Dr^ª Nadia Hage Fialho
Co-Orientador: Prof. Dr. Eduardo Manuel de Freitas Jorge

Salvador
2021

FICHA CATALOGRÁFICA
Sistema de Bibliotecas da UNEB

C972g

Cunha, Neubler Nilo Ribeiro da

A Gestão de dados científicos: uma proposta para o repositório institucional Acesso Aberto da Universidade do Estado da Bahia (UNEB) / Neubler Nilo Ribeiro da Cunha. - Salvador, 2021.

122 fls : il.

Orientador(a): Profª Dra. Nadia Hage Fialho.

Coorientador(a): Profº Dr. Eduardo Manuel de Freitas Jorge.

Inclui Referências

Dissertação (Mestrado Profissional) - Universidade do Estado da Bahia. Departamento de Educação. Programa de Pós-Graduação em Gestão e Tecnologias Aplicadas à Educação - GESTEC, Campus I. 2021.

1..Ciência da informação. 2.Repositórios institucionais. 3..Sistemas de recuperação da informação. 4.Dados de pesquisa. 5.Produção científica.

CDD: 025



UNIVERSIDADE DO ESTADO DA BAHIA

Autorização Decreto nº 9237/86. DOU 18/07/96. Reconhecimento: Portaria 909/95, DOU 01/08-95



Gestão e Tecnologias Aplicadas à Educação

ATA

Programa de Pós-Graduação em Gestão e Tecnologias Aplicadas à Educação - GESTEC SESSÃO DE APRESENTAÇÃO DE DEFESA PÚBLICA

Em 16 de março de 2021, às 09 hora(s), no(a) Webconferência - 03 do Departamento de Educação, Campus I da Universidade do Estado da Bahia - UNEB, realizou-se a Sessão Pública de Defesa da Dissertação, do(a) Mestrando(a) NEUBLER NILO RIBEIRO DA CUNHA, intitulada: "A GESTÃO DE DADOS CIENTÍFICOS: UMA PROPOSTA PARA O REPOSITÓRIO INSTITUCIONAL ACESSO ABERTO DA UNIVERSIDADE DO ESTADO DA BAHIA (UNEB)". O(a) Professor(a) Dr(a) NADIA HAGE FIALHO, orientador(a) e Presidente da Banca Examinadora, iniciou a sessão apresentando os demais examinadores: Prof.(-) Dr.(-) EDUARDO MANUEL DE FREITAS JORGE - UNEB, coorientador, Prof.(a) Dr.(a) SERGIO HENRIQUE DA CONCEICAO, membro interno e Prof.(a) Dr.(a) IVANA APARECIDA BORGES LINS - UFBA, membro externo. A sessão teve a duração de 3 horas e, após a exposição do trabalho e arguição do(a) discente a Banca emitiu o seguinte parecer:

CONCEITO MUITO BOM (9,0)	APROVADO (X)	APROVADO COM RESTRIÇÃO ()	REPROVADO ()
NOTA/PARECER			
Aprovado, sem restrições.			

Para lavrar a presente Ata, eu, Prof. (-) Dr.(-) NADIA HAGE FIALHO, presidente da Banca, encerro a presente ATA, que vai assinada por mim e pelos demais membros da Banca.

DEDICATÓRIA

A

Sebastião Nilo da Cunha, pai, lutador, trabalhador que veio da Roça de Caetité-BA no Centro-Sul Baiano (*in memoriam*).

Nemisia Ribeiro da Cunha, mãe, mulher predestinada e guerreira, nascida no município de Ipupiara-BA no Centro-Sul Baiano (*in memoriam*).

Barbara Coelho Neves, minha esposa, uma mulher Soteropolitana que é uma guerreira, companheira de uma vida desde 2003, e que está nesta luta diária comigo sempre um ajudando e apoiando o outro cada um a sua maneira.

AGRADECIMENTOS

Agradeço a Rosemilia Ribeiro da Cunha, minha irmã pelo apoio e por estar na minha vida desde sempre.

A Nadia Hage Fialho, orientadora sempre tão atenciosa, compreensiva e receptiva, acima de tudo, uma mestre.

A Eduardo Manuel de Freitas Jorge, co-orientador por todo apoio desde o início.

Ao Programa de Pós-Graduação Gestão e Tecnologias Aplicadas à Educação pela oportunidade recebida em cursar o mestrado, pelo apoio, a infra-estrutura, a qualidade e a simpatia dos seus professores, pesquisadores e funcionários.

EPÍGRAFE

“Ensinar não é transferir conhecimento, mas criar as possibilidades para a sua própria produção ou a sua construção”.

Paulo Freire (1996, p.47)

RESUMO

Introdução: A gestão de dados científicos foi a preocupação deste estudo dissertativo na área de Educação. Embora o Brasil tenha apresentado significativa melhoria em relação ao quantitativo de implementação de repositórios digitais em suas instituições de pesquisa, percebe-se que ainda há muito o que se considerar no tocante à gestão dos dados científicos visando diretrizes específicas e parâmetros contemporâneos voltados para universidades, sobretudo quanto ao fomento e critérios de internacionalização para os próximos anos. **Objetivos:** Essa dissertação propõe identificar e descrever os elementos estruturantes de um Plano de Gestão de Dados Científicos, com perspectiva de aplicabilidade em uma universidade pública. Como objetivos específicos, este estudo dissertativo pretende: a) apresentar os principais conceitos sobre a gestão de dados científicos, a partir dos contextos terminológicos e históricos que deram origem às novas práticas de pesquisa, na contemporaneidade; b) descrever os Repositórios de Instituições no Brasil que utilizam Planos de Gestão de Dados Científicos e disponibilizam seus dados; c) identificar mecanismos que apontem elementos do plano de gestão de dados científicos nas universidades brasileiras; d) propor um modelo de plano de gestão de dados científicos para o Repositório Institucional SaberAberto, alinhado às características acadêmicas, científicas e organizacionais da UNEB. **Metodologia:** Empreendeu-se um estudo descritivo, com abordagem qualitativa, de tipo bibliográfica, documental e de campo para atender adequadamente aos objetivos. **Resultados:** Dentre os resultados, destacaram-se definições, conceitos e o processo de gerência desses dados, considerando a ciência como prática social; a importância dos repositórios de acesso aberto, dando ênfase ao RE3DATA como estrutura mundial que possui pré-requisitos para a gestão de dados científicos; informações relevantes, exemplificando os repositórios no Brasil e no mundo que disponibilizam dados de pesquisa; e, finalmente, como produto, a proposição de um modelo de gestão de dados científicos para o Repositório SaberAberto da UNEB. **Considerações:** O estudo conclui destacando a importância do Repositório da UNEB implementar um plano de gestão de dados científicos visando se adequar aos critérios internacionais e ao movimento de acesso aberto tão relevante para os processos e novas metas de internacionalização das universidades públicas.

Palavras-chave: Gestão de Dados Científicos. Repositórios Institucionais. Dados de Pesquisa. Produção Científica. Educação.

ABSTRACT

Introduction: The management of scientific data was the concern of this dissertation study in the area of Education. Although Brazil has shown significant improvement in terms of the amount of implementation of digital repositories in its research institutions, it is clear that there is still a lot to consider regarding the management of scientific data with a view to specific guidelines and contemporary parameters aimed at universities, especially regarding the promotion and internationalization criteria for the coming years. **Objectives:** This dissertation proposes to identify and describe the structuring elements of a Scientific Data Management Plan, with a perspective of applicability in a public university. As specific objectives, this dissertation study intends to: a) present the main concepts about the management of scientific data, from the terminological and historical contexts that gave rise to the new research practices, in contemporary times; b) describe the Institutional Repositories in Brazil that use Scientific Data Management Plans and make their data available; c) to identify mechanisms that point out elements of the scientific data management plan in Brazilian universities; d) propose a model of a scientific data management plan for the Institutional Repository SaberAberto, aligned with the academic, scientific and organizational characteristics of UNEB. **Methodology:** A descriptive study was undertaken, with a qualitative approach, of a bibliographic, documentary and field type to adequately meet the objectives. **Results:** Among the results, definitions, concepts and the management process of these data were highlighted, considering science as a social practice; the importance of open access repositories, emphasizing RE3DATA as a global structure that has prerequisites for the management of scientific data; relevant information, exemplifying the repositories in Brazil and in the world that provide research data; and, finally, as a product, the proposal for a scientific data management model for the UNEB SaberAberto Repository. **Considerations:** The study concludes by highlighting the importance of the UNEB Repository to implement a scientific data management plan aimed at adapting to international criteria and the open access movement so relevant to the processes and new internationalization goals of public universities.

Keywords: Scientific Data Management. Institutional Repositories. Search data. Scientific production. Education.

LISTA DE FIGURAS

	f.
FIGURA 1 CURADORIA DOS DADOS CIENTÍFICOS	52
FIGURA 2 ESTRUTURA DO DOI	58
FIGURA 3 MODELO DO CICLO DE VIDA DOS DADOS DE PESQUISA	72
FIGURA 4 INICIATIVAS DE ACESSO ABERTO NO BRASIL	86
FIGURA 5 SITUAÇÃO DAS REVISTAS DE ACESSO ABERTO BRASILEIRAS	87
FIGURA 6 PRINCIPAIS DOCUMENTOS DISPONIBILIZADOS E INSTITUIÇÕES NO TOP 10 DO OASISBR	89
FIGURA 7 UNIVERSIDADE DO ESTADO DA BAHIA. CAMPUS 1- SALVADOR-BAHIA	97
FIGURA 8 ASPECTOS PRINCIPAIS DO PROCESSO DE GESTÃO DE DADOS CIENTÍFICOS APONTADOS NESTE TRABALHO	98
FIGURA 9 ETAPAS CICLO DE VIDA DOS DADOS DE PESQUISA PAA O SABER ABERTO	100
FIGURA 10 DADOS DE PESQUISA PRODUZIDOS POTENCIALMENTE EM UMA DISSERTAÇÃO	101
FIGURA 11 <i>FRAMEWORK</i> COM DIRETRIZES PARA A ELABORAÇÃO DE UMA POLÍTICA DE GESTÃO DE DADOS CIENTÍFICOS	104
FIGURA 12 REPOSITÓRIO INSTITUCIONAL SABER ABERTO (UNEB) EM DSPACE	105
FIGURA 13 TELA DE VISUALIZAÇÃO DOS DADOS CIENTÍFICOS NO RI SABER ABERTO	107

LISTA DE GRÁFICOS

		f.
GRÁFICO 1	REPOSITÓRIOS DIGITAIS POR PAÍS	55
GRÁFICO 2	TIPOS DE REPOSITÓRIOS REGISTRADOS NO RE3DATA.ORG	58
GRÁFICO 3	SOFTWARE ACESSO LIVRE MAIS UTILIZADOS NO MUNDO	59
GRÁFICO 4	REPOSITÓRIOS POR PAÍS	60
GRÁFICO 5	IDIOMAS	61
GRÁFICO 6	TEMÁTICA DOS ASSUNTOS DOS REPOSITÓRIOS	62
GRÁFICO 7	EVOLUÇÃO DO NÚMERO DE REPOSITÓRIOS	63
GRÁFICO 8	REPOSITÓRIOS DE DADOS DE PESQUISA POR PAÍS	73
GRÁFICO 9	SOFTWARE DE ACESSO LIVRE MAIS UTILIZADOS NO BRASIL	77
GRÁFICO 10	IDIOMAS DOS REPOSITÓRIOS BRASILEIROS	78
GRÁFICO 11	TEMÁTICA DOS ASSUNTOS NOS REPOSITÓRIOS BRASILEIROS	79
GRÁFICO 12	EVOLUÇÃO DO NÚMERO DE REPOSITÓRIOS NO BRASIL	80

LISTA DE QUADROS

		f.
QUADRO 1	TIPOS DE DADOS DE PESQUISA	42
QUADRO 2	MARCOS DO MOVIMENTO DE ACESSO LIVRE	45
QUADRO 3	EVOLUÇÃO DO ACESSO ABERTO – BREVE HISTÓRICO	46
QUADRO 4	DEZ RAZÕES PARA A CRIAÇÃO DE UM PLANO DE GESTÃO DE DADOS DE PESQUISA	71
QUADRO 5	TESES E DISSERTAÇÕES DEFENDIDAS NO BRASIL SOBRE REPOSITÓRIOS DE ARMAZENAMENTO DE DADOS CIENTÍFICOS	93
QUADRO 6	PROPOSTA DO FORMULÁRIO SABER ABERTO/UNEB PARA ARMAZENAMENTO DOS DADOS DE PESQUISA	106

LISTA DE ABREVIATURAS E SIGLAS

ABNT	ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS
AGUIA	AGÊNCIA DE GESTÃO DA INFORMAÇÃO ACADÊMICA
ALPSP	ASSOCIATION OF LEARNED AND PROFESSIONAL SOCIETY PUBLISHERS
BDTD	BIBLIOTECA DIGITAL BRASILEIRA DE TESES E DISSERTAÇÕES
BIREME	CENTRO LATINO-AMERICANO DE INFORMAÇÃO EM CIÊNCIAS DA SAÚDE
CAPE	COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR
CARINIANA	REDE CARINIANA
CNN	CATÁLOGO COLETIVO NACIONAL DE PUBLICAÇÕES SERIADAS
COMUT	PROGRAMA DE COMUTAÇÃO BIBLIOGRÁFICA
CONFOA	CONFERÊNCIA LUSO-BRASILEIRA SOBRE ACESSO ABERTO
CT	CATÁLOGO DE TESES
DCC	DIGITAL CURATION CENTRE
DIADORIM	DIRETÓRIO DE POLÍTICAS DE ACESSO ABERTO DAS REVISTAS CIENTÍFICAS BRASILEIRAS
DOAJ	DIRECTORY OF OPEN ACCESS JOURNALS
DOI	DIGITAL OBJECT IDENTIFIER
DSP	DADOS SUPLEMENTARES DE PESQUISA
EAD	EDUCAÇÃO A DISTÂNCIA
EBLIDA	EUROPEAN BUREAU OF LIBRARY, INFORMATION AND DOCUMENTATION ASSOCIATIONS
ETDS	ELECTRONIC THESES AND DISSERTATION
FAPESP	FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO
IBICT	INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA
IES	INSTITUIÇÃO DE ENSINO SUPERIOR

INPA	INSTITUTO NACIONAL DE PESQUISAS DA AMAZÔNIA
ISSN	INTERNATIONAL STANDARD SERIAL NUMBER
LATINDEX	LÍNEA PARA REVISTAS CIENTÍFICAS DE AMÉRICA LATINA, EL CARIBE, ESPAÑA Y PORTUGAL
MAST	MUSEU DE ASTRONOMIA E CIÊNCIAS AFINS
NCSTR	NETWORKED COMPUTER SCIENCE TECHNICAL REFERENCE LIBRARY
NDLTD	NETWORKED DIGITAL LIBRARY OF THESES AND DISSERTATIONS
NPGA	NÚCLEO DE PÓS GRADUAÇÃO EM ADMINISTRAÇÃO
OA	OPEN ARCHIVES
OAI	OPEN ARCHIVES INITIATIVE
OAIS	OPEN ARCHIVAL INFORMATION SYSTEM
OASISBR	PORTAL BRASILEIRO DE ACESSO ABERTO À INFORMAÇÃO CIENTÍFICA
OCDE	ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO
OJS	OPEN JOURNAL SYSTEM
PGD	PLANO DE GESTÃO DE DADOS DE PESQUISA
PGD	PLANO DE GESTÃO DE DADOS
PGDC	PLANO DE GESTÃO DE DADOS CIENTÍFICOS
PLOS	PUBLIC LIBRARY OF SCIENCE
PMC	PUBMED CENTRAL
PUSAI	POLO UNIVERSITÁRIO SANTO AMARO DE IPITANGA
RA/SBPC	REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA PARA O PROGRESSO DA CIÊNCIA
RDM	RESEARCH DATA MANAGEMENT
RDS	REPOSITÓRIO DIGITAL
REPEC	RESEARCH PAPERS IN ECONOMICS
RI	REPOSITÓRIO INSTITUCIONAL

RI/SABERABERTO	REPOSITÓRIO INSTITUCIONAL SABERABERTO
RIDI	REPOSITÓRIO INSTITUCIONAL DIGITAL DO IBICT
ROAR	REGISTRY OF OPEN ACCESS REPOSITORIES
SAAI	SISTEMA ABERTO DE ARQUIVAMENTO DE INFORMAÇÃO
SBPC	SOCIEDADE BRASILEIRA PARA O PROGRESSO DA CIÊNCIA
SCIELO	SCIENTIFIC ELECTRONIC LIBRARY ONLINE
SEER	SISTEMA ELETRÔNICO DE EDITORAÇÃO DE REVISTAS
SISB	SISTEMA DE BIBLIOTECAS
SL	SOFTWARE LIVRE
SMSI	CÚPULA MUNDIAL SOBRE A SOCIEDADE DA INFORMAÇÃO
SSRN	SOCIAL SCIENCE RESEARCH NETWORK
UEMA	UNIVERSIDADE ESTADUAL DO MARANHÃO
UFBA	UNIVERSIDADE FEDERAL DA BAHIA
UFPB	UNIVERSIDADE FEDERAL DA PARAÍBA
UFSM	UNIVERSIDADE FEDERAL DE SANTA MARIA
UNICAMP	UNIVERSIDADE DE CAMPINAS
UNIPAMPA	UNIVERSIDADE FEDERAL DOS PAMPAS
USP	UNIVERSIDADE ESTADUAL DE SÃO PAULO
WWW	WORLD WIDE WEB
WEB	WORLD WIDE WEB

SUMÁRIO

1	INTRODUÇÃO	16
2	METODOLOGIA	24
2.1	CARACTERIZAÇÃO	25
2.2	REFERÊNCIAS CONCEITUAIS E EMPÍRICA	29
2.3	ETAPAS	31
3	GESTÃO DE DADOS CIENTÍFICOS	33
3.1	REFERENCIAL TERMINOLÓGICO-CONCEITUAL	34
3.2	MOVIMENTO ACESSO ABERTO E SURGIMENTO DOS REPOSITÓRIOS INSTITUCIONAIS	43
4	REPOSITÓRIOS INSTITUCIONAIS COMO FERRAMENTA DE APOIO AO GERENCIAMENTO DE DADOS CIENTÍFICOS	64
4.1	REPOSITÓRIOS INSTITUCIONAIS NO MUNDO QUE DISPONIBILIZAM ACESSO AOS DADOS DA PESQUISA	64
4.2	REPOSITÓRIOS INSTITUCIONAIS NO BRASIL QUE DISPONIBILIZAM ACESSO AOS DADOS DA PESQUISA	73
5	PLANO DE GESTÃO DE DADOS CIENTÍFICOS	94
5.1	GESTÃO DE DADOS CIENTÍFICOS: PROBLEMÁTICAS E DESAFIOS	94
5.2	PLANO DE GESTÃO DE DADOS CIENTÍFICOS: RI/UNEB SABERABERTO	96
6	CONSIDERAÇÕES FINAIS	109
	REFERÊNCIAS	113

1 INTRODUÇÃO

Os dados científicos estão assumindo um novo papel no processo de investigação científica. Tradicionalmente os dados gerados ao longo do processo de pesquisa são descartados ou esquecidos após a conclusão da investigação e da consequente publicação resultante do trabalho (relatórios de pesquisa, artigos, capítulos de livros etc.). Este cenário está mudando, os dados demonstram ser uma fonte primária valiosíssima de recursos científicos que pode ser utilizada por outros pesquisadores na análise de fenômenos distintos e na replicação de experimentos já realizados.

No âmbito acadêmico, a quantidade de dados de pesquisa dispersos em consequência do seu não aproveitamento é um desafio para o gerenciamento das pesquisas e para o reuso em outras pesquisas. Os dados coletados e utilizados em uma pesquisa passam a ser tão importantes quanto a própria publicação da produção científica, em, por exemplo, artigos, livros, teses, dissertações e trabalhos em eventos científicos (NAVES, 1998; VIDOTTI, 2017).

No Brasil, o reconhecimento das universidades públicas como centros de produção do conhecimento é inquestionável: nossas universidades públicas respondem por mais de 95% da produção científica no País (MOURA, 2019). O tema é também objeto de muitos estudiosos que reafirmam, por essa e outras razões, tal reconhecimento (TRINDADE, PRIGENZI, 2002; COSTA, 2008; BORTOLIERO, 2009; CHIARINI, VIEIRA, 2012; FAUSTO, 2015; MULLER, 2015; OLIVEIRA, MORAES, 2016, VANZ et al, 2018; entre outros).

Entretanto, de acordo com Vans et al (2018), para elevar as universidades públicas brasileiras ao patamar das melhores *research-universities* do mundo e inseri-las em políticas de internacionalização, é imperioso empreender ações mais incisivas na resolução da produção científica, a exemplo das questões levantadas neste trabalho sobre a gestão de dados científicos. Como instituições públicas, elas prestam um grande serviço à sociedade, pois é através delas que muitos profissionais são formados, realizam pesquisas científicas que posteriormente retornam para sociedade em forma de serviços, inovações, tecnologias, produtos, dentre outros (BRAGA, 2019; MOURA, 2019).

Weber (2009), uma das maiores especialistas sobre a universidade brasileira, ao refletir, em 2009, sobre marcas históricas que ainda se fazem presentes na educação superior no nosso País, já advertia:

[...] o desafio que representa a aceitação de um modelo misto de universidade que preserve o seu caráter de instituição social capaz de associar as demandas das ciências àquelas provindas do mundo social, cultural, econômico, político e tecnológico da sociedade brasileira (WEBER, 2009, p. 121).

E, mais ainda, a autora dizia não ser possível “[...] omitir o fato de que tem sido preocupação constante, desde o início do século XX, o problema de articular produção de conhecimento e formação superior ao desenvolvimento do país” (WEBER, 2009, p. 134). E, frente a esse panorama, colocou a seguinte questão:

Não seria o caso de aceitar um modelo misto de universidade no qual fosse preservado o seu caráter de instituição social capaz de associar as demandas das ciências àquelas provindas do mundo social e econômico, portanto, comprometida com os presentes desafios da sociedade brasileira? (WEBER, 2009, p. 134).

Martinelli, Corrêa e Schuch Junior (2017), em trabalho apresentado no AVALIES em Santa Catarina, reuniram estudos até então realizados sobre a gestão do conhecimento em universidades brasileiras, revelando um panorama dessa situação, no qual destacamos:

Considerando o fato de a universidade ser o principal local onde são desenvolvidos conhecimentos, é necessário criar neste ambiente modelos de gestão que contribuam para a eficiência, eficácia, efetividade e relevância da produção e disseminação do conhecimento. Leite e Costa (2006) alertam que as iniciativas de estudos ou criação de modelos de gestão do conhecimento, que levem em conta as particularidades do ambiente das universidades, ainda são escassas. Para se criar um modelo de gestão do conhecimento nestas instituições é preciso observar suas especificidades, expressas nos processos de comunicação científica, na natureza da produção do conhecimento científico, na estrutura e comportamento de comunidades científicas e na cultura que permeia o ambiente acadêmico.

Além disso, é importante registrar que a plataforma **Scientific Electronic Library Online** (SciELO), que é compreendida como uma rede, publica um *blog* específico sobre dados abertos denominado *Blog SciELO em Perspectiva*¹.

Esse tema faz parte da nossa trajetória profissional, pois tivemos a oportunidade de trabalhar como parte da equipe de responsáveis pela implementação de repositórios digitais em instituições de ensino superior (IES). Como profissional da informação, esses temas sempre despertaram nosso interesse. A atuação como bibliotecário no Centro Interdisciplinar de Desenvolvimento e Gestão Social (CIAGS)², ligado ao Núcleo de Pós-Graduação em Administração (NPGA), da Escola de Administração, da Universidade Federal da Bahia (UFBA), sob o comando da professora Tânia Fischer, rendeu experiências com plataformas que trabalhavam com dados abertos, a exemplo da revista científica do CIAGS, em *Open Journal* e a Biblioteca Virtual de Gestão Social.

Um dos campos de atuação para o bibliotecário é o apoio à pesquisa, às publicações e à preservação. Nesse contexto se inserem oportunidades de atuação no campo da gestão de dados científicos, como têm apontado os *sites* e *blogs* da área de Biblioteconomia e Gestão da Informação. Exemplos como o Portal Bibliotecário³ e o BiblioVagas⁴ destacam esse campo como uma das possibilidades promissoras para a trajetória e futuro da carreira. Também Silva (2017) chama atenção para a importância e expertise do profissional bibliotecário no momento da elaboração e implementação dos planos de gestão de dados científicos.

Desse modo, a gestão de dados científicos começou a chamar nossa atenção, como campo de investigação, a partir de 6 de maio de 2017, quando a revista *The Economist*⁵ publicou uma capa apontando o dado como o novo petróleo. Entretanto, dados científicos são abundantes no contexto dos campos da ciência e, a rigor, sempre estiveram disponíveis; hoje, ganham notoriedade pela possibilidade de serem tratados com o apoio de tecnologias da informação e comunicação (TIC) e de tecnologias avançadas, a exemplo de modelos de inteligência artificial (IA).

¹ *blog SciELO em Perspectiva*. <https://blog.scielo.org/blog/tag/dados-abertos/#.Xu1Na2hKjIU>

² Informações sobre o CIAGS: <http://www.adm.ufba.br/pt-br/nucleo/centro-interdisciplinar-desenvolvimento-gestao-social-ciags>

³ Matéria sobre Curadoria Digital no Portal do Bibliotecário aponta a tendência do interesse pelo bibliotecário pela gestão de dados científicos. [http://portaldobibliotecario.com/ciencia-da-informacao/curadoria-digital-uma-introducao/;](http://portaldobibliotecario.com/ciencia-da-informacao/curadoria-digital-uma-introducao/)

⁴ Site tem mostrado as tendências das vagas em> [https://www.bibliovagas.com.br/.](https://www.bibliovagas.com.br/)

⁵ A matéria completa está disponível no site da revista: [https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.](https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data)

O gerenciamento e o armazenamento de grandes volumes de dados gerados em pesquisas, já há algum tempo chamam a atenção da comunidade científica de todas as áreas do conhecimento e desafiam instituições de fomento pelo mundo. Como disseminar dados brutos que foram coletados com recursos públicos de pesquisa, que não foram publicados, mas que podem ser essenciais para outros pesquisadores não precisarem ‘inventar a roda’?

Como se observa, a não disponibilidade dos dados de pesquisa compromete o avanço da ciência, a interação entre pesquisadores, a constituição e a consolidação de redes de pesquisa, entre muitos outros problemas; por exemplo, se os dados de pesquisa ficam restritos ao pesquisador ou grupo de pesquisa, dificilmente se chega a uma melhor compreensão até do caminho trilhado para a realização da pesquisa.

As questões em torno da problemática da gestão de dados científicos tiveram início na primeira década de 2000, quando começaram os primeiros movimentos e iniciativas para fazer com que os pesquisadores disponibilizassem seus dados de pesquisa sob o crivo de financiamento, deixando-os sob a gestão da respectiva agência financiadora. A primeira instituição de fomento a exigir que seus pesquisadores fornecessem, junto com sua proposta de pesquisa, um plano de gestão de dados científico foi a *National Science Foundation* (NSF) nos Estados Unidos (EUA), em 2011. Em seguida, no continente europeu, a agência de fomento, denominada *Economic and Social Research Council*, do Reino Unido, em 2012, também passou a exigir um plano que descrevesse como os dados produzidos pelos estudos financiados seriam preservados e divulgados em repositórios públicos (PIERRO, 2018).

A expressão *gestão de dados científicos* é um assunto complexo que envolve termos diversos como acesso livre, acesso aberto à informação, dados de pesquisa, dados científicos, dados suplementares, via verde, via dourado, via diamante, via híbrida⁶, literatura cinzenta, ciclo de dados, *pre-print* e *post-print*⁷, entre outros. Vale

⁶ O Repositório AGUIA caracteriza, por exemplo, **via verde** (Green Road Open Access) para artigo publicado em qualquer revista; nessa perspectiva, o retém os direitos autorais e tem permissão para colocar cópias do artigo (às vezes um *pre-print* ou um *post-print*) em um repositório ou em seu próprio *site*; **via dourada** (Gold Road Open Access) para artigo publicado em uma revista de Acesso Aberto, disponível gratuitamente na web e o autor paga uma taxa de publicação do artigo; são permitidas cópias do artigo arquivadas em outro lugar; não há cobranças para leitores ou bibliotecas, mas existem taxas pelo autor; **via diamante** (Diamond Road Open Access): autor publica artigo em revista de Acesso Aberto, sem cobrança de taxas de autor; são revistas mantidas por voluntários, contribuições e doações de membros, a exemplo de revistas de associações ou sociedades

salientar que tais terminologias se apresentam com definições ora distintas, ora convergentes, mas que se interconectam e são transversais à ciência e à tecnologia. Nesse contexto, este trabalho fará referências às terminologias conforme seus respectivos autores. Contudo, considerando as variadas nomenclaturas para o tema em questão, optou-se pela utilização da expressão ‘gestão de dados científicos’ como a mais adequada para este trabalho e entendendo que é como a maioria de outros estudos têm consolidado o assunto no Brasil.

Atualmente, no cenário internacional há um debate sobre a regulamentação e institucionalização de planos que contemplem a gestão dos dados científicos, como exigência prévia em editais de fomento visando garantir o compromisso do pesquisador na disponibilização desses conteúdos e desses experimentos para uso de outros pesquisadores. Países como Estados Unidos (EUA), Alemanha, Reino Unido e Austrália estão na vanguarda desse processo já há algum tempo. Esses países passaram a solicitar em seus editais a obrigatoriedade de informar como será a gestão e disseminação dos dados científicos.

No Brasil, a Universidade de Campinas (Unicamp) foi a primeira a ter iniciativas em direção a práticas relacionadas à gestão de dados científicos, se espelhando nos exemplos americano e europeu. De acordo com Pierro (2018), a Unicamp foi pioneira na criação de formulários para planos de gestão⁸. Contudo, vale salientar que este procedimento no País é recente e, ainda hoje, não é uma prática comum das agências e instituições de fomento e recursos para a pesquisa.

A principal problemática que chama atenção neste estudo destaca o seguinte exemplo: imagine um pesquisador brasileiro que viaja à Itália para realizar estudos em um laboratório de ponta especializado em substâncias para uma vacina. Enquanto esteve naquele espaço, o pesquisador realizou uma série de testes e amostras que geraram um volume significativo de dados não estruturados e que não

profissionais, organizações sem fins lucrativos, universidades ou agências governamentais. Seu modelo econômico depende de contribuições e doações de membros; **via híbrida** (Hybrid Road Open Access): “parcialmente financiadas por assinaturas, as revistas fornecem apenas acesso aberto para alguns artigos individuais para os quais os autores (ou patrocinador da pesquisa) pagaram uma taxa de publicação”. Disponível em: < <https://www.aguia.usp.br/apoio-pesquisador/acesso-aberto-usp/entenda-o-que-e-acesso-aberto/> >.

⁷ Conforme Repositório AGUIA, **pre-print** corresponde a versão não revisada por pares ou não editada do artigo e **post-print** a versão revisada por pares do artigo, mas não formatado para publicação. Disponível em: < <https://www.aguia.usp.br/apoio-pesquisador/acesso-aberto-usp/entenda-o-que-e-acesso-aberto/> >.

⁸ Os planos de gestão de dados científicos da UNICAMP podem ser acessados na plataforma DMPToom <https://dmptool.org/>.

foram publicados nos canais científicos, mas que contribuíram para um achado que foi publicado em uma revista científica internacional. Então, a questão que surge é: o que foi feito com os dados do estudo? Outros pesquisadores terão acesso a eles para realizarem novos estudos sem precisar começar do início? Se a pesquisa foi realizada com recursos de agência de fomento, existe alguma exigência para a gestão do ciclo desses dados?

A ideia desta dissertação surgiu a partir das seguintes questões: a) como o Brasil se posiciona a respeito da gestão de dados científicos; b) os pesquisadores brasileiros compreendem o retorno social do investimento público realizado em suas pesquisas; c) quais conceitos mais sustentam o debate sobre a gestão de dados científicos no Brasil; d) como as instituições de pesquisa, agências de fomento e universidades brasileiras estão se posicionando quanto ao debate acerca da gestão de dados científicos; e) qual o papel dos repositórios nesse processo de gestão desses dados.

Essas questões motivaram uma reflexão mais detida sobre como esse contexto se expressa na Bahia, especialmente no âmbito das universidades públicas estaduais baianas. Nesse sentido, buscamos ilustrar este estudo dissertativo com a iniciativa de uma dessas universidades, no caso, a Universidade do Estado da Bahia (UNEB) que implantou um repositório institucional – SaberAberto – aqui tomado como referência empírica.

O repositório institucional SaberAberto está inserido e vinculado hierarquicamente ao Sistema de Bibliotecas (SISB) da UNEB. A UNEB é uma universidade *multicampi*, possui 26 bibliotecas, uma em cada *campus* universitário, sendo a maior delas localizada no *Campus I*, em Salvador. Essa biblioteca leva o nome do idealizador da Uneb, Prof. Edivaldo Machado Boaventura. É nela também que está localizado o núcleo de que gerencia o repositório Saber Aberto que atende / pretende atender toda a Instituição. As demais bibliotecas que compõem o sistema estão distribuídas pelos 24 *campi* no interior do estado, além da Biblioteca da Pós-Graduação, localizada no *Campus I* (Salvador) e a Biblioteca da Educação a Distância (EAD), no Polo Universitário Santo Amaro de Ipitanga (Pusai), localizada no município de Lauro de Freitas.

