UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO – UNIRIO
CENTRO DE CIÊNCIAS HUMANAS E SOCIAIS – CCH
PROGRAMA DE PÓS-GRADUAÇÃO EM BIBLIOTECONOMIA – PPGB
MESTRADO PROFISSIONAL EM BIBLIOTECONOMIA – MPB

DISSERTAÇÃO DE MESTRADO PROFISSIONAL EM BIBLIOTECONOMIA

Identificação e mapeamento dos dados científicos da Universidade Federal de Viçosa - MG

Rio de Janeiro, RJ
2021

ANDRÉIA CIPRIANO MACIEL

Identificação e mapeamento dos dados científicos da Universidade Federal de Viçosa - MG

Dissertação apresentada ao Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro como requisito parcial para obtenção do grau de Mestre em Biblioteconomia.

Área de Concentração: Organização e Representação do Conhecimento.

Linha de Pesquisa: Modelo de política de curadoria de dados de pesquisa para o país.

Candidata: Andreia Cipriano Maciel

Orientador: Luís Fernando Sayão

Rio de Janeiro, RJ
2021

CATALOGAÇÃO NA FONTE

Maciel, Andreia Cipriano
MM152i Identificação e mapeamento dos dados científicos da Universidade Federal de Viçosa - MG / Andreia Cipriano Maciel. -- Rio de Janeiro, 2021.
146 f: il; 30 cm

Orientador: Dr. Luís Fernando Sayão.
Dissertação (Mestrado) - Universidade Federal do Estado do Rio de Janeiro, Programa de Pós-Graduação em Biblioteconomia, 2021.

1. Dados científicos. 2. Dados de pesquisa. 3. Gestão de dados científicos. 4. Biblioteca universitária. 5. Small science. I. Sayão, Dr. Luís Fernando , orient. II. Título.

ANDRÉIA CIPRIANO MACIEL

Identificação e mapeamento dos dados científicos da Universidade Federal de Viçosa - MG

Dissertação apresentada ao Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro como requisito parcial para obtenção do grau de Mestre em Biblioteconomia.
Área de Concentração: Organização e Representação do Conhecimento.
Linha de Pesquisa: Modelo de política de curadoria de dados de pesquisa para o país.
Candidata: Andreia Cipriano Maciel
Orientador: Luís Fernando Sayão

Aprovada em: 22 de outubro de 2021.

BANCA EXAMINADORA:

Prof. Dr. Luís Fernando Sayão - Presidente
Universidade Federal do Estado do Rio de Janeiro

Profª. Drª. Luana Farias Sales – Titular Interno
Universidade Federal do Estado do Rio de Janeiro

Profª. Drª. Viviane Veiga – Titular Externo
Fundação Oswaldo Cruz

Profª. Drª. Simone da Rocha Weitzel – Suplente Interno
Universidade Federal do Estado do Rio de Janeiro

Drª Vanessa de Arruda Jorge – Suplente Externo
Fundação Oswaldo Cruz

AGRADECIMENTOS

Primeiramente, agradeço a Deus, pela vida e saúde!

À minha família, meu marido Vinícius, pelo apoio incondicional desde o primeiro momento em que decidi fazer o mestrado. Aos meus filhos, Cícero e Conrado, pela compreensão e carinho durante este trajeto.

À minha sogra Maria, por me auxiliar com o meu filho mais novo, quando resolvi passar uma temporada no Rio de Janeiro para assistir às disciplinas presenciais, sua ajuda foi fundamental neste período.

À Jéssica e Júnior, por cuidarem da minha casa, enquanto estive fora.

Ao meu orientador Luís Fernando Sayão, por conseguir me direcionar com a pesquisa, seu apoio foi crucial, em todas as fases da minha pesquisa.

Aos membros da minha banca, que me apoiaram e contribuíram com o tema proposto e colaboraram grandemente para o andamento e desenvolvimento da pesquisa.

À UFV, que me proporcionou e me liberou durante este período para o aperfeiçoamento do meu trabalho na BBT e, consequentemente, para o meu crescimento profissional.

A todos os professores do PPGB, com os quais eu cursei as disciplinas.

Enfim, muito agradecida a todos que direta ou indiretamente me ampararam durante esta jornada!

RESUMO

MACIEL, Andréia Cipriano. **Identificação e mapeamento dos dados científicos da Universidade Federal de Viçosa – MG.** Orientador: Luís Fernando Sayão. 2021. 146 f. Dissertação (Mestrado em Biblioteconomia) – Programa de Pós-Graduação em Biblioteconomia, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2021.

Esta pesquisa teve como objetivo geral identificar e mapear a tipologia, o acesso, o compartilhamento, o reuso e a gestão dos dados científicos gerados nos ambientes de pesquisa dos Programas de Pós-Graduação da Universidade Federal de Viçosa (UFV). O trabalho procedeu a um levantamento de dados e informações a respeito dos dados científicos processados nos ambientes de pós-graduação da referida universidade. Para que a pesquisa fosse concretizada, foi feita uma revisão bibliográfica sobre "Gestão de dados científicos e suas ramificações". Após identificar o estado da arte, foram elaboradas duas ferramentas de coleta de dados, um questionário *on-line* e uma entrevista estruturada. O primeiro instrumento foi aplicado ao público da pós-graduação, que engloba: pesquisadores, discentes, professores e técnicos administrativos. Posteriormente, foram feitos convites aos respondentes do questionário que tivessem algum vínculo empregatício com a UFV para participarem da entrevista estruturada. Os locais investigados foram os quatro centros de ciência da UFV, compreendendo o Centro de Ciências Agrárias, o Centro de Ciências Biológicas e da Saúde, o Centro de Ciências Exatas e Tecnológicas, o Centro de Ciências Humanas, Letras e Artes e outros setores que direta ou indiretamente trabalham com pesquisa científica. O questionário *on-line* buscou identificar práticas e percepções no que tange aos dados científicos processados por esses pesquisadores. Já a entrevista teve o objetivo de coletar informações/dados a respeito do âmbito administrativo da UFV, ou seja, foi direcionada aos profissionais que exercem cargos de direção/chefia ou outros perfis de atuação e que podiam contribuir com uma visão organizacional e administrativa sobre o tema abordado. Com isso, o mapeamento destes ambientes procurou identificar questões relacionadas a quais tipos de dados criados; se existem acesso, compartilhamento e reuso dos dados científicos gerados; se existe algum tipo de gestão de dados; se as agências de fomento que atuam na UFV têm alguma forma de gestão em relação aos dados científicos, entre outras questões. Seguindo a metodologia desenvolvida, o formulário *on-line* foi enviado para 4.029 contatos de *e-mails* selecionados e foram retornados 183 questionários respondidos. Dos 29 perfis administrativos identificados no formulário, 16 aceitaram participar da entrevista e 7 a concederam. Entre os 183 respondentes, houve um maior volume de respostas do questionário no Centro de Ciências Biológicas e da Saúde 31,7%, seguido pelo Centro de Ciências Agrárias 26,8%, pelo Centro de Ciências Humanas e Artes, 24% e por último o Centro de Ciências Exatas, 17,5%. O formulário também identificou que os tipos de dados científicos produzidos pelos respondentes ficam em torno de 67,7% são textos, 61,2% são planilhas e banco de dados 55,7%. O retorno obtido com as ferramentas foi explanado em gráficos, todos descritos e comentados, expondo, assim, dentro do que foi possível, um panorama geral da comunidade científica no que concerne às práticas e percepções em relação aos dados científicos processados na instituição.

Palavras-chave: Dados científicos. Dados de pesquisa. Gestão de dados científicos. Biblioteca universitária. Small science.

ABSTRACT

MACIEL, Andréia Cipriano. **Identification and mapping of scientific data from the Federal University of Viçosa – MG.** Advisor: Luís Fernando Sayão. 2021. 146 f. Dissertation (Master in Librarianship) – Graduate Program in Librarianship, Federal University of the State of Rio de Janeiro, Rio de Janeiro, 2021.

The general objective of this research was to identify and map the typology, access, sharing, reuse and management of scientific data generated in the research environments of the Federal University of Viçosa Graduate Programs. The work carried out a survey of data and information about the scientific data processed in the postgraduate environments of the aforementioned university. For the research to be carried out, a bibliographic review was carried out on "Management of scientific data and its ramifications". After identifying the state of the art, two data collection tools were developed, an online questionnaire and a structured interview. The first instrument was applied to the postgraduate public, which includes: researchers, students, professors and administrative technicians, later, invitations were made to the respondents of the questionnaire who had some employment relationship with the UFV to participate in the structured interview. The places investigated were the four science centers at UFV, comprising the Center for Agrarian Sciences, the Center for Biological and Health Sciences, the Center for Exact and Technological Sciences, the Center for Human Sciences, Letters and Arts and other sectors that directly or indirectly work with scientific research. The online questionnaire sought to identify practices and perceptions regarding the scientific data processed by these researchers. The interview had the objective of collecting information/data about the administrative scope of the UFV, that is, it was directed to professionals who hold management/management positions or other performance profiles and who could contribute with an organizational and administrative vision about the topic addressed. Thus, the mapping of these environments sought to identify issues related to which types of data were created; whether there is access, sharing and reuse of generated scientific data; whether there is any kind of data management; whether the funding agencies that work at UFV have some form of management in relation to scientific data, among other issues. Following the methodology developed, the online form was sent to 4,029 selected email contacts and 183 answered questionnaires were returned. Of the 29 administrative profiles identified in the form, 16 agreed to participate in the interview and 7 granted it. Among the 183 respondents, there was a greater volume of responses to the questionnaire in the Center for Biological and Health Sciences, 31.7%, followed by the Center for Agrarian Sciences, 26.8%, the Center for Human Sciences and Arts, 24%, and finally the Center for Exact Sciences, 17.5%. The form also identified that the types of scientific data produced by the respondents are around 67.7% are texts, 61.2% are spreadsheets and 55.7% database. The feedback obtained with the tools was explained in graphs, all described and commented, thus exposing, within what was possible, an overview of the scientific community regarding practices and perceptions in relation to the scientific data processed in the institution.

Keywords: Scientific data. Search data. Scientific data management. University library. Small science.

LISTA DE FIGURAS

Figura 1 - Ciclo de vida de dados digitais do *DCC* 38
Figura 2 - Ciclo de vida de dados digitais do *DataOne* 38
Figura 3 - Ciclo de vida de dados digitais do *Research360* 39
Figura 4 - Dados da cauda longa da ciência 42
Figura 5 - Centros de Ciências da UFV 64
Figura 6 - Produções científicas dos últimos 20 anos 67

LISTA DE QUADROS

Quadro 1- Bases específicas e multidisciplinares 18

Quadro 2 - Critérios de busca 18

Quadro 3 - Estrutura do Questionário 22

Quadro 4 - Entrevista 23

Quadro 5 - Objetivos específicos *versus* instrumento de coleta de dados 24

Quadro 6 - Perfil dos gestores participantes da entrevista 95

Quadro 7 - Diretrizes de coleta, tratamento, armazenamento e preservação dos dados científicos 96

Quadro 8 - Existência de documentação formal para política de dados científicos 97

Quadro 9 - Indicação de repositórios de dados 97

Quadro 10 - Obrigações e incentivos a compartilhamento de dados científicos 98

Quadro 11- Ética, privacidade, propriedade intelectual e *copyright* 99

Quadro 12 - Atuações das agências de fomento: planos de gestão de dados 101

Quadro 13 - Atuações das agências de fomento: depósito dos dados em repositórios 102

Quadro 14 - Atuações das agências de fomento: responsabilidade da gestão dos dados 103

Quadro 15 - Política de dados para a instituição 104

LISTA DE TABELAS

Tabela 1- Produções técnico-científicas da pós-graduação de 2019 68
Tabela 2- Projetos de pesquisa e publicações (2018) 68
Tabela 3- Números de pesquisadores, patentes registradas e linhas de pesquisa (2017) 69
Tabela 4- Número de bolsas de iniciação científica (2019) 69
Tabela 5- Repositórios de dados científicos apontados pelos respondentes 81
Tabela 6- Perfis dos gestores 94

LISTA DE GRÁFICOS

Gráfico 1- Escolaridade 73

Gráfico 2- Perfil profissional 73

Gráfico 3- Centro a que sua pós-graduação ou sua pesquisa está vinculada 74

Gráfico 4- Principal financiador 74

Gráfico 5- Termos que melhor descrevem os tipos de dados produzidos 76

Gráfico 6- Tipos de dados que você produz 76

Gráfico 7 - Volume estimado dos dados científicos que você produz em um ano 77

Gráfico 8- Em quais locais você armazena seus dados científicos? 77

Gráfico 9- Você já compartilhou seus dados científicos em acesso aberto? 78

Gráfico 10- Se você respondeu que compartilha seus dados científicos em acesso aberto, onde você compartilha? 78

Gráfico 11- Se você não compartilha, as possíveis razões 79

Gráfico 12- Quão satisfeito você está com a organização dos seus dados 79

Gráfico 13- Familiaridade com a gestão de dados científicos 80

Gráfico 14- Você já elaborou um Plano de Gestão de Dados 80

Gráfico 15- Especifique sua concordância com cada uma das afirmativas a seguir 85

Gráfico 16- Para outros pesquisadores usarem meus dados de pesquisa, eu espero como retorno: 85

Gráfico 17- Onde você acha que seus dados científicos devem ser preservados? 86

Gráfico 18- Serviço de gestão de dados científicos no seu Departamento/Programa, quem é o responsável? 88

Gráfico 19- Conhecimento se os dados produzidos pela sua pesquisa são, de alguma forma, preservados pelo Departamento/Programa? 89

Gráfico 20- Opções de armazenamento dos seus dados que o Departamento/Programa oferece ou indica 89
Gráfico 21- Opções de armazenamento que a agência de fomento oferece ou indica 90
Gráfico 22- Quem você acha que deve oferecer serviços de apoio à gestão dos dados de pesquisa no Departamento/Programa a que você está vinculado? 91
Gráfico 23- Você exerce alguma função administrativa na UFV? 93
Gráfico 24- Sua função 94

LISTA DE ABREVIATURAS E SIGLAS

SIA - Simpósio de Integração Acadêmica (UFV)

PEC - Pró-Reitoria de Extensão e Cultura (UFV)

PPG - Pró-Reitoria de Pesquisa e Pós-Graduação (UFV)

FUNARBE – Fundação Arthur Bernardes (UFV)

ESA - Escola Superior de Agricultura (UFV)

UREMG - Universidade Rural do Estado de Minas Gerais

CEP - Comitê de Ética em Pesquisa com Seres Humanos (UFV)

PPO - Pró-Reitoria de Planejamento e Orçamento (UFV)

RAEX - Registro de Atividades de Extensão (UFV)

PROEXT - Programa de Extensão Universitária (UFV)

SUMÁRIO

1 INTRODUÇÃO……..13
1.1 PROBLEMA E JUSTIFICATIVA ……14
1.2 OBJETIVOS DA PESQUISA ……..16
2 METODOLOGIA……..17
2.1 LEVANTAMENTO DA LITERATURA NAS BASES DE DADOS……...17
2.2 SOBRE OS INSTRUMENTOS DE COLETA DE DADOS……20
2.2.1 Elaboração dos instrumentos de coleta de dados……..21
2.3 POPULAÇÃO E AMOSTRA……...23
2.4 PROCEDIMENTOS DAS ANÁLISES DOS DADOS……25
2.5 LIMITAÇÕES DA DISSERTAÇÃO……26
3 REVISÃO DA LITERATURA ……27
3.1 COMUNICAÇÃO CIENTÍFICA……27
3.2 DADOS CIENTÍFICOS E DADOS DE PESQUISA……...31
3.3 CURADORIA DIGITAL……..35
3.3.1 Curadoria digital no Brasil……39
3.4 CAUDA LONGA DA CIÊNCIA OU *"LONG TAIL SCIENCE"*……41
3.5 POLÍTICA INSTITUCIONAL DE DADOS CIENTÍFICOS……..43
3.5.1 Contexto mundial……44
3.5.2 Contexto brasileiro……...46
3.5.3 Políticas institucionais de dados: definições e estratégias ……...51
3.6 PRINCÍPIOS FAIR……...53
3.7 A BIBLIOTECA UNIVERSITÁRIA NO CONTEXTO DOS DADOS CIENTÍFICOS: DESAFIOS E PERSPECTIVAS……55
4 UNIVERSIDADE FEDERAL DE VIÇOSA……63
5 RESULTADOS OBTIDOS E ANÁLISE DOS DADOS……72
5.1 RESULTADOS E ANÁLISE DOS DADOS REFERENTES AO QUESTIONÁRIO ON-LINE…72
5.1.1 Perfil dos respondentes……72
5.1.1.1 Dados recolhidos……72
5.1.1.2 Análise…… 74
5.1.2 Questões referentes às práticas sobre os dados científicos……...75
5.1.2.1 Dados recolhidos……76
5.1.2.2 Análise ……82
5.1.3 Questões referentes às percepções em relação ao compartilhamento e uso de dados científico ……...84
5.1.3.1 Dados recolhidos……84
5.1.3.2 Análise ……87
5.1.4 Informações organizacionais ……..88
5.1.4.1 Dados recolhidos……88
5.1.4.2 Análise ……91
5.2 APRESENTAÇÃO DOS DADOS OBTIDOS ATRAVÉS DAS ENTREVISTAS……..93
5.2.1 Análise dos dados referentes à entrevista……105
6 CONSIDERAÇÕES FINAIS……..109
REFERÊNCIAS……..112
APÊNDICES ……122
Apêndice 1 – Questionário……...122
Apêndice 2 – Entrevista ……..146

1 INTRODUÇÃO

Reconhecendo que os dados de pesquisa são elementos relevantes para a comunicação científica, para que eles cumpram seu papel no ciclo do desenvolvimento da ciência moderna, é necessário que indagações e investigações sejam levantadas e discutidas (COSTA; LEITE, 2018; SALES; SAYÃO, 2016; WALLIS; ROLANDO; BORGMAN, 2013). Alinhadas à relevância desses novos ativos informacionais, as universidades brasileiras são instituições que necessitam planejar estratégias para que os dados científicos processados em seus ambientes de pesquisa sejam reusáveis e acessíveis à comunidade acadêmica e científica (ALMEIDA, 2019; GOMES, 2019; SANTA ANNA; DIAS; MACULAN, 2019), pois a abertura e o compartilhamento dos processos de pesquisa possibilitam a revisão, a crítica, a autocorreção, o aceite, viabilizando também novas descobertas, ocasionando, assim, novas dinâmicas do conhecimento.

Neste panorama, os processos e fluxos das pesquisas geradas nos laboratórios, nos núcleos de pesquisa e nos departamentos de pós-graduação das universidades podem conter inúmeros registros de dados e produtos de investigações científicas, resultantes de projetos científicos e que, de alguma maneira, não estão devidamente tratados e disponibilizados, tanto para a comunidade local como para o público externo (BORGMAN *et al.*, 2016; RAJASEKAR *et al.*, 2013).

Mesmo com o avanço das tecnologias da informação e comunicação (TICs), muitos locais de pesquisa das universidades brasileiras não conseguiram acompanhar e implementar uma infraestrutura que viabilizasse que os resultados de trabalhos científicos fossem descobertos, referimo-nos, assim, aos ativos informacionais, que não conseguem ser divulgados, quando comparados a objetos convencionais como artigos, teses ou dissertações. Nesse caso, esses ativos informacionais que no andamento de um projeto de pesquisa foram cruciais para se chegar a um produto final, como, por exemplo, uma simulação de algoritmos, um vídeo, um áudio, entre outros, não conseguem ser revelados ao público.

Neste sentido, é válido refletir sobre as etapas de uma pesquisa científica que, no decorrer da sua realização, dependendo da disciplina e/ou do campo estudado, pode ter gerado muitos dados, que, muitas vezes, não fazem parte da divulgação/publicação da pesquisa em questão, ou seja, não são disponibilizados como parte primordial de como foi concluído ou solucionado algum problema de pesquisa.

Nesta perspectiva, é possível que todo esse universo de diversificados tipos de dados científicos possa estar escondido e invisível, o que pode gerar desperdício de recursos,

impedimento de reprodutibilidade dos experimentos, esforços duplicados e, consequentemente, falhas no ciclo da comunicação científica (HEIDORN, 2008; SAYÃO; SALES, 2019). Neste cenário, acrescenta-se também que essa invisibilidade dos dados científicos criados nesses ambientes pode se estender a setores de difusão de conhecimento e informação, como as bibliotecas e os arquivos. Estes setores podem estar completamente alheios aos ciclos, fluxos e produtos decorrentes das investigações científicas que ocorrem nos laboratórios, nos departamentos, entre outros (CARLSON, 2011; SANTA ANNA; DIAS; MACULAN, 2019). Com isso, inevitavelmente as bibliotecas universitárias terão que renovar e acompanhar essas novas formas de compartilhamento e os novos modos de produzir ciência. Assim, pensando em como estas organizações podem contribuir nestes novos processos de produção de conhecimento, a abordagem do estudo foi identificar e mapear a gestão, o acesso, o compartilhamento e reuso dos dados científicos gerados nos ambientes de pesquisa dos Programas de Pós-Graduação da UFV.

Dessa forma, a dissertação foi organizada em seis capítulos. No primeiro capítulo, é apresentado o contexto com uma visão geral da temática, explanando a motivação da pesquisa, o problema, a justificativa e os objetivos. No segundo capítulo, são apresentadas a proposta metodológica da revisão da literatura, da elaboração e aplicação dos instrumentos de coleta de dados e os procedimentos de análise dos dados recolhidos. No terceiro capítulo, são apresentadas a revisão da literatura, que engloba temas como comunicação científica, dados de pesquisa e dados científicos, curadoria digital, cauda longa da ciência ou *long tail science*, política institucional de dados científicos, princípios *fair* e a biblioteca universitária no contexto dos dados científicos. No quarto capítulo, são expostas informações sobre a pesquisa científica e os programas de pós-graduação da Universidade Federal de Viçosa. No quinto capítulo, são apresentados os resultados obtidos e a análise dos dados. Por fim, no último capítulo, são apresentadas as considerações finais da dissertação.

1.1 PROBLEMA E JUSTIFICATIVA

Durante a pesquisa exploratória sobre o tema, notou-se como o campo “dados científicos e de pesquisa” é amplamente rico em suas vertentes. No entanto, foi possível direcionar o planejamento da pesquisa para uma discussão que levantasse questões referentes à atuação de universidades e às bibliotecas universitárias. Assim, de acordo com a literatura analisada, os dados científicos se tornaram ativos informacionais extremamente relevantes em um processo de pesquisa científica, é notório como essa premissa vem cheia de diversidades. Melhor dizendo, para que os dados científicos estejam disponíveis a uma comunidade

científica, serão necessários inúmeros esforços, principalmente em universidades brasileiras. Diante disso, os esforços incluem, primeiramente, o engajamento do profissional bibliotecário, um engajamento que engloba diversas atitudes, como, por exemplo, desenvolver pesquisas que tratem do assunto, reunir-se com profissionais de outras áreas como arquivistas, professores de pós-graduação, profissionais de TI, dentre outros. A união desses profissionais terá como objetivos a elaboração de projetos de gestão e curadoria, a criação de políticas de dados, e outros. Assim, fica a cargo dos bibliotecários também criar estratégias e artifícios para que as bibliotecas universitárias possam conquistar um papel relevante no contexto dos dados científicos.

Dado o contexto, a motivação para a proposta da pesquisa se deu pela publicação anual[1] da UFV de um documento em que são divulgados informações e dados sobre as pesquisas científicas. Esse documento mostra que, no ano de 2017, foram publicados estudos resultantes de pesquisas científicas, num total de 3.412 trabalhos, dos quais 1.367 foram indexados na *Web of Science-Clarivate Analytics*. Esse relatório expressa que a UFV é uma instituição que tem um perfil relevante na pesquisa nacional em vários segmentos, principalmente nas áreas relacionadas às Ciências Agrárias e Ciências Biológicas, podemos incluir, neste cenário, a biblioteca central, que é um órgão estratégico que pode apoiar em questões relacionadas à gestão dos dados científicos. Assim, a partir da conjuntura exposta, foi formulado e declarado o problema questionado por esta pesquisa: quais são as práticas, percepções e procedimentos de gestão de dados científicos da comunidade científica da Universidade Federal de Viçosa?

Com isso, a pesquisa buscou conhecer como são as práticas e as percepções da comunidade científica da UFV no que tange a questões como perfis da comunidade acadêmica e científica, práticas quanto aos dados científicos, percepções em relação ao acesso, reuso e compartilhamento dos dados científicos, ou seja, o estudo objetivou identificar como é o panorama das investigações científicas que os pesquisadores desenvolvem nos departamentos, nos laboratórios, nos grupos de pesquisa, entre outros. Adicionado a este processo de descoberta, o trabalho objetivou também mapear como se dão as dinâmicas dos processos e procedimentos adotados pelas unidades de pós-graduação em relação ao tratamento dos dados científicos.

[1] Relatório de atividade de 2018: ano base 2017.

1.2 OBJETIVOS DA PESQUISA

Objetivo geral

- Identificar e mapear a tipologia, o acesso, o compartilhamento, reuso e a gestão dos dados científicos gerados nos ambientes de pesquisa dos Programas de Pós-Graduação da UFV, para obter subsídios na elaboração de projetos direcionados ao tema na referida instituição.

Objetivos específicos:

OE 1 - Identificar o “estado da arte” sobre dados científicos e dados de pesquisa e suas ramificações.

OE 2 - Identificar as práticas e percepções da comunidade científica em relação aos dados científicos no âmbito da UFV.

OE 3 - Identificar a estrutura organizacional dos Programas de Pós-graduação frente à gestão de dados científicos.

2 METODOLOGIA

A construção da metodologia foi formulada com base na literatura encontrada sobre o tema, assim, a pesquisa teve um planejamento de natureza exploratória e descritiva, “sendo desenvolvida com o objetivo de proporcionar uma visão geral, de tipo aproximativo, acerca de determinado fato” (GIL, 2008, p. 27). A pesquisa exploratória tem o objetivo de propiciar mais informações sobre determinado tema, ocasionando uma delimitação de um tema de trabalho (ASSIS, 2008). Nos estudos exploratório-descritivos, segundo Lakatos e Marconi (2002, p. 85), “podem ser encontradas tanto descrições quantitativas e/ou qualitativas quanto acumulação de informações detalhadas como as obtidas por intermédio da observação participante”. Nesse viés, Gil (2008, p. 27) pontua: “As pesquisas descritivas são, juntamente com a exploratória, as que habitualmente realizam os pesquisadores das ciências sociais [aplicadas, pois são] preocupados com a atuação prática”. Analisando as características da pesquisa descritiva, Assis esclarece:

> A [pesquisa] descritiva visa a observar, registrar, analisar, classificar e interpretar os dados sem interferência, sem manipulação do pesquisador. Entre essas pesquisas, estão as de opinião, as mercadológicas, os levantamentos socioeconômicos e psicossociais, em suma, a maioria das pesquisas desenvolvidas nas Ciências Humanas e Sociais. Envolve técnica padronizada de coleta de dados, realizada principalmente através de questionários e da observação sistemática. (ASSIS, 2008, p. 18)

É com as técnicas da pesquisa descritiva que o pesquisador desenvolve o estudo, a análise e as anotações e transcrições do fato ou o objeto observado, anulando a interação ou influência do pesquisador no tema estudado.

2.1 LEVANTAMENTO DA LITERATURA NAS BASES DE DADOS

Esta foi a etapa para identificar o estado da arte dos dados da pesquisa, foi o passo inicial para a proposta da pesquisa, e com o procedimento da pesquisa bibliográfica, foi possível levantar informações sobre o tema em questão.