Nesse sentido, consideramos importante conhecer como a UNEB tem realizado a gestão dos dados científicos a partir do compartilhamento de informações de forma a permitir a reutilização desses dados ou reprodução de

experimentos e a aplicação de recursos em seu repositório institucional; e, em especial, saber como o repositório SaberAberto, da UNEB tem acompanhado o movimento da gestão de dados científicos.

Diante do exposto, foi possível a discussão da existência de mecanismos que exijam o plano de gestão de dados científicos, para compartilhamento de informações de forma a permitir a reutilização ou reprodução de experimentos e a aplicação de recursos, no repositório da UNEB.

O objetivo geral deste estudo foi:

Identificar e descrever os elementos estruturantes de um Plano de Gestão de Dados Científicos, com perspectiva de aplicabilidade em uma universidade pública.

Como objetivos específicos, este estudo dissertativo pretende:

- a) Apresentar os principais conceitos sobre a gestão de dados científicos, a partir dos contextos terminológicos e históricos que deram origem às novas práticas de pesquisa, na contemporaneidade;
- b) Descrever os Repositórios de Instituições no Brasil que utilizam Planos de Gestão de Dados Científicos e disponibilizam seus dados;
- c) Identificar mecanismos que apontem elementos do plano de gestão de dados científicos nas universidades brasileiras;
- d) Propor um modelo de plano de gestão de dados científicos para o Repositório Institucional SaberAberto, alinhado às características acadêmicas, científicas e organizacionais da UNEB.

Com relação aos limites e às potencialidades deste trabalho, destacamos, aqui, alguns aspectos. Esta dissertação não trata de processos especificamente voltados para o desenvolvimento de pesquisas, não discute metodologias relativas à investigação científica, nem lida com as bases epistemológicas da construção de conhecimentos nem seus processos de validação. O campo abordado por esta dissertação está nos conjuntos de dados reunidos pelo pesquisador antes, durante e depois de concluída sua pesquisa; é esse amplo volume de informações que vai constituir os chamados *dados científicos*, ou seja, quando esses dados já foram, pelo pesquisador e ou por seus pares, qualificados como científicos. É, a partir daí, que a abordagem do tema tratado nesta dissertação tem início. E, ainda assim, esta abordagem não esgota tão amplo, complexo e contemporâneo tema.

A relevância deste trabalho está na sua originalidade, na disseminação de um tema novo, de crescente importância na sociedade contemporânea, ainda não incorporado em muitos ambientes organizacionais do Brasil, inclusive universidades, diz respeito a todas as áreas de conhecimento e, especialmente, promove intensa articulação entre as ciências da informação e as práticas da investigação científica. O tema envolve problemáticas, temáticas e conceitos contemporâneos, questões atreladas ao uso de recursos públicos e ao dispêndio de tempo na realização de pesquisas. Chama ainda atenção para potenciais atrasos científicos, por fazer com que os pesquisadores voltem sempre à posição inicial com relação ao desenvolvimento de estudos e pesquisas.

A quantidade de dados gerada no campo é valiosa e pode contribuir com outros estudos desde que sejam digitalizados, tratados sob novas perspectivas e reaproveitados de maneira inteligente, pois promovem avanços científicos para a sociedade como um todo.

Justifica-se, também, este trabalho, com base na sua relevância social. No contexto social, este estudo contribui para as discussões sobre o custeio de pesquisas e a significância de seus dados gerados a partir do seu financiamento, passando pela sua coleta até chegar às possibilidades de divulgação no meio científico.

Nessa perspectiva, esta dissertação reúne temas que podem interessar segmentos profissionais que já atuam ou que iniciam suas trajetórias funcionais na gestão de dados de pesquisa, a exemplo de pesquisadores e estudantes de quaisquer áreas do conhecimento, segmentos profissionais especializados como bibliotecários, arquivistas e documentalistas, além de outros pesquisadores e profissionais e alunos de áreas como Administração, Ciência da Computação, Ciência da Informação, Comunicação, Educação, Tecnologia da Informação, Sistemas de Informação. Núcleos de estudos, grupos e instituições de pesquisa, agências de fomento e universidades são, especialmente, segmentos institucionais ao alcance das temáticas aqui abordadas, além do público em geral, interessado nas mesmas.

Esta dissertação está composta por seis capítulos: neste **Capítulo 1 – INTRODUÇÃO**, tratamos de apresentar o novo papel que os dados científicos estão assumindo perante as práticas tradicionais de se fazer pesquisa, os motivos que nos levaram à abordagem desse tema, os objetivos do estudo e a estrutura do trabalho.

O **Capítulo 2** traz a **METODOLOGIA** adotada no trabalho, considerando a natureza dissertativa do estudo, sua caracterização, referências conceituais e empíricas, e descrição das etapas desenvolvidas. No **Capítulo 3 – GESTÃO DE DADOS CIENTÍFICOS**, abordamos os contextos terminológicos e históricos que permitem identificar vocábulos, e seus respectivos significados, bem como eventos, de caráter internacional ou nacional, reconhecidos como marcos linguísticos e históricos na construção da identidade do campo. O **Capítulo 4 – REPOSITÓRIOS INSTITUCIONAIS COMO FERRAMENTA DE APOIO À GESTÃO DE DADOS CIENTÍFICOS**, ilustra os repositórios e traça um breve panorama a respeito dos repositórios institucionais no mundo, no Brasil e na Bahia; e apresenta o Repositório Institucional SaberAberto da Uneb. No **Capítulo 5, PLANO DE GESTÃO DE DADOS CIENTÍFICOS**, são destacadas problemáticas e desafios identificados nas experiências internacionais e nacionais referidas nos Capítulos anteriores, bem como elementos reconhecidos como essenciais na composição de Planos de Gestão de Dados Científicos, de modo a subsidiar a formulação de um modelo com vista à sua aplicabilidade no Repositório Institucional SaberAberto da UNEB. E finaliza com o **Capítulo 6, CONSIDERAÇÕES FINAIS**, no qual são cotejados os objetivos geral e específicos com os achados e alcances do trabalho realizado pela dissertação.

2 METODOLOGIA

Este capítulo apresenta o caminho percorrido para abordar o tema-objeto deste estudo. Assim, metodologicamente, a pesquisa desenvolvida se caracterizou por um conjunto de ações próprias de um estudo de natureza dissertativa. Fundamentou-se na exposição de um tema - no caso, a gestão de dados científicos, na modalidade acesso aberto, em repositórios institucionais - e teve, como propósito, identificar argumentos que pudessem subsidiar uma proposta de plano de gestão de dados científicos, com possibilidade de aplicação em um dado repositório institucional de uma universidade pública. O recorte temporal estabelecido correspondeu ao período de 1999 a 2018, respectivamente marcos das declarações em apoio ao movimento Acesso Aberto (Convenção de Santa Fé, em 1999 e

Declaração do México, em 2018), período que também permitiu incorporar referências de estudos desenvolvidos no Brasil, tema recente na realidade do nosso País.

Para um melhor entendimento das escolhas metodológicas realizadas para elaborar o estudo, optou-se por subdividir este capítulo em três seções:

2.1 Caracterização

2.2 Referências conceituais e empírica

2.3 Etapas

2.1 CARACTERIZAÇÃO

Este trabalho orientou-se pela NBR 14.724, da ABNT, segundo a qual cabe apresentar a “[...] a exposição de um estudo científico retrospectivo, de tema único e bem delimitado em sua extensão, com o objetivo de reunir, analisar e interpretar informações”, contemplando a literatura existente e a sua sistematização.

Buscando fundamentar as bases da escolha da metodologia e as implicações do método da pesquisa utilizado, este trabalho será construído em nível descritivo, em uma abordagem qualitativa, compreendendo tais escolhas como as mais apropriadas a esse tipo de estudo. A escolha dos instrumentos e os procedimentos da coleta de dados foi realizada com base nas necessidades dos objetivos específicos.

Neste estudo aplicou-se a pesquisa descritiva qualitativa que segundo Beuren (2004), é caracterizada pelo emprego de instrumentos analíticos com base em impressões do pesquisador, tanto na coleta quanto no tratamento dos dados. Para além dessa dicotomia entre qualitativo, André (2012) fala da necessidade dos pesquisadores as ultrapassarem e tentar encontrar respostas para as suas inúmeras indagações observando uma perspectiva positivista da ciência e um lado também humanista que se entrelaçam nas coletas de dados realizadas.

Comunga-se com Severino (2007) que a abordagem da pesquisa não se refere a uma modalidade de metodologia em particular. Desse modo, a escolha da abordagem foi guiada pelos objetivos e, respectivamente, pelo conjunto de

metodologias.

Dito isso, este estudo possui abordagem qualitativa, pois as análises foram feitas a partir do próprio pesquisador (LAVILLE; DIONE, 1999), mas também possui perspectiva quantitativa, por quantificar alguns aspectos referente ao objeto de análise. Segundo Lakatos e Marconi (2007, p.87):

Como primeiro passo, observamos atentamente certos fatos e fenômenos, passamos a seguir, à classificação, isto é, agrupamentos de fatos e fenômenos da mesma espécie, segundo a relação constante que se nota entre eles.

O caminho percorrido para a realização de uma pesquisa revela o olhar e as crenças do pesquisador. A natureza do objeto de investigação, a temática abordada e os objetivos da pesquisa - a qual se refere a uma investigação de situações dinâmicas, processuais, dentro do ambiente natural em que ocorrem e com um significado histórico-cultural - bem como a possibilidade do envolvimento e interferência, conduziram ao alargamento propiciado pelas pesquisas qualitativas, Bogdan e Biklen (1994, p.11) comentam que:

[...] um campo que era anteriormente dominado pelas questões da mensuração, definições operacionais, variáveis, testes de hipóteses e estatística alargou-se para contemplar uma metodologia de investigação que enfatiza a descrição, a indução, a teoria fundamentada e o estudo das percepções pessoais. Designamos esta abordagem por Investigação Qualitativa.

Ainda esses autores (BOGDAN; BIKLEN, 1994), caracterizam a pesquisa qualitativa como aquela que tem o pesquisador como instrumento principal da pesquisa, o ambiente natural como sua fonte direta de dados, o significado que as pessoas dão às coisas e à sua vida como foco de atenção especial, os dados coletados são predominantemente descritivos e a preocupação está centrada mais no processo do que no produto. Enfim, a análise dos dados não parte de hipóteses pré-estabelecidas, tende a seguir um processo indutivo.

Na abordagem qualitativa, o pesquisador procura aprofundar-se na compreensão dos fenômenos que estudam – ações dos indivíduos, grupos ou organizações em seu ambiente e contexto social – interpretando-os segundo a perspectiva dos participantes da situação enfocada, sem se preocupar com

representatividade numérica, generalizações estatísticas e relações lineares de causa e efeito (TERENCE, 2006).

A pesquisa qualitativa é percebida como um ato subjetivo de construção. Os autores afirmam que a descoberta e a construção de teorias são objetos de estudo desta abordagem (GUNTHER, 2006). A descoberta e a construção de teorias simplesmente constituem o cerne de qualquer ciência. Uma preferência por material textual é uma legítima opção de procedimento, desde que não se contraponha aos princípios elencados no próximo parágrafo. (GUNTHER, 2006). Desse modo, comungamos com Gatti (2004) que boas análises dependem de boas perguntas que o pesquisador venha a fazer, ou seja, da qualidade teórica e da perspectiva epistêmica na abordagem do problema, as quais guiam as análises e as interpretações.

De acordo com a natureza dissertativa do estudo então realizado, este trabalho compreendeu:

- Pesquisa Bibliográfica – consistiu na revisão da literatura da área, publicada em livros ou artigos, bem como disponibilizada na web, com foco na explicitação da terminologia comumente veiculada para referir-se, especialmente, a dados científicos, plataformas digitais, repositórios institucionais.

- Pesquisa Documental – compreendeu o levantamento de (a) documentos, internacionais e nacionais, considerados marcos históricos na evolução das plataformas digitais orientadas para dados científicos; e (b) informações de caráter institucional para a caracterização da Universidade do Estado da Bahia (UNEB) e seu respectivo Repositório Institucional SaberAberto.

Pesquisa de Campo – desenvolveu-se nos limites de um estudo de natureza dissertativa, junto à equipe que concebeu o Repositório Institucional SaberAberto da UNEB e setores universitários administrativos com ele relacionados, com vistas a: (a) aprofundamento de aspectos derivados da pesquisa bibliográfica e ou documental; e (b) caracterização das dimensões acadêmica, científica e organizacional da UNEB como requisitos importantes para consolidação do RI/SaberAberto.

O tipo de observação foi sistemático e os instrumentos de coleta de dados compreendem o formulário de observação para coleta de dados, também chamado de diário de campo.

Para uma definição precisa a respeito do que seja a técnica, peça fundamental na realização da pesquisa, de acordo Lakatos (2010) “[...] são consideradas um conjunto de preceitos ou processos de que se serve uma ciência [...] a habilidade para usar esses preceitos ou normas, na obtenção de seus propósitos. Correspondem, portanto, à parte prática de coleta de dados [...]. Segue as fases que foram realizadas neste estudo:

Fase 1: Reafirmação da problemática e escolha da fundamentação teórica;

Fase 2: Apresentação da Pesquisa de maneira formal;

Fase 3: Solicitação oficial para realização de amostra da pesquisa;

Fase 4: Divisão das atividades do procedimento;

Fase 5: Organização do cronograma e elementos que compunham o questionário e observações;

Fase 6: Estabelecimento do cronograma;

Fase 7: Início da pesquisa nos repositórios;

Fase 8: Análise dos dados recolhidos;

Fase 9: Início do processamento dos dados coletados: comparação, transcrição, análise, descrição (...);

Fase 10: Começa a análise e tabulação dos dados, confronto com as teorias e construção das tabelas e gráficos;

Fase 11: Ajustes metodológicos, conceituais e formatação, preparação para conclusão da pesquisa;

Fase 12: Preparação para a conclusão da pesquisa, pré defesa;

Fase 13: Entrega da dissertação concluída;

Fase 14: Apresentação da produção final: defesa da dissertação.

Isto significa dizer que nessa fase é necessário um bom entrosamento do pesquisador com as tarefas de organização e administração das coletas científicas, o que equivale dizer que sem planejamento e um rigoroso controle dos instrumentos de pesquisa, fica difícil não cometer erros ou deixar lacunas a serem respondidas.

Ainda no entendimento destes autores, seguem-se passos importantes nessa fase que podem ser definidos, com base em Marconi e Lakatos (1999,p.35) como:

- a) Seleção - exame minucioso dos dados, que são submetidos uma verificação crítica, procurando um resultado acurado do que foi coletado;
- b) Codificação - é definida como a técnica operacional que serve para categorizar os dados que se relacionam. Assim, através da classificação destes dados é possível reduzir estas informações em símbolos que podem ser posteriormente tabulados, classificados e contados;
- c) Tabulação - é a disposição dos dados recolhidos em tabelas ou gráficos possibilitando uma maior visualização tanto das inter-relação entre os dados quanto do contexto geral em que esses dados foram gerados.

2.2 REFERÊNCIAS CONCEITUAIS E EMPÍRICA

As referências conceituais contemplam o campo terminológico relativo à gestão de dados científicos, compreendendo o destaque a vocábulos empregados para designar peculiaridades da nomenclatura veiculada pela literatura e documentos consultados.

O campo empírico deste estudo tem, como pano de fundo, a universidade pública e, de modo mais específico, a Universidade do Estado da Bahia (UNEB). Conceitualmente, esta dissertação compreende as universidades como centros de produção do conhecimento.

Neste sentido, as instituições públicas prestam um grande serviço à sociedade, pois é através dela que muitos profissionais são formados, realizam pesquisas científicas que posteriormente retornam para sociedade em forma de serviços, inovações, tecnologias, produtos, dentre outros.

Os acessos ao mundo da cultura são cada vez mais intensamente submetidos a mecanismos alienantes, sem que o Estado assuma qualquer medida no sentido de garantir o acesso efetivamente democrático: pelo contrário, os poderes públicos se fazem cúmplices dos oligopólios midiáticos. A universidade pública é a única instância em que se pode resistir, de alguma maneira e por mais algum tempo, talvez, a este processo que traz na sua própria dinâmica um objetivo destruidor. A universidade pública é a instituição em que a cultura pode ser considerada sem as regras do mercado e sem os critérios de utilidade e oportunidade socialmente introjetados a partir da racionalidade midiática. (SILVA, 2001, p.7)

Uma problematização, ligada às universidades enquanto espaços de produção do conhecimento, diz respeito à existência de grandes centros de pesquisa sem a correspondente demanda social, em especial do sistema produtivo. Neste sentido, explica-se que o Brasil embarcou em um projeto ambicioso de elevar suas universidades públicas ao patamar das melhores *research-universities* do mundo e em políticas de internacionalização, muito antes de empreender qualquer tentativa mais incisiva na resolução de produção científica, a exemplo das questões levantadas neste trabalho sobre a gestão de dados científicos.

As universidades públicas, indiscutivelmente, têm um compromisso com o saber sistematizado, mas têm, também, um sério compromisso com problemas e desafios concretos colocados pela sociedade. A análise do contexto atual em que as universidades se encontram demonstra que a sua atuação, junto aos problemas sociais, tem sido bastante insatisfatória. A luta pela defesa da universidade pública vem enfraquecendo dia-a-dia, pois já não tem mais o apoio da sociedade, que a considera um investimento muito caro, com baixo retorno social. Para reforçar esse quadro, existe todo um 'lob' da privatização que assola o país de ponta a ponta. Não é papel da universidade dar retornos imediatos ou resolver os problemas sociais, mas cabe a ela formar quadros críticos, ou seja, profissionais com competência técnica, científica e social, para o enfrentamento dos desafios e impasses postos pela sociedade. (KAVASAKI, 1997,p.9)

Segundo esta autora, o Brasil possui notadamente alguns dos melhores programas de pós-graduação e um dos piores e mais desiguais sistemas de educação praticamente em todas as regiões. No entanto, dito isto, é preciso pensar nos critérios e elementos que são necessários para garantir visibilidade da produção científica brasileira e como os repositórios podem ser realmente aliados das universidades, também como forma de garantir transparência dos dados brutos de pesquisa, desde que livres de conflito de interesse.

A partir dos anos 50, as universidades públicas passaram a ser consideradas peças-chave na concepção da política científica nacional, pois surgiram, nessa época, as principais lideranças nas comunidades científicas acadêmicas, que criaram as condições para uma política pública de ciência e tecnologia nos anos 60. A principal ideia era a formação de uma infraestrutura técnico-científica e uma massa crítica capazes de desenvolver matérias-primas e aumentar a produção e a produtividade industrial. Entretanto, no mundo real as coisas não funcionaram dessa forma, isto é, na prática, este esquema ajudou a aumentar o número de instituições de educação superior, mas foi incapaz de reforçar a tecnologia local, que permaneceu incipiente e, em geral, sem muita influência na estrutura produtiva. (KAWASAKI, 1997, p.5)

Nesse movimento de produção do conhecimento científico, destaca-se a UNEB neste estudo. De acordo com o *site* oficial da entidade pesquisada, a saber: <<http://www.uneb.br>>, a Universidade do Estado da Bahia (UNEB), sendo reconhecida como instituição de ensino superior sob o Decreto N.º 13.664 de 07 de Fevereiro de 2012, a UNEB, é considerada como a maior instituição pública de ensino superior da Bahia. Esta instituição de ensino superior foi fundada em 1983 e é mantida pelo Governo do Estado por intermédio da Secretaria da Educação (SEC), estando presente geograficamente em todas as regiões do Estado, estruturada no sistema *multicampi*.

2.3 ETAPAS

- Pesquisa Bibliográfica:

Seleção de autores e obras dedicadas à demarcação temporal e espacial sobre a emergência de novas práticas de pesquisa e ao estudo de novas formas de coletar, produzir, armazenar e disponibilizar conhecimentos.

Mapeamento das categorias conceituais utilizadas por esses autores para caracterizar a emergência dessas novas práticas de pesquisa no contexto das plataformas digitais e relativas ao acesso, tratamento, produção e disponibilização de dados científicos.

Identificação das propriedades essenciais de plataformas digitais reconhecidas como repositórios institucionais.

- Pesquisa Documental:

Levantamento dos documentos reconhecidos pela comunidade nacional e internacional como marcos históricos dos eventos inaugurais dos acontecimentos que instauraram uma nova forma de fazer pesquisa.

Levantamento dos documentos institucionais sobre a Uneb disponibilizados no portal da Universidade, com vistas à sua caracterização acadêmica-científica e organizacional: modalidade organizacional *multicampi*; abrangência espacial; quadro docente em termos de titulação e distribuição por *campus* universitário; programas de pós-graduação *stricto-sensu*, centros e grupos de pesquisa por área de conhecimento.

Caracterização do Repositório Institucional SaberAberto/Uneb, conforme informações disponibilizadas no portal da Universidade, com destaque aos seus objetivos.

Pesquisa de Campo:

Visita(s) ao Repositório Institucional SaberAberto/Uneb e outros setores da Universidade eventualmente encarregados da sua manutenção

Reunião com equipe responsável pelo desenvolvimento do Repositório Institucional SaberAberto da Uneb, para identificação de aspectos considerados relevantes à sua consolidação por meio de contato remoto.

A pesquisa bibliográfica e a pesquisa documental foram utilizadas para (i) atender aos três primeiros objetivos específicos que visaram apresentar os principais conceitos que sustentam o debate sobre a gestão de dados científicos no mundo e no Brasil; (ii) descrever os Repositórios de Instituições no Brasil que utilizam Planos de Gestão de Dados Científicos e disponibilizam seus dados; e (iii) identificar elementos para um plano de gestão de dados científicos.

O quarto objetivo foi elaborado para atender uma prerrogativa de uma dissertação no contexto do mestrado profissional, com vistas à aplicabilidade de um modelo de plano de gestão de dados científicos para o RI/UNEB SaberAberto com base nas características acadêmicas, científicas e organizacionais da UNEB. Para alcançar este resultado foi necessário o acúmulo das ações desempenhadas nos três primeiros objetivos específicos bem como cotejar informações e observações coletadas no âmbito da pesquisa de campo, conforme delineada nesta dissertação.

3. GESTÃO DE DADOS CIENTÍFICOS

Este capítulo aborda a base conceitual que se relaciona com a gestão de dados científicos. Desse modo, tem a preocupação de expor termos e expressões referenciados por dicionários ou glossários especializados, que são pertinentes ao entendimento desse objeto, com vistas a uma maior exatidão da terminologia correspondente ao campo e seus significados.

Como já indicado na Introdução, o campo conceitual do tema comporta certo grau de complexidade em face da diversidade e semelhanças entre os termos nele utilizados. Por essa razão – e sem pretender esgotar o assunto – dedica-se, neste capítulo, à apresentação de termos e expressões mais utilizadas pelos pesquisadores na literatura específica, bem como referenciados em dicionários e glossários especializados, sobre a gestão de dados científicos assim como os principais estudiosos.

Nesse sentido, o Capítulo foi, assim, constituído em duas seções:

3.1 Referencial terminológico-conceitual, e

3.2 Movimento acesso aberto e repositórios institucionais.

3.1 REFERENCIAL TERMINOLÓGICO-CONCEITUAL

Dados, informação e conhecimento são conceitos essenciais a esta dissertação. Básicos e polissêmicos, tais conceitos permeiam toda atividade científica, em todas as áreas do conhecimento. Não são conceitos simples; se fazem presentes em campos de abordagens amplas e profundas, sob diversas perspectivas epistemológicas. Por essas razões, delimitamos a sua definição ao recorte mais preciso possível no âmbito do tema aqui tratado. Isto não diminui a sua complexidade nem significa que seja fácil a tarefa para reuni-los, armazená-los, organizá-los, preservá-los, descrevê-los ou compartilhá-los. Na perspectiva desta dissertação, é exatamente este o deslocamento conceitual-metodológico que se opera: como promover a sua gestão, a partir de volumes consideráveis de dados, informações e conhecimentos, em variadas áreas do conhecimento, geradas no fazer pesquisa. Assim, para evitar a dispersão semântica – que, a despeito da sua fundamental importância, não é o objeto deste estudo⁹ -, optamos por compartilhar definições expostas em dicionários da área, como, a seguir, apresentado.

As denominações: *pesquisa e dados científicos*, por exemplo, já ilustram bem a complexidade terminológica com a qual lidamos. Em 2007, relatório da Organização para a Cooperação e Desenvolvimento Econômico (OCDE)¹⁰, por

⁹ Aos interessados sugerimos a leitura da dissertação de mestrado de autoria de Rafael Aparecido Moron Semidão (2014), intitulada DADOS, INFORMAÇÃO E CONHECIMENTO ENQUANTO ELEMENTOS DE COMPREENSÃO DO UNIVERSO CONCEITUAL DA CIÊNCIA DA INFORMAÇÃO: CONTRIBUIÇÕES TEÓRICAS, disponível em: < https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/semidao_ram_me_mar.pdf > e ao artigo Dados, Informação e Conhecimento: elementos de análise conceitual, do mesmo autor, disponível em < <https://brapci.inf.br/index.php/article/download/52967>; ao artigo, publicado em 2015, como ampliação e atualização de artigo anteriormente publicado na Revista Datagrama (*Meios Eletrônicos e a Educação: Uma Visão alternativa*. São Paulo: Editora Escrituras, Coleção Ensaio Transversais Vol. 10, 2001.), de autoria de **Valdemar W. Setzer**, sob o título *Dado, Informação, Conhecimento e Competência*, disponível em < <https://www.ime.usp.br/~vwsetzer/dado-info.html> >; CAPURRO, Rafael; HJORLAND, Birger. O conceito de informação. **Perspectiva. ciênc. inf.**, Belo Horizonte, v. 12, n. 1, p. 148-207, abril de 2007. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362007000100012&lng=en&nrm=iso>. acesso em 18 de junho de 2020. <https://doi.org/10.1590/S1413-99362007000100012>; XAVIER, Rodolfo Coutinho Moreira; COSTA, Rubenildo Oliveira da. Relações mútuas entre informação e conhecimento: o mesmo conceito?. **Ci. Inf.**, Brasília, v. 39, n. 2, p. 75-83, Aug. 2010. Available from <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652010000200006&lng=en&nrm=iso>. access on 18 June 2020. <https://doi.org/10.1590/S0100-19652010000200006>; Glossário de termos técnicos em ciência da informação : Inglês/português / Mariângela Braga Norte. – São Paulo : Cultura Acadêmica ; Marília : Oficina Universitária, 2010, disponível em < <https://www.marilia.unesp.br/Home/Publicacoes/glossario.pdf> >;

¹⁰ Relatório publicado em 2007, título em português: “Princípios e diretrizes para o acesso a dados de pesquisa financiados por fundos públicos”.

exemplo, já advertia para a amplitude conceitual, ao conceituar *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” (OCDE, p.13).

Outras fontes também observam essa complexidade, como detalha, por exemplo, a agência da Universidade Estadual de São Paulo (USP) conhecida como AGUIA - GESTÃO DA INFORMAÇÃO ACADÊMICA¹¹:

Dados são componentes centrais do processo de pesquisa. São registros científicos que embasam os resultados de pesquisa publicados na forma de dissertações, teses, artigos, patentes e trabalhos científicos.

Os dados de pesquisa abrangem um universo muito amplo, que se especifica também em função da área de conhecimento à qual se reportam: documentos textuais, planilhas, estatísticas, cadernos de laboratório, cadernos de campo, diários, questionários, transcrições, arquivos de áudio, vídeo, fotografias, sequências de proteínas ou genéticos, artefatos, amostras, modelos, algoritmos, scripts, arquivos de log, software de simulação, metodologias e fluxos de trabalho, procedimentos operacionais, padrões e protocolos.

Os dados de pesquisa também podem ser de vários tipos, por exemplo, dados observacionais, experimentais, simulações, dados compilados. [...] dados brutos (raw data) ou preliminares, dados derivados, dados referenciais ou canônicos.

Assim, também, o IBICT que, além do acesso à própria Biblioteca, disponibiliza várias bases de dados que possibilitam amplo levantamento a respeito da terminologia adotada na área, a exemplo da Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)¹², o Catálogo Coletivo Nacional de Publicações Seriadas (CCN), o Centro Brasileiro do International Standard Serial Number (ISSN), o Centro Brasileiro do Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal (Latindex), o Diretório de Políticas de Acesso Aberto das Revistas Científicas Brasileiras (Diadorim)¹³, o Portal Brasileiro de Acesso Aberto à Informação Científica (OASISBR), o Portal do Livro Aberto em CT&I, o Programa de Comutação Bibliográfica (Comut), o Repositório Institucional

¹¹ <https://www.aguia.usp.br/bibliotecas/>

¹² <http://www.ibict.br/informacao-para-a-pesquisa/bdtd>.

¹³ <http://www.ibict.br/informacao-para-a-pesquisa/diadorim>

Digital do IBICT (RIDDI) e também Repositórios Digitais (RDs), Revistas no SEER, Diretório Luso-Brasileiro.

Particularmente com relação à preservação digital de documentos eletrônicos brasileiros, ao acesso contínuo e a longo prazo, o IBICT desenvolveu inúmeras outras ações, como a Rede Cariniana¹⁴. Essa Rede contou com o apoio da FINEP, para aderir ao Programa LOCKSS¹⁵ da Stanford University, parceria que resultou essencial para a informação científica no Brasil. As universidades brasileiras Universidade de São Paulo (USP), Universidade Estadual de Campinas (UNICAMP), Universidade Federal da Paraíba (UFPB), Universidade Federal de Santa Maria (UFSM) e Universidade Estadual do Maranhão (UEMA) foram convidadas a participar das primeiras etapas. A Rede Cariniana¹⁶, fruto dessa iniciativa, disponibiliza um glossário de preservação digital, do qual destacamos algumas referências de interesse direto desta dissertação:

Dados

Fatos, medições, gravações, registros ou observações sobre o mundo coletado por cientistas e outros, com um mínimo de interpretação contextual. Dados podem estar em qualquer formato ou meio, sob a forma de escritos, notas, números, símbolos, textos, imagens, filmes, vídeo, gravações de som, reproduções pictóricas, desenhos, desenhos ou outras representações gráficas, manuais processuais, formulários, diagramas, trabalhos fluxogramas, descrições de equipamentos, dados de arquivos, dados algoritmos de processamento, ou registros estatísticos.

Dados abertos (open data)

Por dados abertos em ciência queremos dizer que eles estão disponíveis gratuitamente na internet, permitindo a qualquer usuário baixar, copiar, analisar, reprocessar, capturá-los via softwares ou utilizá-los para qualquer outra finalidade, sem outras barreiras financeiras, legais ou técnicas adicionais além daquelas inseparáveis

¹⁴ <http://cariniana.ibict.br/index.php/inicio>

¹⁵ < <https://www.lockss.org/about/what-lockss> >.

¹⁶ <http://cariniana.ibict.br/index.php/2016-02-19-18-56-47> eu acho que precisa referenciar melhor esses sites

da própria internet. Para este fim os dados relacionados com a ciência publicada devem ser explicitamente colocados no domínio público.¹⁷

Dados científicos (research data)

Dados que são produzidos no contexto de investigação científica ou que de alguma forma são usados em investigação. Estes dados podem ser criados para efeito de processamento científico, como nos dados atmosféricos usados para previsão meteorológica, ou os dados recolhidos de sensores para monitorar o estado de um edifício. Há dados que são obtidos como resultados do processamento automático de objetos (eles próprios representados como dados), como, por exemplo, os histogramas de cor obtidos através do processamento de uma coleção de imagens. Há ainda dados que não sendo produzidos para investigações acabam por ser objeto dela, como as contribuições que os utilizadores de uma rede social fazem, na forma de textos ou outros conteúdos, e que acabam por ser utilizados para estudos sociológicos.¹⁸

Dados de pesquisa

Os dados são fatos, observações ou experiências em que um argumento ou uma teoria é construído ou testado. Os dados podem ser numéricos, descritivos, auditivos ou visuais. Os dados podem ser crus, abstraídos ou analisados, experimentais ou observacionais. Os dados incluem, mas não estão limitados a: cadernos de laboratório, cadernos de campo, dados de pesquisa primária (incluindo dados de pesquisa em formato impresso ou em forma legível por computador), questionários, fitas de áudio, fitas de vídeo, modelos, fotografias, filmes, respostas de teste. Coleções de pesquisa podem incluir slides, artefatos, espécimes, amostras.

Dados dinâmicos

¹⁷ No Glossário da Rede Cariniana consta fonte como Panton Principles: Principles for Open Data in Science. < <https://pantonprinciples.org/> >.

¹⁸ No Glossário da Rede Cariniana consta fonte como RECCAP.

No contexto da programação, ou se a conversa assumir de forma segura qual é a escala de tempo dos dados: dados dinâmicos ou dados transacionais denotam informações que são alteradas de forma assíncrona à medida que as futuras atualizações se tornam disponíveis. O oposto disso são dados persistentes, que são dados acessados com pouca frequência e que não são suscetíveis de serem modificados. Os dados dinâmicos também são diferentes dos dados de transmissão, na medida em que não há fluxo constante de informações. Em vez disso, as atualizações podem vir a qualquer momento, com períodos de inatividade no meio.

Dados estruturados

Dados estruturados, em distinção com dados não estruturados, são dados que estão em conformidade com um esquema definido e fixo. Exemplos: bancos de dados relacionais, planilhas e triplos RDF.

Dados não estruturados

Dados que não podem ser armazenados em um banco de dados relacional e podem ser mais desafiantes para analisar documentos e tweets para fotos e vídeos.

Dados persistentes

Dados persistentes denotam informações que são acessadas com pouca frequência e não são suscetíveis de serem modificadas.

Dados vinculados

Os dados vinculados, também chamados Linked Open Data, são dados nos quais as relações/conexões entre os dados devem estar disponíveis. Isso permite um fácil acesso aos dados. Explicação: Esta coleção conexa de conjuntos de dados inter-relacionados é armazenada na web e disponível através de um formato comum - RDF.

Dissemos, acima, não ser tarefa fácil reunir, armazenar, organizar, preservar, descrever ou compartilhar *dados científicos*. Essa tarefa - que tem mobilizado muitas instituições em todo o mundo, inúmeros especialistas, pesquisadores de várias áreas do conhecimento -, se transformou numa ação contínua, crescentemente alargada, que mudou, radicalmente, o modo tradicional de se fazer ciência.

Como diz Medeiros (2018)¹⁹:

A gestão de dados científicos cobre todo o chamado “ciclo de vida” dos dados, ou seja, desde a sua coleta até o armazenamento de longo prazo, passando por uma série de processamentos de limpeza, curadoria, anotação, indexação e transformação. Grande parte da pesquisa científica de hoje exige algum tipo de análise e processamento de dados. Com isto, o planejamento da gestão dos dados utilizados e gerados em uma pesquisa passou a fazer parte integral da metodologia científica, sendo, inclusive, considerado como um dos itens necessários de boas práticas de pesquisa.

Especialmente com relação à gestão de dados científicos, o *Guia de Gestão da Dados de Pesquisa para Bibliotecários e Pesquisadores*, elaborado por Luana Sales e Luís Fernando Sayão (2015), representa outro importante marco, no Brasil, no processo de preservação de dados científicos. Este Guia também apresenta um GLOSSÁRIO DE TERMOS DE GESTÃO DE DADOS DE PESQUISA, do qual extraímos as seguintes definições:

Dado aberto (open data)

Dados de pesquisa que são disponíveis livremente para reuso e republicação sem restrições de copyright, patentes ou outros mecanismos de controle de propriedade intelectual.

Dado bruto (raw data)

Dado que vem diretamente dos instrumentos científicos ou coletados diretamente da fonte sem sofrer nenhuma manipulação ou processamento. Também chamado de dado cru ou dado primário.

Dado derivado (derivative data)

¹⁹<https://blog.scielo.org/blog/2018/06/22/gestao-de-dados-cientificos-da-coleta-a-preservacao/#.Xu6LEmhKjIU>

Resultado do processamento ou combinação de dados brutos ou de outros dados. Também chamado de dado secundário

Dado de pesquisa (research data)

Unidades de informação criadas ou coletadas no curso da pesquisa científica, e que são frequentemente formatadas de maneira a torná-las adequadas à comunicação, interpretação e processamento por computador. São exemplos de dados de pesquisa: planilhas de estatísticas, uma série de mensagens de email, um registro sonoro de uma entrevista, um registro descritivo de um espécime de rocha, uma coleção de imagens digitais. Dependendo do contexto em que são consideradas, quase todas as coisas podem ser consideradas dados de pesquisa.

Trabalho apresentado por Vidotti, Arakaki e Brandt (2018) no XVIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB 2017²⁰ traz um extenso levantamento sobre o tema, como ilustramos, a seguir: segundo essas autoras e autores, *dados de pesquisa* são, para o Committee for a study on promoting access to scientific and technical data for the public interest "[...] fatos, números, letras e símbolos que descrevem um objeto, ideia, condição, situação, ou outros fatores" (VIDOTTI; ARAKAKI; BRANDT, 2018). Para a National Science Board e a National Science Foundation (2005 apud VIDOTTI; ARAKAKI; BRANDT, 2018), "[...] dados podem ser qualquer informação armazenável em formato digital, como por exemplo, os elementos supracitados, além de: imagens, vídeos, áudios, softwares, algoritmos, equações, animações, modelos e simulações, gerados por diferentes formas, tais como de observação, cálculo ou experimentos. Referindo outro autor, (Pampel et al., 2013 apud VIDOTTI; ARAKAKI; BRANDT, 2018), destacam que dados de pesquisa podem ser “[...] dados digitais sendo uma parte (descritiva) ou o resultado de um processo de pesquisa. Este processo abrange todas as etapas da pesquisa, que vão desde a geração de dados de pesquisa, que podem ser em um experimento nas ciências, um estudo empírico nas ciências sociais ou observações de fenômenos culturais, até a publicação dos resultados da pesquisa.

²⁰ <https://brapci.inf.br/index.php/article/download/58964>

Os dados digitais de pesquisa são produzidos em diferentes tipos de dados, níveis de agregação e formatos de dados, informados pelas disciplinas de pesquisa e seus métodos.”; indicam, também, que instituições como o National Institutes of Health (2003) e a Organisation for Economic Co-Operation and Development (2007 apud VIDOTTI; ARAKAKI; BRANDT, 2018) compreendem dados de pesquisa como “[...] todo aquele material que foi gravado durante a pesquisa, reconhecido pela comunidade científica e que serve para certificar os resultados da pesquisa que se realizam [...] que deve vir de uma única fonte”.

Vidotti et al (2017) também ressaltam que dados de pesquisa

[...] podem estar em variados formatos reconhecidos: numéricos, descritivos, visuais, auditivos ou tácteis, também em formatos de disciplinas ou ferramentas específicas. Ainda, entre as classificações dos dados de pesquisa existentes, destacam-se as baseadas na origem dos dados, nos métodos de coleta ou no objetivo da coleta, e que foram identificadas por Floridi (2005, 2017) como: 1) dados primários, 2) dados secundários, 3) metadados, 4) dados operacionais e 5) dados derivados. Outra classificação com base na origem dos dados refere-se a dados primários e a dados secundários. Além disso, conforme o relatório da National Science Board e National Science Foundation (2005), os dados podem ser categorizados segundo os métodos de coleta em: observacionais, computacionais, experimentais ou registros. A partir dessas categorias, os dados podem ter um valor imediato e duradouro, atingir o valor ao longo do tempo, ter um valor transitório ou ser de fácil geração e que não precisa ser preservado.

Como se observa, *dados de pesquisa* são conceito bastante complexo que pode se manifestar numa multiplicidade de formas, podendo ser descritos ou observados como símbolos, fatos, números e letras que descrevem um objeto, ideia, condição, situação, ou outros fatores, podem incluir também, números, imagens, textos, vídeos, áudio, software, algoritmos, equações, animações, modelos, simulações.

Alguns tipos de dados têm valor imediato e duradouro, enquanto outros adquirem valor ao longo do tempo; alguns dados são capturados num momento específico e irrecuperável, enquanto outros são passíveis de se reproduzir. (SAYÃO; SALES, 2015)

Os dados de pesquisa podem variar conforme o tipo de objeto de pesquisa, métodos de coleta, pesquisadores e as disciplinas. Ao invés de um simples descarte, estes dados estão sendo preservados e compartilhados através de repositórios especificamente construídos para este fim (BORGMAN, 2015).

Para a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), os dados suplementares de pesquisa ou metadados produzidos ao longo da

pesquisa seriam amostras, registros de coleta, formulários, modelos, resultados experimentais, *software*, gráficos, mapas, vídeos, planilhas, gravações de áudio, bancos de dados, material didático e outros. (FAPESP, 2019).

Para melhor entender os tipos de dados de pesquisa, é crucial analisar os tipos de dados de pesquisa disponíveis ou produzidos ao longo de uma trajetória ou percurso das pesquisas. Os tipos de dados e a definição de cada um também foram considerados por Sayão e Sales (2015), conforme *print* a seguir:

Quadro 1: Tipos de dados de pesquisa

<p>QUAIS SÃO OS TIPOS DE DADOS DE PESQUISA?</p> <p>Os dados de pesquisa podem ser caracterizados de várias formas, por exemplo, de acordo com sua natureza, origem ou de acordo com seu status no fluxo de trabalho da pesquisa. Cada uma dessas visões revela uma faceta desses recursos informacionais.</p> <p>CLASSIFICAÇÃO DOS DADOS SEGUNDO A SUA ORIGEM5:</p> <ul style="list-style-type: none"> • DADOS OBSERVACIONAIS <p>São dados obtidos por meio de observações diretas, que podem ser associadas a lugares e tempo específicos, como por exemplo, a erupção de determinado vulcão numa data específica, a fotografia de uma supernova, o levantamento das atitudes de uma comunidade; os dados observacionais – por sua natureza instantânea – guardam uma importância crítica que os qualifica como registros históricos, pois não podem ser coletados uma segunda vez e, portanto, devem ser submetidos a processos de curadoria que os preserve para sempre.</p> <ul style="list-style-type: none"> • DADOS COMPUTACIONAIS <p>São resultados da execução de modelos computacionais ou de simulações, seja, por exemplo, no domínio da física ou para a criação de ambientes virtuais culturais ou educacionais. Para esta categoria de dados a preservação por longo prazo pode não ser necessária, posto que os dados podem ser replicados ao longo do tempo. Entretanto, replicar o modelo ou a simulação no futuro pode exigir um grande número de informações que incluem descrição das dependências de <i>hardware</i>, <i>software</i> e outras dependências técnicas, e ainda os dados de entrada. É preciso notar que algumas vezes é mais conveniente preservar somente os dados de saída.</p>
<ul style="list-style-type: none"> • DADOS EXPERIMENTAIS <p>São provenientes de situações controladas em bancadas de laboratórios, como por exemplo, medidas de uma reação química. Em tese, dados experimentais provenientes “de experimentos que podem ser precisamente reproduzidos não necessitam ser armazenados indefinidamente; porém, na prática, nem sempre é possível reproduzir precisamente todas as condições experimentais, particularmente onde algumas variáveis experimentais não podem ser conhecidas e quando os custos de reprodução do experimento são proibitivos”6. A distinção definida por essa categorização é de grande importância na escolha das estratégias de arquivamento e preservação.</p> <p>CLASSIFICAÇÃO DOS DADOS SEGUNDO A SUA NATUREZA:</p> <ul style="list-style-type: none"> • NÚMEROS, IMAGENS, VÍDEOS ou ÁUDIO, SOFTWARE, ALGORÍTIMOS, EQUAÇÕES, ANIMAÇÕES ou MODELOS e SIMULAÇÕES. <p>CLASSIFICAÇÃO DOS DADOS SEGUNDO A FASE DA PESQUISA7</p> <ul style="list-style-type: none"> • DADOS BRUTOS, CRUS ou PRELIMINARES (RAW DATA em inglês) <p>São dados que vêm diretamente dos instrumentos científicos.</p> <ul style="list-style-type: none"> • DADOS DERIVADOS <p>São resultados do processamento ou combinação de dados brutos ou de outros dados.</p> <ul style="list-style-type: none"> • DADOS CANÔNICOS ou DADOS REFERENCIAIS <p>São coleções de dados consolidados e arquivados geralmente em grandes centros de dados, por exemplo, sequência genética, estrutura química, etc. Muitas áreas de pesquisa fazem uso também de dados produzidos por órgãos do governo. Embora estes dados não tenham sido originalmente coletados para fins de pesquisa, eles se tornam dados de pesquisa uma vez que tenham sido modificados, processados ou expandidos.</p>

Fonte: Sayão e Sales (2015).

O planejamento da gestão vai muito além da duração de um projeto, pois é preciso garantir a disponibilidade dos dados pelo maior tempo possível (MEDEIROS, 2018). Sayão e Sales (2015) descrevem 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. O planejamento da gestão é uma etapa elementar do gerenciamento de dados científico e contribui de maneira assertiva para a etapa seguinte que envolve a extração dos dados. Veremos essas etapas nos capítulos seguintes.

No momento, é importante destacar que, como se observa, o campo evidencia riqueza de vocábulos que se prestam à exatidão conceitual, ainda que sob forte entrecruzamento, demonstrando frequente dificuldade – ou, quiçá, impossibilidade - de demarcação de muitos dos seus termos (VIDOTTI et al, 2017). Esse panorama, que também pode refletir um processo de construção da identidade do próprio campo, foi intensamente influenciado pelo movimento Acesso Aberto, que, por sua vez, também propiciou o surgimento dos chamados Repositórios Institucionais. Trataremos disso na próxima seção.

3.2 MOVIMENTO ACESSO ABERTO E REPOSITÓRIOS INSTITUCIONAIS

Esta seção trata da emergência da perspectiva conhecida como Acesso Aberto (*Open Access*) e o surgimento dos repositórios institucionais, articulando ambos os movimentos em torno da expressão *e-Science (1999-2000)*, que instaura uma concepção de ciência orientada a dados e amplia, consideravelmente, a comunicação científica, o compartilhamento e o desenvolvimento de novas pesquisas.

Nas últimas décadas, pesquisadores e profissionais de várias áreas (COSTA, 2008; SEMELER; BLANK; ROZADOS, 2013; SANTOS JUNIOR, 2010; VIDOTTI, CONEGLIAN; ARAKAKI; BRANDT, 2017) têm abordado e discutido a questão do *open access* (acesso aberto) à literatura científica. Têm, do mesmo modo, discutido a criação de repositórios institucionais em universidades e a edição de periódicos científicos em acesso aberto (COSTA, 2008).

O acesso livre à informação científica, revela-se como uma nova tendência de estudos para Comunicação Científica (SEMELER; BLANK; ROZADOS, 2013). Inúmeras iniciativas consolidam as preocupações da comunidade científica com a filosofia open, ou seja, em direção ao uso de ferramentas, estratégias e metodologias para o desenvolvimento de aplicações de computador, como o *Software Livre*, arquivos abertos e acesso aberto. Essas novas iniciativas surgiram com a Internet e atualmente são muito influentes no que se refere aos recentes processos da comunicação científica (MUELLER, 2006 apud, SEMELER; BLANK; ROZADOS, 2013).

O movimento do acesso livre à informação científica surgiu de uma reação contra os aumentos crescentes da assinatura de periódicos científicos na década de 1990, praticados por editoras comerciais, fato que provocou a chamada crise dos periódicos (ARAÚJO, 2012). A esse respeito, Briquet de Lemos informava em 2005 que uma análise de 123 bibliotecas afiliadas à *Association of Research Libraries*, dos Estados Unidos, mostrou que os gastos com aquisição de periódicos, entre 1986 e 2004, subiram 273% (CAPES, 2012). A crise forçou a busca por alternativas ao tradicional processo de divulgação de resultados de pesquisas e engendrou a entrada da internet no sistema de comunicação científica (GUANAES; GUIMARÃE, 2012). Um dos marcos dessa mudança foi o lançamento do repositório de *pré-prints* criado por Paul Ginsparg, o arXiv, em 1991, para abrigar trabalhos de matemática e física enviados por pesquisadores (ARAÚJO, 2012).

O movimento de acesso aberto é algo consolidado e surgiu como uma alternativa para divulgação e amplificação da visibilidade das publicações científicas na *web* (GUANAES, 2015). Uma publicação científica de acesso aberto, define-se pela capacidade de estar disponível online, de forma gratuita e livre de direitos autorais ou restrições de licenciamento (GUANAES, 2015). Os principais veículos de acesso aberto são as revistas científicas e os repositórios de acesso livre (SEMELER; BLANK; ROZADOS, 2013).

O Quadro 2, a seguir, sintetiza os principais marcos desse processo que mudou radicalmente a comunicação científica e a prática de fazer pesquisa em todo o mundo:

Quadro 2: Marcos do movimento de Acesso Livre

PERÍODO	EVENTO
10/1999	Lançamento da Open Archives Initiative, pela Convenção de Santa Fé
2001	Carta aberta da Public Library of Science (PLoS)
14/02/2002	Iniciativa de Budapeste para o Acesso Aberto ¹
30/10/2002	Carta ECHO Charter
11/04/2003	Declaração de Bethesda - Bethesda Statement on Open Access Publishing ² -27/08/2003 Association of Learned and Professional Society Publishers (ALPSP) ;
22/10/2003	Declaração de Berlim sobre o Livre Acesso ao Conhecimento - —Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities ³ ;
11/2003	Declaração de Princípios do Wellcome Trust em apoio à edição em livre acesso
4/12/2003	Posicionamento do InterAcademy Panel sobre o acesso à informação científica
5/12/2003	Declaração do International Federation of Libraries Association (IFLA) sobre o livre acesso à literatura científica e aos documentos da pesquisa
12/12/2003	Declaração de Princípios da Cúpula Mundial sobre a Sociedade da Informação (SMSI)
15/01/2004	Declaração de Valparaíso
30/01/2004	Declaração da Organization for Economic Cooperation and Development (OECD) sobre o acesso aos dados da pesquisa financiada por fundos públicos
16/03/2004	Princípios de Washington D. C. para o Livre Acesso à Ciência
30/07/2004	Publicação do relatório do comitê do Parlamento Britânico sobre edição científica
12/2004	Declaração de Acesso Aberto, de responsabilidade do Australian Research Information Infrastructure Committee (ARIIC)
13/09/2005	Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica
26/09/2005	Declaração de Salvador sobre Acesso Aberto: A Perspectiva dos Países em Desenvolvimento ⁴ Declaração de Salvador – Compromisso com a Equidade ;
12/2005	Carta de São Paulo
2005	European Bureau of Library, Information and Documentation Associations (EBLIDA) publica a obra Statement Towards an Effective Scientific Publishing System for European Research
05/2006	Declaração de Florianópolis
18/07/2006	Carta Aberta à SBPC
2006	European Research Consortium for Informatics and Mathematics (ERCIM) publica a obra Statement on Open Access
05/2007	Projeto de Lei 1120/2007 09/2007 Canadian Institutes of Health Research (CIHR), lança a política de acesso aos resultados das pesquisas
26/12/2007	Lei que obriga o depósito no Repositório de Acesso Aberto PubMed Central (PMC), de todo resultado de pesquisa financiada pelo National Institute of Health (NIH)
04/2008	Publicado pela UNESCO o livro Open access to knowledge and information: scholarly literature and digital library initiatives; the South Asian scenario, de Anup Kumar Das
2009	Distribuição de kits tecnológicos pelo IBICT – Instituto Brasileiro de Informação em Ciência e Tecnologia e FINEP – Financiadora de Estudos e Projetos para as Universidades Brasileiras para construção de repositórios institucionais
06/2011	Publicação da Lei 14/2011 na Espanha ⁴
12/07/2012	Dez anos a partir da Iniciativa de Acesso Aberto de Budapeste: definindo o padrão para abrir

Fonte: Adaptado de COCCO (2012).

Também Fausto (2013) registrou a evolução do movimento de acesso aberto, em *SciELO em Perspectiva*, com destaque para acontecimentos relevantes que remontam desde 4 de julho de 1971 com o lançamento do Projeto Gutenberg. Fausto (2013) destaca ainda:

Quadro 3: Evolução do Acesso Aberto – Breve Histórico

1989: lançada a revista de acesso livre online *Psycology* por Stevan Harnad, que se tornou revisada por pares em janeiro de 1990.

Agosto de 1989: lançada a revista online de acesso livre *The Public-Access Computer Systems Review*, por Charles W. Bailey Jr, que se tornou revisada por pares em abril de 1992.

Outubro de 1990: Tim Berners-Lee escreve a proposta de seu primeiro servidor de web (lançado em março de 1991). Em 12 de novembro de 1991 publica “*World Wide Web: Proposal for a HyperText Project*”, e em 13 de novembro do mesmo ano escreve sua primeira página web.

1991: surge o ArXiv, lançado por Paul Ginsparg.

Abril de 1991: Lançado o EJournal por Edward M. Jennings.

Setembro de 1990: lançada a revista online revisada por pares e de acesso livre *Electronic Journal of Communication*.

Setembro de 1990: lançada a revista online revisada por pares e de acesso livre *Postmodern Culture*, por Eyal Amiran, Greg Dawes, Elaine Orr e John Unsworth.

1991: a revista online de acesso livre *Surfaces* é lançada por Jean-Claude Guédon.

17 de maio de 1991: os padrões para a World Wide Web são liberados pelo CERN e por Tim Berners-Lee.

Abril de 1993: lançado o Projeto MUSE pela Milton S. Eisenhower Library em parceria com a Johns Hopkins University Press, nos EUA. O Projeto Muse não é de acesso livre, mas foi pioneiro na distribuição online. Permite a pesquisa de texto completo gratuito e a editora JHU Press permitiu que os autores mantivessem seus direitos autorais.

1994: a Fundação Nacional para a Ciência (*National Science Foundation -NSF*) e o *U.S. Federal Agencies* lançam a *Digital Libraries Initiative*.

27 de Junho de 1994: Stevan Harnad propõe o auto-arquivamento (*self-archiving*).

Outubro de 1994: Wayne Marr e Michael Jensen lançam a Rede de Pesquisa em Ciências Sociais (*The Social Science Research Network -SSRN*).

28 de novembro de 1994: a revista impressa *Florida Entomologist*, lançada em 1917, converte-se em acesso aberto. A partir de abril de 1999, todos os números anteriores até 1917 foram disponibilizados em acesso livre.

1996: lançada a *Networked Digital Library of Theses and Dissertations (NDLTD)*, pela Virginia Polytechnic Institute and State University.

Junho de 1996: Brewster Kahle lança o *Internet Archive*.

1997: A base Medline, lançada em 1966 pela *National Library of Medicine* torna-se de acesso aberto, incorporada pelo PubMed.

1997: Início do desenvolvimento do SciELO (*Scientific Electronic Library Online*) pelo projeto da FAPESP (Fundação de Apoio à Pesquisa do Estado de São Paulo) em parceria com a BIREME (Centro Latino-americano de Informação em Ciências da Saúde).

26 de junho de 1997: lançamento do PubMed, incorporando o conteúdo da Medline.

Agosto de 1997: Stevan Harnad lança o *CogPrints*.

27 de março de 1998: Declaração de São José Rumo à Biblioteca Virtual em Saúde.

Junho de 1998: lançada a SPARC (*Scholarly Publishing and Academic Resources Coalition*) pela *Association of Research Libraries* (ARL).

1998: A SciELO (*Scientific Electronic Library Online*) começa a operar publicamente.

1998 Charles Oppenheim lança o SPARC Europa (*Scholarly Publishing and Academic Resources Coalition*).

1998: Vitek Tracz funda a primeira editora comercial em Acesso Aberto do mundo: a *BioMed Central* (BMC).

Outubro de 1998: David Shulenburger propõe o Repositório Nacional Eletrônico de Artigos (*The National Electronic Article Repository* -NEAR).

Fonte: Adaptado de Fausto (2013).

Conforme é possível perceber no Quadro 2, o primeiro evento reconhecido oficialmente, no mundo, para discutir a temática aconteceu em outubro de 1999, nos Estados Unidos. Entretanto, segundo Fausto (2013), o lançamento da primeira revista de acesso livre *online Psycology*, em 1989, é um episódio que merece destaque, conforme pode ser visualizado no Quadro 3. No Brasil, a primeira iniciativa foi o Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica²¹, lançado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), em 13 de setembro de 2005.

O termo Acesso Aberto, consagrado em 2002, na Declaração de Budapest²², foi consensualmente definido como acesso ao “[...] conhecimento científico digital, online, livre de custos, e livre de restrições” (PERRIER, 2018, p.38). São publicações disponibilizadas de forma livre pela Internet e que essas devem permitir a qualquer usuário fazer download, copiar, distribuir e imprimir completamente o seu conteúdo (COSTA; LEITE, 2017).

Com a consolidação do Acesso Aberto, o passo seguinte foi o surgimento do modelo *Open Archives* no início da década de 1990 com a iniciativa de repositório de acesso aberto conhecida como *ArXiv* (SANTOS JUNIOR, 2010). Experiência iniciada em 1991, no Laboratório de Los Alamos, com a implementação do

²¹ Disponível em < <https://livroaberto.ibict.br/Manifesto.pdf> >

²² <https://www.aguia.usp.br/apoio-pesquisador/acesso-aberto-usp/entenda-o-que-e-acesso-aberto/>

repositório *ArXiv*. Outros repositórios surgiram ao longo dos anos subsequentes: CogPrints²³ - Psicologia, linguística e neurociências; *Networked Computer Science Technical Reference Library* (NCSTRL²⁴); *Networked Digital Library of Thesis and Dissertations* (NDLTD²⁵), 170 países; *Research Papers in Economics* (RePEc²⁶). Em 1999, surge o *Open Archives Initiative* (OAI) (SANTOS JUNIOR, 2010).

O OAI foi o marco da união entre comunicação científica e a Internet Convenção de Santa Fé²⁷, realizada em 1999 nos EUA na cidade do Novo México. Na oportunidade, se formulou e se pactuou a adoção da iniciativa de arquivos abertos. A OAI surgiu enquanto protocolo para a publicação científica, sendo criada com o objetivo de desenvolver e promover padrões de interoperabilidade entre repositórios digitais, e propor três princípios básicos: auto-arquivamento, a revisão pela comunidade e a interoperabilidade.

Referente à Convenção de Santa Fé, Varela *et al.* (apud COCCO, 2012) comentam que o protocolo OAI foi uma solução para o alto custo das assinaturas das revistas científicas. Após esta convenção, foram realizados vários eventos em favor do acesso aberto de publicações científicas, conforme é possível observar nos quadros 1 e 2. Desse modo, o OAI surgiu

[...] como uma tentativa na solução para o alto custo das assinaturas de periódicos científicos, a lentidão na disseminação da produção científica, as barreiras impostas pela legislação de direitos autorais; e a necessidade de reduzir o fosso científico, tecnológico e cultural que se instalou entre os países ricos, emergentes e pobres, em função das disparidades educacionais e culturais, para as quais contribuem a centralidade, nas regiões mais prósperas, da produção e distribuição do conhecimento, reduzindo a possibilidade do sujeito, das regiões periféricas, ampliar suas competências cognitivas e de aprendizagem, situação que contribui significativamente para acirrar o quadro de exclusão e desigualdade social [...]. (VARELA et al. apud COCCO, 2012, p.26)

O modelo de referência *Open Archival Information System* (OAIS) ou Sistema de Informação de Arquivamento Aberto, tornou-se, em 2003, uma norma

²³ <http://cogprints.org/> ; <https://sistema.bibliotecas-bdigital.fgv.br/bases/cognitive-sciences-eprint-archive-cogprints>

²⁴ <http://www.cs.cornell.edu/NCSTRL/>

²⁵ <http://www.ndltd.org/>

²⁶ <http://repec.org/>

²⁷ Disponível em: http://www.openarchives.org/sfc/sfc_entry.htm

internacional: a ISO 14721:2003²⁸ foi criada com o objetivo de estabelecer um sistema de arquivamento de informações, por meio de um esquema organizacional composto de pessoas que aceitam a responsabilidade de preservar as informações e disponibilizá-las para uma comunidade designada (SOUZA, et. al. , 2012). Contudo, vale salientar que a referida norma foi revisada em 2012 e em 2018²⁹.

Este modelo é um esquema conceitual que disciplina e orienta um sistema para a preservação e manutenção do acesso à informação digital por longo prazo, sendo utilizado cada vez mais por comunidades como uma estrutura para reunir atividades cooperativas entre diferentes áreas de interesse (SOUZA, et. al. , 2012). Seu objetivo é ampliar a compreensão sobre a preservação digital; definir terminologias e conceitos; ampliar o consenso sobre a preservação e acesso à informação digital; e orientar a identificação e o desenvolvimento de padrões (SOUZA, et. al. , 2012).

A sigla OAIS foi traduzida para o português como Sistema Aberto de Arquivamento de Informação (SAAI). O modelo foi desenvolvido para auxiliar na aquisição, preservação e disseminação do conteúdo digital e está direcionado às organizações que têm a responsabilidade de tornar disponível a informação por um longo tempo (THOMAZ; SOARES, 2004).

OAIS é um modelo de referência que define os processos necessários para que se preserve e se chegue aos objetos de forma efetiva e a longo prazo, estabelecendo entre eles uma linguagem em comum. Como o próprio nome diz, trata-se de um modelo e, portanto, não especifica critérios detalhados sobre os processos de preservação. Contudo, o OAIS descreve as funcionalidades básicas e os tipos de informação requeridos para sua preservação. OAIS identifica as obrigatórias responsabilidades, bem como as interações dos produtores, consumidores e gestores dos documentos (tanto em papel como digitais). Contribui, por assim dizer, para um método que busca descrever a funcionalidade de repositório, proporcionando modelos detalhados de informação e funções arquivísticas (ANGUES, 2019).

²⁸ Disponíveis em: < <https://www.iso.org/standard/24683.html> >; <https://www.iso.org/standard/57284.html> >. https://brapci.inf.br/repositorio/2014/11/pdf_20ae7f8b97_0031823.pdf

²⁹ <https://www.iso.org/standard/57284.html>

Conforme o OAIS amplia seu debate na sociedade, outro conceito toma fôlego na comunidade científica, a e-Science. O fenômeno da *e-Science* é caracterizado por uma ciência orientada a dados. Conforme Anjos, Dias e Rodrigues (2017), a *e-Science* cria oportunidades para o desenvolvimento da comunicação científica, uma vez que os dados de pesquisa são considerados fontes primárias de informação, quando compartilhados e acessíveis motivam o surgimento de novas investigações científicas.

O termo 'e-Science' foi pela primeira vez enunciado em estudos “[...]” do Conselho Nacional de Pesquisa do Reino Unido, para designar o modo então emergente de produção do conhecimento científico, o qual se baseava no uso compartilhado de recursos e tecnologias da informação e da comunicação (VISOLI; DRUCKER, 2017, p.35). Ferreira (2018) informa que o referido termo surgiu em 1999, embora ela mesma advirta que outros autores se referem ao ano de 2000³⁰:

O termo e-Science ou (eScience), originário no Reino Unido e predominante no restante da Europa, foi cunhado por John Taylor, no ano de 1999,¹ quando ocupava o cargo de diretor geral do Research Councils UK, conforme Jankowski (2007). Traduzido para o português como e-Ciência, este termo adquiriu um significado que representa a potência da ciência melhorada com o uso intensivo das TICs e sua ampliação em torno de um esforço colaborativo (FERREIRA, 2018, p. 15).³¹

Na *e-Science* nos deparamos com o apoio das Tecnologias Digitais de Informação e Comunicação (TDIC) para com às atividades de investigação científica, como a colaboração no compartilhamento de dados e divulgação dos resultados (RIBES; LEE, 2010 apud ANJOS; DIAS; RODRIGUES, 2017).

A FAPESP define e-Science assim:

³⁰ Conforme Ferreira (2018), em nota de rodapé (Nota 1, p. 15): “Segundo Tolle e colaboradores (2011, p. 235), o termo foi cunhado no ano 2000”.

³¹ Disponível em < <http://books.scielo.org/id/bc84k/pdf/ferreira-9788523218652.pdf> >

[...] é um nome dado em todo o mundo a pesquisas que são realizadas em todas as áreas do conhecimento e que têm necessidade de lidar com grandes volumes de dados ou de usar métodos computacionais sofisticados e computação de alto desempenho. A pesquisa em eScience aborda todas as etapas de um processo de pesquisa -- desde a criação de ferramentas computacionais que ajudem cientistas a formular problemas de pesquisa, coletar e analisar dados, até a modelagem, a simulação, a divulgação e o reuso dos resultados da pesquisa. Mas não basta, para isso, aplicar técnicas computacionais já existentes, uso de tecnologias padrão, prestação de serviços de assistência de informática ou cessão e utilização de dados. Pesquisa em eScience pressupõe trabalho conjunto e multidisciplinar, em que cientistas da Computação auxiliem pesquisadores de outras áreas a desenvolver pesquisas de forma mais rápida e eficiente e que, essa parceria, a Computação também seja feita de forma inovadora, transformadora [...]. (FAPESP, 2019)

Nesse contexto, conceitualmente entre o que é *e-science/cyberinfrastructure* e o que são os dados científicos, percebe-se que também são denominados dados de pesquisa (COSTA, 2017, p.28). Por ser um tema contemporâneo, apresenta questões conceituais que ainda não passaram pelo processo de reflexão necessário ao seu amadurecimento. Sua discussão tem se ampliado conforme evoluem as tecnologias de informação.

Essas novas plataformas e práticas de pesquisa tem sido identificadas por diferentes autores (COSTA, 2017; FERREIRA, 2018) sob o nome de *e-Science* e compreende, além de outros aspectos, o uso intensivo de dados na pesquisa científica, o uso de infraestrutura tecnológica para a colaboração entre cientistas e os processos de colaboração científica em si, mediados pelas tecnologias segundo (APPEL, 2014).

Atrelada a questão do tratamento técnico da informação na *e-Science* surge a curadoria dos dados científicos. De acordo com Costa (2017, p.66), curadoria de dados pode ser entendida como “[...] a gestão e a preservação de dados a longo prazo, incluindo-se nesse contexto o fato de agregar valor aos dados digitais, bem como viabilizar a criação de novos dados [...]”, de forma colaborativa, a partir dos já existentes.

A curadoria de dados envolve a redução de riscos de obsolescência digital, manutenção, aspectos de preservação e agrega valor, tendo lugar de destaque no ciclo de vida dos dados científicos.

Figura 1: Curadoria dos Dados Científicos



Fonte: Costa (2017).

A Figura 1 ajuda a demonstrar o envolvimento da curadoria digital na gestão de dados de pesquisa desde o seu “[...] planejamento, assegurando a sua preservação por longo prazo, descoberta, interpretação e reuso” (SAYÃO; SALES, 2012, p. 179).

Desse modo, a curadoria digital tem papel relevante no contexto dos repositórios institucionais (RI) de acesso aberto e dados de pesquisa, por apoiarem o tratamento da informação e melhor aproveitamento dos dados.

Os RI são uma alternativa para a comunicação da ‘Ciência Livre’ de barreiras de acesso sendo um serviço de informação científica. Lidam com a produção intelectual de uma instituição e pode ser considerado um tipo de biblioteca digital, mas o contrário não se estabelece pois nem toda biblioteca digital pode ser considerada um repositório institucional. Pode-se, por assim dizer, que os RI visam o controle bibliográfico, a gestão da informação e a devida ampliação e sistematização da comunicação científica.