> A pesquisa bibliográfica é um apanhado geral sobre os principais trabalhos já realizados, revestidos de importância por serem capazes de fornecer dados atuais e relevantes relacionados com o tema. O estudo da literatura pertinente pode ajudar a planificação do trabalho, evitar duplicações e certos erros, e representa uma fonte indispensável

> de informações, podendo até orientar as indagações (LAKATOS; MARCONI, 2002, p. 25).

Nesta etapa, a estratégia foi dividir as bases em específicas e multidisciplinares. No Quadro 1 são apresentados a execução e os resultados de busca nas bases selecionadas.

Quadro 1 - Bases específica e disciplinares

Bases especializadas			
	Documentos encontrados	**Artigos rejeitados**	**Artigos selecionados**
Brapci	**27 registros**	**6 registros**	**21 (artigos e comunicação oral)**
LISA[2]	**35 registros**	**28 registros**	**5 (artigos)**
Bases multidisciplinares			
Web of science	**128 registros**	**109 registros**	**19 (artigos)**
Scopus	**86 registros**	**61 registros**	**25 (artigos)**
Outras fontes de pesquisa			
Portal metabuscador: **OASIS-BR**	**94 registros**	**80 registros**	**14 (artigos)**

Fonte: elaborado pela autora, 2021

Neste processo, foram selecionados 84 documentos, dos quais 72 foram utilizados para compor o desenvolvimento da pesquisa. Pontuamos também que, neste processo, foi necessário estabelecer os critérios de busca, no intuito de obter um bom desempenho no índice de revocação, assim, seguem no Quadro 2 os critérios usados.

Quadro 2 - Critérios de busca

INCLUSÃO	
1.	**Os termos poderiam estar no título, resumo, palavras-chave, indexador da base e/ou texto**
2.	**Tipo de documento: artigo de pesquisa, conferência**
3.	**Idioma: inglês, espanhol e português**
4.	**Conteúdo referente aos dados científicos, dados de pesquisa, gestão de dados, cauda longa da ciência, biblioteca universitária**
5.	**Recorte temporal: 2012 a 2020**
6.	**Acesso ao documento na íntegra**
7.	**Palavras-chave utilizadas:** **Português: dados de pesquisa, dados científicos, gestão de dados, compartilhamento de dados de pesquisa, cauda longa da pesquisa, biblioteca universitária.** **Inglês: *research data, scientific data, research data management, digital curation e long tail of science, data share, research data sharing, library university.*** **Espanhol: Compartir datos de investigación, datos científicos, bibliotecas universitarias, curaduría de datos, datos de búsqueda.**

Fonte: elaborado pela autora, 2021

[2] Na base de dados Library and Information Science Abstracts (LISA), foi usada somente a palavra-chave *long tail of science*

É válido informar que o recorte temporal entre 2012 e 2020, para as buscas nas bases selecionadas, se pautou no fato de que, no Brasil, o assunto trabalhado ainda está em sua fase embrionária nas áreas da Ciência da Informação e Biblioteconomia. Embora, no desenvolvimento do estudo, tenham sido identificadas outras bibliografias anteriores a 2012 ou posterior a 2020.

Destaca-se aqui também as principais características das bases selecionadas para o levantamento bibliográfico. As plataformas possuem as seguintes características:

- **Brapci** (Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação) - O objetivo desta base é subsidiar estudos e propostas na área de Ciência da Informação, fundamentando-se em atividades planejadas institucionalmente. Com esse propósito, a plataforma conseguiu identificar títulos de periódicos da área de Ciência da Informação (CI) e indexar inúmeros artigos, constituindo-se a base de dados referenciais. Ela amplia o espaço documentário, facilita a visão de conjunto da produção na área, ao mesmo tempo que revela especificidades do domínio científico. Atualmente disponibiliza referências e resumos de 19.255 textos publicados em 57 periódicos nacionais impressos e eletrônicos da área de CI. Dos periódicos disponíveis, 40 estão ativos e 17 históricos (descontinuados). A plataforma possui três módulos na sua configuração: o módulo público, para consultas rápidas; o módulo de manutenção, para revisão e correções da base; e o módulo pesquisador, para análises bibliométricas. Para o público, é disponibilizada uma nuvem de *tags* como interface visual para guiar o usuário, sugerindo termos de busca e recuperação. A nuvem é composta pelas etiquetas mais consultadas, isto é, o tamanho da fonte das *tags* representa proporcionalmente a quantidade de utilizações da etiqueta nas buscas dos usuários. Com essas características, a Brapci é uma base de dados cujas condições de efetivação de pesquisas a situam como paradigmática, graças ao desenvolvimento do seu modelo, marcado pelo contínuo aperfeiçoamento na sistematização e organização da literatura periódica da área, em prol da localização e obtenção de artigos de periódicos científicos da área de Ciência da Informação (CI), oferecendo suporte à pesquisa, à organização e à análise de dados. Além disso, tem subsidiado estudos na busca da melhoria na qualidade das publicações periódicas da área de Ciência da Informação e correlatas, contribuindo para socializar saberes editoriais. O acesso e download é gratuito e a base é mantida e administrada pela Universidade Federal do Paraná e pela Universidade Federal do Rio Grande do Sul.

- **Oasisbr** - Inserido no contexto do Movimento em favor do Acesso Aberto à Informação Científica, este portal metabuscador reúne a produção científica nacional em acesso aberto. O portal permite, por meio de uma única interface, a pesquisa simultânea em repositórios digitais, teses e dissertações e periódicos científicos eletrônicos. Por meio do portal é possível

consultar e fazer o download do texto completo, sem nenhum custo, de artigos científicos, teses, dissertações, livros, capítulos de livros, trabalhos apresentados em eventos, entre outros documentos que constituem a produção científica brasileira. O portal também possui parcerias com La Referência (Rede Federada de Repositórios Institucionais de Publicações Científicas), RCAAP (Repositórios Científicos de Acesso Aberto de Portugal), COAR (*Confederation of Open Access Repositories*), dentre outros.

- ***Scopus*** - é o maior banco de dados de resumos e citações da literatura com revisão por pares, engloba revistas científicas, livros, processos de congressos e publicações do setor. Oferece um panorama abrangente da produção de pesquisas do mundo nas áreas de ciência, tecnologia, medicina, ciências sociais, artes e humanidades. A solução *Scopus* disponibiliza ferramentas inteligentes para monitorar, analisar e visualizar pesquisas. Através da assinatura com a Elsevier, o Portal de Periódicos da Capes disponibiliza o acesso à base.

- ***Web of science*** - É um site que fornece acesso baseado em assinatura a vários bancos de dados que fornecem dados abrangentes de citações para muitas disciplinas acadêmicas diferentes. Atualmente é mantido pela *Clarivate Analytics*. É uma base de dados referencial, ou seja, apresenta as informações bibliográficas e resumo dos materiais que disponibiliza, eventualmente mostra o link externo para o texto completo. Além disso, é possível fazer análise de citações, referências, ver *índice h* e outras métricas. O conteúdo da base é multidisciplinar e os tipos de materiais são artigos de periódicos, livros, capítulos de livros, revisões, anais de eventos, entre outros. Por meio do Portal de Periódicos da Capes é possível acessar esta base.

- ***LISA (Library and Information Science Abstracts)*** - É um serviço internacional de resumo e indexação direcionado à profissionais e acadêmicos da área de Biblioteconomia e Ciência da Informação. A base LISA integra mais de 440 periódicos de mais de 45 países em 20 línguas. O acesso é pago e é necessário registro. A base faz parte do sistema Proquest, localizado no estado de Michigam (Estados Unidos da América).

2.2 SOBRE OS INSTRUMENTOS DE COLETA DE DADOS

Objetivando desenvolver ferramentas para identificar o contexto dos ambientes de pesquisa da Universidade Federal de Viçosa, foram criados dois instrumentos de coleta de dados, delineados através de estudos anteriores de Borgmam (2010): Tenopir *et al (*2011); Príncipe *et al* (2014); Costa 2(017); Veiga, (2017) e Vanz *et al* (2018).

Os estudos citados elaboraram ferramentas de coleta de dados e também contribuíram em estabelecer as estratégias para a confecção tanto de questionários como de entrevistas. Os autores buscavam identificar práticas e percepções em temas como ciência aberta, compartilhamento de dados de pesquisa, política nacional e institucional para dados científicos e dados de pesquisa em instituições que desenvolvem pesquisa científica. A construção dos instrumentos dos referidos estudos objetivou mapear dados e informações como: tipos de dados científicos ou dados de pesquisa, comportamento referente ao acesso, compartilhamento e reuso de dados e infraestruturas organizacionais de gestão e curadoria de dados científicos. Esses procedimentos de recolha de dados possibilitaram levantar informações dos cenários dos ambientes de pesquisa, resultando, dessa forma, na viabilização de projetos de infraestrutura de serviços relacionados à gestão e à curadoria de dados de pesquisa.

Assim, a elaboração e a utilização desses instrumentos foram fundamentais no mapeamento das atividades de pesquisa desenvolvidas na UFV, uma vez que é nesse processo que “consiste reconhecer o usuário como ponto central no planejamento e gerenciamento das atividades [relacionadas à gestão e curadoria de dados científicos] e, a partir desse reconhecimento, passar realmente a considerá-lo no desenho e produtos e serviços [que podem ser oferecidos]” (CARVALHO, 2004, p. 102). Dessa maneira, a seguir são descritas as ferramentas criadas para tais fins.

2.2.1 Elaboração dos instrumentos de coleta de dados

Esperando atender aos requisitos da pesquisa exploratória e descritiva, a pesquisa de campo se baseou no quantitativo-descritivo, pois utilizou técnicas de entrevista e questionário. Em relação a estes métodos, Lakatos e Markoni (2002) salietam que:

> O quantitativo-descritivo consiste em investigações de pesquisa empírica cuja principal finalidade é o delineamento ou análise das características de fatos ou fenômenos […] as várias técnicas como entrevistas, questionários, formulários etc. empregam procedimentos de amostragem. Todos eles empregam artifícios quantitativos, tendo por objetivo a coleta sistemática de dados sobre populações […] (LAKATOS; MARCONI, 2002, p. 84).

Dessa forma, para atender aos objetivos 2 e 3 da pesquisa, os métodos de recolha de dados elaborados foram:

- Objetivo 2: Um formulário criado no *Google forms* (Apêndice 1) enviado por *e-mail*, que foi direcionado ao público da pós-graduação, que engloba: pesquisadores, discentes,

professores e técnicos administrativos (estes últimos, no período da pesquisa, estavam matriculados em algum curso de pós-graduação).

- Objetivo 3: uma entrevista estruturada (Apêndice 2) com os funcionários (coordenadores, diretores, editor-chefe, professores e técnicos administrativos) dos programas de pós-graduação, laboratórios de pesquisa, associações de pesquisa, corpo editorial de revistas científicas, entre outros. Através do formulário aplicado por *e-mail*, foi possível identificar os perfis dos entrevistados que tinham alguma função administrativa ou acadêmica. Por meio desta ferramenta, os servidores podiam optar por conceder individualmente a entrevista, ou não. Assim, se o funcionário aceitasse participar da entrevista, era agendado um encontro virtual, sendo usado para tal a ferramenta *Google meet.* Cada entrevista teve, em média, a duração de 15 a 25 min. Na sequência, os instrumentos são mais clarificados em seus objetivos.

- **Questionário *on-line* (formulário)**: este instrumento buscou identificar as práticas e percepções de mestrandos, doutorandos, pós-doutorandos e professores dos programas de pós-graduação em relação aos dados científicos criados e processados em seus ambientes. A ferramenta foi dividida em 4 partes e composta por 36 questões. No Quadro 3 está descrita a composição das partes.

Quadro 3 - Estrutura do Questionário

1 PERFIL DOS RESPONDENTES A primeira parte consiste em coletar dados e informações demográficas dos respondentes, objetivando identificar características como, escolaridade, perfil profissional, em qual Centro/departamento a pós-graduação está inserida e a principal agência de financiamento.
2 QUESTÕES REFERENTES ÀS SUAS PRÁTICAS SOBRE DADOS DE PESQUISA A segunda parte é para averiguar como são as práticas da comunidade acadêmica sobre dados de pesquisa. Nesta parte, são detectadas as práticas de uso, reuso e compartilhamento dos dados científicos dos respondentes.
3 QUESTÕES REFERENTES ÀS SUAS PERCEPÇÕES EM RELAÇÃO AO COMPARTILHAMENTO E USO DE DADOS Na terceira parte da ferramenta, busca-se verificar a percepção dos respondentes no que concerne ao compartilhamento e uso de dados científicos.
4 INFORMAÇÕES ORGANIZACIONAIS A quarta parte é para que os respondentes possam informar sobre as atividades administrativas que eles podem usufruir no tocante à gestão e à curadoria dos dados científicos dos laboratórios, departamentos, núcleos de pesquisa, entre outros ambientes.

Fonte: elaborado pela autora, 2021

- **Entrevista com os coordenadores e dirigentes que têm alguma função administrativa ou acadêmica da Universidade Federal de Viçosa:** a elaboração do instrumento pretendeu identificar dados e informações relacionadas à infraestrutura organizacional dos setores que gerenciam atividades de pesquisa científica da universidade em questão. O Quadro 4 contém a estrutura desenvolvida.
-

Quadro 4 - Conteúdo da Entrevista

O instrumento é composto por um bloco com 12 questões que buscaram mapear: • informações demográficas (setor de atuação) dos entrevistados; • diretrizes de coleta, tratamento, armazenamento e preservação dos dados científicos; • existência de documentação formal para a política de dados científicos; • questões relacionadas à indicação de repositórios de dados; • obrigações e incentivos ao compartilhamento de dados científicos; • questões relacionadas à ética e privacidade e à propriedade intelectual, *copyright;* • questões referentes às atuações das agências de fomento; • relevância de uma política de dados para a instituição.

Fonte: elaborado pela autora, 2021

2.3 POPULAÇÃO E AMOSTRA

Para identificar a população, foi utilizado o banco de dados do sistema de gestão do acervo utilizado pela Biblioteca Central da UFV. Através dos relatórios que esse sistema gerou, foi possível proceder a um levantamento de contatos (*e-mails*) de alunos de mestrado, doutorado, pós-doutorado e professores do *campus* de Viçosa. A escolha do uso do sistema de gestão do acervo se baseia na atualização semestral dos cadastros dos usuários da biblioteca, pois o sistema é conectado com o registro escolar da instituição.

Sendo assim, esta etapa forneceu a proporção da população, pois foi possível localizar o contato de 1.728 mestrandos, 1.345 doutorandos e pós-doutorandos e de 956 professores, resultando um total de 4.029 *e-mails.* Notoriamente, a amostra foi autosselecionada, pois foi composta pelo número de respondentes que aceitaram responder o questionário e consentiram em participar da entrevista.

Logo após identificar a população, iniciou-se a testagem do formulário antes de sua aplicação definitiva. Esta validação buscou identificar possíveis falhas, avaliar o processo de

coleta de dados e analisar os *feedbacks* dos respondentes, ou seja, aprimorar sua estrutura. Para isso, a ferramenta foi aplicada em uma população/universo com características semelhantes à população-alvo. Neste sentido, Assis (2008, p. 28) explica que o pré-teste serve "para que sejam testadas sua validade, seu nível de compreensão e as dificuldades existentes e somente depois, é que se aplica o instrumento na população/universo da pesquisa. A elaboração requer a observação de alguns aspectos como precisão, clareza, estruturação lógica, estética, quantidade de perguntas, organização cronológica das perguntas".

Com isso, do teste piloto do formulário participaram um mestrando, uma doutoranda (ambos da pós-graduação da UFV), um doutorando (o único da área da Ciência da Informação) da Universidade de Coimbra e dois professores universitários (Universidade Federal de Minas Gerais e Universidade Federal de Alfenas). Assim, o pré-teste foi iniciado em 19/05/2021 e finalizado em 26/05/2021. Após as sugestões advindas do teste piloto, o instrumento de coleta de dados sofreu diversas alterações.

Após criar esse banco de dados de *e-mails* e feito o pré-teste, o formulário foi enviado no dia 19/07/2021 e a data final foi em 09/08/2021 . Dos 4.029 *e-mails* enviados, 122 retornaram, pois foram identificados como *spam* ou endereços não encontrados. É válido explanar aqui também que, no dia 02/08/2021, foi reenviado o formulário para a categoria dos professores, pois seu retorno na primeira aplicação foi baixo.

Assim, em relação aos mecanismos de coleta de dados utilizados para conceber a pesquisa descritiva apresentada, foi adotada a coleta de dados quantitativos, mas em determinadas etapas, quando necessário, foram recolhidos dados qualitativos. Devido a isso, para cada objetivo específico, foram utilizadas as ferramentas delineadas para tal tarefa, conforme descrito no Quadro 5.

Quadro 5 - Objetivos específicos *versus* instrumento de coleta de dados

Objetivos específicos	Instrumento de coleta	Fonte de coleta de dados
Objetivo 1: Levantamento e revisão exploratória da literatura.	Levantamento bibliográfico.	As bases de dados BRAPCI, LISA, *Web of science e Scopus;* o Portal metabuscador OASIS-BR; e a plataforma Google Scholar
Objetivo 2: identificar as práticas e percepções da comunidade científica em relação aos dados científicos no âmbito da UFV.	Questionário on-line direcionado aos discentes e professores do *campus* UFV.	Comunidade científica da UFV do Campus de Viçosa.
Objetivo 3: identificar a estrutura organizacional dos Programas de Pós-Graduação frente à gestão de dados científicos.	Entrevista com servidores administrativos e acadêmicos da UFV.	Funcionários com funções administrativas e cargos acadêmicos que consentiram na realização da entrevista.

Fonte: elaborado pela autora, 2021

2.4 PROCEDIMENTOS DAS ANÁLISES DOS DADOS

- **Resultados e análise dos dados do questionário online**

Foi ancorada no conceito de “pesquisa descritiva” (GIL,2004) que pontua o objetivo desse tipo de pesquisa, que é levantar as características de determinada população ou fenômeno, objetiva mapear opiniões, atitudes e crenças de uma população. O levantamento não exige aprofundamento sobre o objeto ou fenômeno estudado. Assim, respaldado no conceito da pesquisa descritiva, a análise dos dados desta ferramenta, buscou identificar e descrever as características da comunidade científica da UFV, ou seja, visou identificar e descrever os perfis dos respondentes, escolaridade, tipos de dados científicos produzidos, e outras questões. Pontua-se que a aplicação da pesquisa teve o intuito de futuramente utilizar o estudo em questões práticas como o desenvolvimento de novos produtos, mudanças de comportamento, melhor dizendo, o estudo forneceu informações e dados estratégicos para fundamentar tomadas de decisões no que se refere a atividades voltadas a dados científicos. Segundo Gil (2004, p. 28), “pesquisas descritivas são […] as que habitualmente realizam os pesquisadores sociais preocupados com a atuação prática. São também as mais solicitadas por organizações como instituições educacionais […]”.

Quanto às ferramentas tecnológicas de apoio, nesta etapa, além do *google forms* usado para criar o questionário on-line, para o tratamento dos dados recolhidos foi utilizado o SQLite, que é uma biblioteca em linguagem C, que implementa um banco de dados SQL (Linguagem de Consulta Estruturada) embutido. Usamos também o *Google sheet* para a geração das análises descritivas e a geração dos gráficos.

- **Resultados e análise dos dados da entrevista**

Para o tratamento e análise dos dados recolhidos através da entrevista, foi aplicada a análise de conteúdo. Defendida por Chizzotti (2000), pontua que ela tem o objetivo de compreender criticamente o sentido das comunicações, seu conteúdo manifesto ou latente, as significações explícitas ou ocultas, e é um método de tratamento e análise de informações colhidas por meio de técnicas de coleta de dados, consubstanciadas em um documento. A técnica se aplica à análise de textos escritos ou de qualquer comunicação (oral, visual, gestual) reduzida a um texto ou documento. O procedimento utilizado foi a análise categorial, ou seja, que classifica o conteúdo, segundo categorias, dessa forma, decompomos os

conteúdos da entrevista, classificando-os em categorias, melhor dizendo, foi criada uma lista em que as questões foram separadas em categorias e em cada categoria foram descritos os diálogos que a pesquisadora teve com os participantes. Apoiando esta técnica, Minayo (2001, p. 70) salienta: “as categorias são empregadas para estabelecer classificações […] trabalhar com elas significa agrupar elementos, ideias ou expressões em torno de um conceito capaz de abranger tudo isso. Esse tipo de procedimento […] pode ser utilizado em qualquer tipo de análise em pesquisa qualitativa”. Diante do exposto, a análise das entrevistas buscou extrair nos diálogos palavras, frases, temas para relacioná-las com a estrutura sistematizada. Neste sentido, “[procuramos] identificar a frequência ou constância dessas unidades para fazer inferências e extrair os significados inscritos no texto a partir dos indicadores objetivos” (CHIZZOTTI, 2013, p. 115).

2.5 LIMITAÇÕES DA DISSERTAÇÃO

Este estudo se limitou a lançar um olhar para obter um panorama geral sobre o tema em questão (identificação e mapeamento dos dados científicos da UFV) e como esse tema é percebido e praticado entre os discentes, docentes e técnicos administrativos. Mesmo quando a proposta foi identificar a qual centro de ensino o respondente está vinculado, foi com o objetivo de verificar o número de participantes de cada centro. Assim, não foi objetivo da pesquisa fazer comparações entre os programas de pós-graduação ou departamentos da UFV, por exemplo, que o programa da pós-graduação em Geografia elaborou mais Planos de Gestão de Dados (PGDs) do que a pós-graduação em Engenharia de Alimentos.

3 REVISÃO DA LITERATURA

No intuito de alcançar o objetivo 1, neste capítulo são apresentados breves relatos e conteúdos teóricos que foram necessários para o andamento da pesquisa. Assim, discute-se a relevância de dados científicos e de pesquisa que se relacionam diretamente com a comunicação científica, com a curadoria digital, com os princípios *fair*, com a *small science*, com a política institucional de dados científicos e com a biblioteca universitária.

3.1 COMUNICAÇÃO CIENTÍFICA

Na trajetória da concepção do termo comunicação científica, é de se notar que na literatura existem variados significados como também termos que se relacionam entre si, como, por exemplo, difusão científica, divulgação científica, popularização da ciência e disseminação científica. No entanto, no presente trabalho foram abordados os conceitos da comunicação científica dentro do contexto da pesquisa científica, isto é, conceituações que remetem a atividades científicas em si.

Assim, Bueno (2010, p. 2) coloca que "a comunicação científica diz respeito à transferência de informações científicas, tecnológicas ou associadas a inovações e que se destinam aos especialistas em determinadas áreas do conhecimento" e ainda acrescenta, "pessoas que, por sua formação específica, estão familiarizadas com os temas, os conceitos e o próprio processo de produção em ciência e tecnologia (C&T)". Neste contexto, Targino (2000) também pontua:

> A comunicação científica é indispensável à atividade científica, pois permite somar os esforços individuais dos membros das comunidades científicas. Eles [cientistas, acadêmicos e pesquisadores] trocam continuamente informações com seus pares, emitindo-as para seus sucessores e/ou adquirindo-as de seus predecessores. É a comunicação científica que favorece ao produto (produção científica) e aos produtores (pesquisadores) a necessária visibilidade e possível credibilidade no meio social em que produto e produtores se inserem, […] a mesma obedece a práticas estabelecidas pela comunidade científica, termo que designa tanto a totalidade dos indivíduos que se dedicam à pesquisa científica e tecnológica como a grupos específicos de cientistas, segmentados em função das especialidades, e até mesmo de línguas, nações e ideologias políticas (TARGINO, 2000, p. 10),

Bernal *apud* Caribé (2015) observa também que a comunicação científica incorpora as atividades associadas à produção, disseminação e uso da informação, desde o momento da concepção da ideia pelo cientista até a informação referente aos resultados alcançados serem aceitos como constituintes do estoque universal de conhecimentos pelos pares.

Assim, dentre as conceituações de comunicação científica apresentadas, percebe-se que seus processos e ciclos evoluíram amparados pelas tecnologias da informação e comunicação (TICs). Seus processos se adaptam aos meios de comunicação vigente desde a era de *Gutenberg* até o momento atual da *big data* e *e-science*. Ela sempre se molda aos canais e processos estabelecidos pela tecnologia utilizada no momento, porque as novas tecnologias estimulam e potencializam canais, fluxos e processos que favorecem e possibilitam a comunicação entre os atores da pesquisa científica, ou seja, a evolução das tecnologias de informação e comunicação interfere nas etapas de criação e produção do conhecimento como um todo.

Neste aspecto, Targino (2000, p. 5) pontua que “no âmbito da execução da própria pesquisa, o impacto das TICs conduz a recursos metodológicos sofisticados e abrangentes que favorecem a manipulação de dados com mais precisão, rapidez e segurança, incluindo indistintamente coleta, tratamento e análise”. Ainda que em certos momentos da história da comunicação científica, em razão do alto custo de manutenções de coleções por instituições de pesquisa, que no caso aqui se refere a revistas em papel e posteriormente a publicações científicas eletrônicas indexadas, não foi permitido o acesso por uma parcela significativa de cientistas, acadêmicos e instituições de pesquisa, melhor dizendo, os custos altos para manter coleções de periódicos afetou o acesso dos pesquisadores à estes objetos de informação. Mas ainda assim, num dado contexto, de uma maneira bem simplista, a ideia é ter um olhar sobre a comunicação científica como fenômeno que se moldou aos meios de comunicação e informação para melhorar o desenvolvimento científico e tecnológico dos países, sendo eles periféricos ou não.

Entretanto, é necessário reconhecer que nas últimas décadas a comunicação científica passou por profundas transformações no que tange à sua abrangência de aberturas e às alterações no modo de acessar o conhecimento científico processado pelas pesquisas científicas da época. Após a chamada “crise dos periódicos”, na década de 1970, que foi a dificuldade de manter as coleções de periódicos, porque cada vez mais as editoras restringiam e limitavam a participação de universidades e centros de pesquisa que tinham poder político e aquisitivo limitado, melhor dizendo, as organizações que tinham poucos recursos políticos e financeiros não conseguiam renovar suas assinaturas, devido à grande elevação dos preços.

Primeiramente, ocorreu nos países em desenvolvimento, como o Brasil. Mueller e Passos (2000) detalham a crise:

> Os orçamentos das bibliotecas não acompanharam os aumentos de preço e demanda, provocando o início do que depois seria conhecido como a crise dos periódicos. A reação das bibliotecas foi

> primeiramente a de tentar driblar a situação cortando outros gastos, mas finalmente o corte de alguns títulos foi inevitável. No Brasil, cuja situação financeira era muito delicada, as consequências foram bastante devastadoras (MUELLER; PASSOS, 2000, p. 19).

Mas a crise teve o seu ápice quando universidades norte-americanas também não conseguiram arcar com os altos preços das assinaturas dos periódicos indexados.

> O gatilho da crise foi a impossibilidade de as bibliotecas universitárias e de pesquisa americanas continuarem a manter suas coleções de periódicos e a corresponder a uma crescente demanda de seus usuários, impossibilidade decorrente da falta de financiamento para a conta apresentada pelas editoras, cada ano mais alta, mais alta mesmo que a inflação e outros índices que medem a economia. (MUELLER, 2006, p. 31)

Dessa forma, no começo da década de 1990, viabilizados pela tecnologia da informação, são criados os periódicos eletrônicos que, para a comunicação científica, foram um ganho em vários aspectos, incluindo rapidez na comunicação, maior interação entre os pesquisadores, ou seja, as vantagens oferecidas foram muitas, o que incluiu rapidez na publicação, abrangência no alcance, possibilidades de interação entre autores e leitores como nunca houve (MUELLER; PASSOS, 2000).