Há certa confusão entre os termos repositório institucional, repositório digital e biblioteca digital. Bibliotecas digitais possuem várias nomenclaturas para designar o mesmo objeto (biblioteca virtual, biblioteca eletrônica, dentre outros) mas, do ponto de vista conceitual, bibliotecas digitais e repositórios digitais são conceitos sinônimos, seria apenas uma nova nomenclatura para um velho conceito. Repositórios institucionais, repositórios digitais, repositórios científicos, repositórios

temáticos, entre outros, são termos que também evidenciam acentuado intercruzamento de significados, a despeito das diferentes nomenclaturas adotadas na literatura e referências documentais. Nesta dissertação, são tomados como sinônimos, com a ressalva de que nosso foco se concentra nos repositórios *institucionais* por caracterizar melhor, a nosso ver, o acolhimento às diversas áreas de conhecimento (diferentemente dos repositórios temáticos, por exemplo) e, assim, contemplar, com mais propriedade, as instituições universitárias.

De acordo com o IBICT (2018), os repositórios digitais (RDs) são bases de dados online que reúnem de maneira organizada a produção científica de uma instituição ou área temática³². Desse modo, assim se caracterizam como RDs institucionais ou RDs 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. (IBICT, 2018)

Registramos, outrossim, que a complexidade do campo terminológico (como visto mais amiúde no Capítulo anterior) também ocorre com relação à definição dos repositórios digitais, de modo que, esclarecemos que, nesta dissertação, seguimos o posicionamento do IBCT tanto sobre o significado atribuído a repositórios digitais como a respeito de sua classificação como institucionais ou temáticos. Destacamos, também, que ambas modalidades participam do movimento Acesso Aberto, disponibilizando dados de forma livre.

Os RI seguem a filosofia preconizada pela Iniciativa dos OAI que defende que todos os materiais de pesquisa devem estar disponibilizados publicamente na WEB, sem restrições de acesso, sobretudo as pesquisas desenvolvidas com recursos oriundos de agências públicas de fomento à pesquisa como CAPES, CNPq, FINEP, entre outros (SANTOS JUNIOR, 2010, p.31).

Inseridos no coração do movimento mundial em favor do acesso aberto à informação científica, os RI constituem de fato inovação no sistema de comunicação

³² Repositórios Digitais - <http://www.ibict.br/informacao-para-a-pesquisa/repositorios>.

da ciência e no modo como a informação – aquela que alimenta e resulta das atividades acadêmicas e científicas – é gerenciada. (LEITE, et al, 2012)

Um repositório institucional de acesso aberto constitui um serviço de informação científica - em ambiente digital e interoperável - dedicado ao gerenciamento da produção científica e/ou acadêmica de uma instituição (universidades ou institutos de pesquisa). Contempla a reunião, armazenamento, organização, preservação, recuperação e, sobretudo, a ampla disseminação da informação científica produzida na instituição.

Alguns exemplos de conteúdos que podem ser armazenados em um repositório digital são arquivos em XML, texto em PDF e DOC, áudios, vídeos, hipermídias, dentre outros. Não diferente de outras áreas do saber, também os campos da Educação têm produzido exponencialmente grandes volumes de conteúdos, provenientes de processos de pesquisa, nos seus diferentes formatos (NEVES, 2018).

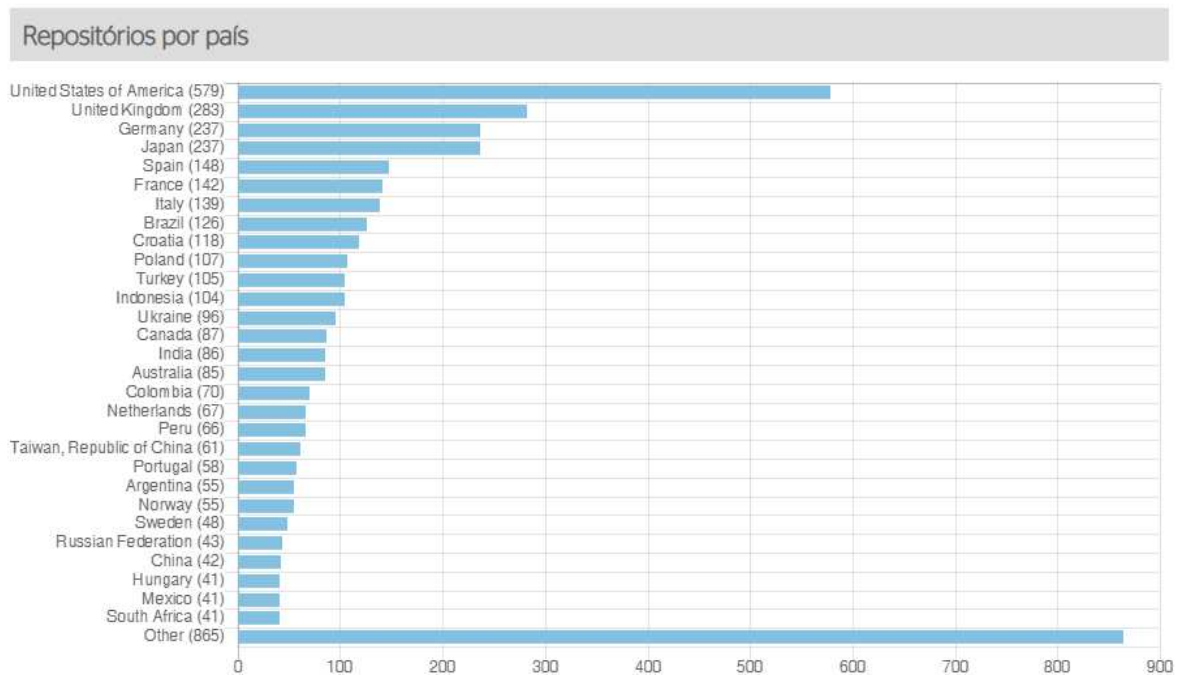
Repositório institucional é um conjunto de serviços que a universidade oferece para os membros de sua comunidade para o gerenciamento e a disseminação de conteúdos digitais, criados pela instituição e membros da sua comunidade. (LYNCH, 2003 apud VOLPATO; RODRIGUÊS; SILVEIRA, 2014).

É essencialmente um compromisso organizacional com a gestão, desses conteúdos digitais, inclusive preservação de longo prazo, quando apropriado, bem como organização e acesso ou distribuição (LYNCH, 2003 apud VOLPATO; RODRIGUÊS; SILVEIRA, 2014).

A forma como algumas universidades no mundo disponibiliza a sua produção científica com acesso livre é através dos seus repositórios digitais que são bases de dados online que reúnem de maneira organizada a produção científica de uma instituição ou área temática. O desenvolvimento de Repositórios Institucionais surgiu em 2002, como uma nova estratégia que permitiu às universidades assumirem o papel de editoras, modernizando os processos de publicação e divulgando a produção acadêmica em conteúdo digital (LYNCH, 2003 apud VOLPATO; RODRIGUÊS; SILVEIRA, 2014). Os RDs armazenam arquivos de diversos formatos e 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.

Apesar dos RDs estarem implantados de forma significativa nas universidades no Brasil, conforme dados do site OpenDOAR destacados no Gráfico 02, são 126 Repositórios Institucionais em funcionamento até o dia de 18/07/2019. Tradicionalmente os pesquisadores só publicam o documento final do seu trabalho de conclusão de curso, seja de graduação, mestrado ou doutorado. Os dados gerados ao longo do processo ou percurso da pesquisa, denominados nesta pesquisa como Dados Suplementares de Pesquisa (DSP) são normalmente descartados ou esquecidos após a conclusão da investigação e da consequente publicação dos artigos científicos resultantes do trabalho. Tipicamente DSP podem ser bases de dados, imagens, vídeos, documentos etc.

Gráfico 1: Repositórios digitais por país



Fonte: Dados do levantamento realizado no OPENDOAR (2019)

Destacamos que a preservação de documentos digitais é uma atividade diferente de preservar documentos analógicos devido às complexidades e especificidades dos digitais, porém de igual relevância social, cultural, informativa e histórica. Inicialmente a preservação de documentos digitais deverá ser a definição de políticas e, a partir desta, elaborar o plano de preservação contendo as estratégias de preservação com eficácia comprovada, e que sejam de conhecimento do acervo. Durante a elaboração do plano de preservação deve-se ter em mente as propriedades significativas dos documentos que se queiram preservar, da mesma

forma, deve-se atentar para os conceitos de forma fixa e conteúdo estável (SANTOS; FLORES, 2015).

O RI serve como um ambiente para reunir acervos, serviços e pessoas para estabelecer um ciclo de vida completo de criação, disseminação, uso e preservação de dados, informações e conhecimentos (VOLPATO; RODRIGUÊS; SILVEIRA, 2014).

Um contraponto a muitas revistas científicas que limitam o acesso ao seu acervo são as iniciativas de criação dos repositórios nas universidades. Cada tipo de repositório digital possui funções específicas e aplicações voltadas para o ambiente no qual será utilizado. A literatura tem conceituado de forma variada a palavra repositório institucional, podendo ser empregada para designar os tipos variados de aplicações de provedores de dados de acesso aberto destinados ao gerenciamento de informação científica. Leite (2009) classifica os repositórios digitais em:

- 1) **repositórios institucionais**: voltados à produção intelectual de uma instituição, especialmente universidades e institutos de pesquisa. Exemplo: e-Prints Soton — repositório de Pesquisa da Universidade de Southampton (<http://eprints.soton.ac.uk/>);
- 2) **repositórios temáticos ou disciplinares**: voltados a comunidades científicas específicas. Tratam, portanto, da produção intelectual de áreas do conhecimento em particular. Exemplo: E-LIS — EPrints in Library and Information Science (<http://eprints.rclis.org/>) e arXiv.org (<http://arxiv.org/>);
- 3) **repositórios de teses e dissertações (Electronic Theses and Dissertation — ETDs)**: repositórios que lidam exclusivamente com teses e dissertações. Muitas vezes a coleta das muitas ETDs é centralizada por um agregador. Exemplo: BDTD/UnB - Biblioteca Digital de Teses e Dissertações da Universidade de Brasília (<http://bdtb.bce.unb.br/tesesimplificado/>) e BDTD (<http://bdtb.ibict.br/>).

Um repositório de dados funciona mediante processos de indexação que agregam valor ao conteúdo disponível ao invés de simplesmente guardá-lo. É importante que os pesquisadores entendam a diferença entre o armazenamento simples e o arquivo de dados científicos corretamente. Desse modo, no armazenamento simples os dados são disponibilizados sem o rigor e a preocupação das etapas que respeitam o ciclo de vida dos dados dentro dos elementos relevantes *e-Science*.

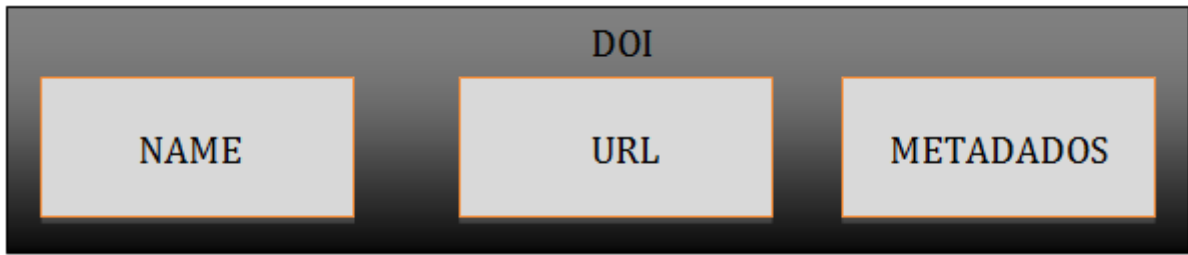
Os repositórios científicos digitais foram criados para facilitar o acesso à produção científica. São bases de dados desenvolvidas para reunir, organizar e tornar mais acessível a produção científica dos pesquisadores (LEITE, et al, 2012).

Por meio do desenvolvimento e da defesa de uma estrutura para o gerenciamento de dados de pesquisa específicos da disciplina, a *Science Europe* está analisando os requisitos mínimos para repositórios de dados de pesquisa que podem ser usados por financiadores e organizações de pesquisa (RE3DATA, 2017). Estes podem ser usados em conjunto com os Protocolos de Dados de Domínio que os pesquisadores individuais de diferentes comunidades científicas poderão usar como modelos para seus planos de gerenciamento de dados.

Em seu recente Workshop em Bruxelas, em janeiro de 2016, o Grupo de Trabalho sobre Dados de Pesquisa da *Science Europe* delineou seus primeiros resultados com requisitos que enfatizam identificadores persistentes, metadados, acesso e uso de dados, legibilidade de máquina e preservação a longo prazo. Os próximos passos incluem a construção de consenso sobre a abordagem e os requisitos, bem como a implementação com registros, sugerindo RE3DATA.

Usando RE3DATA, os repositórios de dados de todos os domínios e de todos os países podem ser pesquisados por assunto, país ou tipo de conteúdo e pesquisados por qualquer combinação de 41 atributos diferentes. Um sistema de ícones torna fácil para um pesquisador identificar as principais características de repositórios, como aqueles que têm acesso aberto, foram certificados, aceitam depósitos de dados e emitem DOI (RE3DATA, 2019).

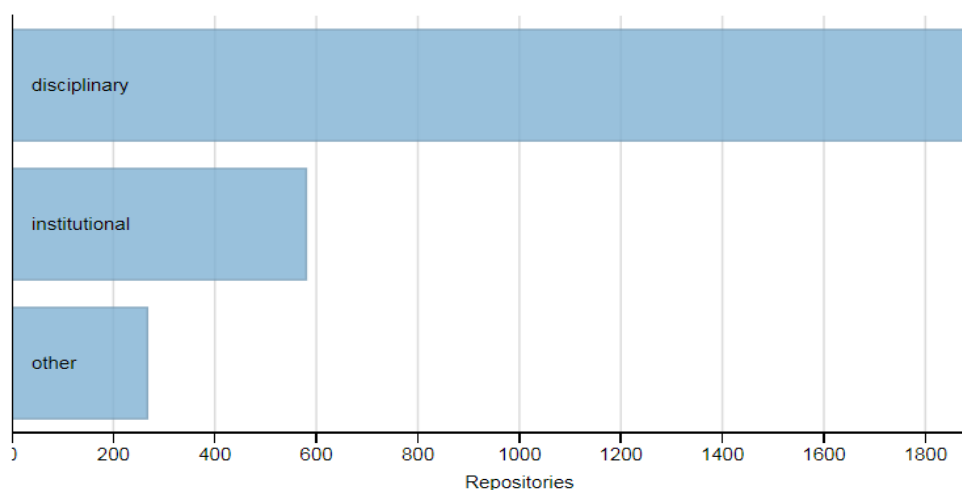
As entradas do registro são selecionadas por um conselho editorial internacional com novas adições e atualizações toda semana. O RE3DATA é um serviço da *DataCite*, uma organização global sem fins lucrativos que fornece identificadores persistentes (DOIs) para dados de pesquisa com o objetivo de ajudar a comunidade de pesquisa a localizar, identificar e citar dados de pesquisa com confiança. O DOI é o *Digital Object Identifier* (DOI). Trata-se de um conjunto alfanumérico atribuído para identificar um objeto de forma exclusiva. Ele está ligado a uma descrição de metadados do objeto, bem como a um local digital, como um URL, onde todos os detalhes sobre o objeto estão acessíveis.

Figura 2: Estrutura do DOI

Fonte: Elaboração própria (2018) com base no DataCite.

Desse modo, o RE3DATA está ligado a Science Europe, uma das maiores autoridades sobre gestão de dados de pesquisa, e se trata de uma associação que promove os interesses coletivos de organizações de financiamento de pesquisa e organizações de pesquisa mundiais (CUNHA; FREITAS; FIALHO, 2019). Sua Assembléia Geral Fundadora aconteceu em Berlim em outubro de 2011 (EUROPEAN COMMISSION, 2013 apud CUNHA; FREITAS; FIALHO, 2019).

O Gráfico 2 a seguir mostra o quantitativo de repositórios registrados na plataforma RE3DATA em junho de 2017. Ao oferecer informações detalhadas sobre mais de 2.000 repositórios de dados de pesquisa, o RE3DATA se tornou a fonte mais abrangente de referência para infraestruturas de dados de pesquisa globalmente.

Gráfico 2: Tipos de repositórios registrados no re3data.org

Fonte: Dados levantados no RE3DATA (2019)

Dados extraídos do OPENDOAR³³, que é o diretório global de repositórios acadêmicos de acesso aberto, permitem a identificação, navegação e busca de repositórios, com base em uma variedade de recursos, como localização, *software* ou tipo de material mantido.³⁴

Os dados apresentados a seguir mostram o panorama e a realidade dos Repositórios Digitais no mundo. No levantamento realizado, identificamos que o sistema mais utilizado no mundo é o *DSpace*³⁵, desenvolvido pelo Instituto de Tecnologia de Massachusetts (MIT). É um *software* de código-fonte aberto que fornece facilidades para o gerenciamento de acervo digital, utilizado para implementação de repositórios institucionais, com suporte a uma grande variedade de tipos de documentos, tais como: livros, teses e dissertações, fotografias, filmes, áudio e outros (DSpace, 2019). Nos dados apresentados no Gráfico 3, observamos que ao *DSpace* correspondem 43% do total, seguidos pelo sistema *EPrints*, desenvolvido pela *School of Electronics and Computer Science na University of Southampton*³⁶, concebido com objetivo de criar um repositório *web* altamente configurável, sendo geralmente aplicado não somente como um repositório aberto para pesquisa de artigos, mas também para outros tipos de arquivo como imagens, dados, áudio, ou qualquer item que possa ser armazenado digitalmente. (MARTINS; SILVA; SIQUEIRA 2018).

Gráfico 3: Software acesso livre mais utilizados no mundo



Fonte: Dados levantados no OPENDOAR (2019)

³³ <https://v2.sherpa.ac.uk/opensoar/>

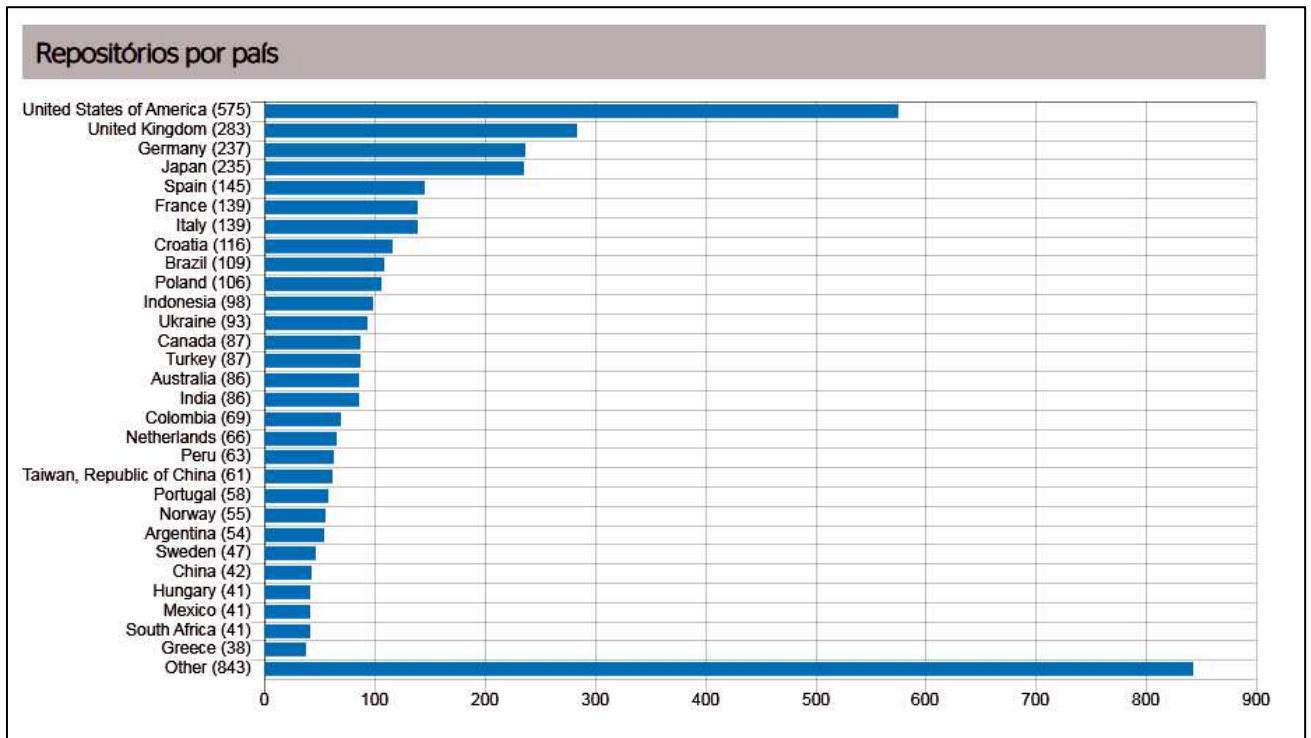
³⁴ Fonte: <http://v2.sherpa.ac.uk/opensoar/information.html>

³⁵ <https://dspace.mit.edu/>

³⁶ <http://eprints.soton.ac.uk/>

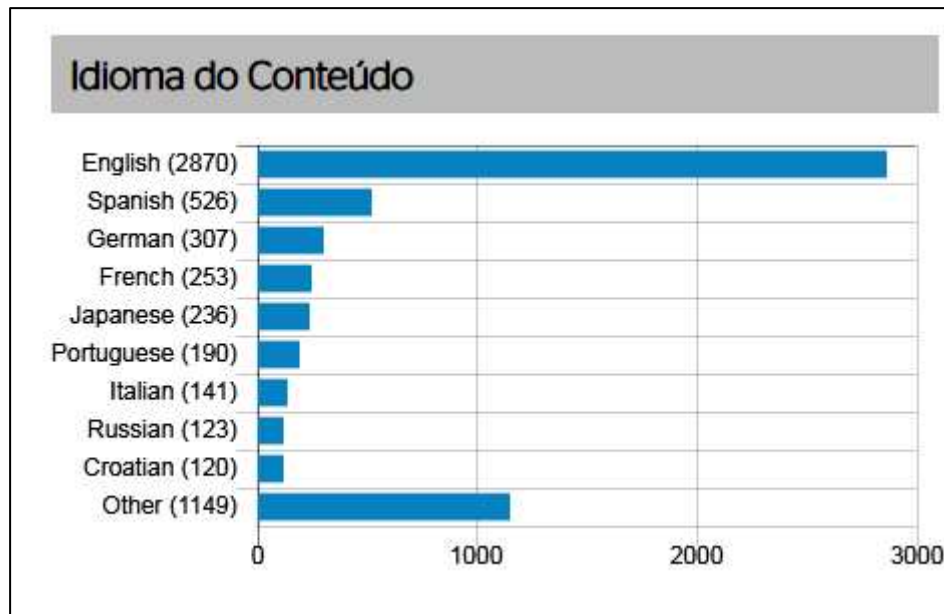
Além dos *softwares* mais utilizados, identificamos os países com maior número de Repositórios Digitais de Acesso Livre, conforme dados do Gráfico 4, liderados por Estados Unidos, Inglaterra e Alemanha, o Brasil ocupa a 9ª posição.

Gráfico 4: Repositórios por país



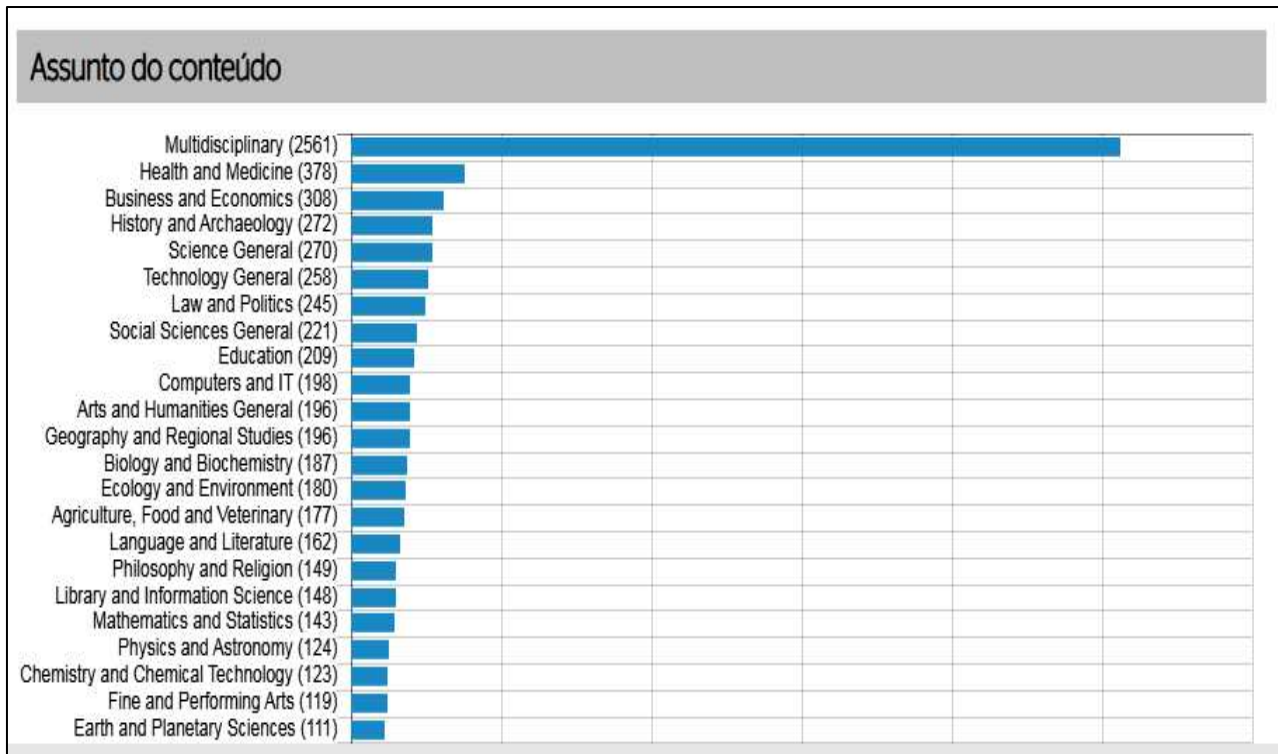
Fonte: Dados levantados no OPENDOAR (2019).

Os idiomas dos conteúdos disponibilizados também são apresentados, conforme Gráfico 5 a maior parte do conteúdo encontrado está no idioma Inglês (com mais da metade do conteúdo), seguido pelo espanhol, alemão, francês, japonês e português. Em razão do exposto, “[...] é prudente refletir sobre a atual estrutura dos dados científicos no Brasil e como essa gestão tem evoluído” (COSTA, 2017). O crescimento dos pesquisadores brasileiros nos repositórios mundiais mostra sinais de se repensar sua implementação com base em planejamentos bem elaborados. Aspecto que demonstra a urgência no estabelecimento de uma política que inclua critérios, regulamentos e documentação claros sobre a gestão dos dados científicos.

Gráfico 5: Idiomas

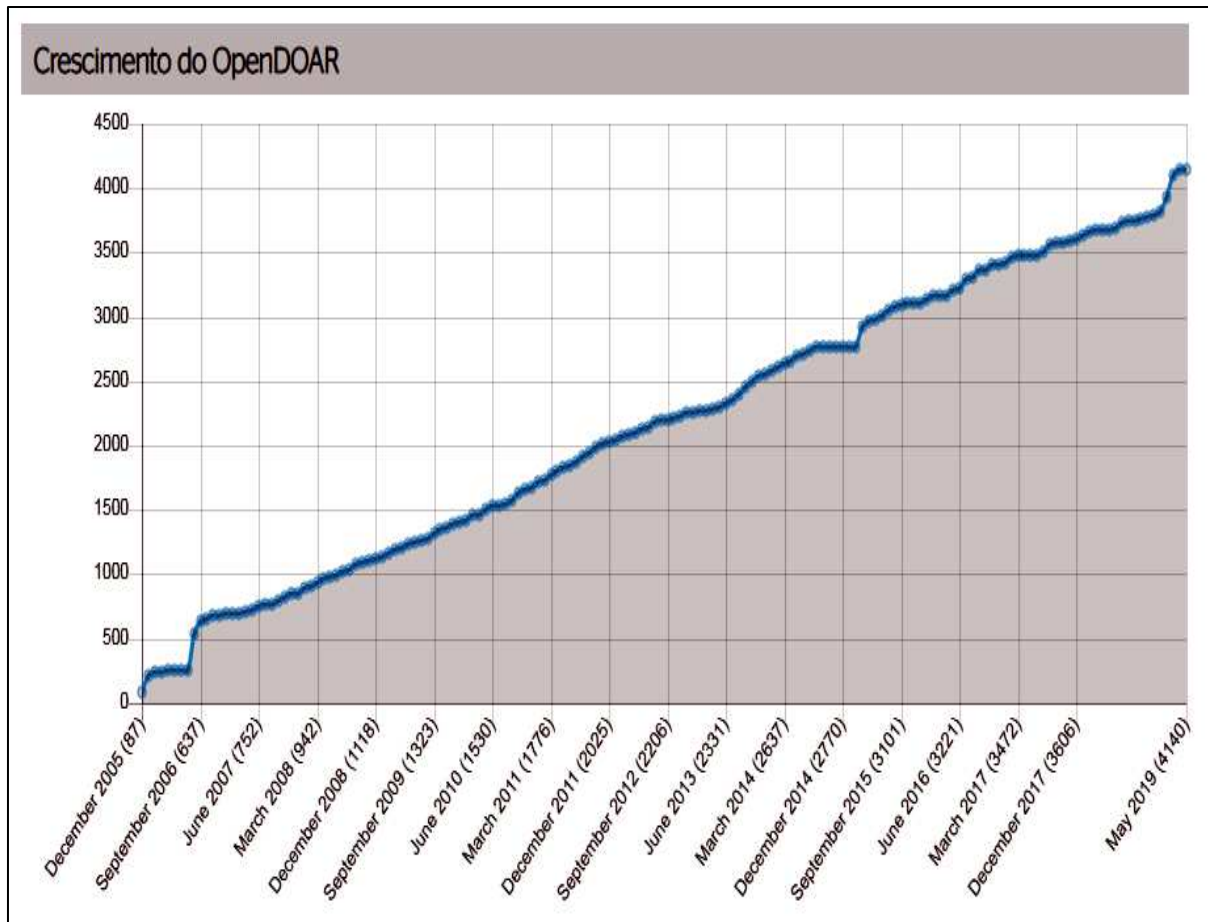
Fonte: (OPENDOAR, 2019)

No Gráfico 6 identificamos os assuntos com maior número de ocorrências. Sendo a área Multidisciplinar com maior número de ocorrências, seguidos da área de saúde, economia e arqueologia respectivamente. Importante destacar que a área Multidisciplinar tem mais da metade das ocorrências e a área de Educação está 9ª colocação em termos de assuntos disponibilizados, segundo os dados obtidos pelo OPENDOAR.

Gráfico 6: Temática dos assuntos dos repositórios

Fonte: (OPENDOAR, 2019)

O Gráfico 7 mostra a evolução do número de Repositórios de Acesso Livre desde o ano de 2015, com 87 registros até o ano de 2019 com 4140 registros até a data deste levantamento,

Gráfico 7: Evolução do número de repositórios

Fonte: (OPENDOAR, 2019)

Este capítulo abordou o referencial terminológico-conceitual aqui considerado como o mais significativo para a problematização desta dissertação. Começou por destacar os conceitos básicos que norteiam a gestão de dados científicos, passando pelas definições e os marcos do movimento de acesso aberto. Apontou ainda os principais modelos de referência para o arquivamento aberto, a e-Science e, por fim, a classificação dos repositórios com ênfase para o RI.

4 REPOSITÓRIOS INSTITUCIONAIS COMO FERRAMENTA DE APOIO AO GERENCIAMENTO DE DADOS CIENTÍFICOS

No capítulo anterior, foram abordados os conceitos básicos relativos à gestão de dados científicos e repositórios institucionais, dois fenômenos que caracterizam as profundas mudanças decorrentes da necessidade de lidar com grandes volumes de dados, informações e conhecimentos. Ao dissertar sobre o tema, verificamos a diversidade terminológica que, de certo modo, imprime razoável complexidade aos estudos dedicados ao campo. Prosseguindo, este Capítulo se dedica ao gerenciamento de dados científicos por via de repositórios institucionais. Nessa perspectiva, procuramos apresentar exemplos de repositórios institucionais no mundo, no Brasil e na Bahia, que disponibilizam acesso a dados de pesquisa.

As seções deste Capítulo estão assim constituídas:

- 4.1 Repositórios institucionais no mundo que disponibilizam acesso aos dados da pesquisa
- 4.2 Repositórios institucionais no Brasil que disponibilizam acesso aos dados da pesquisa

4.1 REPOSITÓRIOS INSTITUCIONAIS NO MUNDO QUE DISPONIBILIZAM ACESSO AOS DADOS DA PESQUISA

Para o levantamento realizado e apresentado neste estudo foi utilizado o RE3DATA para identificar repositórios no mundo que fazem a gestão de dados científicos. Assim, o RE3DATA tanto fez parte do levantamento, a seguir, como foi, também, escolhido como a base de consulta desta seção, uma vez que se trata de um grande repositório mundial, que abriga outros repositórios de características temáticas e institucionais. Por esta razão, ele se constitui um interessante fornecedor de dados para alguns pontos que este estudo pretende discutir, graças ao seu volume e fornecimento de métricas variadas correspondente ao seu conteúdo.

Desse modo, foram destacados os seguintes repositórios:

A) - RE3DATA.ORG

Site: <https://www.re3data.org/about> - Europeu

O RE3DATA é um registro global de repositórios de dados de pesquisa que abrange repositórios de dados de pesquisa de diferentes disciplinas acadêmicas. Apresenta repositórios para o armazenamento permanente e acesso de conjuntos de dados a pesquisadores, órgãos financiadores, editores e instituições acadêmicas (RE3DATA, 2019, tradução livre).

O RE3DATA promove uma cultura de compartilhamento, maior acesso e melhor visibilidade dos dados da pesquisa. O registro foi lançado no outono de 2012 e é financiado pela Fundação Alemã de Pesquisa (DFG) (RE3DATA, 2019, tradução livre).

B) - REPOSITÓRIO CIENTÍFICO DO INSTITUTO POLITÉCNICO DE CASTELO BRANCO

<https://repositorio.ipcb.pt/> - Portugal

O Repositório Científico do Instituto Politécnico de Castelo Branco tem como missão disponibilizar e permitir o livre acesso à produção científica do Instituto Politécnico de Castelo Branco (IPCB). Para cumprir a missão elegeu como objetivos: divulgar a produção científica do IPCB; preservar em suporte digital a sua memória intelectual; partilhar o conhecimento científico e técnico alcançado contribuindo para a geração de novos conhecimentos; aumentar o impacto da produção científica nos meios académico e de investigação. As obras depositadas no Repositório Científico do IPCB estão abrangidas pela licença pública *Creative Commons* e encontram-se em acesso livre (RE3DATA, 2019).

C) - ARQUIVO PORTUGUÊS DE INFORMAÇÃO SOCIAL (APIS)

<http://www.apis.ics.ulisboa.pt/> - Portugal

O Arquivo Português de Informação Social (APIS) é uma infraestrutura científica para as ciências sociais cuja principal atividade consiste na preservação e disseminação de dados de investigação. Com sede no Instituto de Ciências Sociais da Universidade de Lisboa, o arquivo promove a angariação e a partilha de dados de investigação em formato digital com vista à sua consulta pública, análise secundária

e uso pedagógico. Do seu acervo fazem parte um conjunto de bases de dados resultantes de projetos de investigação realizados pela comunidade científica nacional. (RE3DATA, 2019, tradução livre).

D) - DIRETÓRIO DE ACESSO ABERTO (OAD)

http://oad.simmons.edu/oadwiki/Main_Page

O Open Access Directory (OAD) é um compêndio de listas factuais simples sobre o acesso aberto (OA) à ciência e à bolsa de estudos, mantido pela comunidade OA em geral. Reunindo muitas listas relacionadas ao OA em um único local, o OAD facilita a descoberta por todos, as utiliza como referência e as atualiza. Quanto mais fáceis de manter e descobrir, mais eficazmente poderão divulgar informações úteis e precisas sobre a OA. Para ver o que temos, navegue no índice abaixo, navegue na tabela de categorias ou use a caixa de pesquisa perto do canto superior direito. Para ajudar a causa, basta se registrar e começar a editar . Se você tiver alguma dúvida, consulte nossa seção de ajuda ou envie-nos uma linha .

OAD é um *Wiki* e contamos com nossos usuários para manter essas listas precisas, abrangentes e atualizadas. Nosso objetivo é que a própria comunidade de AA mantenha as listas com pouca intervenção dos editores ou do conselho editorial (RE3DATA, 2019, tradução livre).

E) - SERVIÇO NACIONAL DE DADOS DA AUSTRÁLIA (ANDS)

<http://www.ands.org.au/> - Austrália

O ANDS é um esforço colaborativo que visa auxiliar na conexão de vários centros de dados, repositórios e coleções on-line australianos e internacionais para facilitar a descoberta de dados, pesquisas interdisciplinares e fluxos de trabalho entre repositórios (RE3DATA, 2019, tradução livre).

F) - DATA ARCHIVING AND NETWORKING SERVICES (DANS)

<https://dans.knaw.nl/em> - Holanda

DANS (arquivamento de dados e serviços em rede) O DANS, ou Data Archiving and Networking Services, é um instituto holandês dedicado ao armazenamento e compartilhamento de artes, humanidades e ciências sociais. Está pressionando o acesso de dados abertos a dados científicos; Para atingir esse objetivo, estimular outros a adotarem a mesma filosofia, trabalhando de perto com pesquisadores e

gerentes de dados. Ele gerencia uma plataforma de depósito de dados, o EASY, fornece links para grandes gerentes de dados e está envolvido em vários projetos de retro-arquivamento, que são executados em uma tarefa. Seu objetivo final é tornar os dados mais antigos (o que é de difícil acesso) publicamente disponíveis. Finalmente, aproveitando os contatos internacionais, a DANS garante que os pesquisadores holandeses possam obter acesso a bancos de dados de fora da Holanda. (RE3DATA, 2019, tradução livre).

G) - E-PUBLICATIONS@BOND

<https://epublications.bond.edu.au/about.html> - Austrália

E-publications@bond é um repositório de acesso aberto que contém uma grande variedade de pesquisas e publicações acadêmicas de acadêmicos e estudantes da Bond University. O repositório contém saídas em uma infinidade de formatos, incluindo vários dados de pesquisa, galerias de imagens, cartazes, arquivos de áudio e artigos publicados em periódicos revisados por pares de acadêmicos de Bond e outras instituições australianas e internacionais. O repositório é uma representação do crescimento da pesquisa em Bond com a inclusão da coleção de teses e é também uma importante plataforma para material histórico e de arquivo sobre a Universidade. (RE3DATA, 2019, tradução livre).

H) - CURADORIA DE DADOS CIENTÍFICOS NA U.PORTO

<http://sciencedata.up.pt/doc/doku.php> - Portugal

Curadoria de Dados Científicos na página inicial do projeto U.Porto . Trata-se de uma experiência desenvolvida na U.Porto , que teve como objetivo identificar as necessidades de curadoria de dados científicos na universidade e os principais problemas no estabelecimento de uma plataforma multidisciplinar de curadoria de dados. (RE3DATA, 2019, tradução livre).

I) - REPOSITÓRIO DE DADOS DO GOVERNO AUSTRALIANO

<https://data.gov.au/> - Austrália

Fornecer uma maneira fácil de encontrar, acessar e reutilizar conjuntos de dados públicos do governo. O principal objetivo do site é incentivar o acesso público e a reutilização de dados públicos. Foi criado seguindo a Declaração de Governo

Aberto do Governo e como resposta ao Relatório do Grupo de Trabalho do Governo 2.0. (RE3DATA, 2019, tradução livre).

J) - REPOSITÓRIO CIENTÍFICO DE ACESSO ABERTO DE PORTUGAL (RCAAP)

<https://dados.rcaap.pt/> - Portugal

Repositório Científico de Acesso Aberto de Portugal (RCAAP) visa aumentar a visibilidade, acessibilidade e difusão dos resultados da atividade académica e de investigação científica nacional, facilitar o acesso à informação sobre a produção científica nacional em regime de acesso aberto, bem como integrar Portugal num conjunto de iniciativas internacionais neste domínio.

O projeto RCAAP é uma iniciativa da UMIC – Agência para a Sociedade do Conhecimento, IP concretizada pela FCCN – Fundação para a Computação Científica Nacional, disponibilizando mais um serviço avançado sobre a Rede Ciência, Tecnologia e Sociedade (RCTS) gerida pela FCCN. A execução do projeto conta ainda com a participação científica e técnica da Universidade do Minho. (RE3DATA, 2019, tradução livre).

L) - ZENODO

<http://zenodo.org/> - Europeu

Construído e desenvolvido por pesquisadores, para garantir que todos possam participar do Open Science. O projeto OpenAIRE, na vanguarda do acesso aberto e dos movimentos de dados abertos na Europa, foi encomendado pela Comunidade Europeia - CE para apoiar sua nascente política de dados abertos, fornecendo um repositório de todos os recursos para a pesquisa financiada pela CE. O *European Organization for Nuclear Research* (CERN) da Organização Europeia para a Pesquisa Nuclear, um parceiro OpenAIRE e pioneiro em *open source*, acesso aberto e dados abertos, forneceu essa capacidade e o Zenodo foi lançado em maio de 2013. Em apoio ao seu programa de pesquisa, o CERN desenvolveu ferramentas para gerenciamento de Big Data e ampliou as capacidades da Biblioteca Digital para Open Data. Através do Zenodo, essas ferramentas da Big Science poderiam ser efetivamente compartilhadas com a pesquisa de cauda longa. (RE3DATA, 2019, tradução livre).

Um dos principais problemas de gestão de dados científicos está relacionado com a falta de continuidade do registro de dados decorrente da saída de algum membro do projeto ainda em andamento. Por isso é normal que surjam dificuldades de reutilizar um conjunto de dados, uma vez que, sem a devida documentação, torna-se difícil compreender como, quando e por que os dados foram capturados (CORRÊA, 2016).

De acordo com Corrêa (2016, p. 391), gradualmente, o planejamento de gestão de dados científicos está se tornando um requisito dos financiadores e instituições “[...] que pagam para coleta de dados, uma vez que demonstram preocupação com a disponibilidade dos dados no futuro, como se pode verificar as chamadas, por exemplo, o programa Horizonte 2020, em que a Comissão Europeia lançou um projeto piloto chamado *Open Research Data Pilot* [...]” para promover e otimizar a gestão e reutilização de dados de pesquisa gerados pelos projetos que financia.

Como é possível perceber, a gestão de dados, desde o momento da elaboração da pesquisa, tem sido de extrema importância para requisitar financiamento em instituições de pesquisa internacionais. Assim, questiona-se até que ponto os pesquisadores brasileiros têm conhecimento dessa demanda. A literatura científica (LEITE, 2012; CORRÊA, 2016) tem destacado que somente ao se deparar com processos seletivos, em mecanismos científicos estrangeiros, que os pesquisadores nacionais têm verificado a relevância do planejamento da gestão de dados científicos (CUNHA; JORGE; FIALHO, 2019).

A gestão de dados é a reunião de uma série de pequenos passos que se somam às boas práticas. É aconselhável trabalhar em uma prática por vez até que uma correta gestão de dados científicos se torne um hábito constante (CORRÊA, 2016). Por isso, os repositórios científicos digitais têm sido um recurso interessante para a preservação, disseminação e armazenamento de massas de dados. Os repositórios são unidades de informação digitais específicas que funcionam por meio de conjuntos de sistemas complexos e interoperáveis (CUNHA; JORGE; FIALHO, 2019).

Para a efetiva gestão de dados de pesquisa, o planejamento é uma fase essencial. Ele se inicia quando a pesquisa ainda está sendo delineada e deve considerar como os dados serão gerenciados durante o desenvolvimento do projeto e como eles serão compartilhados depois. Dessa forma é necessário formalizar as

ações e compromissos que serão estabelecidos em relação aos dados desde os seus primeiros estágios (SAYÃO; SALES, 2015).

O Plano de Gestão de Dados de Pesquisa (PGD) descreve o ciclo de vida de gestão para todos os dados que serão coletados, de uma forma geral, que dados serão processados, coletados ou gerados; quais as metodologias e padrões que serão utilizados; sob que condições esses dados serão compartilhados e/ou tornados abertos para a comunidade de pesquisa; e como eles serão curados e preservados (VIDOTTI, et al, 2017).

Para a boa gestão dos dados de pesquisa, faz-se necessário também o PGD e agências públicas e privadas de apoio à pesquisa científica dos Estados Unidos, da Europa e Austrália, é parte do planejamento de projeto e integra a documentação de pesquisa. Significa que atualmente, é necessário não apenas explicitar o escopo e objetivo da pesquisa, como também os dados que serão produzidos, como serão tratados, armazenados, preservados e disseminados. Sendo assim, precisamos saber como gerenciar nossos dados de pesquisa hoje, agora mesmo! (DUDZIAK, 2018).

De acordo com Silva (2016, p. 391):

Gradualmente o planejamento de gestão de dados científicos esta se tornando um requisito dos financiadores e instituições que pagam para coleta de dados, uma vez que demonstram preocupação com a disponibilidade dos dados no futuro, como se pode verificar as chamadas, por exemplo, o programa Horizonte 2020, em que a Comissão Europeia lançou um projeto piloto chamado "Open Research Data Pilot" para promover e otimizar a gestão e reutilização de dados de pesquisa gerados pelos projetos que financia.

Para alguns pesquisadores (SAYÃO; SALES, 2015; MEDEIROS, 2018) que escrevem sobre o tema, o PGD é um documento essencial no cotidiano dos pesquisadores e das agências de fomento, haja vista das crescentes exigência feita por diversas agências financiadoras de pesquisa para que um plano de gestão e de compartilhamento de dados faça parte dos pedidos de auxílio, para a liberação de financiamento de projetos de pesquisa.

Quadro 4: Dez razões para a criação de um plano de gestão de dados de pesquisa**SEGUNDO A COMISSÃO DE ENERGIA NUCLEAR DO MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES & COMUNICAÇÕES, AS 10 PRINCIPAIS RAZÕES PARA A CRIAÇÃO DE UM PGD SÃO AS SEGUINTEs:**

1. Ajustar o seu projeto de pesquisa às políticas mandatórias da sua instituição e/ou dos órgãos de fomento à pesquisa;
2. Assegurar a integridade da pesquisa e o seu potencial de replicação;
3. Assegurar que os dados e demais registros de pesquisa sejam acurados, completos, autênticos e confiáveis;
4. Aumentar a sua eficiência como pesquisador – um plano que organize os dados e seu armazenamento permite que você foque na sua pesquisa. Você estará mais capacitado a localizar e usar os seus dados e compartilhá-los com os seus colaboradores;
5. Permitir que os seus dados sejam compreensíveis agora e no futuro – se os dados são bem documentados antes e durante a formação da coleção de dados, eles serão mais facilmente entendidos e reutilizados;
6. Economizar tempo e recursos a longo prazo;
7. Aumentar a segurança dos dados e minimizar os riscos de perda; 29/05/2019 Gestão de dados de pesquisa: o que precisamos saber hoje! - SIBiUSP - Sistema Integrado de Bibliotecas da Universidade de São Paulo - www.sibi.usp.br/noticias/gestao-de-dados-de-pesquisa-o-que-precisamos-saber-hoje/ 3/9
8. Evitar a duplicação de esforços na re-coleta ou re-geração dos dados, possibilitando que outros pesquisadores se beneficiem dos seus dados e os interprete em outros contextos e com novas visões;
9. Aumentar a visibilidade da pesquisa – se os seus dados foram planejados para estarem organizados e corretamente arquivados, eles poderão ser identificados, recuperados e citados, aumentando a visibilidade da sua pesquisa e o seu prestígio como pesquisador;
10. Tornar mais fácil a preservação e o arquivamento – ajustando antecipadamente a geração de dados tomando como referência as práticas, formatos e demais padrões mais adequados ao arquivamento e à preservação de longo prazo, torna a gestão de dados mais fácil e menos custosa; além do mais, tornam os dados mais aderentes aos requisitos dos repositórios e centros de dados.

Fonte: Elaboração própria com base em Dudziak (apud SAYÃO; SALES, 2015).

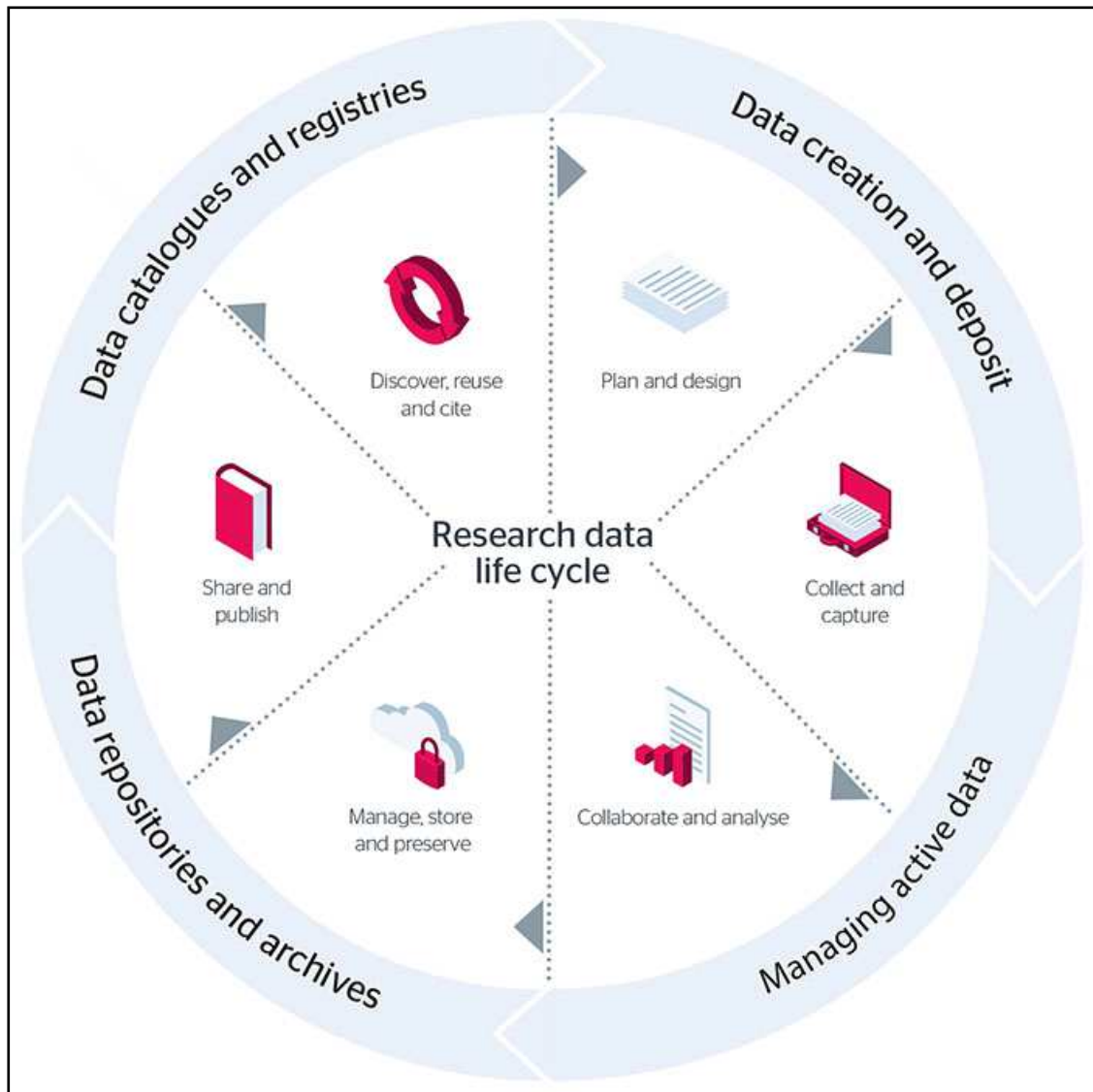
Como é possível perceber, o PGD é essencial para um bom encaminhamento das ações ligadas à gestão de dados científicos tanto para pesquisadores, como para instituições e confere ainda mais transparência e otimiza o investimento público em pesquisa.

Segundo Pierro (2018, p.39), o Horizonte 2020 é o principal programa de apoio à pesquisa e à inovação da União Europeia, que entrou em vigor em 2007 e que lançou “[...] em 2016 um documento descrevendo os passos para a preparação de um plano de gestão de dados, que passaram a ser obrigatórios em todos os projetos submetidos a partir de 2017”. Iniciativas como esta defendem que os pesquisadores financiados com recursos públicos não podem se furtar a assinar e se

comprometer em disponibilizar os dados do seu campo e experimentos, assim como divulgar os resultados de sua pesquisa.

Por outro lado, o Guia do Programa Horizonte da União Europeia tem como prerrogativa importante que a análise dos dados que podem ou não ser divulgados sejam submetidos ao conselho de ética em pesquisa. Pierro (2018) chama a atenção para situações em que a divulgação de dados brutos pode deflagrar problemas éticos. É o exemplo de ensaios clínicos, que utilizam dados pessoais e precisam garantir a privacidade dos pacientes.

Figura 3: Modelo do Ciclo de Vida dos Dados de Pesquisa



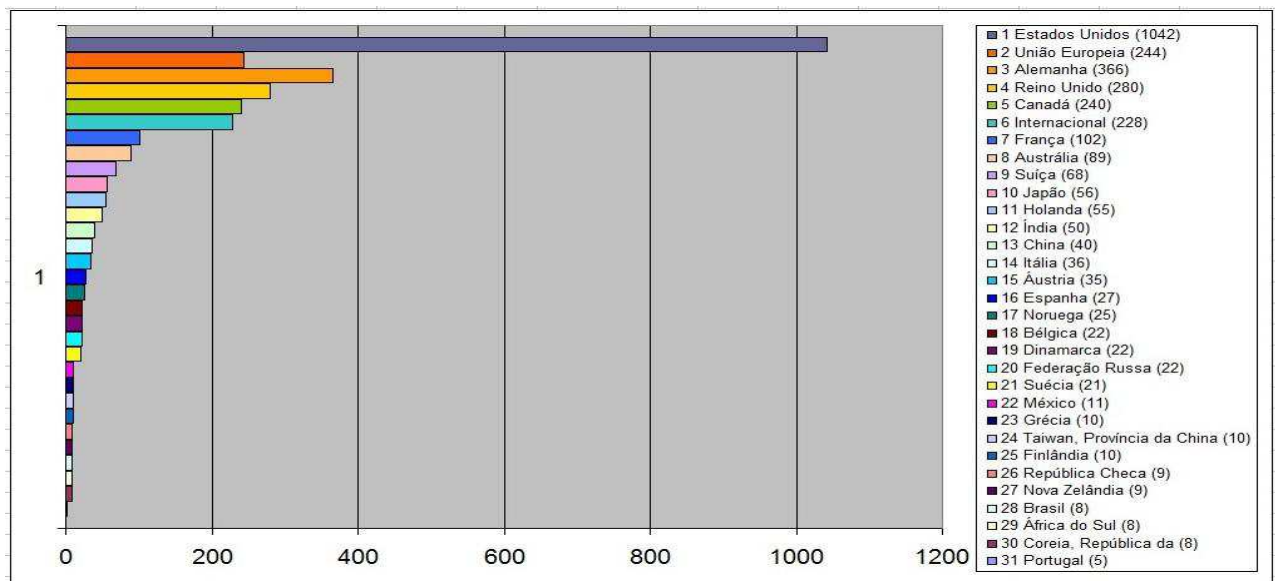
Fonte: Site do JISC, no Reino Unido (apud MEDEIROS, 2018)

De acordo com Medeiros (2018), a figura mostra uma das muitas visões possíveis do ciclo de vida de dados de pesquisa. A visão da JISC privilegia os aspectos de armazenamento, preservação e organização de repositórios de dados científicos. Como é possível perceber, com base em todos os referenciais e no entendimento desta dissertação, o PGD precisa considerar em primeira instância os seguintes aspectos: a) a informação registrada e estruturada (catálogos e registros indexados); b) o processo de criação e depósito (planejamento, coleta dos dados); c) gestão ativa (feedback e análise contínua); e d) a infraestrutura (repositórios e formas eficientes de arquivamento).

4.2 REPOSITÓRIOS INSTITUCIONAIS NO BRASIL QUE DISPONIBILIZAM ACESSO AOS DADOS DA PESQUISA

Segundo as métricas publicadas no macro repositório internacional, o RE3DATA, o Brasil se encontra na 28ª posição, com apenas 8 repositórios de dados de pesquisa em funcionamento, conforme apresentado no Gráfico 8, a seguir:

Gráfico 8: Repositórios de dados de pesquisa por país



Fonte: (RE3DATA, 2019).

Por meio do levantamento realizado neste estudo através do RE3DATA foram identificados os seguintes Repositórios no Brasil que fazem a gestão de dados científicos. São eles:

A) WORLDCLIM - DADOS GLOBAIS DO CLIMA

<http://worldclim.org/>

O WorldClim é um conjunto de camadas climáticas globais (grades de clima) com uma resolução espacial de cerca de 1 quilômetro quadrado. Os dados podem ser usados para mapeamento e modelagem espacial em um SIG ou com outros programas de computador. Fornece Dados Gráficos Estruturados, Formatos de dados científicos e estatísticos, Dados climáticos e Projeções de mudança climática. (RE3DATA, 2019, tradução livre).

B) GLOBE

<http://globe.umbc.edu/>

O GLOBE (Global Collaboration Engine) é um ambiente colaborativo on-line que permite aos pesquisadores de mudanças no solo compartilhar, comparar e integrar estudos locais e regionais com dados globais para avaliar a relevância global de seu trabalho. (RE3DATA, 2019, tradução livre).

C) PROGRAMA INTERNACIONAL DE DESCOBERTA DO OCEAN

<http://www.iodp.org/>

O Programa Internacional de Descoberta do Oceano (IODP) é uma colaboração internacional de pesquisa marinha que explora a história e a dinâmica da Terra usando plataformas de pesquisa oceânica para recuperar dados registrados em sedimentos e rochas do fundo do mar e monitorar ambientes de subsolos. O IODP depende de instalações financiadas por três provedores de plataforma com contribuições financeiras de cinco outras agências parceiras. Juntas, essas entidades representam 26 nações cujos cientistas são selecionados para realizar expedições de pesquisa do IODP conduzidas nos oceanos do mundo. As expedições do IODP são desenvolvidas a partir de propostas científicas baseadas em hipóteses, alinhadas com o plano de ciência do programa *Illuminating Earth's Past, Presente e Futuro*. O plano de ciência identifica 14 questões de desafio nas quatro áreas de mudança climática, vida profunda, dinâmica planetária e riscos

geológicos. Até 2013, sob o nome: *International Ocean Drilling Program*. (RE3DATA, 2019, tradução livre).

D) REPOSITÓRIO DE DADOS PPBIO

<https://ppbiodata.inpa.gov.br/metacatui/>

Repositório de Dados de Levantamentos Biológicos do Programa de Pesquisa em Biodiversidade Repositório de Dados. O Programa de Pesquisa em Biodiversidade (PPBio) foi criado em 2004 com o objetivo de promover estudos de biodiversidade no Brasil, descentralizando a produção científica de centros acadêmicos já desenvolvidos, integrando atividades de pesquisa e disseminando resultados em diversas finalidades, incluindo gestão ambiental e educação. O PPBio contribui com seus dados para a rede DataONE como um nó membro: <https://search.dataone.org/#profile/PPBIO>. (RE3DATA, 2019, tradução livre).

E) REDE IBICT DATAVERSE

<https://repositoriopesquisas.ibict.br>

O Instituto Brasileiro de Informação em Ciência e Tecnologia mantém o *Dataverse Network*. O IBICT está fornecendo um repositório de dados de pesquisa que cuida da preservação de longo prazo e arquivamento de boas práticas, para que os pesquisadores possam compartilhar, manter o controle e obter reconhecimento por seus dados. O repositório suporta o compartilhamento de dados de pesquisa com dados persistentes de *Quote*, permitindo que eles sejam reproduzidos. O *Dataverse* é um grande repositório aberto de dados de todas as disciplinas, criado pelo Instituto de Ciências Sociais Quantitativas da Universidade de Harvard. O IBICT, o repositório *Dataverse*, fornece um meio disponível gratuitamente para depositar e encontrar conjuntos de dados específicos armazenados pelos funcionários das instituições participantes da Rede Cariniana. (RE3DATA, 2019, tradução livre).

F) BASE DE DADOS CIENTÍFICOS DA UNIVERSIDADE FEDERAL DO PARANÁ

<https://bdc.c3sl.ufpr.br/>

O Banco de Dados Científicos da Universidade Federal do Paraná tem como objetivo reunir os dados científicos utilizados nas pesquisas que foram publicadas

pela comunidade da UFPR em teses, dissertações, artigos de periódicos e outros materiais bibliográficos. O BDC se une à RDI / UFPR como um serviço inovador que acompanha a tendência mundial de planejamento, gerenciamento, produção, organização, armazenamento, disseminação e reutilização de pesquisas. A disponibilidade de dados de pesquisa contribui para a transparência e otimização da produção científica por meio da reutilização de conjuntos de dados e a possibilidade de novas análises e abordagens. (RE3DATA, 2019, tradução livre).

G) CEDAP RESEARCH DATA REPOSITORY - DADOS DE PESQUISA

<https://cedap.ufrgs.br/jspui/handle/2050011959/90>

O Centro de Documentação e Pesquisa Digital de Pesquisa (CEDAP) da Universidade Federal do Rio Grande do Sul (UFRGS) tem como objetivo reunir os dados científicos utilizados em pesquisas, classificados como sílabas longas, nas diversas áreas do conhecimento. O Repositório de Dados Científicos da Pesquisa do CEDAP tem como objetivo reunir os dados científicos utilizados nas pesquisas, com a disponibilização de documentação, a fim de proporcionar um ambiente de estudo das metodologias de uso e reutilização dos dados da pesquisa. Mantido em parceria com o Centro de Processamento de Dados (CPD) da UFRGS para o desenvolvimento de políticas, planejamento, gerenciamento, descrição, avaliação, armazenamento, disseminação e reutilização de dados de pesquisa. Criado em junho de 2017. (RE3DATA, 2019, tradução livre).

H) BANCO DE DADOS DE EXPLORAÇÃO E PRODUÇÃO

<http://rodadas.anp.gov.br/en/concession-of-exploratory-blocks/brazil-round-3/bdep-round-3>

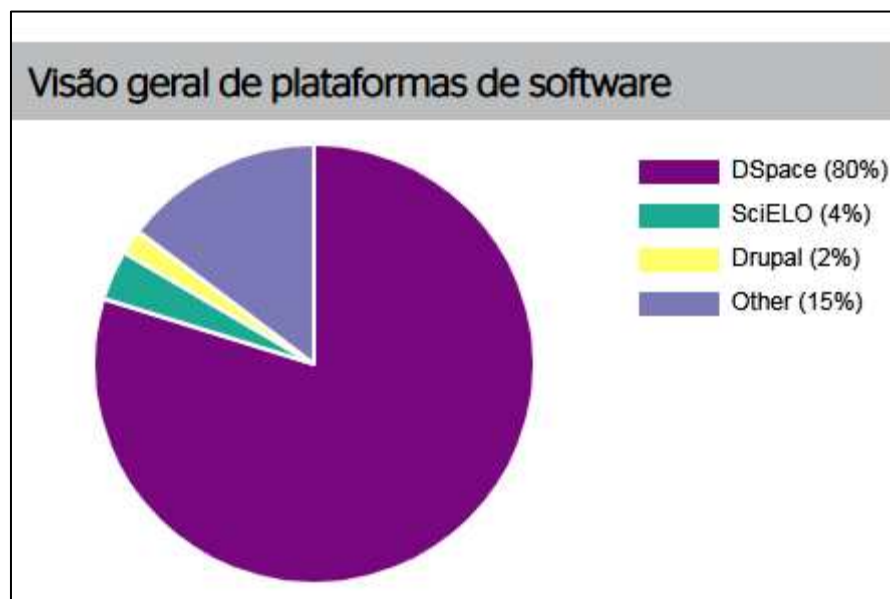
Fundado em maio de 2000, o BDEP armazena, organiza e disponibiliza informações geofísicas, geológicas e geoquímicas. O banco de dados, após processamento e análise, fornece ajuda para as áreas das bacias sedimentares, onde há maior probabilidade de petróleo e gás natural. A aquisição e o gerenciamento de dados dessa coleção garantem ao Brasil o domínio sobre o potencial de conhecimento gerado em hidrocarbonetos. (RE3DATA, 2019, tradução livre).

Vale salientar que estes repositórios contribuíram não somente para o movimento Acesso Aberto, mas também para a implementação de uma base tecnológica que promove a disseminação do conhecimento científico brasileiro. Este conjunto de repositórios estimulam, mesmo que ainda de forma tímida se comparado a países como Estado Unidos, União Européia e Alemanha, a articulação e a formação de redes de estímulo com base técnica, informacional, de comunicação e científica no Brasil.

Os sistemas mais utilizados para gestão de repositórios digitais são relevantes para entendimento do gerenciamento dos dados científicos no Brasil. A realidade brasileira, conforme dados obtidos no OPENDOAR, é relativamente distinta em comparação com os países que mais possuem repositórios de acesso livre. O gráfico a seguir, mostra que o sistema mais utilizado no Brasil é o *DSpace* com 80% do total. Muito disso deve-se à política de incentivo ao uso promovida pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

Ao participar do lançamento do software DSpace, em 2002, o IBICT teve conhecimento de seu uso em alguns países e pela Universidade do Minho, em Portugal, que teria implantado o seu repositório institucional, o RepositoriUM. Como já vinha liderando as ações do acesso livre no Brasil, o IBICT decidiu por customizar o software e distribuí-lo em nível nacional. Assim, a criação da versão brasileira do DSpace, em 2004, representou mais um marco do pioneirismo do IBICT no desenvolvimento e customização de ferramentas para tratamento e disseminação de informações técnico-científicas na Web (INSTITUTO..., 2019).