Porém, nas duas últimas décadas do século XX, a rigidez do regime de proteção de direitos de propriedade intelectual, regimentada pelo mercado editorial, cerceou mais uma vez a circulação da informação científica e do conhecimento científico derivados de pesquisas científicas ao redor do mundo. Neste sentido, Albagli (2015) reitera como as grandes editoras tinham influência em todo o ciclo da pesquisa nas organizações de pesquisa:

> Esse processo repercutiu diretamente nos formatos organizacionais e institucionais de produzir e circular ciência. A partir daí, aumentaram as pressões sobre os ambientes acadêmicos e universitários para o patenteamento e o retorno financeiro dos resultados das atividades de C&T, mobilizando o estabelecimento de aparatos institucionais e legislação pertinentes a esses objetivos (ALBAGLI, 2015, p. 11).

Nesta perspectiva, as editoras científicas aumentaram exponencialmente os preços das assinaturas de periódicos, incluindo também licenças restritivas ao uso de materiais digitais, gerando de alguma forma uma dependência do que era produzido nas investigações científicas ao redor do mundo. Entretanto, diante deste cenário e com o crescimento e a evolução da internet com sua arquitetura *peer to peer*, foi possível criar projetos que sustentaram e implementaram a filosofia do acesso livre, foi amplamente disseminada a cultura livre digital, influenciada pela cultura *hacker*, e novas formas de produção

colaborativa, interativa e compartilhada da informação e do conhecimento foram criadas (ALBAGLI; CLINIO; RAYCHTOCK, 2014). Neste cenário em que o movimento pelo acesso aberto tem seu surgimento, são iniciados o movimento pioneiro *Budapest Open Access Initiative*[3], em 2002, e a Declaração de Berlim, em 2003, ambos os movimentos são bases em nível global da ciência aberta. Relatando sobre o acesso aberto, Sales e Sayão (2012a) pontuam:

> [Esses movimentos] de alcance global sugerem dois caminhos para o acesso livre à informação científica: a via dourada – que incentiva a publicação de periódicos eletrônicos de acesso livre – e a via verde – que incentiva a implantação de repositórios digitais temáticos e institucionais, bem como o autoarquivamento de *e-prints* nesses repositórios (SALES; SAYÃO, 2012a, p. 121).

No contexto brasileiro, o Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica, lançado pelo IBICT, em 2005, é um documento que oficializou seu apoio em prol do acesso livre à informação científica. Desde então, o IBICT vem participando de eventos e projetos que viabilizam estratégias de tornar a informação científica livre, disponível e de forma irrestrita. Assim, nos últimos anos, tem se juntado à Universidade do Minho, redundando no Compromisso do Minho sobre o Acesso Livre à Informação em Países Lusófonos, assinado, entre outros, por representantes do IBICT. Da mesma forma tem participado ativamente das Conferências Luso-Brasileiras de Acesso Aberto (Confoa), seja na organização, na Comissão Científica ou na apresentação de trabalhos ou pôsteres, que alternadamente têm sido realizadas no Brasil ou em Portugal (PINHEIRO, 2014).

Desse modo, a filosofia do acesso livre, possível com as evoluções das TICs, modifica os processos da comunicação científica, pois melhora e diversifica seus canais e fluxos do conhecimento científico. É possível, neste novo contexto, não ter apenas o acesso às publicações do produto final de uma pesquisa, como era antigamente na pesquisa tradicional, mas acompanhar, reutilizar, compartilhar, acessar diversificados objetos de pesquisa, como códigos de *software*, metodologias, simulação de dados, ou seja, é possível ter acesso a uma gama de diferentes objetos de pesquisa. Neste sentido, Chan, Okune, Sambuli (2015, p. 94.) pontuam que “durante a última década, ferramentas e redes abertas permitiram o florescimento dos movimentos de abertura em diversos domínios, iniciativas, em todo o mundo, de acesso aberto e de recursos educacionais abertos, de dados abertos e de governos abertos, de inovações abertas e de desenvolvimento aberto”. Neste panorama, Costa e Leite (2018) também enfatizam e discorrem sobre a filosofia da ciência aberta,

[3] Disponível em: https://www.budapestopenaccessinitiative.org/. Acesso em: 24 nov. 2020.

> A ampliação da noção de abertura e de democratização do acesso trazida por esta filosofia fez emergir um novo modo de pensar, produzir e comunicar a ciência, que recebeu o nome de ciência aberta. Nela, o acesso público e irrestrito não requer apenas as publicações científicas, mas representa mudança paradigmática em todo o processo de produção da ciência, desde os sistemas utilizados para a pesquisa, em que predominam os *softwares* livres, até as metodologias de avaliação do alcance da pesquisa, nas quais discute-se o uso de métricas alternativas para a ciência (COSTA; LEITE, 2018, p. 4457).

percebendo a filosofia da ciência aberta como a responsável por incrementar maior colaboração e interatividade entre pesquisadores no contexto da comunicação científica, seja em um plano local ou global.

Nesta parte da pesquisa da literatura exploratória, buscamos introduzir de modo modesto a definição e a história da comunicação científica e os avanços das TICs incidindo sobre os seus ciclos, como também expôs os movimentos de acesso aberto e da ciência aberta, movimentos que vieram enriquecer e melhorar os processos da comunicação científica.

Com isso, a partir deste breve relato, tenta-se introduzir conceitos e definições que englobam o universo “dados científicos” e suas ramificações. Pontua-se que os termos abordados a seguir têm como pano de fundo a comunicação científica, pois, na produção de conhecimento, na ciência contemporânea, são procedimentos que possibilitam mais rapidez e transparência no modo de produzir ciência.

3.2 DADOS CIENTÍFICOS E DADOS DE PESQUISA

Considerando as transformações pelas quais a pesquisa científica passou nas últimas décadas, no que tange à sua realização e ao seu modo de ser produzida, diversos fatores incidiram nessa alteração. Assim, destaca-se que o principal deles é a evolução das tecnologias da informação. Com o desenvolvimento e a evolução dos meios tecnológicos, foi possível o surgimento da filosofia do *software* livre e do movimento da ciência aberta, assim, o fazer da pesquisa científica tem sido profundamente modificada.

Neste contexto, temos os dados de pesquisa e os dados científicos, que na atualidade são ativos primordiais nos processos de desenvolvimento de uma pesquisa científica. Eles podem auxiliar os pesquisadores a realizar seus trabalhos científicos de forma mais eficaz e com mais agilidade. Ambos os ativos informacionais são objetos essenciais para obter-se qualidade e produtividade, pois são insumos que fornecem informação e conhecimento. Entretanto, explorando a literatura sobre esses conceitos não foi possível estabelecer e apontar de forma precisa as principais diferenças entre eles. Assim, a seguir apresenta-se

primeiramente as definições acerca de dados científicos, logo após, sobre dados de pesquisa. Dessa forma, em seu livro *Gestão de dados científicos*, Silva (2019) declara:

> Os dados científicos compreendem uma área muito extensa (todas as disciplinas) e devido a isso a definição pode variar em relação às abordagens dos diferentes participantes (os cientistas, as instituições, os financiadores, etc.) […] são informações registradas ou produzidas através de qualquer forma ou meio durante o decurso de uma pesquisa. Podem ser de tipos numérico, descritivo ou visual [e podem ser reproduzidas] em formato de papel ou digital. Eles são todas as evidências que um investigador necessita para validar suas conclusões após uma pesquisa (SILVA, 2019, p. 21).

A partir da afirmativa acima, é notório que a tentativa de pontuar diferenças entre os dois termos é uma tarefa sem sucesso. Mas, ainda assim, foi possível demonstrar algumas definições sobre dados de pesquisa encontrados no decorrer da exploração da literatura

Nos últimos 20 anos, foi possível encontrar na literatura da área de Biblioteconomia e Ciência da Informação tentativas de elaborar definições e conceituações referentes ao termo “dados de pesquisa”. Primeiramente, com autores seminais como Borgman (2010), que observa que a noção de dados pode ser diferente em cada área do conhecimento, como também que eles podem ser criados por inúmeros objetivos, como declarado a seguir:

> [...] a noção de dados pode variar consideravelmente entre pesquisadores e ainda mais entre áreas do conhecimento. A constatação de que os dados são gerados para diferentes propósitos, por diferentes comunidades acadêmicas e científicas e por meio de diferentes processos, intensifica ainda mais essa percepção de diversidade. Tipos de dados podem incluir, por exemplo, números, imagens, textos, vídeos, ou 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 fáceis de se recriar (BORGMAN, 2010 *apud* SAYÃO; SALES, 2013).

Nesse viés, percebe-se que a noção de diferenciação de processos e tipos de dados pode aparecer em uma pesquisa científica, assim Simberloff *et al.* (2005) argumentam:

> Que o dado, quanto à sua natureza, em uma coleção, pode ser diverso. Entre os exemplos citam: números, imagens, vídeo, arquivos de áudio, software, informações sobre a versão de um software, equações, animações, algoritmos ou mesmo modelos/simulações. Os autores também argumentam que os dados podem ser diferenciados em função das suas origens. Nesse aspecto, eles podem ser “observacionais, computacionais ou experimentais” (SIMBERLOFF *et al*., 2005 *apud* COSTA, 2017, p. 58).

Esclarecendo a definição do conceito, Sales (2018) também pontua que “dados de pesquisa são objetos complexos, diversificados e heterogêneos, e os objetivos e os métodos usados para produzi-los variam enormemente de acordo com os campos científicos, assim como os critérios para compartilhá-los”. Auxiliando na explanação do termo, Marques (2014, p. 54) também afirma que “[os dados de pesquisa] são a massa de informações primárias que uma vez analisada e interpretada, serve como base para a conclusão do [artigo]”.

Neste sentido, Sayão e Sales (2015, p. 7) conceituam que “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.” Obtendo uma visão dos dados que são gerados no decorrer de um trabalho científico Rice e Southall (2016) *apud* Semeler e Pinto (2019) esclarecem que esses dados podem ser compreendidos como:

> [informações registradas] necessárias para apoiar ou validar as observações, descobertas ou resultados de um projeto de pesquisa. […] o que é coletado, observado ou criado de forma digital, para fins de análise e para produzir resultados de pesquisas originais […] dados de pesquisa são: dados, registros, arquivos ou outras evidências, independentemente do seu conteúdo ou forma (por exemplo, material impresso ou digital), que compreendem observações de pesquisa, achados ou resultados, incluindo materiais primários e dados já analisados (RICE; SOUTHALL, 2016 *apud* SEMELER; PINTO, 2019, p.118).

No contexto da ciência moderna, dados de pesquisa são um termo que não tem somente um significado. O termo carrega em si variadas características como também peculiaridades de cada área do conhecimento, pois cada campo tem seus graus de relevância, neste sentido, Sales e Sayão (2019b) declaram:

> […] é imprescindível considerar a relevância da temática “dados de pesquisa” para a ciência contemporânea, bem como todos os seus desdobramentos, como Gestão de dados, Curadoria Digital, Ciência Aberta, *e-Science* e *Big Data* Científico entre tantos outros; é preciso considerar ainda a interdisciplinaridade e a natureza díspar e heterogênea intrínseca do objeto “dado de pesquisa”, que pode se originar de diferentes áreas de conhecimento, com metodologias, jargões, práticas e fluxos específicos de cada domínio (SALES; SAYÃO, 2019b, p. 33).

No livro *Guia de Gestão de Dados de Pesquisa para Bibliotecários e Pesquisadores*, de Sayão e Sales (2015), que é o primeiro documento das áreas de Biblioteconomia e Ciência da Informação no Brasil, que apresenta elementos estratégicos para uma eficaz gestão de dados de pesquisa, são expostos conceitos, ferramentas e referências a respeito do tema. Assim, no documento citado, são elencadas as classificações dos tipos de dados de pesquisa, que podem ser:

- **Segundo sua origem,** os dados de pesquisa estão divididos em DADOS OBSERVACIONAIS, dados obtidos por meio de observações diretas; DADOS COMPUTACIONAIS, que resultam da execução de modelos computacionais ou de simulações; e DADOS EXPERIMENTAIS, provenientes de situações controladas em bancadas de laboratórios.
- **Segundo sua natureza**, eles podem ser expressos por números, imagens, vídeos, áudios, *software*, algoritmos, equações, animações e simulações.
- **Segundo sua fase da pesquisa**, eles podem ser DADOS BRUTOS, CRUS E PRELIMINARES, dados que vêm diretamente dos instrumentos científicos; DADOS DERIVADOS, que resultam do processamento ou combinação de dados brutos ou de outros dados; e DADOS CANÔNICOS ou DADOS REFERENCIAIS, são coleções de dados consolidados e arquivados geralmente em grandes centros de dados, por exemplo, sequência genética, estrutura química, etc.

Em um estudo mais recente, Sayão e Sales (2020) discutiram um estudo conceitual das definições dos tipos de dados de pesquisa, e no trabalho propuseram uma taxonomia que categorizou os tipos de dados de pesquisa em oito divisões, de acordo com suas características, que são: quanto à origem; quanto ao grau de processamento; quanto à abordagem da pesquisa; quanto à natureza dos dados; quanto ao nível de sensibilidade; quanto à materialidade; quanto à perenidade; e quanto à abertura. Neste estudo, os autores aconselham que a taxonomia desenvolvida pode ser utilizada em modelagens de ferramentas que podem ser criadas para realizar a curadoria dos dados de pesquisa e dados científicos. Assim, a realização das atividades de curadoria, como veremos a seguir, envolve a gestão de todos os ciclos da pesquisa científica, como também a reunião de diversos perfis e expertises para que tal tarefa saia a contento.

3.3 CURADORIA DIGITAL

O compartilhamento e o reuso dos dados científicos vivenciados na contemporaneidade, foram possíveis após o movimento do acesso aberto. Atualmente pode-se ter um amplo acesso a inúmeros objetos digitais resultantes de pesquisas científicas de vários lugares do mundo. Vale lembrar que o panorama exposto já é uma realidade em várias instituições de pesquisa no exterior. Assim, essa dinâmica de acesso, compartilhamento, armazenamento e de reuso dos dados de pesquisa só se torna uma realidade se estes ativos informacionais estiverem devidamente tratados com os princípios e as técnicas de gestão e curadoria digital.

Dessa maneira, para uma melhor compreensão do termo “curadoria digital”, os autores Siebra, Borba e Miranda (2016), que em seu estudo trabalham com as mudanças e contextos do termo “curadoria” ao longo do século XVIII até o século XX. Neste artigo, os autores expõem que o termo “curadoria” tem variados sentidos, porque depende do autor e do contexto em que a expressão é usada. Assim, nos anos 1980 e 1990, surgiu o termo “curadoria dos dados”, que remete à gestão de dados científicos e somente em 2003 a expressão foi incluída no contexto das ciências dos dados e nas tecnologias de rede. No contexto da Ciência da Informação e da Biblioteconomia, autores como Sales e Sayão (2012a) pontuam que:

> A curadoria digital envolve a gestão atuante e a preservação de recursos digitais durante todo o ciclo de vida de interesse do mundo acadêmico e científico […] como todas as atividades envolvidas na gestão de dados, desde o planejamento da sua criação, quando os sistemas são projetados, passando pelas boas práticas na digitação, na seleção dos formatos e na documentação (SALES; SAYÃO, 2012a, p. 184).

No mesmo contexto, Santos (2016), na tentativa de elaborar um panorama de definições entre preservação e curadoria digital, esclarece:

> A curadoria digital é um termo que abarca definições correlatas voltadas à seleção, enriquecimento, tratamento e preservação da informação para o acesso e uso futuro, seja ela de natureza científica, administrativa ou pessoal. Tanto a curadoria quanto a preservação consideram a natureza dos objetos digitais, sua representação e a preocupação com acesso futuro desses ativos digitais. [...] é um termo amplo utilizado para designar as atividades necessárias à gestão da informação em meio digital maiores que a preservação, ela trata da manutenção consciente e orientada à preservação e ao acesso dos ativos digitais resultantes da pesquisa científica (SANTOS, 2016, p. 460).

Siebra (2019) também declara que o papel da curadoria digital é estabelecer uma gestão ativa, intencional e planejada em meio digital da memória individual e coletiva, dos dados científicos, da herança cultural, da produção acadêmica e bibliográfica, o que resulta na sua preservação digital com o intuito de possibilitar o acesso, o uso e o reuso a longo prazo. Sales e Sayão (2012b) resumem que esta atividade assegura a sustentabilidade dos dados a longo prazo e que os recursos estratégicos e metodológicos, que incluem também as tecnologias utilizadas para essa prática, tornam mais fáceis o acesso persistente a objetos digitais confiáveis, tornando possível a checagem da autenticidade desses ativos informacionais.

Em suas buscas para explorar a gestão de dados científicos, Silva (2019) publicou o livro *Gestão de dados científicos*, que trata da gestão, recuperação e preservação dos dados primários da investigação. O material contém informações e recursos para treinamento no campo da gestão de dados científicos, sendo o documento direcionado a bibliotecários, pesquisadores, gestores da informação e demais profissionais interessados na implementação de boas práticas no gerenciamento dos dados resultantes de suas pesquisas. No mesmo documento, Silva (2019) salienta:

> A curadoria digital de dados implica gestão, uma vez que se tem selecionado para a preservação e o armazenamento em longo prazo. Os passos sequenciais do ciclo de vida dos dados são, basicamente, a coleta, avaliação e seleção, o uso direto em pesquisas, a realização de ações de preservação, o armazenamento, o acesso aos dados de uso, sua reutilização e sua transformação (SILVA, 2019, p. 49).

Neste sentido, nota-se que só uma gestão com planos e estratégias bem definidas poderá possibilitar condições necessárias para diminuir os riscos de integridade, perda ou inacessibilidade dos dados. Pois, na pesquisa científica, são produzidos diversos formatos de dados, com centenas de extensões, diversas intencionalidades, várias formas de reprodução e de adaptabilidade e divergentes níveis de processamento (SILVA SEGUNDO; ARAÚJO, 2019).

Dessa maneira, por ser uma atividade naturalmente interdisciplinar, que combina questões tecnológicas, comunicacionais, gerenciais, cognitivas, de geração de conhecimento e informacionais, é compreensível esclarecer que a curadoria digital é uma tarefa que engloba diversas expertises, sendo necessário reunir profissionais de campos de conhecimentos distintos, mas que, ao mesmo tempo, se complementam, como bibliotecários, cientistas de dados, profissionais de tecnologia e de computação (SIEBRA; BORBA; MIRANDA, 2016). Nesta direção, Choudhury *et al.* (2018), em um estudo que discorre sobre um planejamento de

administração de recursos humanos em um projeto de gestão e curadoria de dados de pesquisa, em uma instituição de ensino, pontuam:

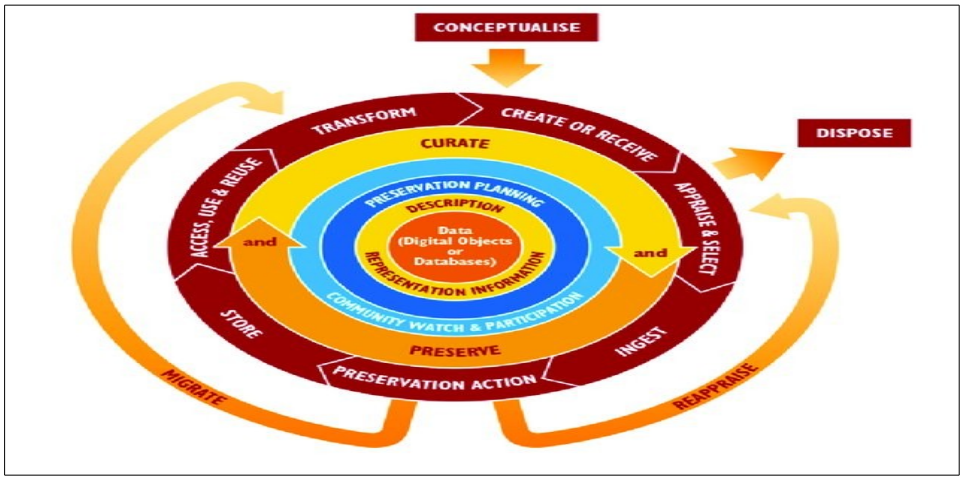
> Por sua própria natureza, a curadoria de dados é uma atividade colaborativa. As parcerias geralmente incluem escritórios de pesquisa, serviços de tecnologia de informação/computação e bibliotecas. Por exemplo, pode haver um comitê gestor de toda a instituição para dados de pesquisa com membros da biblioteca, parcerias com professores para apoiar atividades de pesquisa ou uma força-tarefa de tecnologia de pesquisa. Um modelo de parceria forte apresenta a oportunidade de alavancar as habilidades e talentos existentes em toda a instituição para apoiar a curadoria de dados e permitir que cada parceiro se concentre nos serviços e atividades que a equipe existente e a nova podem fornecer em apoio à curadoria de dados (CHOUDHURY *et al*., 2018, p. 14, tradução nossa).

Assim, em alguns países, existem organizações em que a atividade de curadoria digital de dados científicos já tem uma longa trajetória de atuação. Os projetos de curadoria digital que serão apresentados a seguir têm como elemento principal os modelos de ciclos de vida dos dados, e estes ciclos padronizam todo o processo da atividade de curadoria.

Confirmando esta visão, Silva, (2019, p. 50) endossa: “a importância dos modelos de ciclo de vida é que proporcionam uma estrutura para a consideração das muitas operações que devem ser realizadas em um registro de dados durante seu uso. Muitas ações de curadoria digital se tornam mais fáceis quando preparadas antecipadamente, mesmo antes da criação do registro ou no mesmo momento”. Para uma melhor compreensão, a seguir destaca-se três modelos de ciclos de vida encontrados na literatura:

- ***Digital Curation Centre*** **(DCC)**: projetado em 2004 no Reino Unido pela *University of Edinburgh*, oferece um modelo para o ciclo de vida da curadoria que reflete uma visão de alto nível dos estágios necessários para o sucesso do processo de curadoria e de preservação de dados de pesquisa. No projeto, foi desenvolvido um modelo de ciclo de vida que propicia uma visão gráfica dos estágios exigidos para que os dados sejam preservados e cuidados adequadamente. A Figura 1 mostra o ciclo de vida de dados digitais do **DCC**.

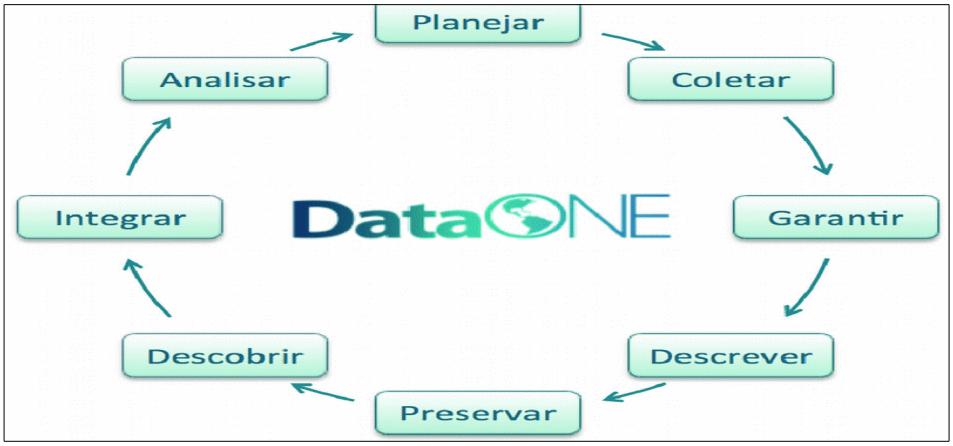
Figura 1 - Ciclo de vida de dados digitais do DCC



Fonte: Digital Curation Centre[4]

- ***Data Observation Network for Earth (DataOne)*:** O projeto é sustentado pela *National Science Foundation (NSF)*. A iniciativa mantém um forte braço educacional, que empenha esforços para fornecer materiais que instruem os pesquisadores a gerenciar seus dados de forma adequada, com isso, para guiar sua posição pedagógica, o *DataOne* elaborou o próprio ciclo de vida dos dados (SILVA SEGUNDO, ARAÚJO, 2019). O *DataOne* é um projeto comunitário que disponibiliza o acesso a dados de variados repositórios que fazem parceria com o projeto e apoia e dá suporte a pesquisas avançadas e à descoberta de dados ambientais e da terra (CÓRDULA; ARAÚJO, 2019). A Figura 2 mostra o ciclo de vida de dados digitais do DataOne.

Figura 2 - Ciclo de vida de dados digitais do *DataOne*



Fonte: DataOne

[4] Disponível em: https://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf. Acesso em:16 mar. 2020.

- ***Research360****:* este projeto é a junção da *United Kingdom Office for Library and Information Networking (UKOLN)*, da *University of Bath* com o *DCC* da *University of Edinburg*. O objetivo da união dessas organizações foi elaborar um mecanismo que possibilitasse que a gestão de dados científicos fosse integrada a todo o ciclo de vida dos dados de uma pesquisa científica. Neste modelo, a instituição tem o papel principal durante todas as etapas da gestão dos dados, por isso o elemento *Institution* aparece no núcleo do modelo (SILVA SEGUNDO, ARAÚJO, 2019). A Figura 3 mostra o ciclo de vida de dados digitais do *Research360*.

Figura 3 - Ciclo de vida de dados digitais do *Research360*

Fonte: ResearchGate[5]

Na tentativa de expor o contexto do tema no território brasileiro, na próxima subseção serão incluídas algumas organizações que desenvolvem tais atividades.

3.3.1 Curadoria digital no Brasil

O panorama no Brasil em relação à curadoria digital ainda é bem incipiente no que tange a planejamentos e execução de projetos desta natureza. Neste aspecto, Resende (2019) afirma:

> São poucas as iniciativas das atividades de curadoria digital, sendo, efetivamente, implementadas nas comunidades científicas brasileiras. Percebe-se que há grande interesse das instituições em desenvolver a

[5] Disponível em: https://www.researchgate.net/figure/The-Research360-Institutional-Lifecycle-Research-Concept_fig1_273052937. Acesso em:16 mar. 2020.

> curadoria digital de dados científicos, mas ainda não se tem conhecimento de uma política sólida, com definições claras sobre como isso poderá se tornar atividade obrigatória nas instituições (RESENDE, 2019, p. 54).

Assim é que, no Brasil, algumas instituições têm colocado as bibliotecas de pesquisas acadêmicas para realizar atividades de curadoria digital dos dados resultantes de trabalhos científicos. Nestes locais, exercem gerência e disponibilização de serviços, que englobam dados científicos. No momento, são estas as organizações que fazem este trabalho pioneiro no país:

- A plataforma CarpeDIEN de Informações em Energia Nuclear, do Instituto de Engenharia Nuclear (IEN/CNEN), vinculada ao Conselho Superior de Pesquisas Científicas da Espanha (CSIC). Desde 2014, faz a integração de instituição de pesquisas na área nuclear e tem mais de 45 instituições brasileiras participantes, entre elas, destacam-se a USP, a Fiocruz e a EMBRAPA. A finalidade desta plataforma é promover a preservação, curadoria e disseminação da memória digital técnico-científica produzida pelas instituições parceiras.

- O Centro de Documentação e Acervo Digital da Pesquisa (CEDAP), órgão auxiliar da Faculdade de Biblioteconomia e Comunicação (FABICO), da Universidade Federal do Rio Grande do Sul (UFRGS), que tem como finalidade fornecer suporte às pesquisas científicas institucionais. Oferece vários serviços para armazenamento, acesso e compartilhamento de dados científicos.