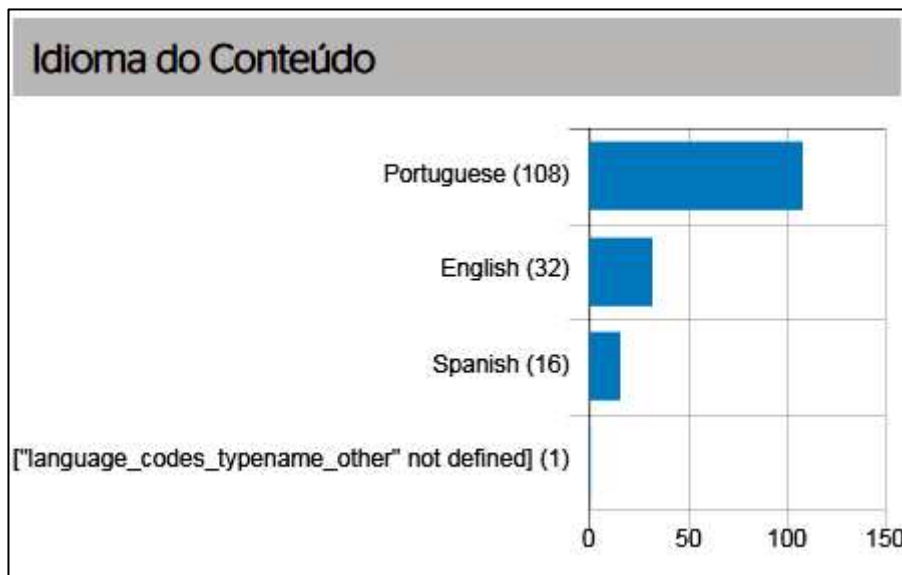
Gráfico 9: Software de acesso livre mais utilizados no Brasil



Fonte: Dados deste estudo levantados no OPENDOAR (2019)

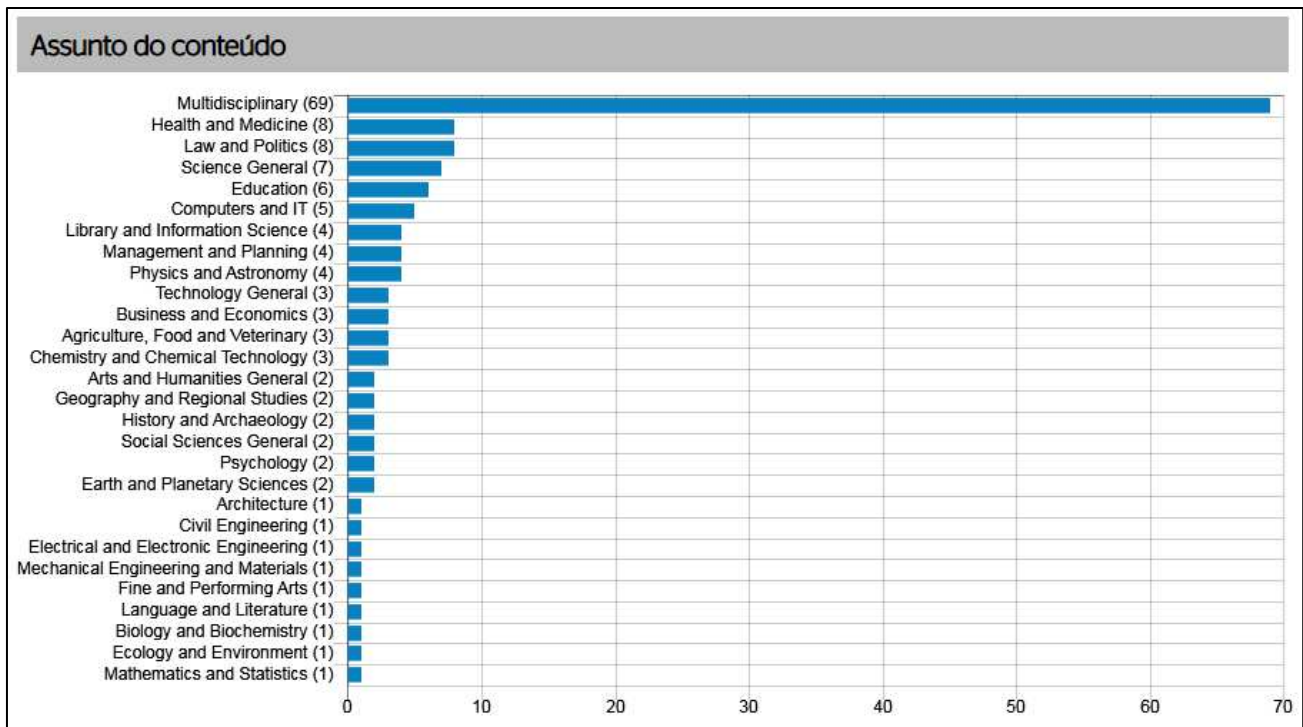
No gráfico a seguir, é possível perceber a existência de destaque para os repositórios com conteúdo em língua portuguesa. Em razão do exposto, “[...] é prudente refletir sobre a atual estrutura dos dados científicos no Brasil e como essa gestão tem evoluído” (COSTA, 2017). O acesso dos pesquisadores brasileiros aos repositórios em língua portuguesa mostra sinais de crescimento desde sua implementação. Aspecto que demonstra a urgência no estabelecimento de uma política que inclua critérios, regulamentos e documentação claros sobre a gestão dos dados científicos.

Gráfico 10: Idiomas dos repositórios brasileiros



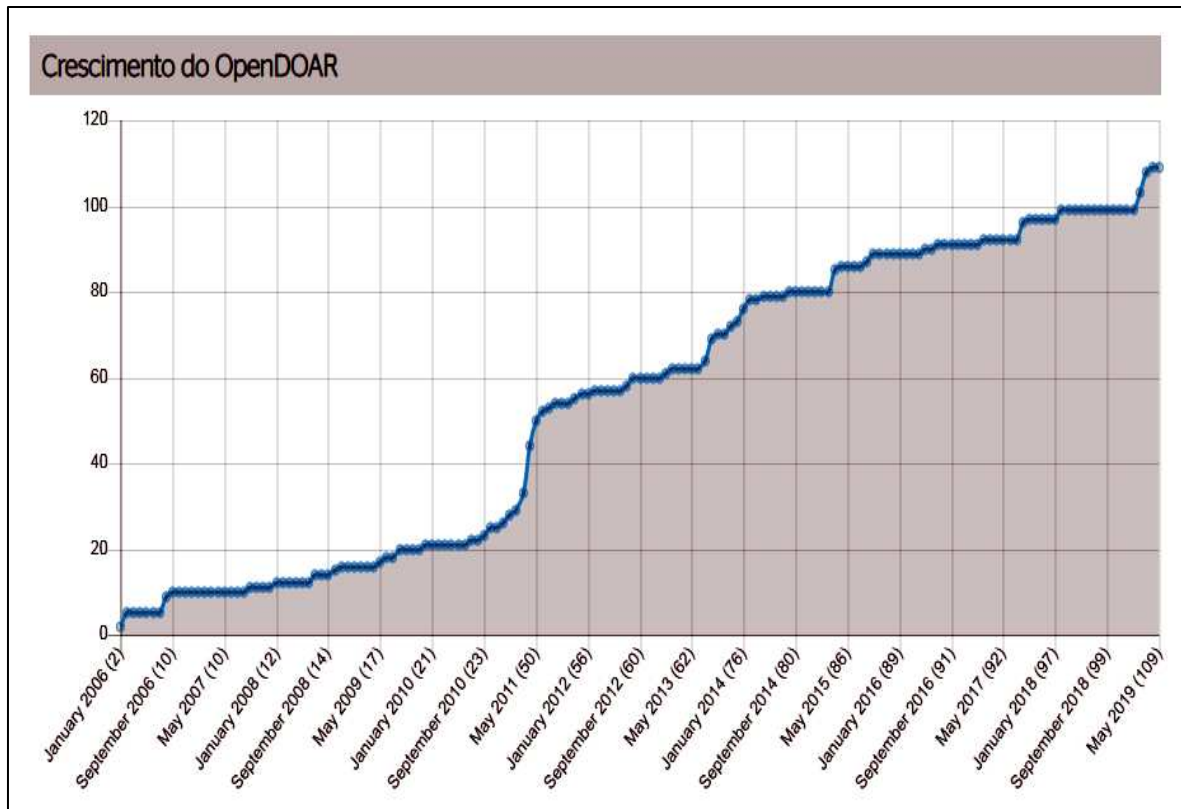
Fonte: Dados deste estudo levantados no OPENDOAR (2019)

No Gráfico 11 a seguir constam informações sobre o tipo de assunto e áreas por repositórios no Brasil com destaque para quantidade de registros na área Multidisciplinar (69). Em seguida aparecem as áreas de Saúde e Medicina e Políticas e Legislação, cada uma com oito registros. A área de Educação aparece em seis repositórios, no exemplo brasileiro.

Gráfico 11: Temática dos assuntos nos repositórios brasileiros

Fonte: Dados deste estudo levantados no OPENDOAR (2019)

No Gráfico 12, a seguir, podemos observar a evolução do número de Repositórios de Acesso Livre no Brasil, onde, em 2006, (ano que delimitamos como inicial neste levantamento), o país contava com 2 (dois) repositórios e em maio de 2019 havia crescido para a quantidade de 109 (cento e nove) repositórios.

Gráfico 12: Evolução do número de repositórios no Brasil

Fonte: Dados deste estudo levantados no OPENDOAR (2019)

A inserção do Brasil na imensa onda de mudanças promovida pelo movimento Acesso Aberto, ainda que relativamente recente, já acumula passos seguramente significativos; inclusive, há registros que já sugerem iniciativas institucionais, relativas ao compartilhamento de informações científicas, no Brasil, desde julho de 2002, quando a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) disponibilizou o catálogo de teses (CT) com referências e resumos das teses/dissertações defendidas em programas de pós-graduação do país, com o objetivo de facilitar o acesso a estas informações³⁷.

No cenário brasileiro, esse movimento se avoluma com o Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica³⁸, lançado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), em 2005, ao qual se seguem, no mesmo ano, a Declarações de Salvador: Sobre acesso aberto: a perspectiva dos países em desenvolvimento; e Compromisso com a equidade; a Declaração de Florianópolis (Psicologia) e a Carta de São Paulo: Declaração de apoio ao acesso

³⁷ https://sdi.capes.gov.br/banco-de-teses/02_bt_sobre.html

³⁸ disponível em < <https://livroaberto.ibict.br/Manifesto.pdf> >

aberto à literatura científica. Ano seguinte, na 58^a. Reunião Anual da Sociedade Brasileira para o Progresso da Ciência (SBPC, 2006), associações, universidades e institutos manifestam-se em favor da divulgação de dados científicos e elaboram Carta Aberta à SBPC³⁹, apoiados pelos participantes do encontro aberto 'Acesso à informação científica: questões políticas, tecnológicas e diferenças disciplinares', integrado à programação da RA/SBPC 2006 e realizado no dia 18 de julho de 2006. Na referida Carta Aberta à SBPC, os signatários consideraram:

1. a necessidade de a comunidade científica brasileira ter maior acesso à informação científica;
2. a importância do registro e da disseminação da produção científica brasileira;
3. que todos os resultados de pesquisas financiadas com recursos públicos devem ser de livre acesso a todos os brasileiros;
4. que o compartilhamento da informação e do conhecimento é condição necessária, mas não suficiente, para a redução das desigualdades sociais;
5. as iniciativas realizadas em todo o mundo em favor do acesso aberto à informação (Anexo 1);
6. as iniciativas realizadas no país em favor do acesso aberto à informação científica (Anexo 2);
7. que no país existem ações concretas que implementam repositórios e publicações de acesso aberto, como: a) Scielo (Bi-reme/Fapesp), integrando mais de 300 periódicos científicos nacionais e estrangeiros; b) Biblioteca Digital de Teses e Dissertações (BDTD, IBICT), integrando 30 instituições de ensino superior (IES), com um

³⁹ Cadernos SBPC. Registro dos debates na 58^a Reunião Anual, 2006, p.48-50. Disponível em < <http://www.sbpcnet.org.br/livro/58ra/pdf/3%20Pol%EDtica%20de%20C&T.pdf>

acervo de 21 mil teses e dissertações eletrônicas, em texto integral; c) Sistema Eletrônico de Editoração de Revistas (IBICT), utilizado por cerca de 90 periódicos científicos nacionais; d) DSpace (IBICT), com comprovada adoção por mais de 1.500 instituições em todo o mundo e por seis instituições nacionais;

8. que o Brasil ocupa, hoje, o quarto lugar entre os países que mais construíram repositórios de acesso livre, ficando atrás apenas dos Estados Unidos, do Reino Unido e da Alemanha;
9. que a posição brasileira no ranking mencionado, acompanhando a evolução dos países desenvolvidos quanto à construção e manutenção de repositórios digitais, poderá tornar o país mais autônomo com relação aos editores científicos comerciais, reduzindo os gastos com o acesso à informação científica;
10. que no Brasil existem instituições que detêm competência técnica para a absorção e o desenvolvimento de ferramentas de software para a construção e manutenção de repositórios digitais e publicações eletrônicas de acesso aberto, em conformidade com os padrões internacionais, os quais fazem parte do modelo Open Archives;
11. que, em função do uso desse modelo, diversos repositórios brasileiros são indexados por ferramentas de buscas e diretórios de repositórios de acesso aberto internacionais, como o Directory of Open Access Journals (DOAJ), o Registry of Open Access Repositories (ROAR), o Google Scholar etc.

Em face desse contexto, a Carta Aberta à SBPC também encaminhou solicitação às associações científicas, particularmente as filiadas à SBPC, no sentido de que discutissem “a questão do acesso aberto” e que considerassem posicionar-se sobre o Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica (IBICT, 2005).

Especialmente à SBPC, na Carta Aberta, os representantes das associações, universidades e institutos signatários solicitaram que:

- encaminhe recomendação aos vários órgãos governamentais atuantes nas áreas de ciência, tecnologia e educação, em especial às agências de fomento, para que integrem esforços a fim de:
- sensibilizar os dirigentes dessas instituições quanto à importância do acesso livre à informação científica;
- promover as ações preconizadas pelo movimento brasileiro de acesso aberto à informação científica, as quais constam dos documentos listados no Anexo 3⁴⁰;
- promover e apoiar a construção de repositórios digitais com vistas ao registro e à disseminação da produção científica das instituições brasileiras de ensino superior e de pesquisa;
- ampliar as fontes de informação existentes nas bibliotecas das IES, em apoio ao ensino de graduação e pós-graduação, além da pesquisa.

Outra iniciativa pioneira, no Brasil, deu-se com o Edital Chamada FINEP/PCAL/XBDB nº. 002/2009, que teve como objetivo:

[...] apoiar projetos de implantação de repositórios institucionais (RI) nas instituições públicas (federais, estaduais e municipais) de ensino e pesquisa e sua integração ao Portal Oásis.Br, com vistas a possibilitar o registro e a disseminação da produção científica destas instituições e proporcionar maior visibilidade à sua produção científica. (INSTITUTO..., 2009)

Segundo Elisangela Mota Pires (2011), o mencionado Edital apoiou 33 instituições⁴¹ (22 universidades federais, 2 universidades estaduais e 1 universidade

⁴⁰ O Anexo 3 contém os cinco documentos: 1. Manifesto brasileiro em favor do acesso livre à informação científica; 2. Declarações de Salvador: Sobre acesso aberto: a perspectiva dos países em desenvolvimento; Compromisso com a equidade; 3. Carta de São Paulo: Declaração de apoio ao acesso aberto à literatura científica 4. Declaração de Florianópolis; 5. Definição de publicação de acesso aberto, conforme a Declaração de Bethesda.

municipal, 1 centro universitário, além de 5 outras instituições de pesquisa e 1 museu):

1. Centro de Tecnologia da Informação Renato Archer (CTI),
2. Centro Tecnológico de Minas Gerais (CETEC/MG),
3. Centro Universitário Fundação Santo André (FSA),
4. Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT),
5. Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO),
6. Instituto Nacional de Tecnologia (INT),
7. Museu Paraense Emílio Goeldi (MPEG),
8. Universidade de Brasília (UNB),
9. Universidade de São Paulo (USP),
10. Universidade Estadual de Ponta Grossa (UEPG) e
11. Universidade Federal da Bahia (UFBA),
12. Universidade Federal da Grande Dourados (UFGD),
13. Universidade Federal de Alagoas (UFAL),
14. Universidade Federal de Goiás (UFG),
15. Universidade Federal de Ouro Preto (UFOP),
16. Universidade Federal de Pelotas (UFPEL),
17. Universidade Federal de Pernambuco (UFPE),
18. Universidade Federal de Santa Catarina (UFSC),
19. Universidade Federal de São Carlos (UFSCAR),
20. Universidade Federal de Sergipe (UFS),
21. Universidade Federal de Uberlândia (UFU),
22. Universidade Federal de Viçosa (UFV),
23. Universidade Federal do Acre (UFAC),
24. Universidade Federal do Maranhão (UFMA),
25. Universidade Federal do Mato Grosso (UFMT),
26. Universidade Federal do Rio Grande (FURG),
27. Universidade Federal do Rio Grande do Norte (UFRN),

⁴¹ <http://repositorio.furg.br/bitstream/handle/1/5775/Reposit%c3%b3rios%20institucionais%20-%20caracter%c3%adsticas%20atrav%c3%a9s%20da.pdf?sequence=1>

Pires, Elisângela Mota Repositórios institucionais: características através da visibilidade [Trabalho de conclusão de curso] / Elisângela Mota Pires ; orientadora, Angélica Conceição Dias Miranda. – Rio Grande, 2011.

28. Universidade Federal do Rio Grande do Sul (UFRGS),
29. Universidade Federal do Tocantins (UFT),
30. Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM),
31. Universidade Federal Fluminense (UFF),
32. Universidade Municipal de São Caetano do Sul (USCS)
33. Universidade Tecnológica Federal do Paraná (UTFPR).

Da pesquisa realizada por Pires (2011) participaram 22 das 33 instituições então contempladas com *kits* tecnológicos. Desse conjunto, a autora identificou os *links* de 21 repositórios⁴² dessas 22 instituições. Reexaminando, para efeito desta dissertação, os *links* deste conjunto de instituições, verificamos que: (a) dez continuam ativos e com o mesmo radical informado pela pesquisa de Pires⁴³; (b) sete mudaram o radical de identificação do repositório, indicando que o link anterior talvez já não esteja ativo⁴⁴; (c) dois direcionam para bibliotecas digitais⁴⁵; (d) o repositório da USP tem link ativo⁴⁶; e (e) em dois casos não foi possível ter acesso ao repositório, nem pelo *link* referido na pesquisa de Pires (2011) nem outro, que identificamos, em consulta à *web*⁴⁷.

Ricardo Saraiva e Bianca Amaro (2012), comentando os mencionados Editais, referem que o Ibict iniciou o projeto de Repositórios Institucionais com um projeto-piloto em quatro universidades federais, em 2009, ano em que também lançou o Edital chamada FINEP/PCAL/ XBDB n.º 01/2009 que, segundo os autores, contemplou “27 instituições, com *kits* tecnológicos”, ampliando, no ano seguinte, a distribuição dos referidos *kits* para mais “7 instituições”, via Edital chamada

⁴² A autora declarou que, à época, não tinha sido localizado o *link* do repositório da USP.

⁴³ INMETRO: <http://repositorios.inmetro.gov.br/>; MPEG: <https://repositorio.museu-goeldi.br/jspui/>; UEPG: <http://ri.uepg.br:8080/riuepg/>; UFF: <https://app.uff.br/riuff/>; UFG: <http://repositorio.bc.ufg.br/>; UFSC: <http://repository.bu.ufsc.br:8080/xmlui/>; UFU: <http://repositorio.ufu.br/>; UFVJM: <http://acervo.ufvjm.edu.br:8080/jspui/>; UFRGS: <https://lume.ufrgs.br/>; UTFPR: <http://riut.utfpr.edu.br/jspui/>.

⁴⁴ CETEC/MG: <https://diretorio.rcaap.pt/handle/1/60>; CTI: <https://diretorio.rcaap.pt/handle/1/700>; FURG: <http://repositorio.furg.br/>; INT: <https://diretorio.rcaap.pt/handle/1/184>; UFMA: <https://repositorio.ufma.br/jspui/>; UFPEL: <http://guaiaca.ufpel.edu.br:8080/>; UNB: <https://repositorio.unb.br/>.

⁴⁵ FSA: <https://www.fsa.br/biblioteca/>; UFGD: <https://portal.ufgd.edu.br/setor/biblioteca/repositorio>.

⁴⁶ USP: <https://www.aguia.usp.br/bibliotecas/digitais-sistemicas/producao-usp/>

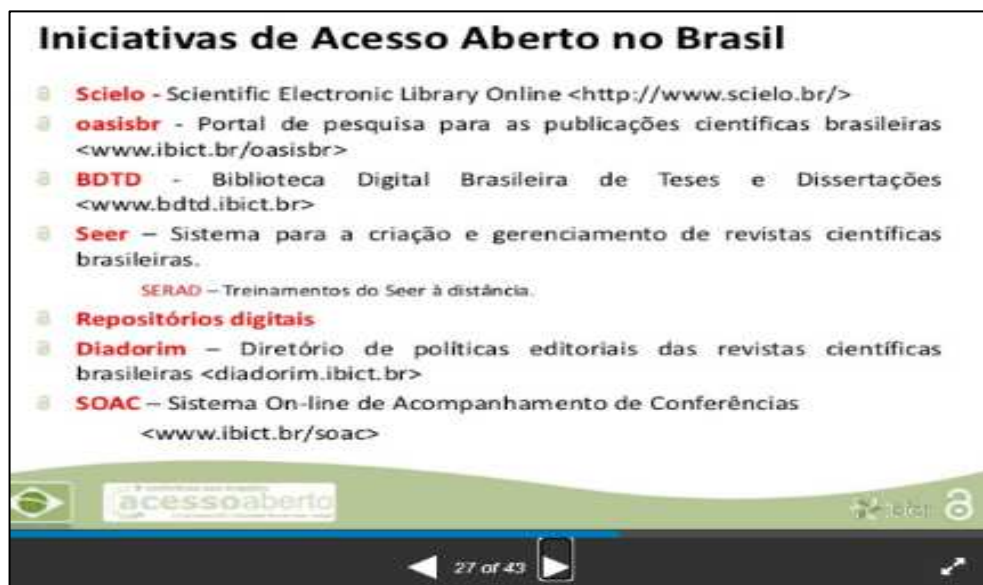
⁴⁷ USCS: <http://repositorio.uscs.edu.br/> ou <http://repositorio.uscs.edu.br/handle/123456789/286> (sem acesso); UFAC: <http://repositorios.ufac.br:8080/repositorio/> ou <https://portal.ufac.br/biblioteca/pesquisa/pesquisar.action>.

FINEP/PCAL/ XBDB n.º 02/2010⁴⁸. Entretanto, cabe a ressalva de que, em nota de 20/04/2009, o IBICT esclareceu que o primeiro edital foi revogado⁴⁹ e substituído:

Com o propósito de aperfeiçoar o texto do edital 001/2009, de forma a atender aos objetivos dos projetos PCAL e XBDB e, desta forma, ampliar o número de instituições elegíveis, foi revogado, no dia 17 de abril de 2009, o referido edital. Na mesma data foi publicado um novo edital, o Edital Finep PCAL/XBDB n.º 002/2009, o qual contém as correções citadas e um novo cronograma para entrega e avaliação das propostas.

A participação dos pesquisadores Saraiva e Amaro, na CONFERÊNCIA LUSO-BRASILEIRA SOBRE ACESSO ABERTO (CONFOA), em 2012, realizada em Lisboa⁵⁰, acabou se configurando como um importante passo do Brasil perante o movimento Acesso Aberto, à época já em crescente escala mundial. Naquela Conferência, dedicando-se aos *10 anos de Acesso Aberto no mundo lusófono*, os referidos pesquisadores apresentaram o panorama dos repositórios e dos contextos científicos relativos aos países Portugal, Cabo Verde, Moçambique, Guiné-Bissau, São Tomé e Príncipe, Angola, Timor Leste e Brasil. Com relação ao contexto científico, o Brasil já apresentava importantes características, a exemplo das iniciativas de Acesso Aberto apontadas na Figura 4 a seguir:

Figura 4: Iniciativas de Acesso Aberto no Brasil



Fonte: Acesso Aberto.

Nota: <https://www.slideshare.net/ConfOA/10-anos-de-acesso-aberto-no-mundo-lusfono>

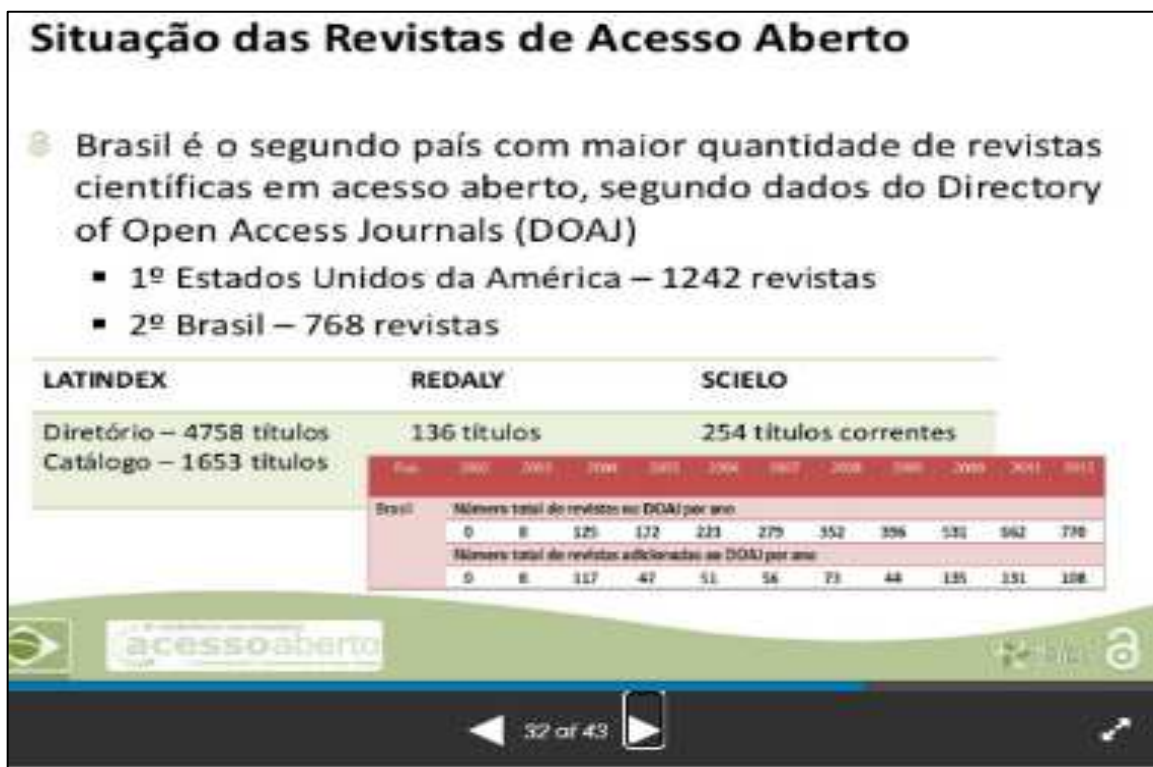
⁴⁸ <https://www.slideshare.net/ConfOA/10-anos-de-acesso-aberto-no-mundo-lusfono>

⁴⁹ <http://www.ibict.br/sala-de-imprensa/noticias/item/1327-revogado-edital-finep-pcal-xbdb-n-001-2009>

⁵⁰ <https://www.slideshare.net/ConfOA/10-anos-de-acesso-aberto-no-mundo-lusfono>

Outro ponto importante nesse contexto que acreditamos ser influenciado por estas iniciativas é a quantidade de produções brasileiras em diretórios e catálogos de Acesso Aberto. O Brasil se destaca como um dos países em situação de produção e disponibilização de conhecimento aberto. Nesse quesito ocupa a segunda posição no ranking em quantitativo de revistas científicas em acesso aberto conforme dados do Directory of Open Access Journals (DOAJ) na figura a seguir:

Figura 5: Situação das revistas de acesso aberto brasileiras



Fonte: Acesso Aberto.

Nota: <https://www.slideshare.net/ConfOA/10-anos-de-acesso-aberto-no-mundo-lusfono>

Outro estudo que destaca as iniciativas do IBICT e da FINEP no fomento de repositórios nacionais foi desenvolvido por Tiago Rodrigo Marçal Murakami e Sibeles Fausto (2013), os quais também mencionam os editais de chamada da FINEP, referindo-se ao Edital chamada FINEP/PCAL/XBDB n. 001/2009 e ao Edital chamada FINEP/PCAL/XBDB n. 002/2010⁵¹. Segundo Murakami e Fausto (2013),

⁵¹ Também aqui vale a ressalva quanto a revogação do FINEP/PCAL/XBDB n. 001/2009 e a sua substituição, de conformidade com nota do Ict, publicada em 20/04/2009, anteriormente referida. Disponível em: < <http://www.ibict.br/sala-de-imprensa/noticias/item/1327-revogado-edital-finep-pcal-xbdb-n-001-2009> >

foram contempladas 34 instituições⁵² com um kit tecnológico, além de capacitação e apoio técnico do IBICT para a implantação de seus repositórios. O referido kit tecnológico era composto por:

[...] um servidor instalado com um pacote de aplicativos livres (Linux, Apache e PHP) e os softwares DSpace (para repositórios) e SEER (Sistema Eletrônico de Editoração de Revistas) - versão traduzida do Open Journal System (OJS) e distribuída pelo IBICT para o fomento de revistas científicas em acesso aberto.

Ainda segundo Saraiva e Amaro (2012⁵³), foi prevista a entrega de outros 17 *kits* tecnológicos, no segundo semestre de 2020, para 9 universidades e 8 institutos de pesquisa: Universidade Federal de Lavras (UFLA), Universidade Federal do Recôncavo da Bahia (UFRB), Universidade Federal dos Pampas (UNIPAMPA), Universidade do Estado do Amazonas (UEA), Universidade Federal de Santa Maria (UFSM), Universidade Federal do Rio de Janeiro (UFRJ), Universidade Federal do Piauí (UFPI), Universidade Federal Rural de Pernambuco (UFRPE), Universidade Federal da Fronteira do Sul (UFFS), Museu de Astronomia e Ciências Afins (MAST), Observatório Nacional, Instituto Nacional de Pesquisas da Amazônia (INPA), Laboratório Nacional de Computação Científica (LNCC), Centro de Tecnologia Mineral (CETEM), Instituto de Desenvolvimento Sustentável Mamirauá (IDSM), Centro Brasileiro de Pesquisas Físicas (CBPF), Fundação Biblioteca Nacional (BN). Nesta distribuição, seriam contempladas 4 universidades e 2 institutos de pesquisa das regiões Norte e Nordeste: a UFRB, a UEA, a UFPI, a UFRPE, o INPA e o IDSM.

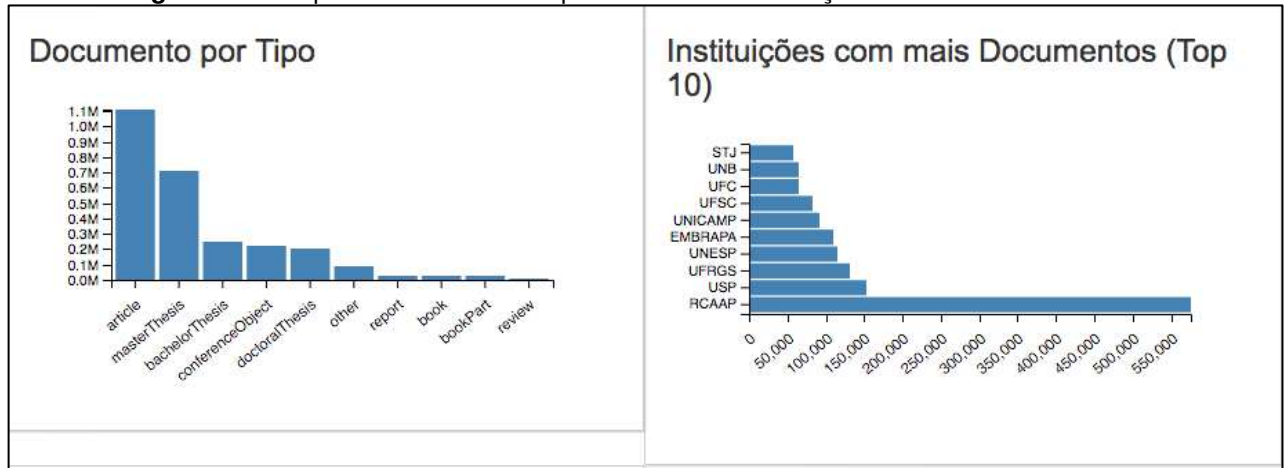
Outra importante iniciativa foi o portal brasileiro OASISBR (implementado em 2008 e com nova versão lançada em julho de 2015), no qual estão reunidas publicações científicas em acesso aberto. O OASISBR é “um mecanismo de busca multidisciplinar que permite o acesso gratuito à produção científica de autores

⁵² Não há consenso sobre o número de instituições contempladas nem com relação aos editais correspondentes: é possível que a menção a 34 instituições deva-se à soma das que foram contempladas no Edital chamada FINEP/PCAL/ XBDB n.º 02/2009 (27) – que substituiu o Edital FINEP/PCAL/ XBDB n.º 01/2009 - e aquelas contempladas no Edital chamada FINEP/PCAL/ XBDB n.º 02/2010 (7). Ainda assim, não fica esclarecida a situação, levando-se em conta que, como visto, Pires (2011), informa o apoio a 33 instituições (FURG; USP; UFG; UFMT; UFOP; UFPEL; UEPG; UFSC; USCS; UFV; UFAC; UFVJM; UFMA; UTFPR; UFRN; UFS; UFT; CETEC MG; INMETRO; UFPE; UFAL; UFSCAR; UFGD; INT; UFU; UNB; CTI; FSA; MPEG; UFBA; UFF; IBICT. (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2010).

⁵³ SARAIVA, R.; AMARO, B. *10 Anos de acesso aberto no mundo lusófono*. In: CONFERÊNCIA LUSO-BRASILEIRA SOBRE ACESSO ABERTO (CONFOA), 3., Lisboa, 2012. Disponível em: < <http://www.slideshare.net/ConfOA/10-anos-de-acesso-aberto-no-mundo-lusfono> >.

vinculados a universidades e institutos de pesquisa brasileiros [...] é possível também realizar buscas em fontes de informação portuguesas”⁵⁴.

Figura 6: Principais Documentos disponibilizados e Instituições no TOP 10 do OASISbr



Fonte: Dados fornecido pelo OASISBR (2020)

Nota: oasisbr.ibict.br

Alice Aquino Zanin (2017), em pesquisa realizada, cujos resultados foram publicados em artigo intitulado *Recursos educacionais abertos e direitos autorais: análise de sítios educacionais brasileiros*⁵⁵, identificou 242 repositórios no Brasil (84 no OpenDoar; 147 no ROAR; e 11 no MIRA); destes, restaram, ao seu estudo, 118 repositórios, após descarte de 100 repositórios por duplicidade e 24 inativos.

Estudo realizado por Simone da Rocha Weitzel (2019)⁵⁶, publicado sob o título *O mapeamento dos repositórios institucionais brasileiros: perfil e desafios*, têm indicado que as iniciativas do Brasil acompanharam o movimento Acesso Aberto e têm sido elogiadas pela comunidade internacional (ALPERIN; FISCHMAN; WILLINSKY, 2008; HARNAD; CARR; GINGRAS, 2008). Nesse sentido, a autora considerou como ações de destaque as empreendidas pelas seguintes entidades e redes:

- Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (Bireme)

⁵⁴ <http://oasisbr.ibict.br/vufind/>

⁵⁵ Revista Brasileira de Educação v. 22 n. 71 e227174 2017. Disponível em < <https://www.scielo.br/pdf/rbedu/v22n71/1809-449X-rbedu-22-71-e227174.pdf> >.

⁵⁶ <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2019v24n54p105/38092> - P. 108

- Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)
- Programa Scielo
- Rede Scielo da Bireme
- Rede Scielo Livros

Destacou, ainda, a autora Zanin (2017), que:

As lideranças das universidades e institutos de pesquisa foram envolvidas no processo a fim de garantir o acesso às tecnologias e recursos requeridos tanto para publicação digital quanto para a implementação de repositórios institucionais conforme relatos de Costa, Kuramoto e Leite (2013, p. 134). Os autores consideram que as ações relativas ao Acesso Aberto no Brasil iniciam a partir de 2003 quando seus representantes participaram da 7.ª ELPUB – The International Conference on Electronic Publishing – ocorrido em 2003 em Portugal.

O levantamento realizado por Weitzel visou o os repositórios institucionais brasileiros até o período de maio de 2017 e utilizou, como fontes, OpenDOAR⁵⁷, ROARMAP⁵⁸, ROAR⁵⁹, Lista de repositórios do IBICT⁶⁰, resultado do edital da FINEP que apoiou a implantação de repositórios no país⁶¹, lista de usuários do DSPACE⁶² bem como repositórios listados no The Ranking Web of World Repositories⁶³ (2017, p. 107). As variáveis estudadas foram:

- a) número de artigos e volume total de itens para identificar os repositórios alinhados ao Acesso Aberto Verde;
- b) tipos de financiamento recebidos além do apoio da própria instituição como indicadores de sua manutenção e permanência;
- c) verificação da qualidade dos repositórios por meio do indicador gerado pela fonte The Ranking Web of World Repositories que avalia visibilidade e impacto.

⁵⁷ VER NOTA ANTERIOR

⁵⁸ <https://roarmap.eprints.org/>

⁵⁹ <http://roar.eprints.org/>

⁶⁰ <http://www.ibict.br/informacao-para-a-pesquisa/repositorios-digitais>

⁶¹ Edital FINEP/PCAL/XBDB nº. 002/2009-confirmar se foi este o Edital

⁶² <https://wiki.lyrasis.org/pages/viewpage.action?pageId=104565826>

⁶³ https://repositories.webometrics.info/en/Latin_America/Brazil

Dentre as conclusões, a autora ressaltou:

Foi verificado que cerca de 54,5% dos repositórios concentram 97,5% do total de artigos dentre os 101 repositórios identificados no país. Um terço dos repositórios receberam financiamento direto ou indireto do governo brasileiro e representam 20% do total de artigos depositados. O total de 45,5% dos repositórios foram qualificados no The Ranking Web of World Repositories. Apesar dos dados apontarem para a concentração de repositórios com artigos científicos cumprindo o estabelecido para a estratégia do Acesso Aberto Verde foram observadas dificuldades em relação ao apoio institucional e qualidade (visibilidade e impacto). O estudo recomenda que o governo brasileiro estabeleça diretrizes nacionais e políticas específicas para fomentar as boas práticas em repositórios no país com o objetivo de fortalecer o desenvolvimento do Acesso Aberto Verde. (WEITZEL, 2019)

Os autores Murakami e Fausto (2013) traçaram o panorama dos repositórios institucionais de instituições de ensino superior no Brasil⁶⁴, tendo, como fontes, os diretórios ROAR, OpenDOAR, Diretório Luso-Brasileiro de Periódicos⁶⁵. Dentre as conclusões, identificaram:

[...] 49 repositórios que em agosto de 2013 disponibilizavam 396.881 itens, sendo as instituições federais as com maior povoamento e o repositório LUME o primeiro em volume de itens; a região Sudeste com o maior número de repositórios e volume de itens disponibilizados; o *DSpace* o *software* predominante, com maior utilização da versão 1.6.2 e o padrão de metadados *Dublin Core* em todas as aplicações desse *software*.

Esses resultados indicaram, com relação ao total então pesquisado, os repositórios com maior quantidade de trabalhos disponibilizados: o Repositório Digital da Universidade Federal do Rio Grande do Sul (LUME/UFRGS), com 18,8%; a Biblioteca Digital da Universidade Estadual de Campinas (UNICAMP), com 13,6%; a Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo (USP), com 10,4%; o Acervo Digital da Universidade Estadual de São Paulo (UNESP), com

⁶⁴ MURAKAMI, T.; FAUSTO, S. Panorama atual dos Repositórios Institucionais das Instituições de Ensino Superior no Brasil. **InCID: Revista de Ciência da Informação e Documentação**, v. 4, n. 2, p. 185-201, 20 dez. 2013. Disponível em: < <http://www.revistas.usp.br/incid/article/view/69327> >.

⁶⁵ <http://projecto.rcaap.pt/index.php/lang-pt/sobre-o-rcaap/servicos/directorio>
O serviço de diretório pretende disponibilizar num ponto único uma lista completa e exaustiva dos repositórios e revistas científicas de Acesso Aberto disponibilizadas em Portugal e no Brasil. A informação relativa a Portugal é mantida pelo RCAAP sendo a informação relativa ao Brasil mantida pelo IBICT. O registo de um repositório ou revista no diretório pode ser realizado no diretório. O registo das políticas das editoras e revistas portuguesas para o auto-arquivo em repositório institucional deverá ser efetuado através do serviço Blimunda.

9,4%; e o Repositório Institucional da Universidade Federal de Santa Catarina (UFSC), com 8,3%.

Neste estudo, Murakami e Fausto (2013) também apresentaram a distribuição por instituição mantenedora⁶⁶; o quantitativo de repositórios institucionais por natureza da instituição; e o volume de itens por natureza institucional, tendo sido verificado que:

Em relação à natureza da instituição, 65.3 % dos repositórios são federais (32 repositórios), 18.4% são particulares (9 repositórios), 14.3% são estaduais (7 repositórios) e 2% são municipais (1 repositório) [...] em relação à quantidade de itens povoados por natureza da instituição [...], 50.2% estão em repositórios federais (199.227), 41.6% em repositórios estaduais (164.939) e 8.2% em repositórios particulares (32.498) [...] o repositório municipal possui 217 itens.

Os citados autores também levantaram a distribuição geográfica dos trabalhos disponibilizados nesses repositórios, correspondendo 217.673 à região Sudeste (54.8%); 127.992 à região Sul (32.2%); 24.731 à região Nordeste (6.2%); 23.624 à região Centro-Oeste (6%); e 2.861 à região Norte (0.7 %).

Na literatura nacional ainda são poucos os estudos sobre Repositórios Institucionais que tratam do armazenamento de dados de pesquisa. São poucos também os Repositórios de Dados de Pesquisa existentes no País.

Existem algumas iniciativas, já pontuadas nesta dissertação, mas com relação a produção de trabalhos sobre o tema [repositório de armazenamento de dados científicos] temos poucos registros no Brasil conforme mostrou este levantamento das teses e dissertações na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD). O quadro a seguir mostra os 3 (três) registros encontrados na BDTD:

⁶⁶ Os autores esclarecem que adotaram esse critério pois “[...] uma instituição pode manter mais de um repositório”. (2013)

Quadro 5: Teses e dissertações defendidas no Brasil sobre repositórios de armazenamento de dados científicos

TIPO	AUTOR	TÍTULO	ANO
Tese (doutorado)	Costa, Máira Murrieta	Diretrizes para uma política de gestão de dados científicos no Brasil	2017
Dissertação (mestrado)	Roriz, Carlos André Cursino	Expedições científicas : a coleta de dados e materiais científicos por estrangeiros no Brasil : análise da gestão realizada pelo CNPq no período de 1991 a 2001	2003
Dissertação (mestrado)	Silva, Ronaldo Ferreira da	Plataformas computacionais de e-Science para gestão de dados científicos: estudo de caso de um programa de pesquisa da Amazônia	2018

Fonte: Dados do estudo (2020).

Nota: BDTD - <https://bddd.ibict.br/vufind/>

Vale a pena salientar que repositórios implementados especificamente para armazenamento de dados de pesquisa, como os levantamentos desse estudo mostraram, ainda estão em fase de crescimento no Brasil.

Como foi possível perceber neste capítulo, os repositórios digitais no mundo e no Brasil são uma realidade de interesse da comunidade acadêmica e condensam grande volume de dados produzidos na *E-Science*. De acordo com Costa (2017, p.69), o crescente uso dos RI “ [...] traz à tona pelo menos duas grandes vertentes a serem trabalhadas, a primeira diz respeito a infraestrutura tecnológica necessária para permitir o compartilhamento de dados [...] a segunda trata da preservação desses dados para a posteridade”. Desse modo, por sua vez, estas questões trazem consigo pontuações relevantes que esta dissertação compartilha como a do armazenamento, das regras de acesso e até mesmo das regras de reutilização dos dados.

5. PLANO DE GESTÃO DE DADOS CIENTÍFICOS

Do exposto nos Capítulos anteriores, destacamos, aqui, as problemáticas e desafios na gestão de dados científicos, conforme mencionados na literatura e nos registros documentais consultados; também elencamos, neste Capítulo, as características recomendadas ou que pensamos que devem prevalecer em um plano de gestão de dados científicos.

Este capítulo foi desenvolvido conforme as seguintes seções:

5.1 Gestão de dados científicos: problemáticas e desafios

5.2 Plano de Gestão: proposição para aplicabilidade em repositório universitário

5.1 GESTÃO DE DADOS CIENTÍFICOS: PROBLEMÁTICAS E DESAFIOS

A situação atual da pesquisa em Educação foi recentemente registrada e organizada por Moura e Lopes (*apud* NEVES, 2017) e tem demonstrado momento de expansão no quadro de pesquisadores, de eventos científicos e de programas de pós-graduação. Conforme relata Moura e Lopes (*apud* NEVES, 2017) o Portal da CAPES, em 2015 apresentou 45 Programas de Pós-Graduação em nível Stricto Sensu, com doutorados e mestrados acadêmicos, sendo todos avaliados com notas 6, 5, 4 e 3. Até o final de 2016, estas notas permanecem, pois, a avaliação passou a ser quadrienal.

Eles podem ser qualquer informação armazenável em formato digital como imagens, vídeos, áudios, *softwares*, algoritmos, equações, animações, modelos e simulações, gerados por diferentes formas, tais como de observação, cálculo ou experimentos. (VIDOTTI, 2017).

Alguns tipos de dados têm valor imediato e duradouro, enquanto outros adquirem valor ao longo do tempo; alguns dados são capturados num momento específico e irrecuperável, enquanto outros são passíveis de se reproduzir.

Identificamos como uma das principais problemáticas para implementação da gestão de dados científicos, o idioma dos sistemas de gestão de RI. Esses sistemas funcionam como ferramenta de apoio ao gerenciamento de dados científicos e por serem desenvolvidos em outros países não apresentam o módulo “admin” e “user” em português.

Outra problemática se trata da infraestrutura e dos recursos financeiros nas instituições que pretendem implementar a gestão de dados científicos. Elas precisam se adaptar com uma estrutura de Tecnologia de Informação, pessoal especializado e capacitado no contexto da gestão de dados científicos que contemple competências (de/com):

- Informação digital;
- Curadora digital;
- Habilidades com repositórios
- Gestão da qualidade;
- Representação da informação;
- Aspectos éticos, legais e de responsabilidade em pesquisa; dentre outros.

Também a situação dos investimentos em educação e na ciência nos últimos anos tem ficado aquém das necessidades. Com orçamentos reduzidos e diminuição de bolsas de pesquisa para que instituições, como o IBICT, possa continuar com a realização de adaptações e traduções dos sistemas de apoio à gestão de dados científicos. Esse aspecto tem sido muito prejudicado nos dois últimos anos.

Associado a esse aspecto temos também cortes significativos que as universidades têm vivenciado nos últimos três anos. Esse ponto tem inviabilizado uma série de iniciativas que eram realizadas em parceria e convênios entre as universidades públicas e centros de informação avançada, como o IBICT. Fica muito prejudicado o trânsito de pesquisadores para cursos e implementações dos sistemas *in-loco*.

Com base na literatura (SAYÃO; SALES, 2017; VIDOTTI et al, 2017, SILVA, 2019), os maiores desafios para a importância da implementação de plano de gestão de dados científico são da ordem da conscientização da instituição e de seus pesquisadores.

5.2 PLANO DE GESTÃO DE DADOS CIENTÍFICOS: RI/UNEB SABERABERTO

Com base na definição do ResearchData, da Universidade de Oxford (2013), entende-se nesta proposta que a gestão de dados científicos é um termo geral capaz de cobrir a organização, a estrutura, o armazenamento e o cuidado da informação gerada durante o processo de pesquisa. Desse modo, este estudo entende a relevância da elaboração de um plano de gestão de dados científicos para o RI da UNEB, denominado SaberAberto.

A UNEB possui sua estrutura e abrangência de suas atividades diretamente relacionadas à missão social que desempenha no estado da Bahia. Sendo assim, esta instituição possui 29 Departamentos instalados em 24 *campi*: um sediado na capital do estado, onde se localiza a administração central da instituição, e os demais distribuídos em 23 importantes municípios baianos de porte médio e grande. A UNEB tem como missão a produção, difusão, socialização e aplicação do conhecimento nas diversas áreas do saber.

Assim, objetiva a UNEB a formação integral do cidadão e o desenvolvimento das potencialidades econômicas, tecnológicas, sociais, culturais, artísticas e literárias da comunidade baiana, sob a égide dos princípios da ética, da democracia, das ações afirmativas, da justiça social, pluralidade étnico-cultural e demais princípios do Direito Público.

Atualmente, a Universidade disponibiliza mais de 150 opções de cursos e habilitações nas modalidades presencial e de educação a distância (EaD), nos níveis de graduação e pós-graduação, oferecidos nos 29 Departamentos instalados em 24 *campi*: um sediado na capital do estado, onde se localiza a administração central da instituição, e os demais distribuídos em 23 importantes municípios baianos de porte médio e grande.

Vale destacar, nessa seara, o expressivo crescimento na oferta de cursos *stricto sensu* (mestrados e doutorados) nos últimos anos, em Salvador e outras cidades, promovendo a interiorização da pós-graduação pública, gratuita e de qualidade. Além dos *Campi*, a UNEB está presente na quase totalidade dos 417 municípios do estado, por intermédio de programas e ações extensionistas em convênio com organizações públicas e privadas, que beneficiam milhões de

cidadãos baianos, a maioria pertencente a segmentos social e economicamente desfavorecidos e excluídos.

Figura 7: Universidade do Estado da Bahia. Campus 1- Salvador-Bahia



Fonte: Site da UNEB (2020).

Sendo assim, são oferecidos programas que atendem diversos segmentos da sociedade como a) Alfabetização e capacitação de jovens e adultos em situação de risco social; b) educação em assentamentos da reforma agrária e em comunidades indígenas e quilombolas; c) projetos de inclusão e valorização voltados para pessoas deficientes, da terceira idade, GLBT, entre outros, são algumas das iniciativas que aproximam a universidade da sociedade. Entretanto:

Não se pode também confundir democratização do acesso com atendimento a carências da sociedade, que na maioria das vezes existem como efeito de opções políticas mais amplas assumidas pelos governos. A ideia de que a universidade pública destina-se aos carentes é equivocada ou demagógica. Ela destinasse a todos aqueles que cumprem os requisitos de ingresso. Ocorre que as universidades públicas mais conceituadas têm uma história caracterizada pela manutenção de um elevado padrão de ensino e de exigências, o que já está presente, pelo menos em parte, nos exames de ingresso. Essas exigências estão largamente distanciadas do preparo que o aluno do colégio público recebe. Juntando-se a isto o número cada vez mais elevado de postulantes, o resultado é o afunilamento que, na prática, se traduz nas diferenças socioeconômicas. A universidade não resolve por si mesma e imediatamente as carências sociais, embora possa contribuir, na esfera que lhe é própria, para o encaminhamento de soluções. Está inscrita na própria ideia de autonomia universitária uma relação mediada com a sociedade (SILVA, 2001,p.4).

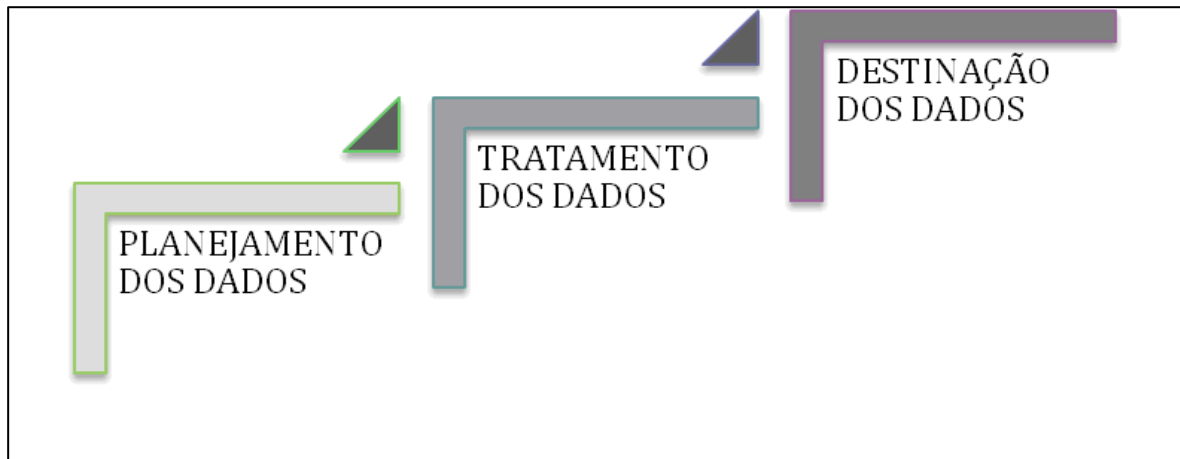
Se considerarmos que a Bahia tem quatrocentos e quinze municípios, deduzimos que a UNEB se apresenta como uma oportunidade de curso superior a

mais da metade deles e pela inclusão da Capital e cidades importantes de suas regiões, no seu *rol* de atendimento, podemos afirmar que esta Universidade se oferece a significativa parcela de nossa população. E possui notória participação no cenário científico brasileiro e produtora de volume expressivo de dados científicos de maneira que precisa repensar seu repositório para que possa atender não somente a salvaguarda e a divulgação de sua produção, mas também a disponibilização de mecanismos que garantam a gestão de dados científicos.

Nesse sentido, diante de sua dinâmica social e aberta, a UNEB demanda de uma estrutura de RI que atenda todas as suas necessidades relacionadas ao ensino, à pesquisa e à extensão. Como produto desta dissertação, procuramos apresentar uma proposta para elaboração de um plano de gestão de dados científicos que possa estar integrado e disponível a partir do seu repositório institucional.

Na nossa proposta, o processo de gerenciamento de dados científicos espera que os sistemas de gestão de RI estejam configurados para que possam se sustentar em três aspectos principais conforme sinaliza a Figura 8 a seguir.

Figura 8: Aspectos principais do processo de gestão de dados científicos apontados neste trabalho



Fonte: Cunha, Jorge e Fialho (2019).

- 1) **Planejamento dos dados científicos:** Planejamento do cuidado a ser dispensado aos dados – muitos financiadores atualmente requerem planos para gestão de dados como parte das propostas;

- 2) **Tratamento dos dados científicos:** Como os dados serão tratados ao longo de todo o ciclo de vida do projeto;
- 3) **Destino dos dados científicos:** O que acontecerá com os dados no longo prazo – o que será feito com eles depois que o projeto terminar. (CUNHA; JORGE; FIALHO, 2019)

Tomando como ponto de partida estes aspectos do processo de implementação, elaboramos nosso plano com base em alguns dos direcionamentos tratados na oficina sobre como elaborar um plano de gestão de dados científicos (SILVA, 2018). Também foram relevantes os fatores já mencionados nos capítulos anteriores desta dissertação. Por fim, considerou-se também os direcionamentos de outros planos já executados em outras universidades públicas brasileiras, a exemplo da USP, UNICAMP e do CARPE DIEN do Instituto de Engenharia Nuclear e Centro de Informações Nucleares. Também nos espelhamos em alguns documentos oficiais de instituições que já implementaram seus planos de gestão de dados científicos ou PGD:

- POLÍTICA DE DADOS E INFORMAÇÕES SOBRE BIODIVERSIDADE DO INSTITUTO CHICO MENDES DE CONSERVAÇÃO DA BIODIVERSIDADE (Instrução Normativa Nº 2, de 25 de novembro de 2015);
- POLÍTICA DE ACESSO A DADOS E INFORMAÇÕES CIENTÍFICAS DO INSTITUTO DE PESQUISAS JARDIM BOTÂNICO DO RIO DE JANEIRO (Portaria JBRJ Nº 077/12, 19/07/2012.);
- POLÍTICA DE DADOS DO PROGRAMA DE PESQUISA EM BIODIVERSIDADE – PPBIO (Portaria Nº 693, de 20/08/2009);

Fundamentados em Sayão e Sales (2015), entendemos que o plano do RI Saber Aberto da UNEB deve contemplar todo o ciclo de vida dos dados (FIGURA 9), tendo início na etapa de planejamento, coleta, assegurar a qualidade, descrição, descoberta, integração e análise.

Figura 9: Etapas Ciclo de Vida dos Dados de Pesquisa para o Saber Aberto



Fonte: Adaptado de Sayão e Sales (2015).

A ideia é que o plano seja um documento sucinto, assim como a recomendação do plano em uso na USP que sinaliza o limite de até duas laudas como sugestão. Ele deve ser tratado como uma carta de intenções que considere o que realmente é necessário para a preservação, compartilhamento e reuso dos dados (SAYÃO; SALES, 2015).

Resolvemos também utilizar o limite de duas páginas considerando que o plano de gestão de dados científicos possa fazer parte do elenco de exigências dos pedidos de auxílio à pesquisa. Nesse contexto, entendemos, com base nas melhores práticas já em execução (USP e UNICAMP) que o número de páginas restrito contribui para a versatilidade tanto do pesquisador, quanto da equipe gestora do RI.

O Plano de Gestão de Dados Científico do SaberAberto da UNEB precisa contemplar todo o ciclo de dados produzidos durante a realização de um trabalho científico, como o exemplo apresentado na figura a seguir.

Esta proposta considera significativo que o pesquisador possa determinar durante o preenchimento do plano, a categoria dos dados quanto a sua classificação de valor. Nesse ínterim existem três tipos de dados: preliminares, públicos e sensíveis:

- **Dados preliminares** são aqueles que já se encontram inseridos no repositório, sendo relacionados a trabalhos ainda não publicados e que, por este motivo, são considerados preliminares, e serão de acesso restrito durante o período de embargo.
- **Dados públicos** são aqueles sem restrições de acesso pelo usuário.
- **Dados sensíveis** são aqueles que, se liberados ao acesso público, podem resultar em efeitos adversos aos ecossistemas ou populações humanas locais, e por isso são passíveis de restrições de acesso pelo usuário.

Figura 10: Dados de pesquisa produzidos potencialmente em uma dissertação



Fonte: Elaboração própria (2020).

No exemplo hipotético desta dissertação, os dados científicos podem ser gerados a partir do projeto de pesquisa. Desse modo, o documento precisa atender dados gerados provenientes de um fluxo de trabalho de observação, computacional

ou experimental. Foi a partir dessa categorização que decidimos apontar os campos que determinam os aspectos mais relevantes para potencializar o arquivamento e a preservação.

O Plano de Gestão de Dados Científicos do SaberAberto da UNEB, assim como nos demais planos em uso nas instituições de pesquisa, irá descrever “o ciclo de vida de gestão para todos os dados que serão coletados, processados e gerados por um projeto de pesquisa” (SAYAO; SALES, 2017, p.15), mas consideramos na nossa proposta que

[...] alguns projetos podem usar apenas parte do ciclo de vida, por exemplo, um projeto envolvido com meta-análise pode se concentrar nas etapas descobrir, integrar e analisar e desconsiderar as outras etapas, ou seja, alguns projetos podem não seguir de forma linear o caminho delineado pelo modelo (SAYAO; SALES, 2017, p.13).

De uma forma resumida, o documento formal constituído nesta dissertação como proposta para o RI da UNEB, visa estabelecer um compromisso entre o pesquisador e a instituição de origem do trabalho/pesquisa. Nosso plano de gestão de dados científicos deverá permitir que os dados dos pesquisadores da UNEB possam ser encontrados, acessíveis, interoperáveis e reutilizáveis por outros em suas investigações (VIDOTTI, et al, 2017).

Também consideramos as recomendações do *Digital Curation Centre* (DCC)⁶⁷ ou Centro de Curadoria Digital para entender a expertise da curadoria da informação digital com foco nas competências em gestão de dados de pesquisa.

Nesse contexto, os componentes, que julgamos imprescindíveis neste primeiro volume do plano do Saber Aberto da UNEB, são os descritos por Silva (2019):

- 1) Descrição dos dados e metadados;
- 2) Possibilidade de atualização;
- 3) Organização;
- 4) Aquisição;
- 5) Processamento;
- 6) Análise;

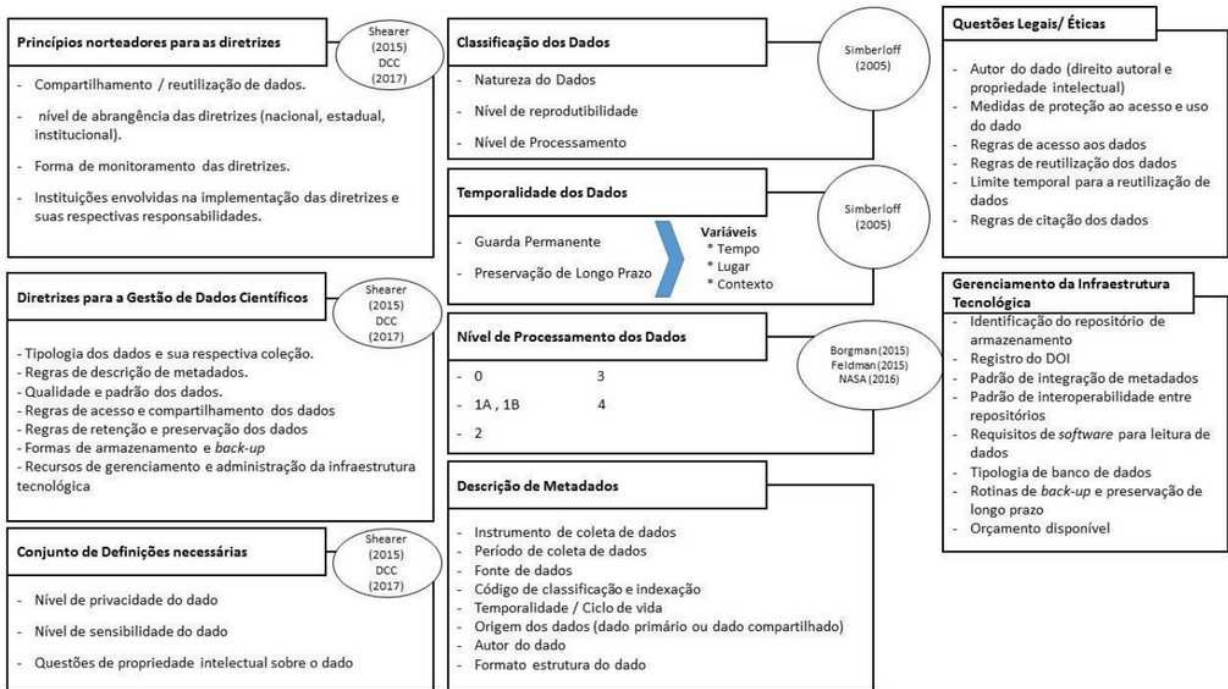
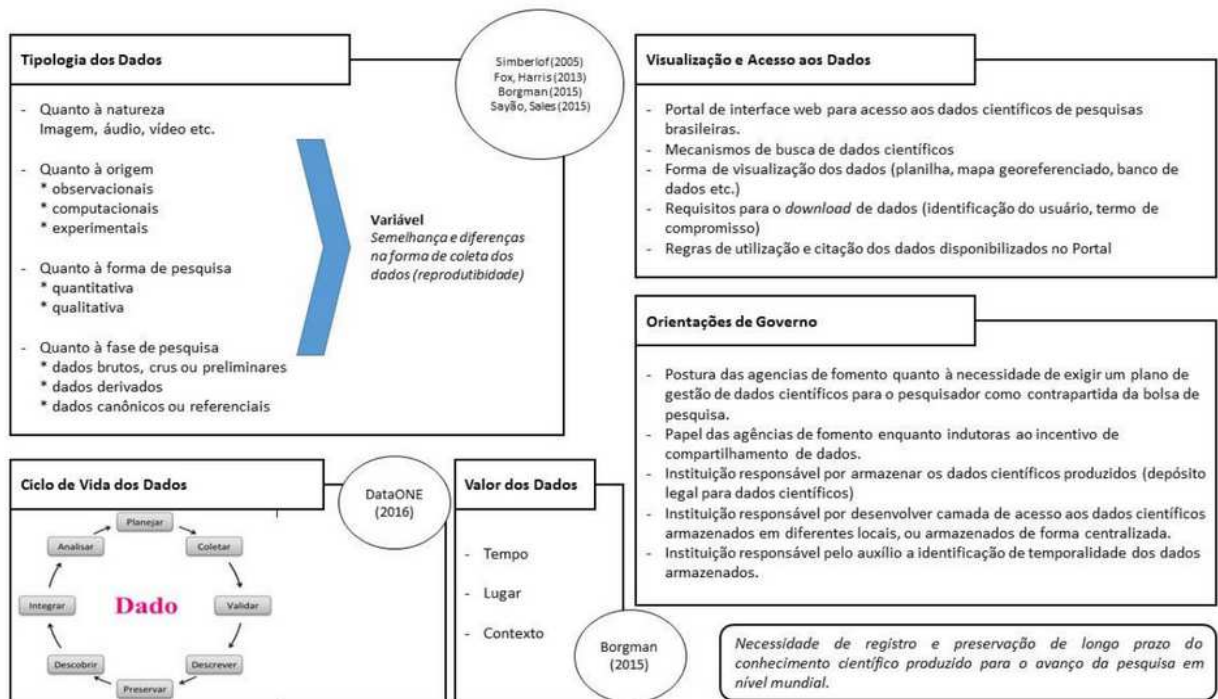
⁶⁷ <https://www.dcc.ac.uk/resources/data-management-plans>.

- 7) Preservação;
- 8) Publicação;
- 9) Identificadores;
- 10) Citação de dados;
- 11) Cópia de dados;
- 12) Ética
- 13) Propriedade intelectual;
- 14) Cópia de segurança;
- 15) Armazenamento a longo prazo;
- 16) Acesso e reutilização;
- 17) Armazenamento a curto prazo de gestão;
- 18) Recursos;
- 19) Pessoal;
- 20) Consideração para compartilhar dados;
- 21) Formas de compartilhar os dados.

Entendemos que o nosso documento precisa estar atento a estes componentes que serão tratados durante todo o desenvolvimento do projeto e, também, após a sua conclusão.

A proposta ora apresentada nesta dissertação seguirá o mesmo fluxo dos autores abordados na fundamentação empírica e teórica. O *framework* elaborado por Costa (2017, p. 222 e 223) está alinhado com as ideias que achamos pertinentes ser consideradas neste momento do nosso projeto:

Figura 11: *Framework* com diretrizes para a elaboração de uma política de gestão de dados científicos



Fonte: Costa (2017).

Concordamos com Costa (2017) que a UNEB precisará realizar uma campanha de conscientização junto aos seus pesquisadores para a importância do seu dado coletado. E a descrição adequada ajudará no armazenamento, preservação e recuperação destes dados à posteriori. Vale salientar que o Plano de Gestão de Dados Científicos (PGDC) constará na Política de Uso do RI Saber Aberto após submetido ao Grupo Gestor da UNEB.