- O Repositório da Produção Científica CRUESP, do Conselho de Reitores das Universidades Estaduais de São Paulo – USP, UNICAMP e UNESP, que tem como objetivo reunir, preservar e proporcionar acesso aberto, público e integrado à produção científica de docentes, pesquisadores, alunos e servidores das universidades participantes. Integrados ao CRUESP, estão os repositórios institucionais de produção intelectual (científica, artística, acadêmica e técnica) da USP, a Biblioteca Digital da Produção Intelectual da Universidade de São Paulo (BDPI), a Biblioteca Virtual da FAPESP, a Biblioteca Digital da Produção Intelectual e Científica da UNICAMP e o Repositório Institucional da UNESP (RESENDE, 2019).

Destaca-se no trecho acima algumas organizações brasileiras que iniciaram serviços de curadoria de seus dados científicos, em contrapartida, existem universidades brasileiras ou instituições de pesquisa que ainda não conseguiram perceber o grande potencial que os dados científicos processados em seus laboratórios podem propiciar. Assim, a seguir, discute-se o

conceito “cauda longa da ciência” ou “*long tail science”*, que são os dados científicos gerados em pequenos laboratórios e que, por incontáveis razões, não conseguem visibilidade.

3.4 CAUDA LONGA DA CIÊNCIA OU “*LONG TAIL SCIENCE*”

Os dados científicos no contexto da comunicação científica contemporânea se tornaram elementos primordiais nos processos da produção científica. Inúmeros estudos demonstram como a *big science* tem contribuído para o crescimento e o avanço da ciência, principalmente em campos de conhecimento como a física de partículas, genoma, sensoriamento remoto, entre outros. Para a estrutura da *big science*, são alocados investimentos estratosféricos, que são direcionados a infraestruturas que englobam toda a cadeia de gestão e curadoria dos dados científicos processados por grandes laboratórios (COSTA; CUNHA, 2014).

Dessa forma, nesses laboratórios de ponta, os dados científicos têm uma estrutura uniforme, são homogêneos, a geração e a coleta são feitas através de instrumentos automatizados, sua gestão é centralizada e institucionalizada, existe uma infraestrutura de compartilhamento e o reuso é imediato e globalizado (SALES; SAYÃO, 2019a). Inversamente a esta realidade, existem os pequenos e modestos laboratórios situados em institutos de pesquisa e universidades, mas que têm incontáveis projetos de pesquisa e que também podem gerar uma gama de dados científicos, só que não conseguem visibilidade em razão de seus poucos recursos financeiros e políticos.

Os estudos desses espaços de pesquisa foram iniciados por Heidorn (2008), quando o autor chamou a atenção para a problemática dos inúmeros dados científicos que são processados em instituições acadêmicas e/ou de pesquisa e que não recebem tratamento adequado, tornando-se invisíveis, pois são dados que nunca foram publicados ou disponibilizados para o resto da comunidade científica. Mas esses ativos informacionais em algum momento nas etapas da investigação científica foram cruciais para obter-se um resultado final, seja na criação de um produto, na criação de uma patente, na programação de um *software* ou até mesmo na elaboração de uma metodologia.

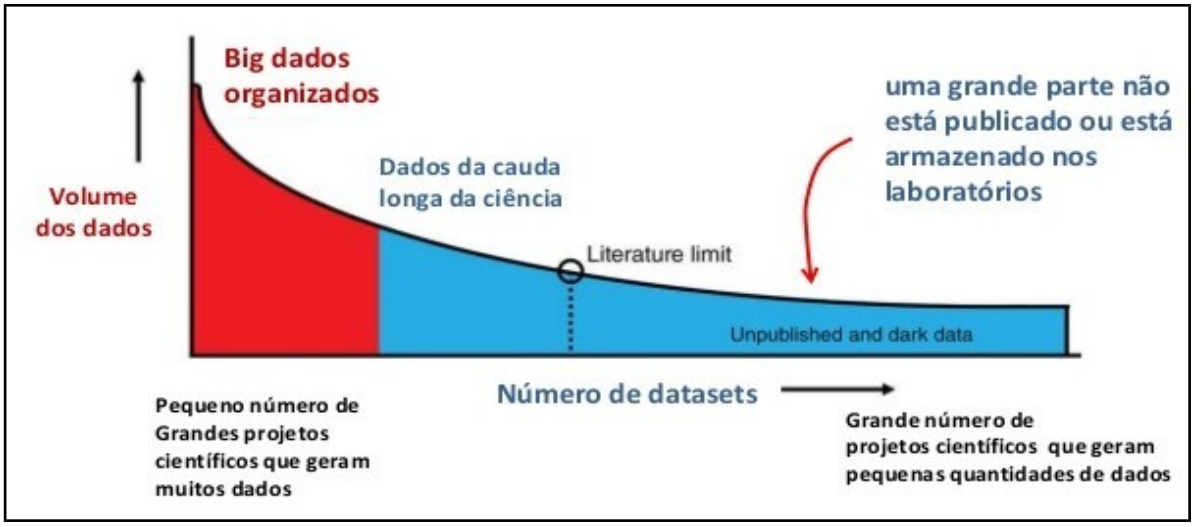
Com isso, devido a algumas barreiras, seja de cunho político, orçamentário ou comportamental, esses objetos informacionais se mantêm invisíveis. Neste escopo, Heidorn (2008), para descrever as atividades de pesquisa que se referem ao tamanho, escala, prazos, volume, financiamento econômico e variedade de tipos de dados, utilizou a metáfora da

“cauda longa da ciência” ou “*long tail science*”, que foi originada do *e-commerce*. Neste contexto, Sales e Sayão (2019b) complementam que:

> [o] conceito estatístico da “cauda longa” (*long tail*), muito utilizado nas áreas de marketing, especialmente, de produtos digitais [e que] essa forma específica de distribuição estatística é que uma pequena parte da população tem muitas ocorrências – a chamada “cabeça”; enquanto uma grande parte tem poucas ocorrências – conhecida como a `cauda longa´ da distribuição (SALES; SAYÂO, 2019a, p. 155).

A Figura 4 mostra como são distribuídas as diferenças entre a *“big science”* e a *“small science”*.

Figura 4 - Dados da cauda longa da ciência



Fonte: Adaptada por Sales e Sayão (2019a)

Assim, na cabeça há um extenso volume de dados, derivados da *big science,* e na cauda, estão as pequenas ciências, que são formadas por projetos menores. Nos pequenos laboratórios é que é processada a *small science*, melhor dizendo, é a característica do perfil das pesquisas nas universidades, mas os dados processados nestes ambientes são heterogêneos e de ampla complexidade e de diferentes domínios, o que se torna uma grande massa de dados, dispersos, sem critérios de procedimentos de coleta, de gestão e curadoria. Sales e Sayão (2019a) apontam que a característica das pequenas coleções de dados é a diversidade, onde reside sua riqueza informacional, pois são dados heterogêneos em formatos, tecnologias, tipos e que foram gerados segundo padrões e fluxos específicos, o que torna sua gestão bem complexa e desafiadora.

Dessa forma, esse tema tem se tornado muito importante para algumas instituições de pesquisa, é válido supor que, aparentemente, existe um certo consenso de que é necessário elaborar estudos e projetos que fomentem a visibilidade desses ativos informacionais. Assim,

a identificação e o mapeamento de como o fenômeno *small science* circula e como é processado em uma organização de pesquisa, melhor dizendo, a compreensão e a investigação deste fenômeno possibilitam elaborar planos e estratégias para planejar ações que integrem atividades e projetos que vislumbrem a gestão e a curadoria desses dados científicos (ALMEIDA, 2019).

Com isso, para uma melhor compreensão deste fenômeno, são destacados dois projetos que foram encontrados na literatura. O primeiro relata um estudo de dez anos do *Center for Embedded Network Sensing (CENS),* da *National Science Foundation*, onde foi feita uma pesquisa sobre práticas de compartilhamento de dados dos pesquisadores do local. O intuito foi levantar informações e dados que encontrassem sustentabilidade de um projeto de infraestrutura de coleta, gerenciamento e descoberta de dados nos laboratórios do CENS (WALLIS; ROLANDO; BORGMAN, 2013).

No Brasil, também foi identificado um projeto que tem como objetivo compreender os fenômenos da pesquisa na cauda longa. Este projeto é conduzido no âmbito da Comissão Nacional de Energia Nuclear - Centro de Informação Nuclear (CNEN/CIN) e do Instituto Brasileiro de Informação em Ciência e Tecnologia – Coordenação de Ensino e Pesquisa (IBICT/COEP) (SALES; SAYÃO, 2018).

Diante do exposto, nos processos para criar infraestrutura que possibilite a gestão e a curadoria da "*small science*", é necessário que se respeitem a multiplicidade dos dados e seus domínios, como também que se tenha uma visão holística de todos os elementos onde a pesquisa científica é desenvolvida. Para isso, a preocupação é desenvolver infraestrutura que possa estabelecer uma política institucional dos dados científicos, ou seja, a política tem que se adequar aos ambientes de pesquisas existentes. Assim, serão abordados no próximo bloco alguns elementos para a implementação de políticas institucionais de dados científicos no contexto mundial e brasileiro.

3.5 POLÍTICA INSTITUCIONAL DE DADOS CIENTÍFICOS

Nesta Seção, serão abordados o surgimento e a implantação de políticas institucionais de dados científicos, tanto do contexto mundial como brasileiro nos últimos 25 anos.

3.5.1 Contexto mundial

A partir dos anos 2000 tivemos diversas iniciativas que possibilitaram o acesso e o reuso de dados científicos que temos na atualidade, foram contribuições significativas no que tange a políticas de dados dentro do contexto da gestão e curadoria de dados científicos, como também da informação científica. São estudos e relatórios que discutem diretrizes e planejamentos de acesso aos ativos informacionais resultantes de pesquisas científicas. Elenca-se cronologicamente alguns deles.

- Entre 2002 e 2003, foi criado pela *National Science Foundation* o relatório *Revolutionizing Science and Engineering Through Cyberinfrastructure,* que marca um movimento de pesquisadores que se reuniram para obter recursos financeiros no intuito de criar uma nova infraestrutura para novos modelos de práticas de pesquisa (ATKINS *et al.*, 2003).
- Em 2004, em Paris, membros da Ciência e Tecnologia de países pertencentes à OCDE se reuniram com a Rússia, Israel, África do Sul e China para a criação da *Declaration on Access to Research Data from Public Funding* (OECD, 2004).
- No mesmo ano, foi implantado o *Open Knowledge Foundation* (OKF), que tinha o objetivo de fomentar e promover o acesso aos conteúdos de dados abertos. Para alavancar e executar os planos de atuação da fundação, foram colocados em prática os projetos *Compreensive Knowledge Archive Network* e o *Open Data Common* (OPEN KNOWLEDGE FOUNDATION, 2020).
- Em 2007, a OCDE, no intuito de dar prosseguimento à declaração de 2004, elabora o documento *Principles and guidelines for access to research data from public funding* (OECD, 2007), que estabelece diretrizes que auxiliam no acesso a dados científicos resultantes de financiamento público e “os princípios e as diretrizes contidos nesse documento pretendem servir de orientação aos governos, às organizações de financiamento, às instituições de pesquisa e aos próprios pesquisadores” (SILVA, 2019, p. 12).
- Também no ano de 2007, a *Research Information Network* fez um levantamento das políticas e das práticas vigentes dos principais projetos de pesquisa do Reino Unido. Neste trabalho, foram selecionados organizações de pesquisa, conselhos de pesquisa, universidades, departamentos do governo. Na época, o levantamento buscava obter uma visão dos objetos mais tradicionais, como artigos em periódicos, atas de congressos, entre outros. Porém o mapeamento identificou projetos de pesquisa que já tinham infraestrutura

de gestão e curadoria dos dados científicos (RESEARCH INFORMATION NETWORK, 2007).

- No mesmo ano, a Comissão Europeia publica e divulga o relatório *Scientific Information in the Digital Age,* em que são expostas tecnologias digitais que podem melhorar o acesso a publicações de pesquisa, como também aos dados científicos que agregam inovação aos projetos de pesquisas. No mesmo documento, a Comissão enfatiza que dará subsídios à União Europeia para impulsonar o acesso à informação científica e à preservação dos dados científicos (EUROPEAN COMMISSION, 2007).
- Em 2010, a mesma Comissão organizou o relatório *A Surfboard for Riding the Wave*, que foi resultado de um trabalho de um grupo de especialistas que identificou acesso, uso, reuso, integridade e qualidade dos dados para o ano de 2030, sendo que "o objetivo principal [do relatório] era conhecer os benefícios e os custos do andamento de uma infraestrutura global de dados confiável e estável que permitisse aos pesquisadores a utilização, reutilização e exploração dos dados científicos de imediato, ao máximo benefício da ciência e da sociedade" (SILVA, 2019, p. 13).
- Em 2011, o projeto *Opportunities Data Exchange* (ODE) lançou o *Report on integration of data and publications,* documento que reuniu o pensamento atual e as opiniões de várias e diversas fontes para revelar oportunidades de apoiar um registro acadêmico mais conectado e integrado. No relatório, quatro cenários foram considerados: os do pesquisador, que gera ou reutiliza os dados primários; os do editores, que fornecem os mecanismos para comunicar as atividades de pesquisa; e os cenários das bibliotecas e entradas de dados, que mantêm e preservam as evidências que sustentam a comunicação científica e o registro publicado (REILLE *et al*, 2011).
- Em 2012, a *Royal Society* realizou um estudo sobre o uso da informação científica, o relatório intitulado *Science as an open enterprise*, que aponta as vantagens do acesso aberto aos dados científicos tanto para a utilização das comunidades científicas como para a sociedade em geral. (THE ROYAL SOCIETY, 2012).
- Em 2013, o Projeto SIM4RDM de 6 sócios europeus - JISC (Reino Unido), HEA (Irlanda), ThIIF (Hungria), ThordForsk (Noruega), CSC (Finlândia) e SURF (Holanda) - publicou o relatório *European Landscape Stude of Research Data Management,* em que são apresentadas as atividades de organizações de financiamento, institutos de pesquisa, editores dos países da União Europeia para a capacitação das comunidades científicas na utilização de serviços e sistemas de dados científicos. "O relatório examina se tais programas ou políticas [das organizações de pesquisa analisadas] incluem as intervenções

de apoio aos pesquisadores, que incentivam a obtenção dos conhecimentos, as habilidades e o apoio necessário para a gestão de dados" (SILVA, 2019, p. 16).

3.5.2 Contexto brasileiro

O contexto brasileiro em relação à política institucional de dados ainda está em fase inicial, porém temos algumas instituições que têm trabalhado no intuito de desenvolver plataformas que oferecem acesso, compartilhamento e reuso de dados científicos. Neste cenário, apresenta-se estudos e trabalhos científicos que contribuem com a problemática e que são fontes relevantes para o tema em questão. Lista-se, a seguir, as principais plataformas e alguns trabalhos científicos:

- Repositório de Dados do Programa de Pesquisas Ecológicas de Longa Duração (PELD), criado em 1999, é uma iniciativa pioneira no sentido de obter informações relevantes para a conservação da biodiversidade e uso sustentável dos recursos naturais dos ecossistemas brasileiros. As informações coletadas no PELD incluem longas séries temporais de dados sobre os ecossistemas e sua biota associada (COSTA; CUNHA, 2019).
- Portal da Biodiversidade, o Repositório de Dados do Programa de Pesquisa de Biodiversidade da Amazônia Ocidental (PPBIO13)[6]. Foi criado em 2004 com o objetivo de desenhar uma estratégia de investimento em CT&I que aponte prioridades, integre competências em diversos campos do conhecimento e dissemine informações sobre biodiversidade, que possam ser utilizadas para diferentes finalidades (COSTA; CUNHA, 2019). Esta plataforma integra atividades de pesquisa e dissemina resultados em diversas finalidades, incluindo gestão ambiental e educação, trabalhando com dados ecológicos complexos (PAGANINE; AMARO, 2020).
- Portal GEOINFO[7] de infraestrutura de dados espaciais da EMBRAPA, lançado em 2018 pela Embrapa Territorial e Unidades parceiras, é uma plataforma desenvolvida para o compartilhamento de dados espaciais com a sociedade de forma gratuita e aberta. A plataforma foi criada a partir da adesão da Embrapa à Infraestrutura Nacional de Dados Espaciais (INDE[8])
- Base de Dados Científicos da Universidade Federal do Paraná[9]. Em dezembro de 2017 foi feito o lançamento da BDC/UFPR que reúne os dados científicos utilizados nas pesquisas

[6] Portaria MCT nº 693, de 20 de agosto de 2009. Institui, no âmbito do Programa de Pesquisa em Biodiversidade - PPBio, a Política de Dados.

[7] Informação disonível em: https://www.embrapa.br/busca-de-solucoes-tecnologicas/-/produto-servico/4390/geoinfo--infraestrutura-de-dados-espaciais-da-embrapa

[8] Em nível federal, há apenas o Decreto nº 6.666 de 27 de novembro de 2008, referente à informação geoespacial.

[9] As diretrizes da plataforma estão disponíveis em ttps://portal.ufpr.br/documentos/BDC/diretrizes_BDC.pdf.

publicadas pela comunidade da UFPR em teses, dissertações, artigos de periódicos e outros materiais bibliográficos.[10]

- Portal da Biodiversidade[11], criado em meados de 2011 e 2012, o portal tem como objetivo disponibilizar à sociedade brasileira dados e informações sobre a biodiversidade brasileira, gerados ou recebidos pelo Ministério do Meio Ambiente e pelas instituições a ele vinculadas. Atualmente, estão disponíveis as bases de dados de alguns dos sistemas mantidos pelo Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio), principalmente nos seus Centros de Pesquisa e Conservação, pelo Jardim Botânico do Rio de Janeiro (JBRJ) e outros parceiros. A plataforma de dados e informação é capaz de agregar consumidores e fornecedores de informação biológica vinculando com outros sistemas de informação e iniciativas de sistemas externos, como o *Global Biodiversity Information Facility* (GBIF), Sistemas de Informação sobre a Biodiversidade Brasileira (SIBBr), INDE, DataOne e Scielo (CORRÊA, 2019).
- Decreto nº 6.666, de 27 de novembro de 2008, que "institui, no âmbito do Poder Executivo federal, a Infraestrutura Nacional de Dados Espaciais - INDE, e dá outras providências".

> Percebe-se a criação de uma legislação específica para a área, que culminou com a padronização de marcos geodésicos, entre outras atividades, que, enfim, culminaram com a criação de padrões de interoperabilidade para esse tipo de informação. As instituições brasileiras envolvidas na INDE são o Instituto Brasileiro de Geografia e Estatística (IBGE) e a Secretaria de Planejamento e Investimentos Estratégicos do Ministério do Planejamento, Orçamento e Gestão (COSTA; CUNHA, 2019, 30 p.).

- Em 2013, Sayão e Sales, contribuindo com o tema, lançam o artigo *Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país.* Este documento descreve os elementos necessários para a elaboração de um modelo de curadoria e gestão de dados, expõe pontos fundamentais para uma política de dados que inclui aspectos políticos, infraestrutura organizacional e tecnológica, pesquisa em curadoria digital, desenvolvimento de coleções, formação de especialistas, sustentabilidade econômica, implicações sociais, éticas e legais e oferecimento de serviços (SAYÃO; SALES, 2013).
- Em 2016, o IBICT lançou o *Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã,* que dá continuidade e amplia sua política de apoio ao acesso aberto/livre à informação científica no Brasil. Nessa direção, o IBICT estende sua visão

[10] Informação disponível em: https://bdc.c3sl.ufpr.br/

[11] A política de dados e informações está disponível em: portaldabiodiversidade.icmbio.gov.br/portal/portal/politicaDadosPd

sobre o acesso aberto e reconhece os dados de pesquisa como um recurso imprescindível para as ações de Ciência Aberta, Ciência para todos, Ciência Cidadã (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2016).

- Em 2017, o trabalho de doutoramento de Costa, intitulado *Diretrizes para uma política de gestão de dados científicos no Brasil,* argumenta que o Brasil carece de uma política explícita que norteie as ações do Estado em termos de gestão e preservação de dados científicos, bem como diretrizes para sua reutilização. Neste estudo é apresentado um *framework c*om itens considerados de extrema relevância para a elaboração de um conjunto de diretrizes que venham a servir de elementos norteadores para a elaboração de uma política para a gestão de dados científicos no Brasil (COSTA, 2017).
- Em 2018, a Fiocruz promoveu um amplo debate com ênfase na abertura de dados de pesquisa, ancorando a discussão na filosofia da Ciência Aberta. Para dar início ao diálogo, a instituição elaborou o termo de referência *Gestão e abertura de dados para pesquisa na Fiocruz.*

> [o documento] sistematiza um conjunto de princípios e diretrizes que representam uma proposta inicial para a abertura de dados para pesquisa na Fiocruz. Com base na escuta e nos debates a serem promovidos em Câmaras Técnicas, Fórum das Unidades Regionais, seminários, grupos focais, pretende-se construir uma visão coletiva consensual a ser submetida a consulta pública à sociedade brasileira (FIOCRUZ, 2018, p. 2).

- Em 2018, no intuito de desenvolver diretrizes e princípios para uma política nacional para a gestão de dados de pesquisa, o Ministério da Ciência, Tecnologia, Inovações e Comunicações instituiu a Portaria n° 6.072, de 23 de novembro de 2018, que possibilitou a criação do Grupo de Trabalho para a elaboração de proposta de Política Nacional para Ciência Aberta e que posteriormente foi atualizada pela Portaria n° 1.515, de 04 de abril de 2019, determinando que “o Grupo de Trabalho tem 180 dias prorrogáveis justificadamente para a conclusão de seus trabalhos” (COSTA; CUNHA, 2019, p. 302).
- 2020 – Publicação dos *Princípios e diretrizes orientadoras para políticas institucionais de dados científicos abertos em apoio à ciência aberta*. O documento foi criado no âmbito do Compromisso 3 do 4º Plano de Ação Nacional do Brasil. Esse compromisso teve o objetivo de “Estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil”, e ficou conhecido como o Compromisso pela Ciência Aberta. Foi coordenado pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa) e teve a parceria dos seguintes órgãos governamentais e da sociedade civil: Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC), Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Fundação Oswaldo Cruz (Fiocruz),

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Rede Nacional de Ensino e Pesquisa (RNP), Open Knowledge Brasil (OKBR), Comissão Nacional de Energia Nuclear (CNEN), Arquivo Nacional, Instituto de Pesquisa do Jardim Botânico do Rio de Janeiro (JBRJ). O Compromisso foi dividido em nove Marcos e o Marco 3 tinha como propósito "definir princípios e diretrizes para políticas institucionais em apoio à Ciência Aberta". O Marco 3 foi coordenado pelo Ibict em parceria com a Embrapa e a Fiocruz, cujo resultado foi o documento aqui citado. Trata-se de um conjunto de recomendações que podem servir de modelo para a construção ou o aprimoramento de políticas de abertura dos dados científicos em instituições de pesquisa (PAGANINE *et al.*, 2020).

Diante deste breve cenário de relatórios, documentos e plataformas de dados, nota-se um desenvolvimento considerável do tema nos últimos 20 anos. Em relação ao contexto internacional, a literatura aponta que o planejamento e a execução para a criação de infraestrutura de dados científicos têm integrado diversos setores da pesquisa científica, ou seja, para encontrar soluções viáveis de implantação de serviços de dados, foi necessário a conjunção de diferentes atores no processo, "[é] um contexto que abrange diferentes atores internos e externos à instituição (grupos de pesquisa, departamentos de tecnologias da informação, bibliotecas e outras seções da instituição, agências de apoio à pesquisa, empresas da iniciativa privada etc.)" (MARIN-ARRAIZA; VIDOTTI, 2019, p. 265)

Neste contexto, Costa (2017) salienta que uma das características que difere é que no movimento em países como EUA e Reino Unido, as iniciativas sempre partiam das necessidades das comunidades científicas em direção ao governo. Esse ponto destaca que, em alguns países, as comunidades científicas têm trabalhado juntamente com instâncias governamentais, com o objetivo de desenvolver uma legislação que possibilite e consolide a criação de políticas nacionais que moldem todos os ciclos das pesquisas científicas.

> Inúmeros governos e agências de fomento começam a elaborar políticas públicas relacionadas com a gestão de dados científicos […] Geralmente essas políticas visam a ampliar a eficiência da pesquisa, motivar a reutilização de dados, acelerar as ações cooperativas entre pesquisadores e suas entidades […] as iniciativas mais maduras tanto em termos de infraestrutura tecnológica, como de diretrizes para a gestão de projetos de dados científicos, concentram-se nos Estados Unidos e no Reino Unido, sendo que o Reino Unido aparece na vanguarda, tendo lançado o primeiro programa em 2001. Em contrapartida, nos EUA, o início é marcado pela publicação do Relatório *Atkins* em 2003. (COSTA, 2017, p. 72)

Neste panorama, o Brasil, em comparação com os países desenvolvidos, sua relevância no tema é bem modesta. As iniciativas citadas acima podem, com certeza, estar abrindo caminhos para novos empreendimentos de solução para a criação de política de dados científicos, mas, em muitos quesitos, o Brasil está relativamente vagaroso.

> Considerando que o Reino Unido iniciou os programas de *e-Science* em 2001 e os EUA em 2003, portanto, há pelo menos 15 anos, o Brasil encontra-se extremamente atrasado no que diz respeito tanto ao entendimento das agências de fomento sobre o tema, bem como quanto ao envolvimento dos pesquisadores para com a complexidade da questão (COSTA, 2017, p. 219).

Entretanto, apesar de avançar pouco no tema, algumas áreas do conhecimento, como a Biblioteconomia, a Ciência da Informação, a Ciência da Computação, entre outras, já desenvolvem pesquisas expressivas sobre o tema, ou estratégias que contribuem com a temática. Distingue-se, neste contexto, o Programa FAPESP de Pesquisa em eScience, que tem sido muito atuante no que tange à política para a gestão de dados científicos, pois a fundação exige que o projeto, submetido por edital de Chamada de Propostas de Pesquisa, apresente “a explicação de como será sua política de gestão de dados, incluindo a especificação do tipo de dado gerado, a forma e o tempo que serão disponibilizados, o modo de preservação e os tipos de cuidados tomados em relação a questões de privacidade e ética” (FAPESP, 2019).

Em adição às ações da FAPESP, na literatura revisada identificou-se possíveis ações que podem direta ou indiretamente contribuir com a questão, como, por exemplo:

- A reunião de várias instâncias federais de pesquisa científica para elaborar uma possível política nacional de dados, tendo como exemplo a criação do Grupo de Trabalho para a elaboração de proposta de Política Nacional para Ciência Aberta, em 2018, como já exposto acima; e
- A mobilização de comunidades científicas em prol do compartilhamento e reuso de dados, sendo um exemplo o projeto RDP Brasil, uma parceria entre a Rede Nacional de Ensino e Pesquisa (RNP), o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), a Universidade Federal do Rio Grande do Sul (UFRGS) e a Universidade Federal do Rio Grande (FURG), que, juntos, executam um projeto de mapeamento de serviços e potenciais usuários de Acesso Aberto a Dados de Pesquisa (AADP) em instituições brasileiras. Essas organizações planejam desenvolver um portal *web* para reunião da comunidade nacional interessada em AADP e fazem um levantamento dos serviços e das soluções tecnológicas existentes para compartilhamento de dados, com isso, planejam uma

proposta de solução tecnológica que atenda aos usuários identificados no estudo (GABRIEL JUNIOR *et al.*, 2019).

Neste cenário, temos algumas bibliotecas de pesquisa e universitárias juntamente com outros setores da instituição que elaboram estratégias de divulgação que mostram a relevância do compartilhamento, acesso e reuso dos dados científicos, respeitando sempre as questões de privacidade, sensibilidade, ética, moral e propriedade intelectual.

Nestes ambientes, os profissionais envolvidos estão se mobilizando no intuito de inserir a comunidade de pesquisa na relevância do tema. Ou seja, estão criando uma consciência para que estas ações potencializem o crescimento e o avanço da ciência, neste caso, podemos destacar a criação do Grupo de Trabalho em Ciência Aberta, coordenado pela Vice-Presidência de Ensino, Informação e Comunicação da Fiocruz, e o Grupo de Trabalho e Pesquisa de Dados de pesquisa da Rede de Bibliotecas Fiocruz, ambos os grupos objetivando mostrar a necessidade de se conhecer o universo dos dados de pesquisa da instituição para apoiar ações e decisões estratégicas, posicionando-se diante desse movimento de abertura dos dados de pesquisa (VEIGA *et al.*, 2019).

3.5.3 Políticas institucionais de dados: definições e estratégias

Na atualidade, quando se pensa em desenvolver um projeto de plataforma de dados científicos, a ferramenta principal para traçar diretrizes para tal atividade é a política institucional de dados, ferramenta que, se bem elaborada, mostra um compromisso institucional. Neste sentido, Correa (2019) pontua:

> A política é um conjunto de orientações em conformidade com os objetivos de uma instituição. Essas orientações tomam como base os valores fundamentais e a vocação da instituição e servem para definir estratégias, táticas e planos operacionais. […] ela define objetivos estratégicos de longo prazo para gerenciar os dados em todos os aspectos de um projeto, agência ou organização […]. Em uma política de dados, são definidos os limites de responsabilidade, autoridade e possibilidades de acesso e utilização a um conjunto de dados. […] pode também ser utilizada para tratar de questões estratégicas, como aspectos jurídicos, aquisição, administração e custódia dos dados, entre outras ações (CORREA, 2019, p. 162).

Na elaboração de uma política de dados, é fundamental que sejam formados grupos de competências e que estes grupos sejam vinculados a uma instituição. Pensando em um ambiente universitário, Sales e Sayão (2017) afirmam: “A política de um repositório [abrange] os compromissos que a instituição se obriga em relação aos seus principais

stakeholders pesquisadores, curadores, financiadores, coletores de metadados […] e com relação ao ciclo de vida das coleções de dados que [serão] gerenciados". Outro fator importante no processo de preparação de uma política é avaliar as necessidades da comunidade acadêmica em relação aos dados científicos, neste sentido, os autores Huang, Cox e Sbaffi (2020) mencionam:

> Em um nível prático, as necessidades dos pesquisadores atuam como um importante condutor para RDM, [...] devem prestar mais atenção a isso antes de implementar novos serviços. [...] Entender as necessidades primeiro, em vez de implementar os novos serviços em uma análise de necessidades, ajudará a evitar a baixa taxa de uso dos serviços (HUANG; COX; SBAFFI, 2020, p. 12, tradução nossa).

Dando apoio a este fator, Zeeland e Ringersma (2017, p. 155) declaram que "uma política [...] precisa atender às práticas e requisitos de dados de uma organização. As informações específicas do contexto, como serviços de armazenamento e arquivamento disponíveis, precisam ser levadas em consideração, juntamente com as necessidades reais de gerenciamento de dados dos pesquisadores" (tradução nossa).

A formulação de uma política tem que ser criada respeitando o contexto da instituição, sendo necessário adquirir informações e dados dos ambientes de pesquisa. É essencial que as políticas e estratégias sejam desenvolvidas com referência ao contexto institucional. E para construir uma imagem precisa deste contexto, é necessário engajar-se em um período de coleta e análise de requisitos. Os agentes envolvidos neste processo precisam estar cientes da escala e da natureza dos dados a serem gerenciados, realizando levantamentos que envolvam os pesquisadores, sendo possível entender os principais problemas que eles enfrentam, possibilitando, assim, identificar quaisquer lacunas na infraestrutura e planejar o suporte de serviços de dados desejado (RANS; JONES, 2013).

Para além de uma proposta de projeto de política de dados preocupada apenas com questões tecnológicas e interoperabilidade, é recomendável que a política faça observações e detalhes acerca de questões de normas de compartilhamento, reutilização de dados, mecanismos de fomento ao compartilhamento (SHEARER, 2015). Neste aspecto, Briney, Goben e Zilinski (2017) argumentam:

> Políticas de dados institucionais exemplares devem cobrir a propriedade de dados de pesquisa, administração e expectativas, bem como fornecer definições claras, identificar reivindicações de acesso e propriedade aos dados, especificar os períodos de retenção e definir responsabilidades de todas as partes interessadas nos dados […] (BRINE; GOBEN; ZILINSKI, 2017, p. 65, tradução nossa).

Contudo, a definição de uma política institucional de dados necessita atender às práticas e requisitos da organização, dessa forma, a seguir são assinalados alguns fatores que podem auxiliar na formulação de uma política:

- **Dinamismo e flexibilidade**: uma política de dados pode ser modificada diante de desafios não vislumbrados em sua concepção, diferentes tipos de projetos e parcerias potencialmente vantajosas, mantendo sua orientação no foco estratégico da organização;
- **Custo:** deve-se considerar o custo da obtenção dos dados *versus* seu custo de acesso. O custo pode ser uma barreira tanto para os utilizadores adquirirem certos conjuntos de dados, quanto para os provedores fornecerem dados no formato ou extensão requeridos;
- **Propriedade dos dados**: a propriedade dos dados deve ser claramente definida na política, devendo respeitar os direitos de propriedade intelectual que possam existir em diferentes níveis (CORRÊA, 2019, p. 163);
- **Quanto ao nível de abrangência:** se for uma política institucional, faz-se necessário refletir se ela irá atingir apenas as pesquisas produzidas no âmbito institucional ou se também contemplará pesquisas produzidas por redes de pesquisadores presentes em diferentes instituições. É essencial levar em conta que as dificuldades para a obtenção de recursos financeiros no âmbito científico e tecnológico bem como a própria complexidade da pesquisa culminam com a formação de redes de pesquisadores (COSTA, 2017, p.225).

Neste escopo, a política de dados das instituições tem que se preocupar em oferecer gestão e acesso ininterrupto de seus serviços, como também fomentar práticas adequadas para o reuso dos dados científicos. “O FAIR oferece um conjunto de princípios para aprimorar a utilidade dos dados e marca um refinamento importante dos conceitos necessários para dar maior valor aos dados e aumentar sua propensão para reutilização [...]” (MONTEIRO; SANTANA, 2019, p. 349). Assim, no próximo tópico, discute-se sobre os princípios *Fair.*

3.6 PRINCÍPIOS FAIR

Em 2014, na conferência internacional ‘*Jointly designing a data FAIRPORT*’, no intuito de promover e aumentar o reuso e compartilhamento de todos os tipos de dados, foram reunidas diversificadas expertises que atuam na área da pesquisa científica, e o ponto central do evento foi planejar estratégias para a criação de uma infraestrutura mundial que

contribuísse para a publicação, compartilhamento e reuso de dados resultantes de investigações científicas.

A partir desse encontro, foram estabelecidos quatro princípios e práticas que orientassem a publicação de dados e que pudessem ser facilmente localizáveis, acessíveis, interoperáveis e reutilizáveis, ou seja, *FAIR (Findable, Accessible, Interoperable, Re-usable)*, tanto por máquinas quanto humanos. Neste contexto, HENNING *et al. (*2019, p. 399) definem que: “O acrônimo *FAIR* [...] visa a implementar um conjunto de metadados definidos tanto para uso por mecanismos computacionais automatizados, quanto para uso por pessoas.”

Com isso, abaixo estão apresentados os princípios orientadores *FAIR* e seus conceitos:

- **Localizável (Findable)**

  a) dados e metadados precisam ter um identificador único persistente;

  b) os dados devem ser descritos com metadados ricos;

  c) ter o identificador persistente para o conjunto de dados descritos nos metadados; e

  d) metadados e dados devem ser recuperáveis por meio de repositórios confiáveis.

- **Acessível (Accessible)**

  a) dados e metadados devem ser recuperados pelo seu identificador, usando protocolos de comunicação padrão;

  b) os protocolos devem ser gratuitos, abertos e suportar autenticação e autorização; e

  c) os metadados devem estar acessíveis mesmo quando os dados não mais estiverem disponíveis.

- **Interoperável (Interoperable)**

  a) dados e metadados devem ser codificados usando padrões de representação acordados; e

  b) dados e metadados devem usar vocabulários alinhados aos princípios *FAIR* e incluir referências relevantes.

- **Reutilizável (Reusable)**

  a) dados e metadados precisam estar associados a atributos relevantes;

  b) dados e metadados devem ser liberados com licenças de uso claramente definidas;

  c) metadados e dados devem estar associados à sua proveniência de forma detalhada;

  d) dados e metadados devem atender aos padrões da comunidade (HUDSON, 2018).

Em 2017, o *Dutch Techcentre for Life Sciences* (DTL), apoiado pelos governos da Holanda e da Alemanha, criou a iniciativa *GO FAIR*, que tem a missão de disponibilizar um modelo de implantação dos princípios *FAIR* que pode ser usado na criação de uma

infraestrutura mundial de interoperabilidade de dados.

> Essa iniciativa busca o desenvolvimento de um ambiente global compartilhado, voltado para a pesquisa e inovação baseadas em dados. Está no seu centro incentivar a criação de redes regionais que se comprometam, de forma participativa e colaborativa, a seguir os princípios *FAIR* como estratégias de implementação (GOFAIR BRASIL SAÚDE, 2019).

No Brasil, o contato com o movimento foi através do documento *The Internet of FAIR data and services*, que foi apresentado na Conferência Luso-Brasileira de Ciência Aberta em 2017, no Rio de Janeiro. Recentemente, em 2018, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) aderiu à iniciativa, criando o *GO FAIR*-Brasil, "que tem a responsabilidade de difundir, apoiar e coordenar no Brasil as atividades relacionadas à adoção das estratégias de implementação dos Princípios *FAIR*, definidas pela iniciativa *GO FAIR*" (ENCONTRO DE GESTÃO E PRESERVAÇÃO DE DADOS DE PESQUISA EM HUMANIDADES, 1., 2019).

Com base no universo dos dados científicos e nas atividades e processos necessários para tornar real seu acesso, compartilhamento e reuso, é válido neste cenário discutir os desafios e perspectivas que as bibliotecas universitárias terão que enfrentar. Dessa forma, discute-se sobre este assunto na próxima seção.

3.7 A BIBLIOTECA UNIVERSITÁRIA NO CONTEXTO DOS DADOS CIENTÍFICOS: DESAFIOS E PERSPECTIVAS

Ao redor do mundo, de acordo com estudos analisados, essas instituições têm caminhado lado a lado com pesquisadores, agências de fomentos e órgãos de programas de pós-graduação, objetivando desenvolver serviços de gestão e curadoria de dados científicos (HEY; TREFETHEN, 2005; MARTÍNEZ-URIBE; MACDONALD, 2008; LUCE, 2010; TENOPIR *et al*., 2017; ZEELAND; RINGERSMA, 2017; MARCUM; GEORGE *apud* LIMA; FARIAS, 2019), isto é, as bibliotecas universitárias estão estabelecendo e determinando sua relevância no contexto dos dados científicos. Costa e Cunha, no início de 2010, percebendo o movimento desse novo caminho que as bibliotecas universitárias poderiam traçar, iniciaram a discussão sobre *e-science*, dados de pesquisa, curadoria, gestão de dados, tendo destacado que "para as bibliotecas universitárias a evolução gradual da *e-science* provoca desafios profundos e, ao mesmo tempo, proporciona às bibliotecas uma oportunidade de redefinir seus papéis e agregar valor ao seu portfólio de serviços" (LUCE *apud* COSTA; CUNHA, 2014, p. 196.). Na literatura estudada, é possível encontrar algumas indagações de qual será o papel da biblioteca universitária/científica brasileira no contexto

dos dados científicos (COSTA; CUNHA, 2014; SAYÃO; SALES, 2015; SANTA ANNA; DIAS; MACULAN, 2019; SEMELER; PINTO, 2019; TARTAROTTI; DAL'EVEDOVE; FUJITA, 2019).

Nesta linha, é válido analisar se as bibliotecas universitárias do Brasil estão preparadas para esta nova área de atuação. Levando em conta que as bibliotecas, sendo elas universitárias, científicas ou de outros tipos, são organizações que sempre conseguiram acompanhar e moldar suas atividades e serviços de acordo com as transições tecnológicas e comunicacionais, melhor dizendo, são uma entidade milenar e sempre conseguiram transformar seus serviços, acompanhando a evolução do público ou sociedade que dela se utiliza. Complementando esta reflexão, Carvalho (2004) reforça:

> Sujeitas aos reflexos das vicissitudes que interpretam o contexto onde estão inseridas, as bibliotecas, não somente as universitárias, durante o seu percurso, sempre estiveram envolvidas num processo de vencer desafios gerados por mudanças significativas em suas funções. Fatores como invenção da técnica de impressão, crescimento do volume e da importância da informação e comunicação, reconhecimento da importância do compartilhamento de recursos e do valor dos documentos não impressos e a busca da informatização dos seus serviços e produtos têm levado as bibliotecas universitárias a buscar formas mais apropriadas para o seu gerenciamento, mas também formas mais apropriadas para o gerenciamento do conhecimento que acumulam [em seus estoques informacionais] e para o atendimento de seus usuários [...] (CARVALHO, 2004, p. 77).

Sendo assim, o caminho a percorrer para se tornar uma organização relevante nesta nova tendência incluirá muitos desafios de aprendizagem. Considerando que as atividades e tarefas de catalogação, organização, recuperação, entre outras atividades biblioteconômicas que fazem parte do dia a dia do profissional da informação, mas que são direcionadas a elementos mais tradicionais como o livro, revistas, entre outros, capacitam o bibliotecário em atividades e tarefas que concernem ao desenvolvimento de serviços que envolvem a curadoria e gestão de dados científicos. Neste sentido, Semeler e Pinto (2019) apontam:

> As práticas tradicionais são retomadas com o objetivo de serem aplicadas a conjuntos de dados. Por exemplo, a catalogação e a organização de materiais; a preservação e a curadoria; serviços de usuários e de referência; consultoria e treinamentos são repensados para serem aplicados junto às novas práticas, que envolvem o gerenciamento, a curadoria e o compartilhamento de dados (SEMELER; PINTO, 2019, p. 124).

Entretanto, todo o rol de habilidades do bibliotecário será direcionado a objetos informacionais, que necessitam de estudos e pesquisas mais aprofundadas, pois são ativos informacionais até então desconhecidos em ambientes das bibliotecas universitárias brasileiras. Melhor dizendo, dependendo do campo de conhecimento em que o dado científico é gerado e processado, o bibliotecário terá que pesquisar, levantar informações a respeito do dado criado, como por exemplo, o metadado utilizado no domínio, qual repositório o bibliotecário pode indicar ou sugerir, entre outras tarefas.

Neste escopo, Levy *apud* Carvalho (2004, p. 66) enfatiza que "[a] descoberta de novos campos de conhecimento bem como a constante renovação dos já consolidados, [...] provocam uma rápida obsolescência das competências adquiridas no percurso profissional e exige criatividade e percepção para saber usar e reciclar as informações disponíveis". Nessa lógica, Álvaro *et al. apud* Costa e Cunha (2014, p. 196) ressaltam que "a *e-science*, [a gestão e curadoria de dados científicos] pode fornecer um campo potencial para os bibliotecários se ramificarem para além dos limites das práticas tradicionais de biblioteca […] a *e-science* não é prática comum e, em função disso, a Biblioteconomia deverá prosseguir neste novo território com cautela".

Contribuindo com esta visão, Semeler e Pinto (2019, p. 122), explicitando também uma nova designação ao bibliotecário, salientam que "se acredita que [o] bibliotecário de dados deve apreender e assimilar novas habilidades ligadas aos jargões disciplinares da gestão dos dados de pesquisa conforme o domínio dos cientistas com quem atua". Abordando a complexidade das tarefas da curadoria e gestão de dados, Sales e Sayão (2017, p. 24) declaram que "em comparação com os repositórios de *e-prints*, as funções, as descrições, os padrões e os controles são mais numerosos e complexos. Essa complexidade, no entanto, varia de acordo com os ambientes disciplinares considerados e com a política adotada pela instituição".

Objetivando ser mais pontual na análise, destaca-se neste cenário o Serviço de Referência das bibliotecas universitárias, que é um setor da instituição que lida diretamente com o público universitário e que tem por finalidade "permitir que as informações fluam eficientemente entre as fontes de informações e [para] quem precisa de informações … [é] pelo serviço de referência [...] que o conceito de prestação de serviços que se coaduna, cuidadosa e especificamente, com as necessidades do indivíduo" (GROGAN, 1995, p. 23). É um setor que está sempre lançando iniciativas para que a comunidade acadêmica possa se inteirar do que a biblioteca tem para oferecer de serviços e atividades de aprendizagem, seja com cursos e treinamentos do *software* que a instituição utiliza para empréstimos, seja com treinamentos de gerenciadores bibliográficos, cursos para utilizar o Portal Capes, entre outros.

Diante disso, o Serviço de Referência pode se tornar um ambiente apropriado para iniciar ações que viabilizem projetos que possam informar e instruir informações a respeito de acesso, compartilhamento, acesso e reuso de dados científicos, isto é, no setor de Referência é possível criar ambientes para que os pós-graduandos possam interagir com a temática em questão. Em conformidade com a ideia, os autores Santa Anna, Dias e Maculan (2019) reforçam:

> Por constituir-se um tema incipiente, mas de extrema importância, e também, considerando o investimento das universidades nos processos de pesquisa, torna-se iminente a necessidade de se garantir o acesso a esses dados, o que aponta para novas atividades e competências por parte das bibliotecas e dos bibliotecários, principalmente dos que atuam no serviço de referência (SANTA ANNA; DIAS; MACULAN, 2019, p. 3).

Neste sentido, é necessário que estes setores das bibliotecas universitárias estudem formas para ampliar a perspectiva de não serem apenas transmissores de saber, mas que também se tornem mediadores, coadjuvantes, animadores de um processo de pesquisa-aprendizagem, para que de fato possam contribuir e apoiar a autogestão de processos de aprendizagem, parceria cognitiva e construção cooperativa [do conhecimento] (MORIN, 2000).

Diante do exposto, para que o setor de Referência inicie atividades que incluem serviços que abordem acesso, compartilhamento e reuso de dados científicos, é necessário elaborar recursos e instrumentos para averiguar o contexto da universidade em relação à conjuntura dos dados científicos produzidos, o que se relaciona diretamente com as práticas e ações que os pesquisadores realizam nos ambientes de pesquisa, quer seja nos laboratórios, nos departamentos, entre outros. É nesta etapa que a biblioteca poderá vislumbrar o contexto da instituição perante a forma como os dados científicos se apresentam, reforçando, assim, ser este o primeiro passo da estratégia.

> o primeiro passo a ser realizado pelas bibliotecas é conhecer a forma de tratamento dos dados nos devidos departamentos e unidades acadêmicas que estão diretamente ligadas à produção das pesquisas. Além disso, as bibliotecas precisam reconhecer outras iniciativas adotadas pela instituição, de modo a consolidar um panorama ou diagnóstico sobre o tratamento de dados em nível institucional. A partir disso, planos emergentes de ação podem ser realizados para que as bibliotecas contribuam com a gestão dos dados (COSTA; CUNHA, 2014 *apud* SANTA ANNA; DIAS; MACULAN, 2019, p. 8).

Neste seguimento, Carvalho (2004, p. 165) enfatiza: “Se [a biblioteca] pretende atender às expectativas, às necessidades e aos desejos de seus usuários, ou seja, de reconhecê-los como elemento norteador de suas atividades, as bibliotecas universitárias […] precisam começar por mapear as características e demandas desses usuários”. Nesta perspectiva, Santa Anna, Dias e Maculan (2019) ressaltam:

> Os bibliotecários, ao planejar serviços de informação, devem ficar atentos à realidade e ao contexto institucional, observando as necessidades da comunidade que visa a atender. Tais necessidades podem ser identificadas a partir da realização de diagnósticos organizacionais realizados junto aos departamentos e unidades responsáveis […] (SANTA ANNA; DIAS; MACULAN, 2019, p. 11).

Em outras palavras, é nesta etapa do mapeamento das práticas científicas que irá se consubstanciar a elaboração de um planejamento para tal atividade, pois só esse levantamento irá mostrar o universo, os recursos disponíveis e os perfis das pesquisas que estão sendo desenvolvidas na instituição, ou seja, das práticas de pesquisa mais frequentes (CARLSON, 2011).

Após realizar o mapeamento dos contextos dos ambientes de pesquisa da universidade e com os resultados, compreender as informações e os dados coletados, o bibliotecário, juntamente com as chefias, podem planejar serviços que introduzam o público da pós-graduação: docentes, discentes, técnicos administrativos, profissionais de TI ao tema em questão. Neste momento é que a equipe do setor de referência irá analisar qual área ou disciplina ou departamento está inteirado ou não da temática, pois, identificando esta característica, a equipe terá mais subsídios para determinar em qual departamento e/ou centro poderá iniciar a empreitada. Perante isso, para a próxima etapa da iniciativa, que foi elaborada a partir da literatura analisada (CORRÊA, 2017; KOLTAY *apud* SANTOS, 2019; SAYÃO, 2020), elencamos algumas iniciativas institucionais que o setor de referência pode articular.

- elaborar webconferência para introdução do tema e sempre dividir em programas de pós-graduação, pois com a divisão é possível, separar em domínios ou em cursos;
- elaborar aulas para que possa ser administrada a disciplina de metodologia científica ou disciplinas afins, com conteúdo introdutório sobre o assunto;
- criar guias sobre o tema e disponibilizar na página da biblioteca;
- fornecer um catálogo de dados institucionais ou repositório de dados, esquematizando em áreas do conhecimento;
- criar grupos de discussão que integrem pesquisadores, professores da pós-graduação, bibliotecários, arquivistas e profissionais de TI da instituição; e

- elaborar campanhas para as redes sociais, vinculando a biblioteca ao tema.

Contudo, as atividades sugeridas acima, à primeira vista, são incipientes, porém são tarefas que têm o potencial de incluir a biblioteca e a universidade na lógica de como “o pesquisador deve começar a pensar sobre seus dados, [ou seja] como um produto que necessita ser preservado adequadamente, desde um projeto de pesquisa até sua finalização [e] que somente através de uma eficiente gestão de dados podemos perceber o potencial completo dos dados científicos” (SILVA, 2016, p. 404).

Em outras palavras, estas ações são os primeiros passos que a instituição pode dar para inserir a comunidade acadêmica no tema em questão, assim, nota-se que a gestão e a curadoria de dados científicos são um campo com inúmeras vertentes, podendo a biblioteca avaliar qual ramificação poderá seguir, mas é preciso que a equipe responsável por tal tarefa esteja atenta ao levantamento das práticas científicas realizadas previamente.

Diante do exposto, as iniciativas e inovações que poderão ser incluídas nos serviços e atividades de uma biblioteca universitária, que integre gestão e curadoria de dados, dependerão da capacidade de o profissional da informação acompanhar, atualizar e se capacitar para contribuir com tais tarefas. Nesta perspectiva, pensando em uma via de mão dupla, Silva *et al*. (2019) apontam:

> outro desafio que surge para o profissional bibliotecário é o de disseminar a importância da gestão de dados e capacitar os pesquisadores que querem fazer, mas não sabem por onde começar. Surge, assim, a necessidade de complementar a formação do profissional bibliotecário, uma vez que para capacitar, o bibliotecário também precisa estar capacitado (SILVA *et al.*, 2019, p. 306).

A informação no contexto da gestão e curadoria de dados científicos, dada a complexidade do assunto, requer desse profissional uma educação continuada para que sua atuação e posicionamento reflitam em serviços que colaborem em pesquisas e soluções para projetos viáveis de implantação de curadoria e gestão de dados. As bibliotecárias Lima e Farias (2019) salientam:

> o desafio imposto aos bibliotecários requer um perfil profissional engajado, capaz de mediar, estabelecer contatos, fazer ligações e alianças com os potenciais pesquisadores/produtores de dados, pelo fato de ser um profissional que deveria sair do seu lugar comum e espaço físico de trabalho para ir até os laboratórios, departamentos, escritórios e gabinetes para a captação dos dados a fim de dar-lhes um sentido e inseri-los dentro de um contexto, possibilitando o tratamento, a curadoria, a preservação digital e a recuperação desses dados (LIMA; FARIAS, 2019, p. 190).

Dessa forma, nota-se que se o bibliotecário tiver interesse em atuar nesta nova tendência da área, são necessárias uma constante educação e especialização e a aquisição de novos conhecimentos que, consequentemente, irão ser incorporados aos já adquiridos. Por este ângulo, Silva *et al.* (2019) enfatizam:

> esses profissionais precisam ampliar suas competências para acompanhar as mudanças no âmbito da pesquisa científica, além de compreender o mundo complexo e heterogêneo dos dados de pesquisa, para que possam desenvolver serviços que atendam às necessidades dos pesquisadores em suas instituições, abrangendo as dimensões tecnológicas, legais e éticas (SILVA *et al.*, 2019, p. 312).

Nesta visão, Santos (2019) elenca alguns papéis e competências que o bibliotecário deverá ter para que sua atuação no tema em questão seja relevante:

- Conhecer os fluxos de pesquisa da instituição;
- Empregar a curadoria digital dos dados de pesquisa;
- Empregar a competência em dados, no uso da tecnologia e infraestrutura de dados de pesquisa, para o pensamento crítico na resolução de problemas, promover cursos e treinamentos;
- Incentivar e planejar a Gestão de dados de Pesquisa, incluindo *advocacy* e orientação aos pesquisadores;
- Definir formatos e padrões de codificação de dados de pesquisa e metadados descritivos, administrativos, técnicos, estruturais, preservação, disciplinares, de uso e proveniência;
- Documentar os versionamentos dos dados de pesquisa e todas as tecnologias empregadas;
- Analisar e interpretar dados de pesquisa quantitativos e qualitativos, com o uso da competência em dados (saber entender a tabulação, ler gráficos e/ou análise estatística);
- Identificar os processos de levantar tecnologias para coleta de dados, limpeza, processamento, análise, semântica, simulação, modelamento, mineração, visualização e reuso;

- Reconhecer e promover treinamentos sobre pacotes estatísticos, localizar conceitos, práticas em linguagem de programação, algoritmos e tecnologias da ciência de dados;
- Promover consultoria sobre dados de pesquisa;
- Promover treinamentos em competência em dados de pesquisa e gerenciamento;
- Auxiliar no desenvolvimento de políticas de dados de pesquisa e planos de gestão de dados de pesquisa;
- Incentivar a adoção de políticas de dados abertos, quando apropriado;
- Projetar a arquitetura da informação nos repositórios de dados de pesquisa de acordo com as particularidades de cada área;
- Gerenciar repositórios de dados de pesquisa;
- Criar representações visuais de dados (*data storytelling*, a história dos dados de pesquisa);
- Planejar, avaliar e orientar estratégias para segurança da informação de dados de pesquisa, para gerar ambientes confiáveis;
- Estruturar a conversão de dados e a interoperabilidade, lidar com os riscos e possíveis perdas ou corrupção de informação causados pela mudança de formatos de dados e promover *backups* capazes de ser lidos e interpretados continuamente; e
- Promover a citação de dados, aplicando identificadores persistentes aos dados de pesquisa.

Deste modo, na tentativa de introduzir algumas iniciativas que remetem ao universo dos dados científicos em uma biblioteca universitária, este estudo procedeu a coleta das práticas e perspectivas sobre o tema na Universidade Federal de Viçosa (UFV), assunto a ser tratado a seguir.