Figura 12: Repositório Institucional Saber Aberto (UNEB) em DSpace



Fonte: Saber Aberto UNEB (2020)

Nota: <http://www.saberaberto.uneb.br/>

A intenção é que o Plano de Gestão de Dados Científicos do Saber Aberto fique disponível no RI e possa ser acessado pelos pesquisadores por meio da aba “Comunidades” no menu superior do repositório. O RI da UNEB foi implementado por meio da plataforma DSpace. Esse tipo de repositório atende de maneira satisfatória o armazenamento dos resultados finais das pesquisas (TCC, Teses, Dissertações, Artigos, Journals, etc.). Entretanto, neste caso, para colocar em prática o nosso PGDC do Saber Aberto, necessitará de implementação

interconectada e articulada com um repositório de dados de pesquisa capaz de proporcionar visibilidade e compartilhamento desses dados brutos de pesquisa, além de dar “[...] crédito ao autor dos dados, preservação digital, memória científica e transparência, segurança dos dados, disponibilidade, curadoria digital, serviços inovadores, reuso dos dados, redes de repositórios e indicadores de qualidade e produtividade da instituição” (SAYÃO; SALES, 2016, p.21).

Neste sentido, recomendamos o uso do RE3DATA, considerando a compatibilidade entre as plataformas técnicas, a interoperabilidade e as perspectivas de preservação dos dados. Os usuários/pesquisadores disponibilizarão seus dados por meio de autoarquivamento. A seguir, um modelo do formulário que deverá ser preenchido no RI Saber Aberto pelos investigadores da UNEB.

Quadro 6: Proposta do Formulário Saber Aberto/UNEB para armazenamento dos dados de pesquisa

IDENTIFICAÇÃO	
Campo	Descrição
Autor	Autor do documento descrito
ORCID do autor	Nº do ORCID
Orientador	Nome do orientador do documento.
Instituição	Instituição a qual o autor do recurso está filiado no momento da publicação
Data de criação	Data criação ou produção do conteúdo
Data de disponibilização	Data que o item fica disponível para o público
Resumo	Resumo
Palavras-chave	Assuntos
Agência de Fomento	Agência de fomento/financiamento.
Nº do processo	Número do processo
DESCRIÇÃO DOS DADOS	
Tipo	Quanto à natureza
Tipo	Quanto à origem
Tipo	Quanto à forma da pesquisa
Tipo	Quanto à fase da pesquisa
Formato	Formato do arquivo
PROTEÇÃO DOS DADOS	
Idioma	Indicação do idioma dos dados
Faz parte de	Referências físicas
Tipo de licença	Informações de licença.

Fonte: Elaboração própria (2020)

Estes são os principais pontos que o pesquisador precisará fornecer junto ao formulário para que se possa inserir as informações necessárias que permitirá identificar sua posição no ciclo de vida dos dados. Esses pontos essenciais do Plano de Gestão de Dados Científicos do Saber Aberto “[...] são a base para que um repositório atenda a disponibilização de dados [...]”, segundo Vidotti et al (2017, p.233), abrangendo também os elementos de um repositório de dados científico.

A seguir, na figura 13, demonstramos uma provável visualização da tela dos dados científicos disponibilizados no repositório institucional da UNEB.

Figura 13: Tela de visualização dos dados científicos no RI Saber Aberto

Informações do Item

Título:

Autor(es):

Palavras-chave:

Data:

Editora:

Resumo:

Descrição

URL:

Coleção

Tipo de licença:

Arquivos neste item:

Arquivo	Descrição	Tamanho	Formato	
Procedimentos Coleta.txt		258 B	Text	Ver/Aberto
table.txt		79,45 kB	Text	Ver/Aberto
TesteAgencia.rar		43,01 MB	Unknown	Ver/Aberto

Aspectos éticos:

Fonte: Elaboração própria (2020)

A figura acima trata-se de uma proposta de como ficaria disponível o registro no RI DSpace da UNEB após implementado o formulário de coleta de dados por

autoarquivamento do plano de dados científicos. De acordo com Vidotti et al (2017, p. 236), destacamos que eventuais mudanças da política do repositório podem alterar os metadados apresentados sempre que for necessário.

O processo contínuo de avanço tecnológico está alterando a forma como a Ciência é desenvolvida. A gestão dos dados científicos, como recursos informacionais, se torna cada vez mais importante para o desenvolvimento de novos estudos, além de contribuir com as instituições envolvidas e interessadas no desenvolvimento da ciência.

Este capítulo procurou apresentar as principais problemáticas e desafios envolvidos na implementação do plano de gestão de dados científicos, para atender ao terceiro objetivo específico de contribuir com o RI/UNEB na perspectiva da estruturação de um plano de gestão de dados científicos, para compartilhamento de informações de forma a permitir a sua reutilização ou reprodução de experimentos e a aplicação de recursos.

6 CONSIDERAÇÕES FINAIS

Esta dissertação tratou da Gestão de Dados Científicos (GDC) tradução do termo inglês *Research Data Management* (RDM). Esse estudo compreendeu referências conceituais relativas a temas como comunicação científica, o acesso aberto ao conhecimento científico, *Open Archives Initiative*, *Open Archival Information System*, *e-science*, curadoria de dados, o movimento de acesso aberto, os repositórios com suas variações tipológicas, e os tipos de dados de pesquisa. Ainda na perspectiva de elucidar o constructo, apresentamos os mecanismos identificados para composição de um modelo, tendo em vista a proposição de um Plano de Gestão de Dados Científicos para o Repositório Institucional AcessoAberto da Universidade do Estado da Bahia (UNEB).

Pesquisadores, das mais variadas áreas do conhecimento, em diversos países e instituições estão dando especial atenção aos dados de suas pesquisas. No contexto da pós-graduação *stricto sensu*, as avaliações estão voltadas ao número de patentes, pesquisas e publicações ou ao grau de internacionalização (VOSGERAU, ORLANDO, MEYER, 2017).

Diante deste novo cenário no fluxo da comunicação científica, os bibliotecários e professores universitários têm sido atores significativos, auxiliando os pesquisadores em um nível mais amplo do processo de pesquisa, em vez de se concentrar unicamente em meios formais de comunicação científica (CORRÊA, 2016).

Nesta dissertação apresentamos um panorama dos repositórios digitais no mundo e no Brasil, considerando suas principais características de suporte, configuração, relação de países com maior número de repositórios digitais de acesso livre, idiomas e áreas de conhecimento predominantes. Buscamos identificar mecanismos e subsídios teóricos empíricos para nos apoiar na composição de um plano de gestão de dados científicos a partir das particularidades do RI SaberAberto da UNEB.

Teve a intenção de problematizar a questão do processo de gestão dos dados científicos, considerando as definições sobre os termos científicos e os repositórios temáticos e institucionais. Foi dada ênfase ao principal sistema de convergência de repositórios para gerência de dados científicos, o RE3DATA. Também foi exposto

alguns elementos do quadro metodológico que foram utilizados para realizar este estudo.

Frente a aspectos relacionados às diretrizes – de ordem da tipologia dos dados, do ciclo dos dados, visualização, compartilhamento e armazenamento – destacadas por Corrêa (2016), elaboramos uma proposta de plano de gestão de dados científicos e como ficariam visíveis e disponíveis para acesso no RI da UNEB.

A origem do desejo de se problematizar a situação dos pesquisadores e sua relação com os repositórios institucionais na área de educação foi uma preocupação que encontrou respaldo nas relações profissionais desempenhadas na nossa trajetória acadêmica. Desse modo, se identificou como problemática que a partilha de dados de investigação tem sido fundamental no processo de avaliação e reprodução de descobertas científicas e deve ser pensada previamente pelo pesquisador e realizada de modo a permitir a sua localização, análise, reutilização e citação.

Com base no estudo realizado, a principal consideração final destaca que os maiores desafios para a importância da implementação de plano de gestão de dados científico são da ordem da conscientização da instituição e de seus pesquisadores. Desse modo, recomendamos que a UNEB promova debates envolvendo todos os atores de interesse no contexto da pesquisa científica universitária.

A pesquisa realizada nos proporcionou conhecimento em torno de como a UNEB tem realizado a gestão dos dados científicos a partir do compartilhamento de informações de forma a permitir a reutilização desses dados ou reprodução de experimentos e a aplicação de recursos em seu repositório institucional; e, em especial, saber como o repositório SaberAberto, da UNEB tem acompanhado o movimento da gestão de dados científicos. Considerando este aspecto, é possível pontuar que o objetivo do estudo foi alcançado.

Assim, o objetivo principal que procurou - identificar e descrever os elementos estruturantes de um Plano de Gestão de Dados Científicos, com perspectiva de aplicabilidade em uma universidade pública - foi atendido neste trabalho dissertativo com base nos seus objetivos específicos. A metodologia foi uma guia utilizada de forma rigorosa para atender tais objetivos.

Reconhecemos que nem todas as questões que emergiram em torno do tema central desta dissertação foram respondidas, ficando como futuras problematizações algumas das seguintes pontuações: a) como o Brasil se posiciona a respeito da

gestão de dados científicos; b) os pesquisadores brasileiros compreendem o retorno social do investimento público realizado em suas pesquisas; c) quais conceitos mais sustentam o debate sobre a gestão de dados científicos no Brasil; d) como as instituições de pesquisa, agências de fomento e universidades brasileiras estão se posicionando quanto ao debate acerca da gestão de dados científicos; e) qual o papel dos repositórios nesse processo de gestão desses dados.

Vale salientar ainda que este estudo desdobrou em novos questionamentos que poderão ser tratados em futuros trabalhos científicos, a saber:

Como os pesquisadores da área de educação têm tido contato com a discussão sobre a gestão dos dados científicos? Como os pesquisadores usam dados de outros pesquisadores e/ou geram dados em suas pesquisas? Estes pesquisadores julgam relevante disponibilizar estes dados? Quais os pesquisadores que têm maior conhecimento sobre a importância dos dados científicos? Os dados científicos mais acessados por pesquisadores brasileiros são do Brasil ou estrangeiros? Os grupos de pesquisa fornecem seus dados científicos? Estes grupos conhecem a relevância desses dados para os futuros estudos? Como os bibliotecários têm participado desse debate na área de Educação? No Brasil, como a gestão de dados científicos têm corroborado com a e-Science? O que motivou as instituições de pesquisa brasileiras a começarem a se preocupar com a gestão de dados científicos? Quais os principais encaminhamentos e estágio de preocupação das instituições de fomento com o potencial dos repositórios digitais para a preservação dos dados da pesquisa?

É da pesquisa que provém parte significativa do financiamento das universidades; dessa forma, o produto da educação superior se sobrepõe aos processos, fazendo da pesquisa o principal componente do reconhecimento docente (VOSGERAU, ORLANDO, MEYER, 2017).

Portanto, o presente trabalho descreveu a importância da gestão dos dados científicos para o avanço da ciência, apontando a situação da UNEB. Essas laudas que trataram sobre os dados de pesquisa, parte central deste trabalho, procurou destacar as nossas principais inquietações e aspectos observados durante nossa trajetória profissional e acadêmica. Nesse contexto, demonstrou a importância do planejamento, tratamento e destinação dos dados dos pesquisadores da Educação.

Assim, espera-se que esta dissertação se junte a outros textos que promovem discussões sobre o momento científico do país, por abordar uma questão relevante

para o desenvolvimento e inovação. Ressalta-se ainda a necessidade de adequação que as instituições de pesquisa precisam realizar para participarem dos projetos que visam internacionalização. Vale salientar que, hodiernamente, a gestão de dados científicos é um aspecto essencial para a internacionalização.

REFERÊNCIAS

- ANGUES, I. . **A importância do modelo OAIS enquanto meio de preservação digital**. 2019. Disponível em: <https://dadosdepesquisa.rnp.br/a-importancia-do-modelo-oais-enquanto-meio-de-preservacao-digital>. Acesso em: 28/04/2019.
- ANJOS, R. L. dos; Dias, G. A.; Rodrigues, A. A.. **Dados científicos: as práticas de gestão dos pesquisadores brasileiros na ciência da informação**. XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017. Disponível em: http://enancib.marilia.unesp.br/index.php/XVIII_ENANCIB/ENANCIB/paper/viewFile/266/1269. Acesso em: 05 maio 2019
- APPEL, Andre Luiz. **A e-Science e as atuais práticas de pesquisa científica**. Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação da UFRB, 2014.
- APPEL, A. L.; Maciel, M.L.; Albagli, S.. **A e-Science e as novas práticas de produção colaborativa de conhecimento científico**. Revista Internacional de Ciencia y Sociedad Volume 3, Número 1, 2016, , ISSN 2340-9991. Disponível em: <http://repositorio.ibict.br/bitstream/123456789/928/1/470-1274-1-PB.pdf> Acesso em: 12/03/2019
- ARAÚJO, F. C. S.. **Acesso livre: uma nova crise no horizonte?**. 2012. Disponível em: http://www.periodicos.capes.gov.br/?option=com_pnews&component=Clipping&view=pnewsclipping&cid=159&mn=0. Acesso em: 30/04/2019.
- BORGMAN, C. L. **Big data, little data, no data: Scholarship in the Networked World**. London: MIT Press, 2015. (Ebook). Disponível em: <http://www.scielo.br/pdf/hcsm/v4n3/v4n3a10>
- BOGDAN, R.; BIKLEN, S. . *Investigação Qualitativa em Educação*. Uma Introdução à Teoria e aos Métodos. Porto: Porto Editor, 1994.
- BEUREN, I. M. et al. Como elaborar trabalhos monográficos. 1ª ed. São Paulo: Atlas, 2003.
- CAVALCANTI, M. T., Sales, L. F.. Gestão de dados de pesquisa: um panorama da atuação da União Européia. **Biblos**: Revista do Instituto de Ciências Humanas e da Informação, v. 31, n. 1, p. 73-98, jan./jun. 2017.
- COCCO, A. P.. Repositórios institucionais de acesso aberto: análise do cenário nos países ibero-americanos. 2012. 195 f. **Dissertação** (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de Santa Catarina, Florianópolis, 2012. Disponível em: Disponível em: <https://repositorio.ufsc.br/handle/123456789/100623>. Acesso em: Acesso em 03/09/2018

CONVENÇÃO DE SANTA FÉ PARA A INICIATIVA DE ARQUIVOS ABERTOS. 2000. Disponível em: http://www.openarchives.org/sfc/sfc_entry.htm. Acesso em: 08/03/2019

CORRÊA, F. C. O papel dos bibliotecários na gestão de dados científicos. **RDBCI: Rev. Digit. Bibliotecon. Cienc. Inf.**, V.14, N.3.

COSTA, M. M.; CUNHA, M. B. O bibliotecário no tratamento de dados oriundos da e-science: considerações iniciais. **Perspectivas em Ciência da Informação**. V. 19, n. 3. Belo Horizonte. Jul./Set. 2014. p. 189-206. Disponível em: <<http://www.scielo.br/pdf/pci/v19n3/a10v19n3.pdf>>. Acesso em: 06 jul. 2017.

COSTA, M. M. . Diretrizes para uma política de gestão de dados científicos no Brasil. 2017. . 288f. **Dissertação** (Mestrado em Ciência da Informação) - Programa de PósGraduação em Ciência da Informação da Universidade de Brasília, Brasília-DF, 2017. Disponível em: file:///C:/Users/1743719/Downloads/2017_Ma%C3%ADraMurrietaCosta.pdf. Acesso em: 01 fev 2020

COSTA, S. . Abordagens, estratégias e ferramentas para o acesso aberto via periódicos e repositórios institucionais em instituições acadêmicas brasileiras. **Liinc em Revista**, v.4, n.2, setembro 2008, Rio de Janeiro, p. 218 - 232 – Disponível em: <http://revista.ibict.br/liinc/article/view/3175/2840>. Acesso em: 04/05/2019.

COSTA, S. . Abordagens, estratégias e ferramentas para o acesso aberto via periódicos e repositórios institucionais em instituições acadêmicas brasileiras. **Liinc em Revista**, v.4, n.2, setembro 2008, Rio de Janeiro, p. 218 - 232 – Disponível em: <http://revista.ibict.br/liinc/article/view/3175/2840>. Acesso em: 04/05/2019

CUNHA, N. N. R. da; JORGE, E. M. de Freitas; FIALHO, N. H. . Gestão de dados científicos: repositório institucional para acesso aberto a trabalho científicos completos na área de educação. **Anais...**, II International Conference on Convergence in Information Science, Technology and Education. 2nd CONCITEC. Salvador: UFBA, 26 a 28 de setembro de 2019. Disponível em: <<http://concitec2.com/pages/gts.html>>. Acesso em: maio 2020.

DA SILVA, F. C. C. . O papel dos bibliotecários na gestão de dados científicos. **RDBCI: Revista Digital De Biblioteconomia E Ciência Da Informação**, 14(3), 387-406. 2016. Disponível em: <https://doi.org/10.20396/rdbci.v14i3.8646333>. Acesso em: maio 2020.

DECLARAÇÃO BETHESDA. Bethesda Statement on Open Access Publishing. 2003. Disponível em: <<http://www.earlham.edu/~peters/fos/bethesda.htm>>. Acesso em 10 set. 2018.

DUDZIAK, E. A. **Gestão de dados de pesquisa: o que precisamos saber hoje!** 2018. Disponível em: <<https://www.sibi.usp.br/?p=17574>> Acesso em: 01/03/2019.

_____. **Competências do Bibliotecário na Gestão de Dados de Pesquisa, Comunicação Científica e Acesso Aberto.** Disponível em: <http://www.sibi.usp.br/?p=5804>. Acesso em: 01/04/2019

_____. **Dados de Pesquisa agora devem ser armazenados e citados .** Disponível em: <http://www.sibi.usp.br/?p=6189>. Acesso em: 01/03/2019.

DSPACE. Disponível em: <https://dspace.mit.edu/>. Acesso em: 01-08-2019

ECO, U.. **Apocalípticos e integrados.** São Paulo: Perspectiva, 1993.

EUROPEAN COMMISSION. **Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020.** v. 1.0. Dec. 2013.

FALCÃO, J. T. da R.; RÉGNIER, Jean-Claude. Sobre os métodos quantitativos na pesquisa em ciências humanas: riscos e benefícios para o pesquisador. **Revista Brasileira de Estudos Pedagógicos**, Brasília, v. 81, n. 198, p. 229-243, maio/ago. 2000

FAUSTO, S. **Evolução do Acesso Aberto:** breve histórico. SciELO em Perspectiva. São Paulo, 2013 Disponível em: <https://blog.scielo.org/blog/2013/10/21/evolucao-do-acesso-aberto-breve-historico/>. Acesso em: 30 abr. 2019.

FERREIRA, A. F. . **A Difusão do Conhecimento na Universidade do Estado da Bahia (UNEB):** Uma Proposta De Implantação Do Repositório Institucional. 2016. 201f. Disponível em: <<http://www.saberaberto.uneb.br/handle/20.500.11896/623> >, Acesso em: 26/06/2018

FERREIRA, V. B. . **E-science e políticas públicas para ciência, tecnologia e inovação no Brasil.** Salvador: EDUFBA, 2018.

FREITAS, H. M. R. **Análise léxica e análise de conteúdo:** técnicas complementares, seqüenciais e recorrentes para exposição de dados qualitativos. Porto Alegre: Sagra Luzzatto, 2000.

FREIRE, P. **Pedagogia da autonomia:** saberes necessários à prática educativa. São Paulo: Paz e Terra, 1996.

GESTÃO de dados de pesquisa: o que precisamos saber hoje! SIBiUSP 2018. Disponível em: <http://www.sibi.usp.br/?p=17574> . Acesso em: 01/03/2019

GUANAES, P. C. V., GUIMARÃE, M. C. S. . Acesso livre: uma nova crise no horizonte? **Revista Gestão e Conexões.** 2012. Disponível em: <http://www.periodicos.ufes.br/ppgadm/announcement/view/20>. Acesso em: 07/07/2019.

HAIR JR., J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. **Análise multivariada de dados.** 5.ed. Porto Alegre: Bookman, 2005.

<http://www.brapci.inf.br/index.php/article/view/0000027234/f7fb2ef061bd7c765769402292caab07>

http://www.periodicos.capes.gov.br/?option=com_pnews&component=Clipping&view=pnewsclipping&cid=159&mn=0. Acesso em: 30/04/2019.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA - IBICT. Manifesto brasileiro de apoio ao acesso livre à informação. Disponível em: <<http://www.ibict.br>. Acesso em 01/09/2018.

_____. **Repositórios Digitais**. Brasília, 2018. Disponível em: <<http://www.ibict.br/informacao-para-a-pesquisa/repositorios-digitais>>. Acesso em 01/09/2018).

_____. Sistema para Construção de Repositórios Institucionais Digitais (DSpace). Disponível em: <http://www.ibict.br/tecnologias-para-informacao/dspace>. Acesso em: 05-06-2019)

KAWASAKI, C. S. . Universidades públicas e sociedade: uma parceria necessária. **Blog Scielo**. Disponível em: <http://www.scielo.br/scielo> > Acesso em 23/03/2016

KAPLAN, A. . **A conduta na pesquisa: metodologia para as ciências do comportamento**. São Paulo: EPU/EDUSP, 1975.

KINDLING, M. . **et al**. O panorama dos repositórios de dados de pesquisa em 2015: uma análise re3data. Revista D-Lib, Março / abril de 2017 Volume 23, Número ¾. Disponível: <http://www.dlib.org/dlib/march17/kindling/03kindling.html>

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos metodologia científica**. 6.ed. São Paulo: Atlas, 2007.

LAKATOS, E. M. ; MARCONI, M. de A. . **Metodologia científica**. São Paulo: Atlas, 1988.

LAKATOS, E. M. . **Fundamentos de metodologia científica**. São Paulo: Atlas, 2010.

_____. **Metodologia do trabalho científico**: procedimentos básicos, pesquisa bibliográfica, projeto e relatório, publicações e trabalhos científicos/ Marina de Andrade Marconi, Eva Maria Lakatos. – 7 ed. – 5 reimp. –São Paulo: Atlas, 2010.

LANCASTER, L. C.; STILLMAN, D. . O Y da questão: como a geração Y está transformando o mercado de trabalho. São Paulo: Saraiva, 2011.

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: 08/03/2019

- LEITE, F. C. L. . **Como gerenciar e ampliar a visibilidade da informação científica brasileira**: repositórios institucionais de acesso aberto. Brasília: Ibict, 2009. Disponível em: <<http://livroaberto.ibict.br/bitstream/1/775/4/Como%20gerenciar%20e%20ampliar%20a%20visibilidade%20da%20informa%C3%A7%C3%A3o%20cient%C3%ADfica%20brasileira.pdf>>. Acesso em 01/09/2018
- MARTINS, D.; SILVA, M.; SIQUEIRA, J. Comparação entre sistemas para criação de acervos digitais: análise dos softwares livres DSpace, EPrints, Fedora, Greenstone e Islandora a partir de novas dimensões analíticas. **InCID: Revista de Ciência da Informação e Documentação**, v. 9, n. 1, p. 52-71, 1 jun. 2018.
- MEDEIROS, C. B. Gestão de Dados Científicos – da coleta à preservação [online]. **SciELO em Perspectiva**. Disponível em: <https://blog.scielo.org/blog/2018/06/22/gestao-de-dados-cientificos-da-coleta-a-preservacao/>. Acesso em: 10 set. 2018
- MEDEIROS, J. da 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, setembro, 2012, Rio de Janeiro, p. 311-322 – Disponível em : <http://revista.ibict.br/liinc/article/download/3368/2969>. Acesso em: 09/05/2019
- MONTEIRO, E. C. de S. de A. ; SANT'ANA, R. C. G. ; SANTARÉM SEGUNDO, J. E. . E-science semântica: integração dos dados na comunicação científica. **Informação em Pauta**, Fortaleza, v. 1, n. 1, p. 9-29, jan./jun. 2016.
- MUELLER, S. P. M. . A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, Ago 2006, Volume 35 Nº 2 Páginas 27 – 38.
- MURAKAMI, T.; FAUSTO, S. Panorama atual dos Repositórios Institucionais das Instituições de Ensino Superior no Brasil. **InCID: Revista de Ciência da Informação e Documentação**, v. 4, n. 2, p. 185-201, 20 dez. 2013. Disponível em: <<http://www.revistas.usp.br/incid/article/view/69327>>.
- NEVES, B. C..**Tecnologia e mediação**. Curitiba: CRV, 2017.
- _____. **Gestão de dados científicos**: definições, conceitos e processo visando internacionalização. In: Fórum de Educação Superior, Inovação e Internacionalização, Salvador, Anais [...] Saçvadpr: Governo do Estado da Bahia, 2018 . Disponível em: <https://virtualeduca.org/encuentros/bahia2018/>. Acesso em: 08/03/2019
- OLIVEIRA, A. C. S.; SILVA, E. M. Ciência aberta: dimensões para um novo fazer científico. **Informação & Informação**, v. 21, n. 2, p. 05-39, 2016. Disponível em: <<http://www.brapci.ufpr.br/brapci/v/a/22211>>. Acesso em: 30 jul. 2017.
- OPENDOAR. **Diretório de Repositórios de Acesso Aberto**. Disponível em: http://v2.sherpa.ac.uk/view/repository_visualisations/1.html. Acesso em: 17/07/2019.

PIERRO, B. de. Uma estratégia para dados: Pesquisadores são estimulados a gerenciar e compartilhar as informações científicas que produzem. **Revista pesquisa**, Edição 267 mai. 2018. Disponível em: https://revistapesquisa.fapesp.br/wp-content/uploads/2018/05/036-039_Gestao-de-Dados_267.pdf. Acesso em: 17/07/2020.

PLANO DE GESTÃO DE DADOS – FAPESP. Disponível em: <http://www.fapesp.br/gestaodedados/#documentos> Acesso em: 20/04/2019.

PROGRAMA FAPESP DE PESQUISA EM ESCIENCE. Disponível em: http://www.fapesp.br/publicacoes/2015/folder_escience.pdf . Acesso em: 20/04/2019.

RE3DATA - **REGISTRO DE REPOSITÓRIOS DE DADOS DE PESQUISA**. Disponível em: <https://www.re3data.org>. Acesso em: 17/07/2019.

REPOSITÓRIO CIENTÍFICO DO INSTITUTO POLITÉCNICO DE CASTELO BRANCO. Disponível em: <https://repositorio.ipcb.pt>. Acesso em: 07-05-2019.

RODRIGUES, E.; et al. Os repositórios de dados científicos: estado da arte. Relatório D-24: **RCAAP**. 2010. Disponível em: https://www.researchgate.net/publication/235754139_Os_repositorios_de_dados_cientificos_estado_da_arte/download . Acesso em: 25 dez. 2018 .

SANTOS JUNIOR, R. dos. **Repositórios institucionais de acesso livre no Brasil: estudo delfos**. DISSERTAÇÃO (MESTRADO). 2010. Disponível em: file:///C:/Users/1743719/Downloads/repositorios_institucionais_santosjunior.pdf. Acesso em 02/09/2018.

SANTOS, E. O. **Mídias e tecnologias na educação presencial e a distância**. Rio de Janeiro: LTC, 2016.

SANTOS, H. M. dos. ; FLORES, D. . **Repositórios digitais confiáveis para documentos arquivísticos: ponderações sobre a preservação em longo prazo**. Perspectivas em Ciência da Informação, v.20, n.2, p.198-218, abr./jun. 2015. Disponível em: <http://www.scielo.br/pdf/pci/v20n2/1413-9936-pci-20-02-00198.pdf>. Acesso em : 08/04/2019

SARAIVA, R.; AMARO, B. *10 Anos de acesso aberto no mundo lusófono*. In: **Conferência Luso-brasileira Sobre Acesso Aberto (CONFOA)**, 3., Lisboa, 2012. Disponível em: <<http://www.slideshare.net/ConfOA/10-anos-de-acesso-aberto-no-mundo-lusfono> >.

SAYÃO, L. F. SALES, L. F. Curadoria geral: um novo patamar para a preservação de dados digitais de pesquisa. **Informação & Sociedade**. V. 22, n. 3. João Pessoa. Set./Dez. 2012. p. 179- 191. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/viewFile/12224/8586>>. Acesso em: 10 jul. 2017.

_____. **Dados abertos de pesquisa: ampliando o conceito de acesso livre.**

RECIIS – Rev. Eletron. de Comun. Inf. Inov. Saúde. 2014 jun.; 8(2) – p.76-92
[www.reciis.icict.fiocruz.br] e-ISSN 1981-6278 | DOI:10.3395/receis.v8i2.934.pt
Disponível em:

<https://www.reciis.icict.fiocruz.br/index.php/receis/article/view/611/1252>. Acesso em: 10 jul. 2020.

_____. **Guia de Gestão de Dados de Pesquisa para Bibliotecários e Pesquisadores.** – Rio de Janeiro : CNEN/IEN, 2015. Disponível em:

<http://www.cnen.gov.br/component/content/article/75-cin/material-didatico-cnen/160-guia-de-gestao-de-dados-de-pesquisa> Acesso em: 05/05/2019.

SAYÃO, L. F.; TOUTAIN, L. M. B. B. . **Implantação e gestão de repositórios institucionais:** políticas, memória, livre acesso e preservação. Salvador, BA: EDUFBA, 2009. 365 p. ISBN 9788523206550 (broch.).

SCHÖPFEL , J. . **Acesso aberto à informação científica em países emergentes.**

Revista D-Lib. Março / abril de 2017. Volume 23, Número 3/4 . Disponível em:

<http://www.dlib.org/dlib/march17/schopfel/03schopfel.html>. Acesso em:13/04/2020

SEMELER, A. Ribas; Blank, V. A. , Rozados, H. B. F. . **Acesso aberto ao conhecimento científico e as métricas informacionais:** metainscrições

científicas para identificar a organização intelectual em periódicos científicos. XXV Congresso Brasileiro de Biblioteconomia, Documento e Ciência da Informação – Florianópolis, SC, Brasil, 07 a 10 de julho de 2013 . Disponível em:

<https://portal.febab.org.br/anais/article/view/1375> . Acesso em:13/04/2019

SETZER, V. W. Dado, Informação, Conhecimento e Competência. **DataGramZero - Revista de Ciência da Informação** - n. zero dez/1999. Disponível em:

<https://www.ime.usp.br/~vwsetzer/datagrama.html>. Acesso em: 06 set. 2018.

SILVA, F.C.C. O papel do bibliotecário na gestão de dados de pesquisa. **RDBCI**, v.14, n.3, 2016. Disponível em:

< <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8646333>> Acesso em: 06 set. 2018.

SEVERINO, A. J. . **Metodologia do Trabalho Científico.** 21 ed. rev. e ampliada. São Paulo: Cortez, 2000.

SILVA, L. A. . O uso pedagógico de mídias na escola: práticas inovadoras, **Reduc**, v.1, 2001. Disponível em: <http://www.educacao.al.gov.br/reduc/edicoes/1a-edicao/artigos> Acesso em:23/12/15

SOUZA, A. H. L. R. de et. al.. O modelo de referência OAIS e a preservação digital distribuída **Ci. Inf.**, Brasília, DF, v. 41 n. 1, p.65-73, jan./abr., 2012. Disponível em: <file:///C:/Users/1743719/Downloads/1352-2016-1-PB.pdf>. Acesso em 10/03/2019.

THOMAZ, Katia P.; SOARES, Antonio José. A preservação digital e o modelo de referência Open Archival Information System (OAIS). **DataGramZero - Revista de Ciência da Informação** - v.5 n.1 fev/04 . Disponível em: <https://www.brapci.inf.br/index.php/article/view/0000007749/6f3e125671cfe45b7fddce5de6cf37f8/>, Acesso em: 01 maio 2019.

TERENCE, A. C. F. ; ESCRIVÃO FILHO, E. . **Abordagem quantitativa, qualitativa e a utilização da pesquisa-ação nos estudos organizacionais**. XXVI ENEGEP - Fortaleza, CE, Brasil, 9 a 11 de Outubro de 2006 ENEGEP 1 2006 ABEPRO . Disponível em: http://www.abepro.org.br/biblioteca/enegep2006_tr540368_8017.pdf. Acesso em: 08 de junho 2019

TRISKA, R. ; Café, Lígia Arquivos abertos: subprojeto da Biblioteca Digital Brasileira. **Ci. Inf.**, Dez 2001, vol.30, no.3, p.92-96. ISSN 0100-1965.

UNIVERSIDADE DE OXFORD. **Research Data Oxford**. Oxford: The University Oxford, 2013. Disponível em: <<http://researchdata.ox.ac.uk/>>. Acesso em: jun de 2017.

VANZ, S. A. de S. ; et. al.. Acesso aberto a dados de pesquisa no Brasil: práticas e percepções dos pesquisadores: relatório 2018. Disponível em: <<http://hdl.handle.net/10183/185195>>.

VARELA, A. ; BARBOSA, M. L. A.; GUIMARAES, I. B. . Dos processos analógicos às tecnologias digitais contemporâneas de recuperação da informação: caminhos cognitivos na mediação para o acesso ao conhecimento. In: SAYÃO, Luíz Fernando *et al.*(Orgs.). **Implantação e gestão de repositórios institucionais**: políticas, memória, livre acesso e preservação. Salvador : EDUFBA, 2009.

VIDOTTI, S. A. B. G.. et. al. Repositório de Dados de Pesquisa para Grupo de Pesquisa: Um Projeto Piloto. **Informação & Tecnologia** (ITEC), Marília/João Pessoa, v.4, n.2, p.221-242, jul./dez. 2017. Disponível em: <http://www.periodicos.ufpb.br/ojs/index.php/itec/article/view/40198>

VOLPATO, S. M. B.; Rodrigues, L. C.; Silveira, A. . Inovação no acervo e no acesso de informações: o sistema de repositório institucional nos Tribunais de Contas do Brasil. **Perspect. ciênc. inf.**, Dez 2014, vol.19, no.4, p.160-181. ISSN 1413-9936 . Disponível em: <<http://www.scielo.br/pdf/pci/v19n4/a10v19n4.pdf>>. Acesso em 03/09/2018

VOSGERAU, D. S. R.; ORLANDO, E. de A.; MEYER, P. Produtivismo acadêmico e suas repercussões no desenvolvimento profissional de professores universitários. **Educ. Soc.**, Campinas , v. 38, n. 138, p. 231-247, jan. 2017 . Disponível em