4 UNIVERSIDADE FEDERAL DE VIÇOSA

Neste capítulo são expostas informações da Universidade Federal de Viçosa, descrevendo sua trajetória desde sua criação e sua atuação na área da pesquisa científica, ou seja, serão apresentados como são organizados os setores de pós-graduação. Esta instituição tem atualmente três campi: Viçosa, Florestal e Rio Paranaíba - todos localizados no estado de Minas Gerais, porém o estudo foi aplicado somente no campus de Viçosa.

História, apresentação e pesquisa[12]

O *campus* Viçosa originou-se da Escola Superior de Agricultura e Veterinária (Esav), criada pelo Decreto nº 6.053, de 30 de março de 1922, do então Presidente do Estado de Minas Gerais, Arthur da Silva Bernardes. A Esav foi inaugurada em 28 de agosto de 1926 por seu idealizador, Arthur Bernardes, que, na época, ocupava o cargo de Presidente da República.

Em 1927, foram iniciadas as atividades didáticas com a instalação dos cursos Fundamental e Médio e, no ano seguinte, do Curso Superior de Agricultura. Em 1932, foi a vez do Curso Superior de Veterinária. No período de sua criação, o professor Peter Henry Rolfs, da Universidade da Flórida (Estados Unidos da América), foi convidado para organizar e dirigir a Esav.

Visando ao desenvolvimento da Escola, em 1948, o Governo do Estado a transformou em Universidade Rural do Estado de Minas Gerais (Uremg), que era composta pela Escola Superior de Agricultura, Escola Superior de Veterinária, Escola Superior de Ciências Domésticas, Escola de Especialização (Pós-Graduação), Serviço de Experimentação e Pesquisa e pelo Serviço de Extensão. Devido à sólida base e ao bom desenvolvimento, a Uremg adquiriu renome pelo país, o que motivou sua federalização pelo Governo Federal em 15 de julho de 1969, quando passou a ser chamada de Universidade Federal de Viçosa (UFV).

Inicialmente, a instituição se destacou nas Ciências Agrárias, com os cursos de Agricultura e de Veterinária. Hoje, além de continuar sendo referência na área, também tem excelência no ensino, pesquisa e extensão nas Ciências Biológicas, Exatas e Humanas.

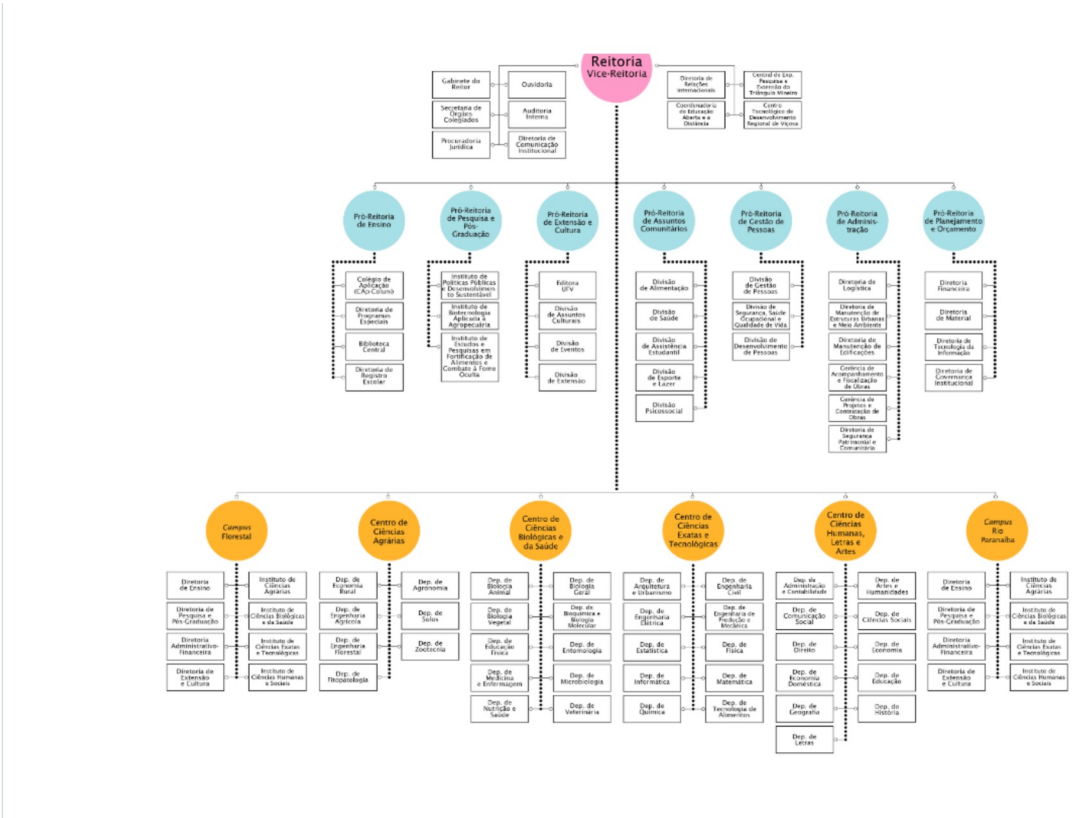
Apresentação

O campus Viçosa tem quatro Centros de Ciências, que são os órgãos responsáveis por administrar as atividades de ensino, pesquisa e extensão em uma ou mais áreas de conhecimento, respeitando as normas legais, estatutárias, regimentais e as resoluções dos

[12] Disponível em: https://www.ufv.br/historia/

órgãos competentes. A administração dos Centros de Ciências cabe ao Conselho Departamental, à Câmara de Ensino e à Diretoria de Centro. Aos Centros de Ciências, estão vinculados os Departamentos e os cursos de graduação (Figura 5).

Figura 5 - Centros de Ciências da UFV



Fonte: UFV (2021)

Pesquisa

Em relação aos cursos de pós-graduação, a UFV é pioneira no Brasil no oferecimento de cursos de pós-graduação, com mais de 50 anos de tradição. São diversos programas de pós-graduação que oferecem treinamento em níveis de mestrado, doutorado e mestrado profissional. Assim, as pós-graduações *stricto sensu* no campus de Viçosa se dividem nos centros conforme descrito a seguir:

- Centro de Ciências Agrárias (CCA): Agroecologia, Ciência Florestal, Economia Aplicada, Engenharia Agrícola, Extensão Rural, Fitopatologia, Fitotecnia, Defesa Sanitária Vegetal (Profissional), Meteorologia Aplicada, Solos e Nutrição de Plantas, Tecnologia de Celulose e Papel (Profissional), Zootecnia e Zootecnia (Profissional).
- **C**entro de Ciências Biológicas e da Saúde (CCB): Biologia Animal, Biologia Celular e Estrutural, Bioquímica Aplicada, Botânica, Ciência da Nutrição, Ciência da Saúde

(Profissional), Ecologia, Educação Física, Entomologia, Fisiologia Vegetal, Genética e Melhoramento, Medicina Veterinária e Microbiologia Agrícola.

- Centro de Ciências Exatas e Tecnológicas (CCE): Agroquímica, Arquitetura e Urbanismo, Ciência da Computação, Ciência e Tecnologia de Alimentos, Engenharia Civil, Engenharia Química, Ensino de Física (Profissional), Estatística Aplicada e Biometria, Física Aplicada,e Matemática Multicentro em Química de Minas Gerais.
- Centro de Ciências Humanas, Letras e Artes (CCH): Administração, Economia, Economia Doméstica, Educação, Letras, Patrimônio Cultural, Paisagens e Cidadania (Profissional).

No que concerne à pesquisa e à pós-graduação, ambas são administradas pela Pró-Reitoria de Pesquisa e Pós-Graduação (PPG). Este órgão implementa e executa políticas com o objetivo de ampliar a produção científica e intelectual e de fortalecer a pós-graduação na UFV. A PPG também apoia e auxilia a ter visibilidade a elaboração de projetos institucionais de pesquisa, a busca de oportunidades de financiamento e de prêmios, a proteção à propriedade intelectual e a transferência de tecnologia.

Além do apoio logístico, a PPG também oferece suporte para o funcionamento de comitês e comissões que visam a adequar laboratórios de pesquisa, quanto à legalidade dos aspectos éticos e de biossegurança para consolidação de grupos de pesquisa, e laboratórios multiusuários, para o fortalecimento da iniciação científica e para a divulgação e registro dos projetos de pesquisa (UNIVERSIDADE FEDERAL DE VIÇOSA, 2018).

Trabalhando em conjunto com a PPG, destacam-se neste contexto setores que auxiliam nos processos de andamento dos projetos de pesquisa, que são:

- Comissão Permanente de Propriedade Intelectual (CPPI), que atua como Núcleo de Inovação Tecnológica (NIT) da UFV, vinculada à PPG, promove a disseminação da cultura de proteção à propriedade intelectual e a transferência de tecnologia, com foco na inovação no âmbito da UFV;
- Comitê de Ética em Pesquisa com Seres Humanos (CEP): o CEP analisa eticamente todo projeto que envolve pesquisa com seres humanos, direta ou indiretamente, sendo coordenado por um pesquisador responsável vinculado à UFV ou indicado pela Comissão Nacional de Ética em Pesquisa (Conep);
- Comissão de Ética no Uso de Animais (Ceua): trata-se de uma comissão institucional, constituída de um colegiado interdisciplinar, autônomo, de caráter consultivo, deliberativo e educativo, com a finalidade de garantir a

utilização ética de animais em atividades de ensino, pesquisa científica, extensão, treinamento, trabalho de iniciação científica, trabalho de conclusão de curso (TCC), entre outros.

- Comissão Interna de Biossegurança (CIBio): é responsável pelo monitoramento e vigilância das atividades com Organismos Geneticamente Modificados (OGM) e seus derivados e por fazer cumprir as normas de biossegurança. A CIBio é encarregada de obter licenças junto à Comissão Técnica Nacional de Biossegurança (CTNBio) e tem por finalidades assessorar, analisar e emitir pareceres quanto aos aspectos técnicos de biossegurança de todos os procedimentos científicos a serem desenvolvidos na UFV, que envolvam a manipulação de OGMs.

Tendo em vista que o estudo proposto buscou identificar e mapear informações e dados a respeito dos dados científicos, neste aspecto é válido elencar os institutos de pesquisa[13] da UFV.

- **Instituto de Biotecnologia Aplicada à Agropecuária (Bioagro):** desenvolve pesquisas na área de biotecnologia para a criação de produtos e processos biotecnológicos. A estrutura do Bioagro tem duas edificações: o Bioagro e o Instituto Nacional de Ciência e Tecnologia (INCT), instalado no edifício anexo ao Bioagro. Essa estrutura tem 26 laboratórios, além de três laboratórios associados, e congrega aproximadamente 600 usuários, entre pesquisadores, técnicos e estudantes de diversos departamentos da UFV, além de pesquisadores visitantes de outras instituições de ensino e/ou pesquisa nacionais e internacionais.
- **Instituto de Estudos e Pesquisas em Fortificação de Alimentos e Combate à Fome Oculta da UFV (Ipaf-UFV)**: tem a missão de gerar, integrar e transferir conhecimento científico e tecnológico e contribuir para a formação de pessoal em áreas estratégicas relacionadas à fome oculta e à desnutrição, causadas, principalmente, pela deficiência de micronutrientes.
- **Instituto de Políticas Públicas e Desenvolvimento Sustentável (IPPDS):** tem função gerenciadora e executiva do Programa Institucional de Desenvolvimento de Pesquisas, Ensino, Extensão e Prestação de Serviços em Ciências Humanas e Sociais Aplicadas, com foco em Políticas Públicas e Desenvolvimento Sustentável.
- **Instituto Nacional de Ciência e Tecnologia em Interações Planta-Praga (INCTIPP)**: foi estruturado para atender à necessidade de se avançar no conhecimento

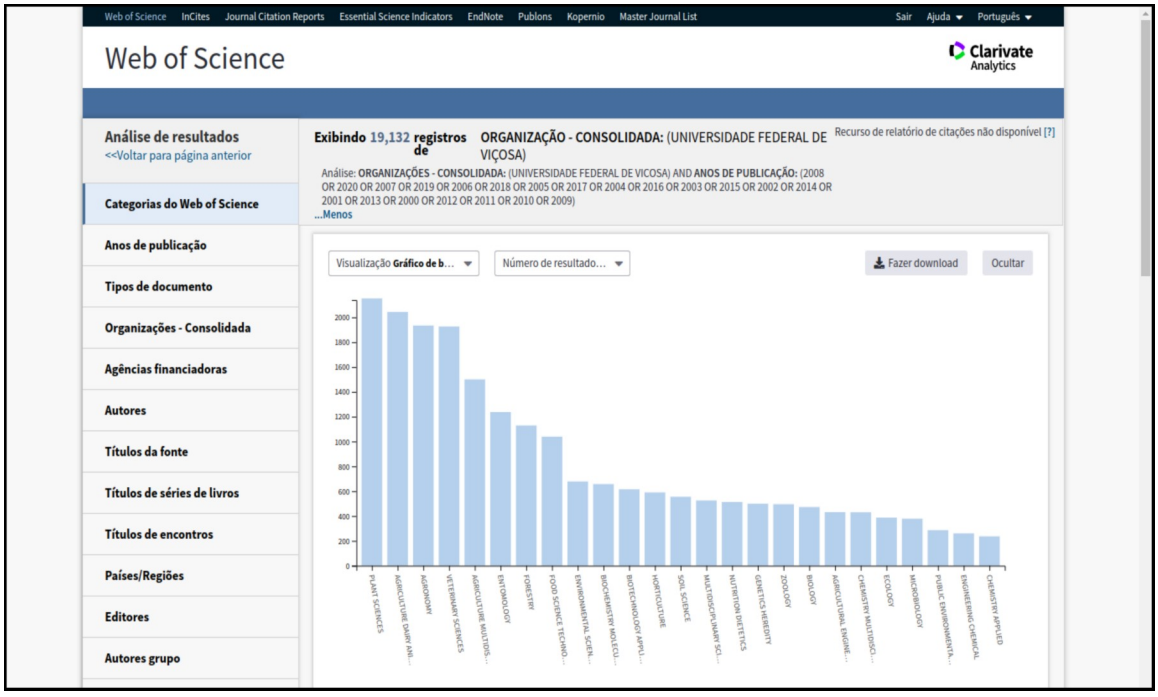
[13] Relatório de atividade de 2018: ano base 2017.

científico a respeito de bases moleculares e funcionais das interações entre plantas e pragas relevantes para a agricultura brasileira e de se intensificarem colaborações multidisciplinares e multi-institucionais que contribuam para melhor adequação de procedimentos na coleta de dados, a fim de torná-la o mais eficiente possível.

- **Instituto Nacional de Ciência e Tecnologia de Ciência Animal (INCTCA)**: coordenado pelo Departamento de Zootecnia da UFV, tem o objetivo de desenvolver novas metodologias e produzir informação biológica para apoiar atividades científicas e tecnológicas inovadoras sobre ciência animal, a fim de melhorar os atuais resultados de eficiência de produção, diminuir as perdas e os impactos ambientais e maximizar o potencial produtivo em todas as áreas da ciência animal no Brasil.

Buscando levantar um panorama dos números da UFV em relação a publicações, bolsas concedidas, projetos de pesquisa, números de pesquisadores, linhas de pesquisa e patentes registradas, expõe-se a seguir tais informações. A Figura 6 sinaliza as produções científicas dos últimos 20 anos, utilizando a *Web of science/Clarivate analytics*, podendo ser percebido um crescimento significativo entre 2000 e 2020.

Figura 6 - Produções científicas dos últimos 20 anos



Fonte: Web of science/Clarivate analytics (2021)

A Tabela 1 mostra as produções técnico-científicas de todos os Centros da UFV, no ano de 2019.

Tabela 1 - Produções técnico-científicas da pós-graduação de 2019

CENTROS	TIPOS DE PUBLICAÇÕES/2019 [14,15]																	
	APE	APN	TCE	TCN	TPE	TPN	LPE	LPN	CLE	CLN	JFO	MNT	REL	TMD	TMD	TDD	BOL
TOTAL	1117	625	241	902	86	255	3	45	65	176	63	-	-	-	-	-	-
CCA	331	229	77	215	13	71	2	18	21	43	22	-	-	-	-	-	-
CCB	482	123	74	416	7	21	-	9	14	54	5	-	-	-	-	-	-
CCE	313	96	57	158	42	92	-	13	18	27	23	-	-	-	-	-	-
CCH	51	177	33	113	24	71	1	5	12	52	13	-	-	-	-	-	-

Fonte: Disponível em: https://www3.dti.ufv.br/ppo/integra/relatorio/producaocientifica

A Tabela 2 mostra os projetos de pesquisa e publicações referentes ao ano de 2018.

Tabela 2 - Projetos de pesquisa e publicações (2018)

Projetos de pesquisa e publicações				
Centros				
	Reg.[1]	And.[2]	Concl.[3]	Publ.[4]
CCA	438	1047	440	1159
CCB	420	978	449	901
CCE	343	735	350	1108
CCH	201	392	213	762
Total	**1402**	**3152**	**1452**	**3930**

Fonte: UFV. Relatório de atividade de 2018

LEGENDA: 1 Registrados; 2 Em andamento; 3 Concluídos; 4 Publicações

A Tabela 3 traz a evolução do número de pesquisadores, patentes registradas e linhas de pesquisa.

[14]Disponível em: https://www3.dti.ufv.br/ppo/integra/relatorio/producaocientifica.

[15] *APE- Artigos em Periódicos Estrangeiros; APN - Artigos em Periódicos Nacionais; TCE - Comunicações em Congressos Estrangeiros; TCN - Comunicações em Congressos Nacionais; TPE - Trabalho Completo em Congressos Estrangeiros; TPN - Trabalhos Completos em Congressos Nacionais; BOL - Boletins;CLE – Capítulos de Livros Estrangeiros; CLN – Capítulo de Livros Nacionais; JFO - Artigos em Jornais e Folhetos; LPE - Livros Publicados no Exterior; LPN- Livros Publicados Nacionais; MNG - Monografias e Trabalhos Finais de Cursos; MNT - Manuais Técnicos; REL - Relatórios Técnicos; TDD - Tese de Doutorado Defendida; TMD - Teses de Mestrado Defendidas; TDL - Texto didático de uso local e OUT - Outras publicações.*

Tabela 3 - Quantitativo de pesquisadores, patentes registradas e linhas de pesquisa (2017)

Área	Pesquisadores					Patentes Registradas					Linhas de Pesquisa				
	2013	2014	2015	2016	2017	2013	2014	2015	2016	2017	2013	2014	2015	2016	2017
TOTAL	**517**	**780**	**577**	**562**	**433**	**25**	**13**	**13**	**23**	**19**	**293**	**353**	**361**	**401**	**247**
Ciências Agrárias	170	217	157	162	131	3	1	-	4	5	111	88	71	89	82
Ciências Biológicas e da Saúde	131	204	153	159	114	13	2	9	13	10	64	74	71	71	48
Ciências Exatas e Tecnológicas	147	223	158	142	107	9	10*	4**	6	4	81	102	91	150	81
Ciências Humanas, Letras e Artes	69	136	109	99	81	-	-	-	-	-	37	89	128	91	36

Fonte: UFV. Relatório de atividade de 2018

A Tabela 4 destaca o número de bolsas de iniciação científica[16] oferecidas no ano de 2019.

Tabela 4 - Número de bolsas de iniciação científica (2019)

Bolsas Ofertadas - Viçosa

Centro	Processo de Selecao	Bolsas
Centro de Ciências Agrárias	CNPQ/PIBIC	61
Centro de Ciências Agrárias	CNPQ/PIBIC-AF	1
Centro de Ciências Agrárias	CNPQ/PIBITI	2
Centro de Ciências Agrárias	FUNARBIC	2
Centro de Ciências Agrárias	PIBICFAPEMIG	35
Centro de Ciências Agrárias	SICOOBUFVCREDI	2
Centro de Ciências Biológicas e da Saúde	CNPQ/PIBIC	105
Centro de Ciências Biológicas e da Saúde	CNPQ/PIBIC-AF	1
Centro de Ciências Biológicas e da Saúde	CNPQ/PIBITI	5
Centro de Ciências Biológicas e da Saúde	FUNARBIC	3
Centro de Ciências Biológicas e da Saúde	PIBICFAPEMIG	49
Centro de Ciências Exatas e Tecnológicas	CNPQ/PIBIC	85
Centro de Ciências Exatas e Tecnológicas	CNPQ/PIBIC-AF	1
Centro de Ciências Exatas e Tecnológicas	CNPQ/PIBITI	5
Centro de Ciências Exatas e Tecnológicas	FUNARBIC	3
Centro de Ciências Exatas e Tecnológicas	PIBICFAPEMIG	50
Centro de Ciências Humanas, Letras e Artes	CNPQ/PIBIC	52
Centro de Ciências Humanas, Letras e Artes	CNPQ/PIBIC-AF	1
Centro de Ciências Humanas, Letras e Artes	CNPQ/PIBITI	1
Centro de Ciências Humanas, Letras e Artes	FUNARBIC	4
Centro de Ciências Humanas, Letras e Artes	PIBICFAPEMIG	39
Total		**507**

Fonte: Disponível em: https://paineis.ufv.br/pesquisa/ppg

[16] Disponível em: https://paineis.ufv.br/pesquisa/ppg

Neste cenário, inclui-se a Biblioteca Central Professor Antônio Secundino São José[17], órgão suplementar da UFV ligado administrativamente à Pró-Reitoria de Ensino. Está localizada no centro do campus universitário de Viçosa, ocupa um edifício moderno e funcional de quatro andares, com área total de 12.816,59 m^2. Disponibiliza aos usuários mais de 2.050 postos de estudos, que incluem salas de uso individual e em grupo, além de sala de videoconferência, coleções especiais, coleções de obras raras, multimídia, mapoteca, espaço para leitura de lazer, espaço para pesquisa às bases de dados e periódicos eletrônicos, *hall* para exposições diversas e um auditório. O órgão também conta com 11 bibliotecas setoriais, que somam uma área total construída de mais de 1000m^2 à disposição dos estudantes nos diversos departamentos,

O acervo bibliográfico, localizado na Biblioteca Central e nas bibliotecas setoriais do campus de Viçosa, atingiu em dezembro de 2018 os seguintes quantitativos: 188.041 livros, 37.850 teses e dissertações e 426.439 periódicos. O acervo inclui ainda 4.935 materiais especiais (mapas, “slides”, fitas de vídeo, “CD-ROM”, DVDs, disquetes e outros), além de 11.624 boletins, 140 normas técnicas e 1.185 obras de referência.

A Biblioteca Central oferece diversos serviços como Achados e Perdidos, Atendimento às Bibliotecas Setoriais, Catálogo *On-line*, Empréstimos, Empréstimos entre Bibliotecas, Ficha Catalográfica, Nada Consta, Normalização, Recebimento de Doações, Renovação/Reserva, Solicitação de Documentos, Treinamento de Usuários, Uso da Rede WI-FI e Visitas Orientadas. Entre esses serviços expostos, a biblioteca também oferece uso e acesso aos sistemas abaixo:

- Biblioteca depositária da ONU;
- Acesso ao Sistema Brasileiro de Informação do Café – SBICafé (disponível em www.sbicafé.ufv.br);
- Portal de Periódicos da CAPES, bem como bases de dados referenciais;
- Biblioteca-base do COMUT (Programa de Comutação Bibliográfica);
- Integra a Biblioteca Digital de Teses e Dissertações (BDTD) do IBICT/CNPq, com o Repositório Locus, que conta também com 9.682 artigos científicos, disponibilizados na integra.

Em adição às informações mencionadas acima, a Universidade Federal de Viçosa também é destaque em *rankings* educacionais brasileiros e mundiais. A citada universidade já esteve, por três vezes, entre as 100 melhores instituições de ensino superior do mundo na área das Ciências Agrárias e Florestais, de acordo com o *QS World University Rankings*, e entre as

[17]Disponível em: http://www.bbt.ufv.br/?page_id=143

melhores de Minas Gerais, segundo o ranking da *Times Higher Education* e o *Brics & Emerging Economies (*UFV, 2020).

Ressalta-se, mais uma vez, que a apresentação dessas informações ficou limitada a órgãos e setores que convergem em suas atividades e tarefas. Contudo, a exposição das informações acima teve o intuito de auxiliar e reforçar a justificativa criada anteriormente, pois através do cenário apresentado pode-se substanciar que a pesquisa científica na UFV tem um status relevante na região mineira, brasileira e mundial, afirmando, assim, ser uma instituição pertinente para o planejamento e execução da pesquisa.

5 RESULTADOS OBTIDOS E ANÁLISE DOS DADOS

Neste capítulo serão apresentados os resultados obtidos através das respostas do formulário *on-line* e da entrevista com os funcionários da UFV, sendo que o primeiro instrumento de coleta teve o objetivo de mapear as práticas e percepções dos pesquisadores quanto a seus dados científicos. Já o segundo instrumento buscou conhecer a estrutura organizacional de gestão dos dados científicos dos programas de pós-graduação, isto é, os resultados aqui expostos correspondem aos objetivos 2 e 3, respectivamente. O formulário *on-line* foi enviado para 4.029 contatos de *e-mails* selecionados, retornaram 183 questionários respondidos. Dos 29 perfis administrativos identificados no formulário, 16 aceitaram participar da entrevista e 7 a concederam. Assim, a seguir são apresentados os dados e as respectivas análises.

5.1 RESULTADOS E ANÁLISE DOS DADOS REFERENTES AO QUESTIONÁRIO *ON-LINE*

Buscando alcançar o objetivo 2, esta etapa da pesquisa descreve o cenário da comunidade científica da UFV quanto às práticas e percepções em relação aos dados científicos criados. Serão apresentados os resultados obtidos dos dados levantados e suas respectivas análises. Esclareça-se que, nesta análise dos dados resultantes do questionário, foram utilizadas algumas citações indiretas não inseridas na revisão da literatura selecionada, mas que serviram para reforçar os dados encontrados.

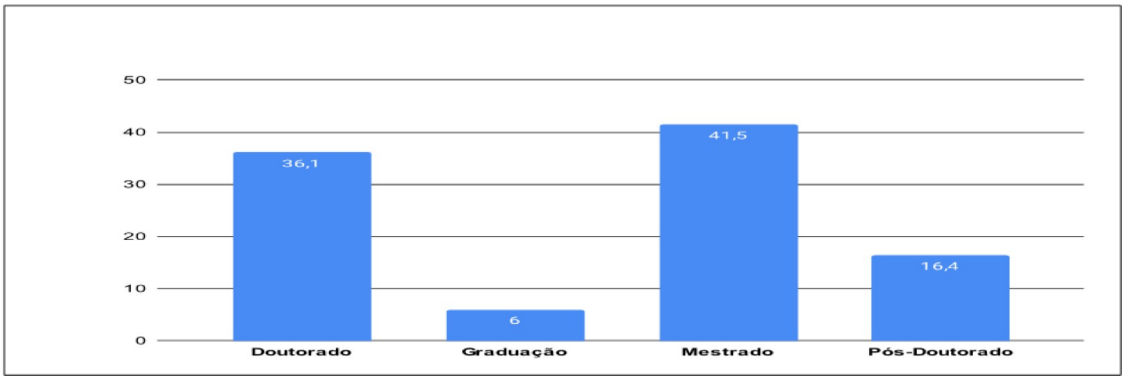
5.1.1 Perfil dos respondentes

A primeira parte do questionário consistiu em coletar dados e informações demográficas dos respondentes, objetivando identificar características como escolaridade, perfil profissional, em qual Centro/departamento a pós-graduação está inserida e a principal agência de financiamento.

5.1.1.1 Dados recolhidos

O Gráfico 1 mostra a escolaridade dos respondentes. De um total de 183, cerca de 41,5% são mestrandos e 36,5% estão cursando o doutorado.

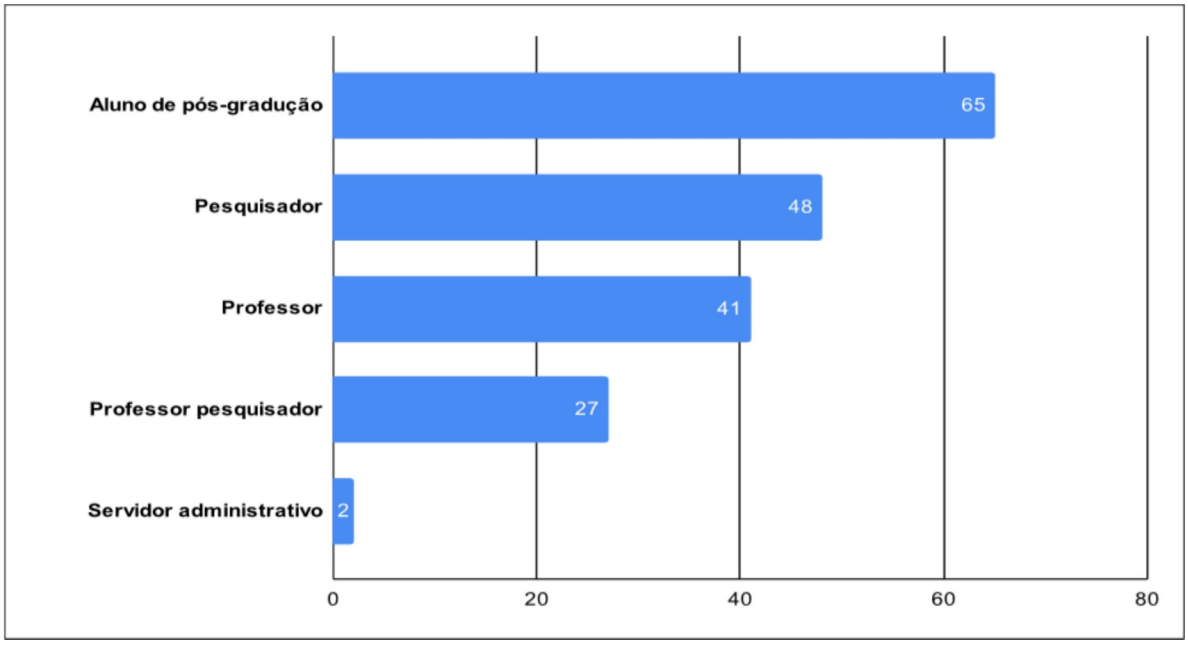
Gráfico 1 - Escolaridade



Fonte: Elaborado pela autora, 2021.

Quanto ao perfil profissional, o Gráfico 2 mostra que a maioria se declarou aluno de pós-graduação.

Gráfico 2 - Perfil profissional[18]

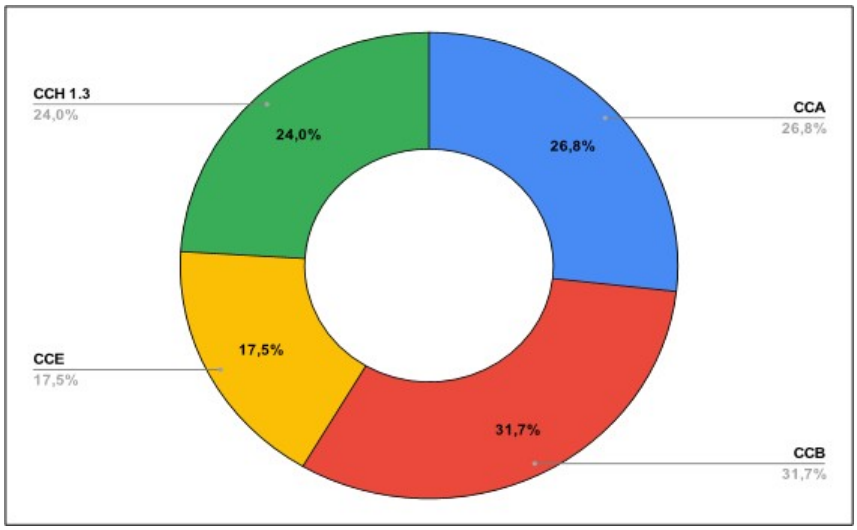


Fonte: elaborado pela autora, 2021

O Gráfico 3 mostra a representatividade dos centros de ciências da UFV aos quais a pós-graduação ou a pesquisa do respondente está vinculada. Entre os 183 respondentes, há maior volume de respostas no Centro de Ciências Biológicas e da Saúde, 31,7%, seguido pelo Centro de Ciências Agrárias, 26,8%, pelo Centro de Ciências Humanas e Artes, 24%, e pelo Centro de Ciências Exatas, 17,5%, respectivamente.

[18] Em relação à questão que gerou este gráfico, foi necessário ajustá-la, por conter a opção “outros”, o que gerou vários perfis, em que as palavras eram diferentes, mas os significados eram os mesmos.

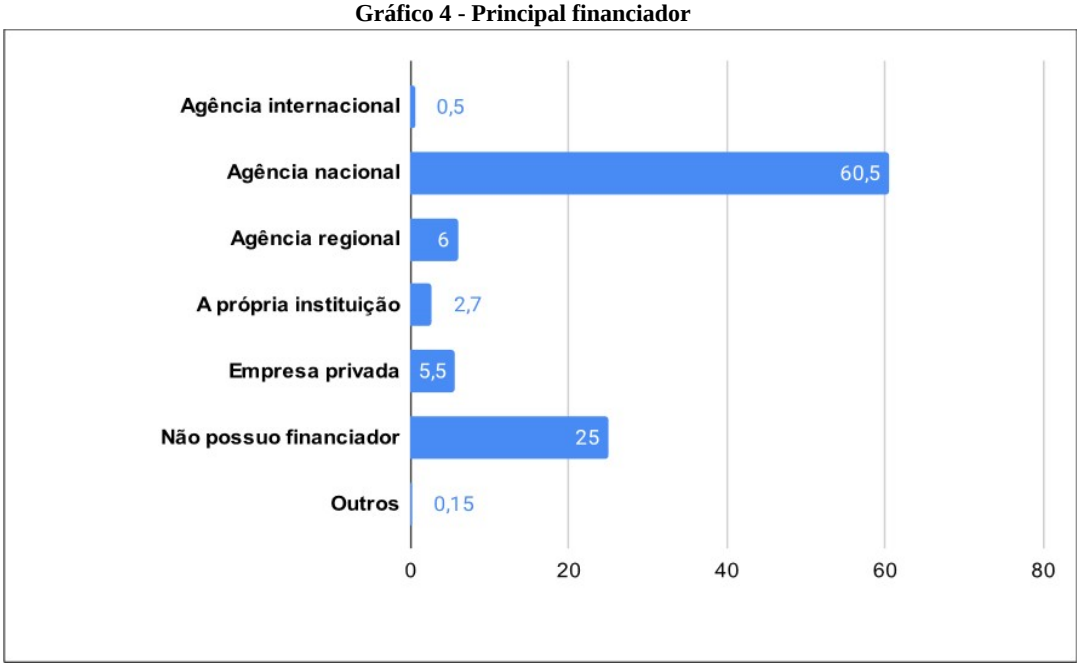
Gráfico 3 - Centro a que a sua pós-graduação ou sua pesquisa está vinculada



Fonte: Elaborado pela autora, 2021.

O Gráfico 4 mostra os tipos de agência de financiamento que atuam na UFV, ou seja, quem são os principais financiadores a que os pesquisadores pedem fomento. Em torno de 60,5% dos respondentes recebem financiamento de agências nacionais, percentual bem significativo, enquanto que 25% não têm financiador.

Gráfico 4 - Principal financiador



Fonte: elaborado pela autora, 2021.

5.1.1.2 Análise

Este mapeamento foi crucial para obter informações sobre qual público teve mais interesse em responder ao questionário, ou seja, quem mais se interessou pelo tema abordado.

Foi possível identificar os perfis e o nível de escolaridade como também fazer o levantamento de qual Centro teve maior número de respondentes. Vale ressaltar que, dentro desse universo de 183 respondentes, 89,9% responderam ter interesse em participar de eventos criados pela UFV sobre o tema abordado.

Assim, por meio dessa sondagem, foi possível adquirir informações que podem dar subsídios para as tomadas de decisões no que se refere a iniciativas que a biblioteca pode planejar em relação a serviços de apoio referentes a dados científicos e oferecer aos alunos de pós-graduação. Com as informações levantadas, a biblioteca pode direcionar seu trabalho primeiramente ao público que teve maior participação, melhor dizendo, pode elaborar planejamentos de atividades de apoio e assistência que obedeçam a uma sequência de atendimentos, incluindo, em primeiro lugar, o Centro de Ciências Biológicas e da Saúde, em segundo, o Centro de Ciências Agrárias, em terceiro, o Centro de Ciências Humanas e Artes, posteriormente, o Centro de Ciências Exatas. Esta observação, captada na primeira parte do questionário, fornece um panorama sobre quais caminhos o setor de Referência e Atendimento ao Público da Biblioteca Central da UFV pode traçar. Reforçando esta ideia, os autores Santa Anna, Dias e Maculan (2019, p. 4) pontuam que as bibliotecas precisam avançar e se envolver mais ativamente para fornecer serviços, informações *on-line* e desenvolver práticas educacionais. Neste escopo, percebem também o papel educativo do serviço de referência e as iniciativas que podem ser desenvolvidas para que os serviços de apoio à pesquisa se direcionem à gestão dos dados produzidos pelas pesquisas acadêmicas.

Identificou-se também que, entre as agências de fomento que atuam na UFV, a maioria delas é nacional. As agências de fomento desempenham papel relevante nas pesquisas desenvolvidas na UFV, implicando que, em um projeto que envolva educação em dados, estas organizações precisam estar inseridas, dialogando com os bibliotecários, com os coordenadores de programas de pós e com outros atores, ou seja, elas têm um papel importante neste cenário, pois elas podem ajudar a elaborar atividades que auxiliem para uma melhor compreensão do tema entre os pesquisadores.

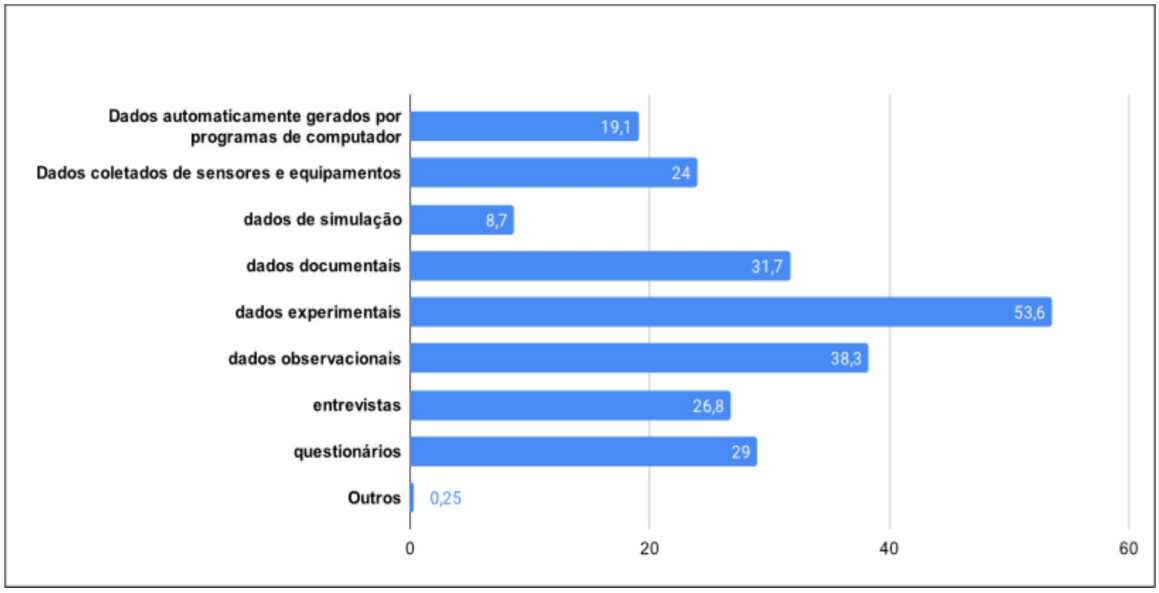
5.1.2 Questões referentes às práticas sobre os dados científicos

A segunda parte buscou averiguar as práticas dos pesquisadores em relação a dados científicos. Esta parte consistiu em detectar as práticas de uso, reuso e compartilhamento dos dados científicos dos respondentes.

5.1.2.1 Dados recolhidos

O Gráfico 5 demonstra como os respondentes classificam os termos que melhor descrevem seus tipos de dados. Para 53,3%, são dados experimentais e para 38,3%, são dados observacionais.

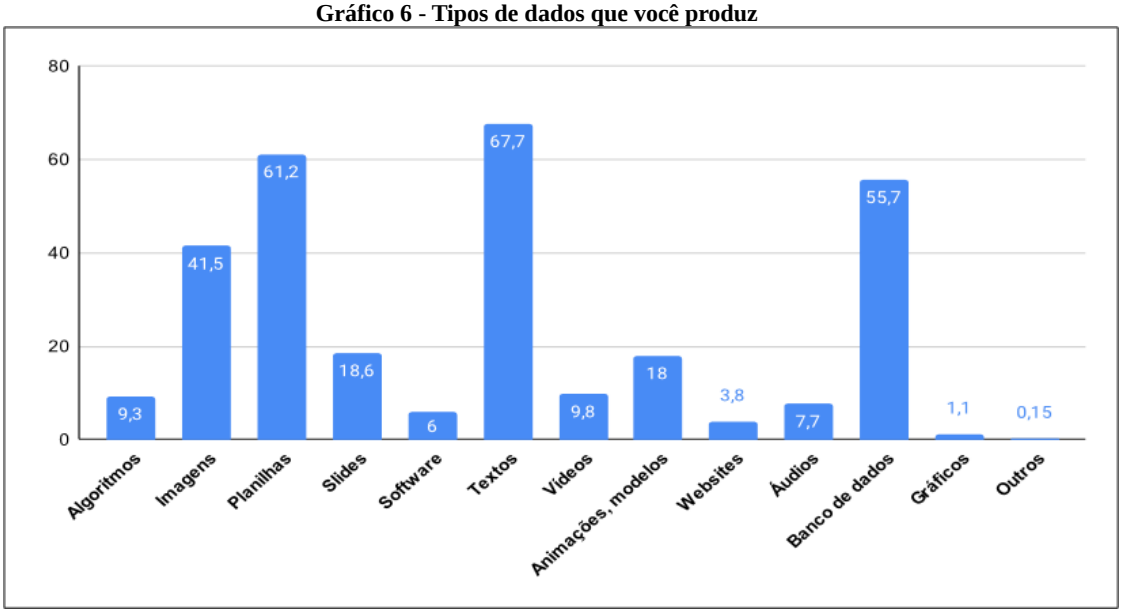
Gráfico 5 - Termos que melhor descrevem os tipos de dados produzidos



Fonte: Elaborado pela autora, 2021.

Buscando verificar os tipos de dados produzidos pelos respondentes, sendo possível que pudessem escolher mais de uma opção, no Gráfico 6 vê-se a abrangência dos tipos de dados produzidos na UFV. Assim, temos em destaque: textos (67,7%), seguidos de planilhas (61,2%) e de banco de dados (55,7%).

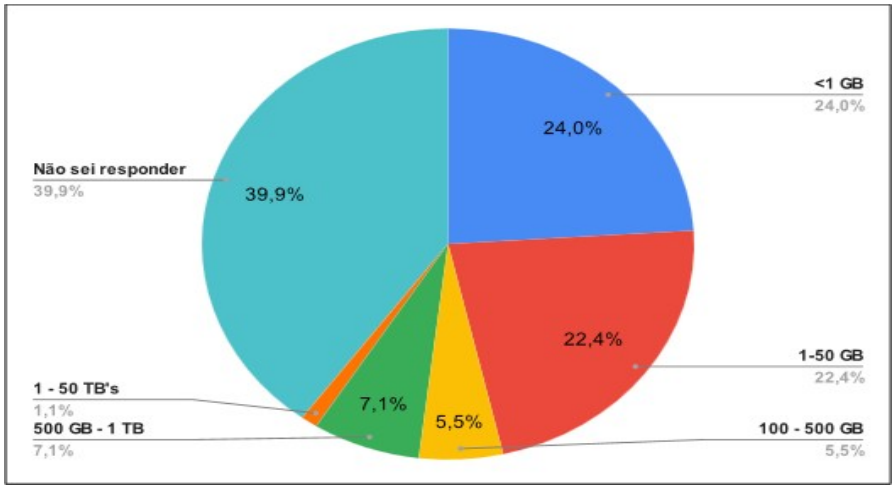
Gráfico 6 - Tipos de dados que você produz



Fonte: Elaborado pela autora, 2021.

O Gráfico 7 dimensiona o volume de dados produzidos por ano e por percentual de respondente. Para 24% o volume estimado é de 1 GB; para 22,4%, de 1-50 GB, e 39,9% não souberam responder.

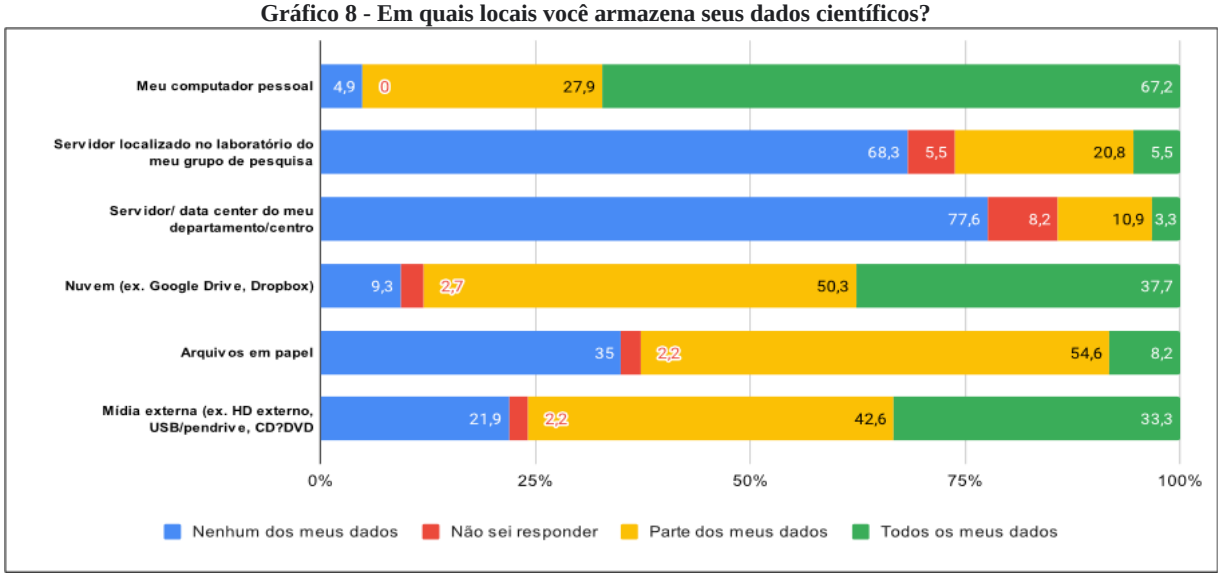
Gráfico 7 - Volume estimado dos dados científicos que você produz em um ano



Fonte: Elaborado pela autora, 2021.

Em seguida, os respondentes indicam os locais de armazenamento dos dados científicos utilizados. O Gráfico 8 mostra que a maioria (67,2%) guarda todos os dados em seus computadores pessoais e 37,7% em nuvem. Destaca-se também que 77,6% dos respondentes não utilizam servidor ou *data center* do Departamento ou Centro.

Gráfico 8 - Em quais locais você armazena seus dados científicos?

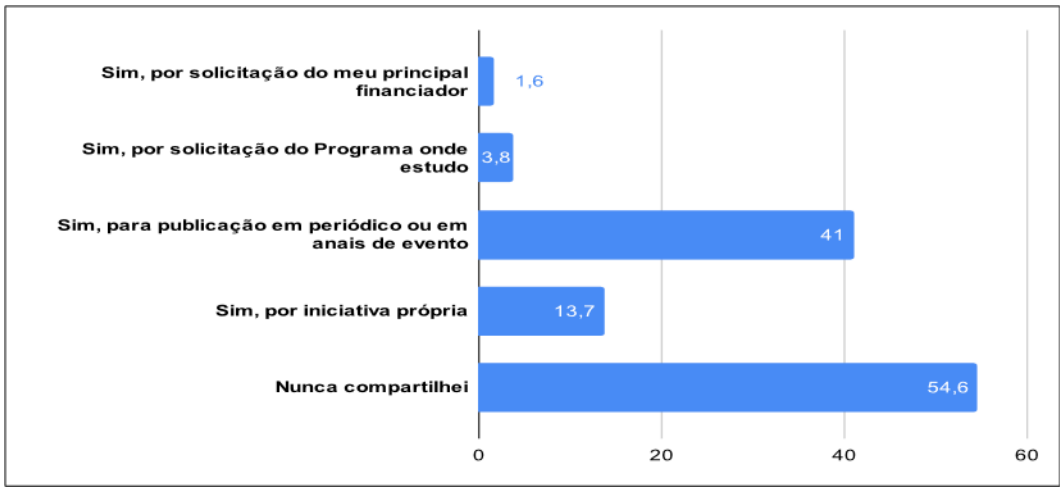


Fonte: Elaborado pela autora, 2021.

Buscando verificar se os respondentes tiveram a oportunidade de compartilhar seus dados científicos em acesso aberto e em quais circunstâncias, vê-se no Gráfico 9 que 41%

compartilharam através de publicação em periódicos ou em anais de eventos, outros 13,7% por iniciativa própria, porém, cerca de 54,6% nunca compartilharam.

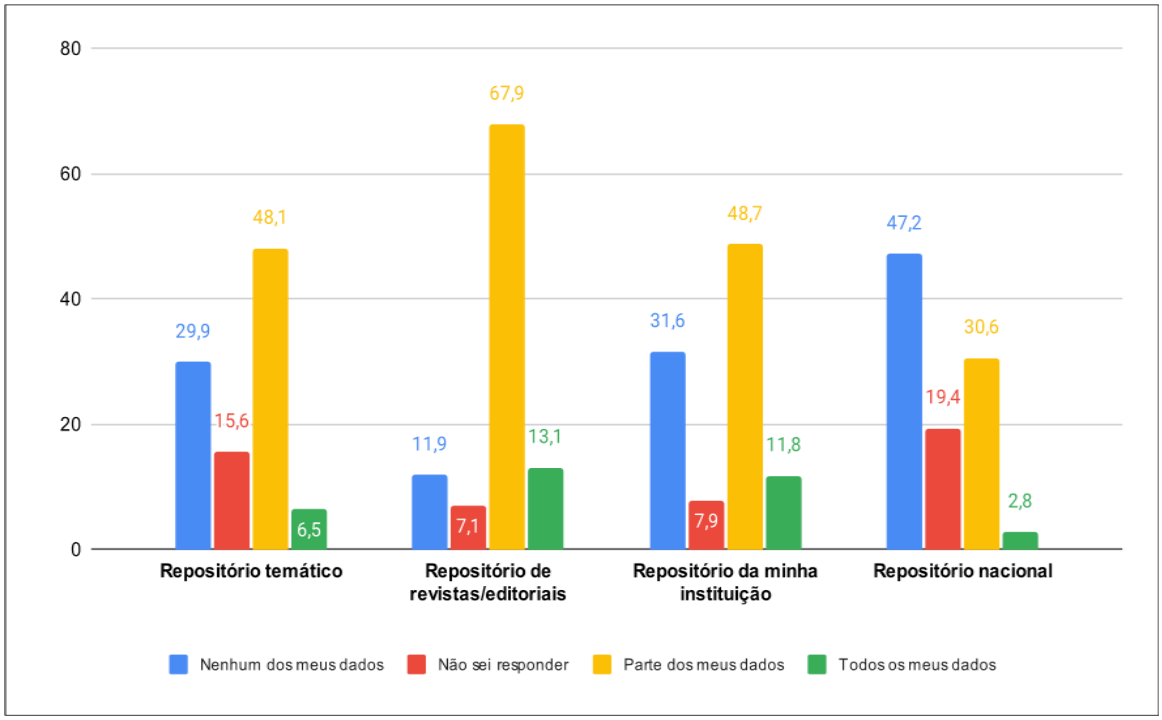
Gráfico 9 – Você já compartilhou seus dados científicos em acesso aberto?



Fonte: Elaborado pela autora, 2021.

No Gráfico 10, é possível observar em que tipos de repositórios os respondentes depositam seus dados, como também identificar o percentual dos dados compartilhados em acesso aberto. Um baixo percentual de respondentes compartilha todos os seus dados nos diversos tipos de repositórios apresentados. Mas quando se trata de compartilhar somente parte dos dados, 67,9% dos respondentes apontam os repositórios de revistas.

Gráfico 10 – Se você respondeu que compartilha seus dados científicos em acesso aberto, onde você compartilha?



Fonte: Elaborado pela autora, 2021.

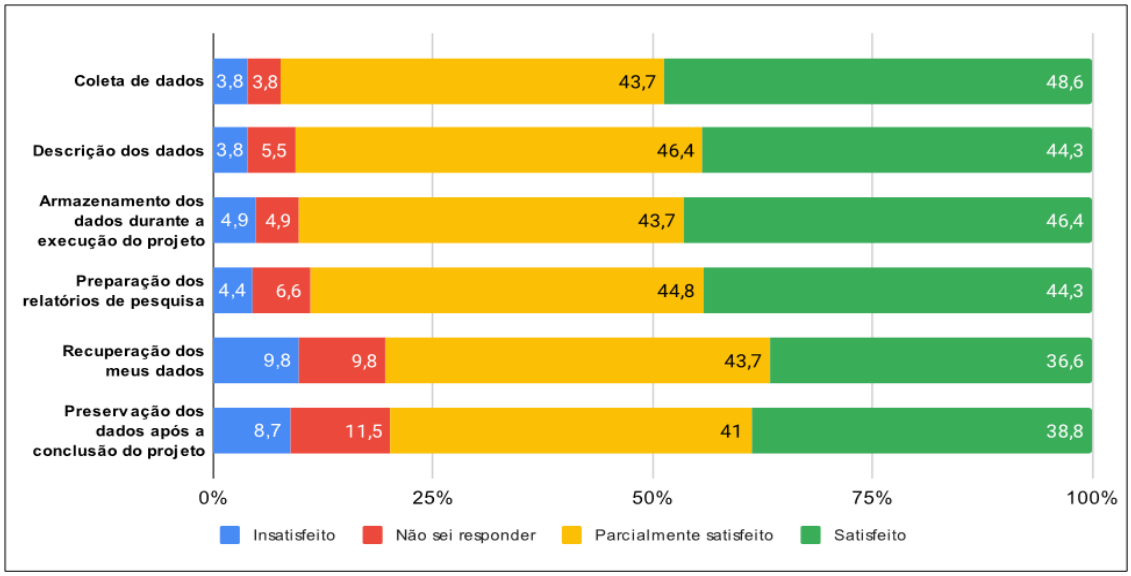
Quando os respondentes são questionados por que razão não compartilham os dados, cerca de 61,7% apontam que precisam publicar os resultados da pesquisa antes de compartilhá-los, outra razão que aparece é que o compartilhamento não lhe é exigido, em torno de 52,2% (Gráfico 11).

Gráfico 11 - Se NÃO compartilha, as possíveis razões

Fonte: Elaborado pela autora, 2021.

No Gráfico 12, os respondentes apontam o grau de satisfação concernente à organização dos seus dados científicos.

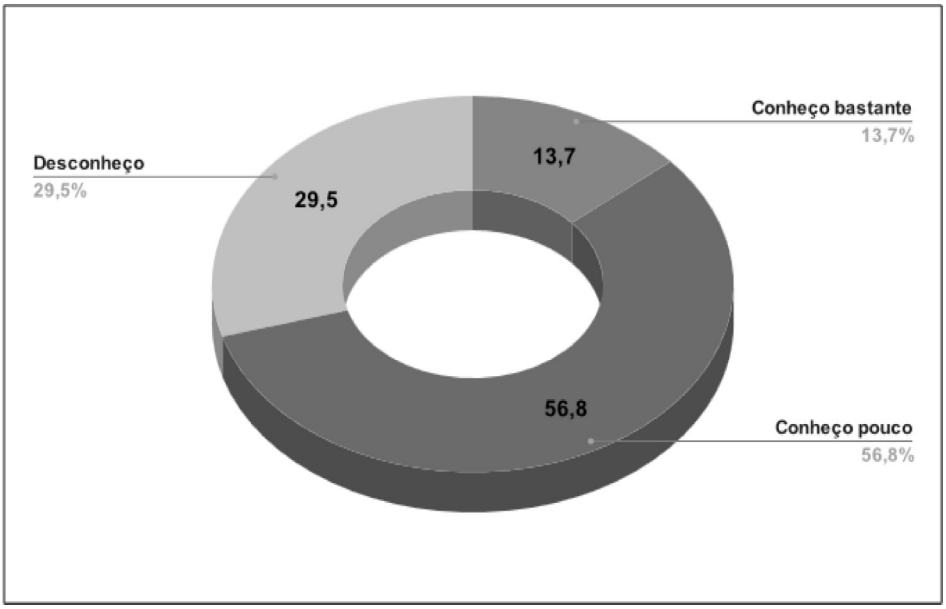
Gráfico 12 – Quão satisfeito você está com a organização dos seus dados?



Fonte: Elaborado pela autora, 2021.

Na tentativa de investigar se os respondentes têm alguma familiaridade com a gestão de dados científicos, 56,8% informaram que conhecem pouco, 29,5% que desconhecem, e apenas 13,7% responderam que conhecem bastante.

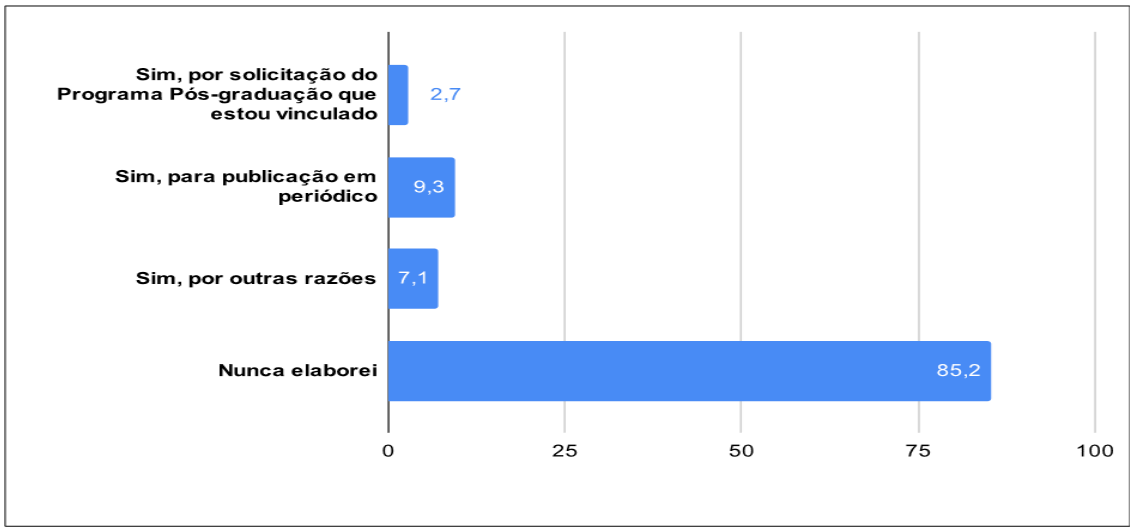
Gráfico 13 - Familiaridade com a gestão de dados científicos



Fonte: Elaborado pela autora, 2021.

Ainda com relação à gestão de dados, o Gráfico 14 mostra que 85,2% nunca elaboraram um Plano de Gestão de Dados e que apenas 9,3% já o fizeram para a publicação em periódico.

Gráfico 14 – Você já elaborou um Plano de Gestão de Dados?



Fonte: Elaborado pela autora, 2021.

No decorrer do questionário, os respondentes são indagados se utilizam algum repositório de dados para acessar dados científicos de outros pesquisadores, 27,3% respondem

que sim, em oposição a 72,7% dos que afirmam não. Na Tabela 5, os participantes nomeiam os repositórios utilizados, tendo sido descritos bases de dados, redes sociais acadêmicas, repositórios institucionais, dentre outros.

Tabela 5 - Repositórios de dados científicos* apontados pelos respondentes

Archive.org	1
Arxiv	1
AtoM	1
(BID) Banco Interamericano de Desenvolvimento	1
BDTD (Biblioteca Digital Brasileira de Teses e Dissertações)	7
Biblioteca de Base de Dados da Universidade de Michigan	1
Biodiversity Heritage Library	1
BLAST (Basic Local Alignment Search Tool - NCBI - NIH)	1
Dropbox	1
Earth System Grid Federation (ESGF)	1
FMI (Fundo Monetário Internacional)	1
GenBank	1
GitHub	5
Google Drive	1
Google Earth Engine	1
Google Scholar	5
IBGE (Instituto Brasileiro de Geografia e Estatística)	1
IEEE (Instituto de Engenheiros Eletricistas e Eletrônicos)	1
IPEA (Instituto de Pesquisa Econômica Aplicada)	1
JSTOR	1
Kaggle	1
Kegg (Kyoto Encyclopedia of Genes and Genomes)	1
Library Genesis	1
Linkedin	1
LOCUS Repositório Institucional da UFV	2
Mendeley	1
Mendeley data	1
Mycobank	1
NASA data catalog	1
National Center for Biotechnology Information (NCBI)	4
PBDB (Paleobiology Database)	1
Periódicos Capes	9
PLoS (Public Library of Science	1
PubMed	4
Repositório da Universidade de Lisboa	1
Repositório digital do Instituto Butantan	1
Repositórios de Universidades Federais	5
ResearchGate	4
RI da UFPB	1
Sci-hub	3
Scielo	4
Science Direct	1
Scopus	3
The World Bank	1
UniProt	1
Web of Science	1
Zenodo	1
Total*	**47**
***As respostas refletem o que os participantes entendem por repositório de dados.**	

Fonte: Elaborada pela autora, 2021.

5.1.2.2 Análise

Na segunda etapa do questionário *on-line*, foi possível mapear que tipos de dados são processados pela comunidade científica dos programas de pós da UFV. Observa-se também como essa comunidade descreve os principais tipos de dados produzidos. Com este levantamento, foi possível obter a abrangência dos dados criados e processados pelos respondentes e, o mais importante, constata-se a diversidade de tipologia dos dados científicos existente. Assim, essa etapa da pesquisa pode ser desdobrada em planos de ação que podem ser elaborados pela biblioteca e por outros setores que tiverem interesse. Melhor dizendo, em razão dessa análise da tipologia dos dados, obteve-se informações relevantes quando for desenvolver atividades ou projetos relacionados à gestão e à curadoria de dados, pois, reconhecer a heterogeneidade dos dados é um dos pontos cruciais nesse tipo de atividade. Assim, a pesquisa observou que em torno de 53,6% trata-se de dados resultantes de experimentos, ou seja, são dados experimentais e com estes tipos "os procedimentos [são] realizados em condições controladas com a finalidade de provar o estabelecimento de hipótese sobre um determinado fenômeno. Se um experimento está esboçado para ser replicável, esses dados podem ser mais fáceis de se reutilizar e se preservar." (SILVA, 2019 p, 28). São dados resultantes de experimentos controlados em bancadas de laboratórios, como por exemplo, parâmetros de uma reação química (SAYÃO; SALES, 2015). Por volta de 38,3% são dados observacionais, oriundos de observações científicas, com estes tipos de dados "os pesquisadores medem tantas variáveis quanto lhes seja possível, a fim de elucidar as possíveis relações de causa e efeito. São um tipo de registro que tem caráter único e que, diferentemente dos experimentos, não podem voltar a ser reproduzidos, portanto, são insubstituíveis" (SILVA, 2019 p, 27). Neste sentido, Sayão e Sales (2015) pontuam que são dados capturados através de observações diretas e que podem ser associadas a lugares e tempos específicos, por ter natureza instantânea e possuem uma importância crítica que os qualifica como registros históricos.

Em relação aos tipos de dados, identifica-se que 67,7% são textos, seguidos de 61,2% de planilhas. Por meio dessas informações, é plausível um planejamento de estratégias que solucionam questões referentes a formatos de arquivo e uso de *software.* Quando se pensa em um projeto de gestão dos dados ou em outras tarefas relacionadas, é possível que se formulem meios para que estes dados científicos possam ser aceitos para intercâmbio, reuso e preservação. "É necessário o uso adequado de formatos de arquivos e de *software* para garantir que os dados possam ser identificados de acordo com as normas internacionais de maneira única e que sejam acessíveis para usos futuros" (SILVA, 2019, 40 p.).

Nesta etapa de mapear as práticas sobre os dados científicos, observa-se que em torno de 39,9% dos respondentes não conseguiram opinar sobre o volume de dados, mostrando seu desconhecimento quanto ao volume de armazenamento de seus dados. Já em relação aos locais de armazenamento, eles conseguem explanar de forma mais contundente suas preferências. Os respondentes afirmam que todos os seus dados são armazenados em computadores pessoais e, em direção oposta, 76,6% pontuam que nenhum dos seus dados é armazenado em dispositivos do Programa ou no Departamento. O que indica uma falta de estrutura institucional que contenha diretrizes que possam sinalizar ações que mudem essa postura da comunidade científica. Segundo Vanz *et al.* (2018, p. 21), “percebe-se no Brasil a falta de instrumentos e diretrizes uniformes para a preservação e segurança de dados de pesquisa, a matéria-prima do empreendimento científico.” Em outras palavras, os riscos de manter os dados somente em PCs potencializam atitudes que não favorecem o compartilhamento como também inviabilizam que outros pares e a comunidade científica possam acessar e reusar os dados criados. Combinando com essa teoria, Heidorn (2008), na Seção 3.4 desta pesquisa, percebe a problemática de inúmeros dados científicos que criados e processados nos laboratórios ou núcleos de pesquisa, não são publicados ou disponibilizados, tornando-se invisíveis para os pares da comunidade científica.

No tocante ao compartilhamento dos dados em acesso aberto, 54,6% dos respondentes nunca praticaram tal ação e os principais motivos são: primeiro, precisam publicar os resultados da pesquisa antes de compartilhá-los; segundo, alegam que isso nunca lhes foi exigido. Esse quadro evidencia que na mentalidade dos participantes da pesquisa ainda prevalece o modelo tradicional da ciência, eles veem seus dados como um produto competitivo, e quanto mais eles retêm a disponibilidade, mais garantem o crédito pelo uso de seus dados.

Com isto, é necessário que as estratégias que podem contornar essa resistência incluam desenvolver projetos que estimulem, entre a comunidade científica da UFV, a importância de compartilhar dados científicos, entretanto é substancial que nestes projetos também sejam abordadas questões como privacidade, anonimização e confidencialidade dos dados.

Nessa etapa da pesquisa, detectou-se que os respondentes têm pouca familiaridade com gestão de dados científicos, acentuando, mais uma vez, que a introdução sobre o tema nos ambientes de pesquisa é algo primordial, isto é, essa informação aponta que existe uma necessidade de que a biblioteca direcione uma estruturação de ações que tenham elementos que possam trazer mudanças, mesmo que graduais, quanto à concepção do tema referido.

Neste escopo, observa-se que de forma esmagadora 85,2% nunca elaboraram um PGD, o que confirma o desconhecimento sobre a gestão de dados. Nesta direção, na seção 3.7 desta dissertação, Santos (2019) elenca alguns papéis e competências do bibliotecário numa instituição de pesquisa. Este novo perfil inclui o auxílio no desenvolvimento de políticas de dados de pesquisa e planos de gestão de dados.

Já no que se refere ao uso de repositórios, alguns respondentes apontaram nomes que entendem como repositórios, muito deles informaram bases de dados, redes sociais e repositórios institucionais, que ainda não têm infraestrutura de dados científicos. Esse desconhecimento dos pesquisadores da UFV em relação ao conceito de repositórios de dados não é diferente de outros pesquisadores brasileiros. Vanz *et al* (2018, p. 28) no relatório *Acesso aberto a dados de pesquisa no Brasil: práticas e percepções dos pesquisadores*, identificou que os pesquisadores ao serem solicitados para especificar o nome do repositório utilizado, a grande maioria dos respondentes informou bases de dados como Scielo e *Web of Science*; portais como o Portal de Periódicos da Capes; redes sociais acadêmicas como *Mendeley, ResearchGate* e o Academia.edu; repositórios institucionais; revistas científicas, o que denota o desconhecimento por parte dos pesquisadores brasileiros do que seja um repositório de dados de pesquisa.

Diante disso, a biblioteca pode proporcionar serviços de apoio que integrem atividades que apresentem os repositórios de dados, ou seja, expor suas funcionalidades, estruturas de dados, tipos de dados que agregam e outras características, além de que, salientar como os repositórios se tornaram relevantes para o desenvolvimento da ciência moderna, isto é, realçar que “os repositórios assumiram o centro das atividades de procura para a produção do conhecimento e chegam a tempo, pois a quantidade de dados científicos que nascem digitalmente aumenta rapidamente.” (SILVA, 2019, p. 89).

5.1.3 Questões referentes às percepções em relação ao compartilhamento e uso de dados científicos

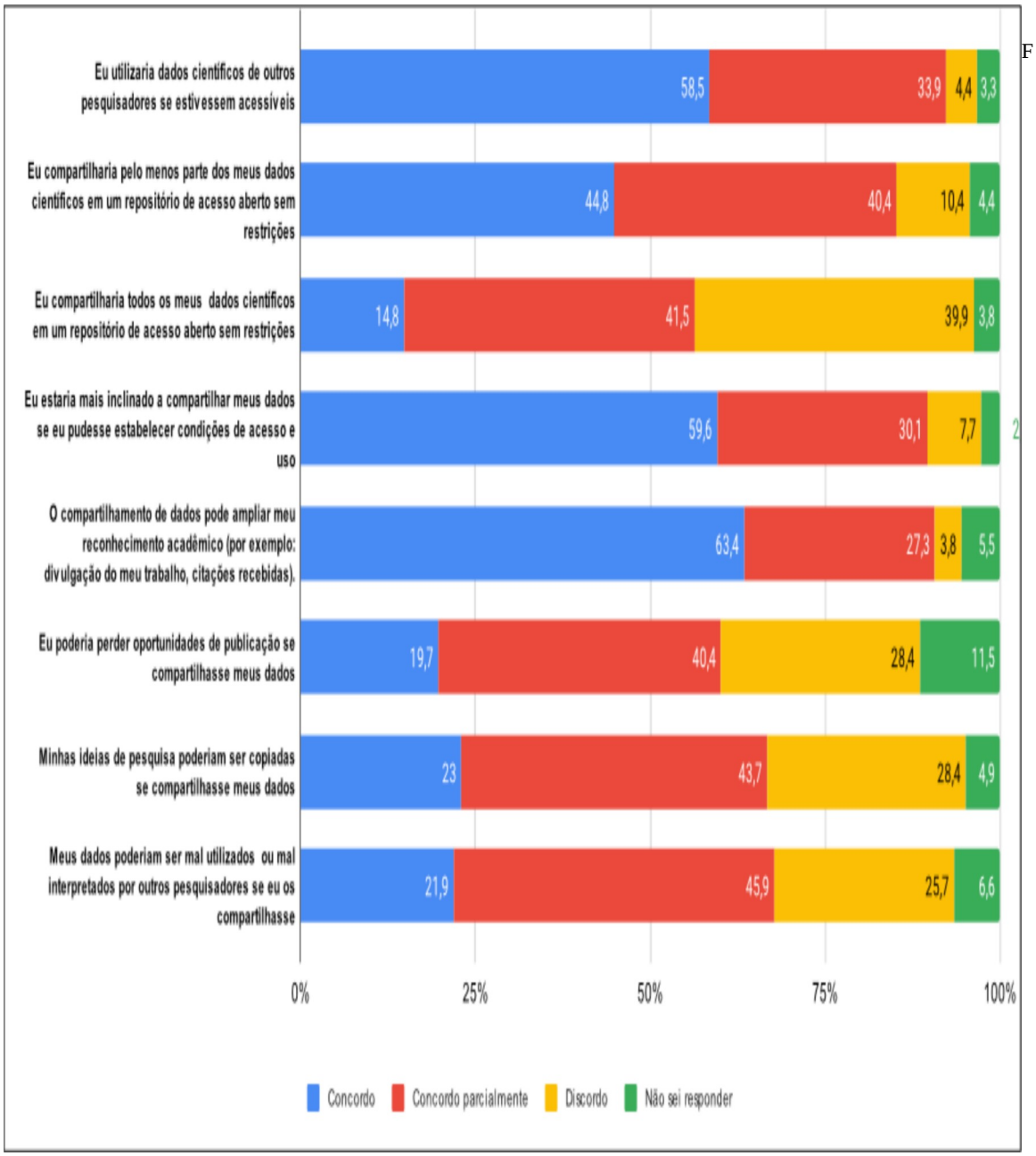
A terceira parte da ferramenta buscou verificar a percepção dos respondentes no que concerne ao compartilhamento e uso de dados científicos.

5.1.3.1 Dados recolhidos

A questão que gerou o Gráfico 15 apresenta algumas afirmativas sobre o compartilhamento e uso de dados de outros pesquisadores, ou seja, 63,4% dos respondentes afirmam que concordam que o compartilhamento de dados pode ampliar seu reconhecimento

acadêmico e, quase o mesmo percentual, 58,5% indicam que concordariam e que utilizariam dados científicos de outros pesquisadores se estivessem acessíveis.

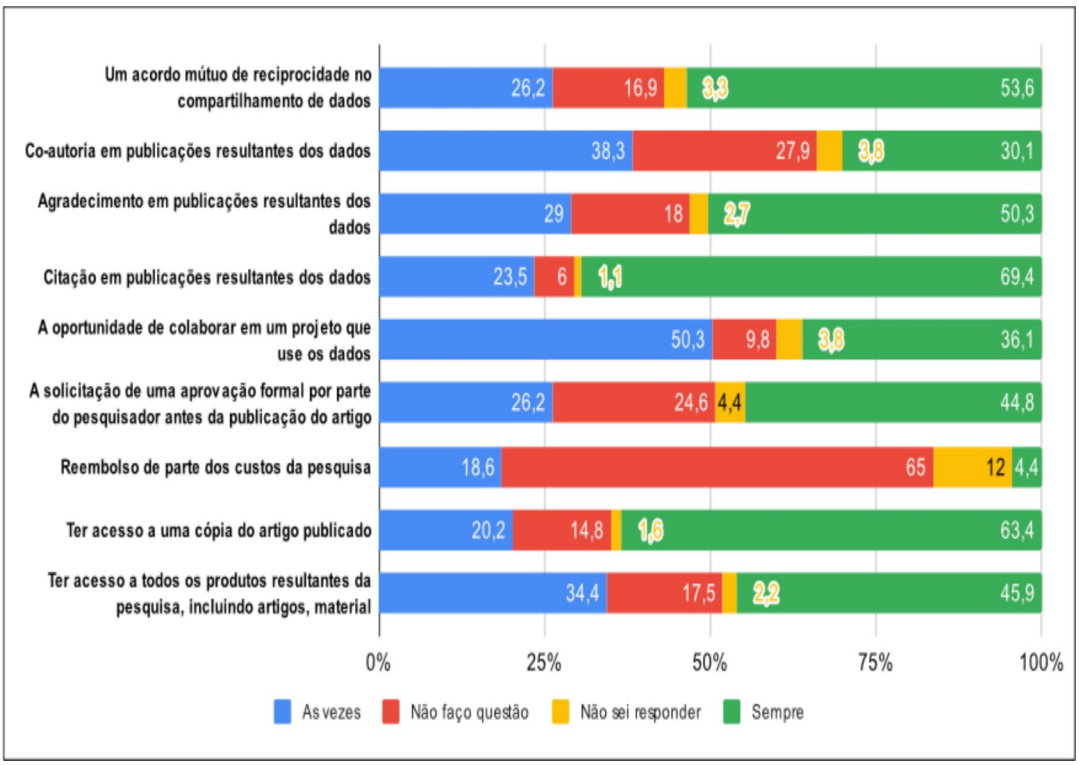
Gráfico 15 - Especifique a sua concordância com cada uma das afirmativas a seguir



Fonte: Elaborado pela autora, 2021.

O Gráfico 16 mostra o que os respondentes almejam ou esperam como retorno quando compartilham seus dados científicos: 69,4% sempre esperam como retorno citação em publicações resultantes dos dados; 63,4% sempre esperam ter acesso à cópia do artigo publicado.

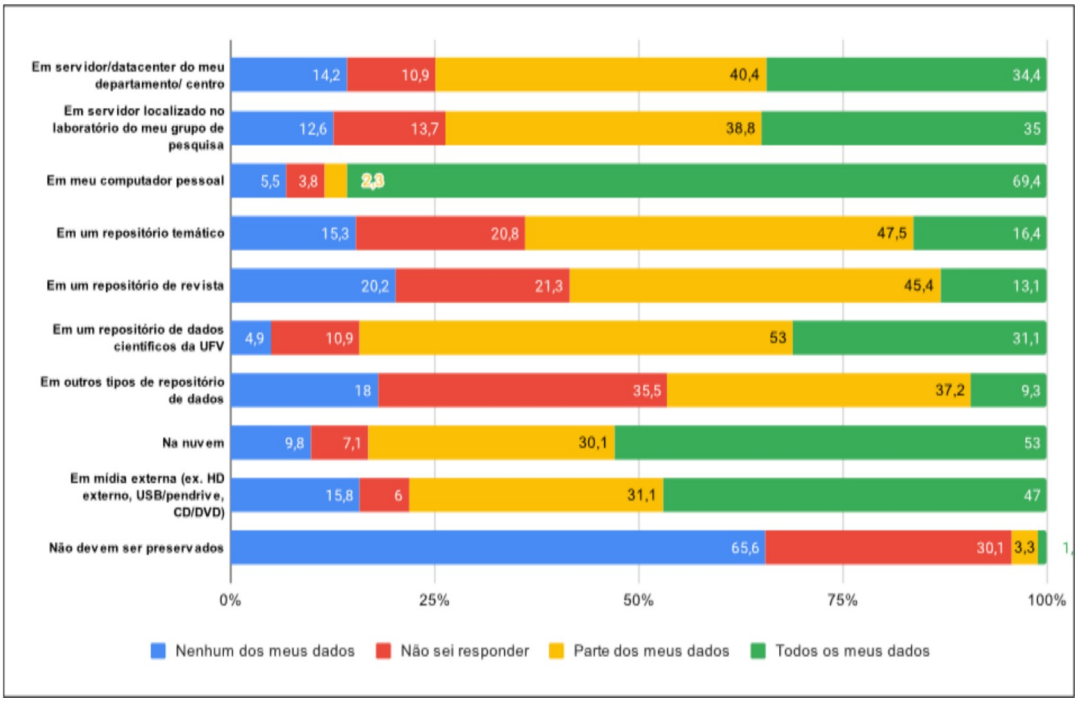
Gráfico 16 - Para outros pesquisadores usarem meus dados de pesquisa, eu espero como retorno:



Fonte: Elaborado pela autora, 2021.

Quando os respondentes são indagados em quais locais seus dados científicos podem ser preservados, em torno de 69,4% julgam que todos os seus dados podem ser preservados em computadores pessoais; 53% consideram que parte dos dados pode ser preservada em um repositório de dados científicos da UFV (Gráfico 17).

Gráfico 17 - Onde você acha que seus dados científicos devem ser preservados?



Fonte: Elaborado pela autora (2021).

5.1.3.2 Análise

A terceira etapa do questionário trata da postura dos respondentes quando se fala em compartilhamento dos seus dados e o uso de dados de outros pesquisadores. Este levantamento pontua que uma parte significativa, em torno de 63,3%, concorda que o compartilhamento dos dados pode ampliar o reconhecimento acadêmico. Essa informação n demonstra que os respondentes têm a percepção de que o compartilhamento de dados pode trazer inúmeros benefícios, tanto para o pesquisador como para a comunidade científica, isto é, de acordo com o levantamento, através do compartilhamento, eles conseguem uma amplitude em quesitos como acessibilidade, visibilidade e confiabilidade, possibilitando, assim, crescimento acadêmico e profissional. Os respondentes também apontam que utilizariam dados científicos de outros pesquisadores se eles estivessem acessíveis, indicando que os pesquisadores da UFV são suscetíveis ao tema, melhor dizendo, um número representativo compreende que o compartilhamento e o uso de dados de outrem são fatores importante no andamento de suas pesquisas.

Entretanto, como exposto anteriormente, os respondentes pouco conhecem sobre um plano de gestão de dados e muitos tampouco elaboraram tal documento, ou seja, o PGD é um dos requisitos para viabilizar o compartilhamento de dados. Contudo, conclui-se que o público em questão não tem informações e conhecimento a respeito, denotando, assim, ser necessário que seja dado aos respondentes um apoio, tanto em ferramentas já disponíveis que já fazem uso do PGD, como também em instrumentos de como elaborá-los, pois já existem diversos tipos de PGDs.

Analisando sob esta perspectiva, esta assistência não pode partir apenas da biblioteca, é imprescindível que os responsáveis por esta atividade se reúnam com os professores e/ou coordenadores dos programas de pós-graduação e façam uma seleção de ferramentas e instrumentos que coadunem com o domínio dos cursos e, neste contexto, atentar também para questões de interesses públicos, princípios éticos e autoria. Conforme visto na Seção 3.7 (A biblioteca universitária no contexto dos dados científicos: desafios e perspectivas), Silva *et al* (2019) enfatizam que os profissionais bibliotecários envolvidos em atividades e serviços de gestão de dados precisam ampliar suas competências para acompanhar as mudanças no âmbito da pesquisa científica, além de compreender o mundo complexo e heterogêneo dos dados de pesquisa, para que possam desenvolver serviços que atendam às necessidades dos pesquisadores, abrangendo as dimensões tecnológicas, legais e éticas.

Com relação aos retornos esperados, quando os respondentes cedem seus dados para outros pesquisadores, eles sempre esperam que sejam citados, isto é, em torno de 69,3%

acham que esta ação pode resultar em visibilidade. Eles também esperam um acordo de reciprocidade, o que supõe e dá vestígios de que existem anseios de cooperação e colaboração entre os respondentes, assim, esse levantamento assinala que os retornos esperados conciliam com o entendimento de Vanz *et al*. (2018) ao afirmarem que a comunidade científica brasileira tem como principais motivações para realizar compartilhamento de dados, visibilidade, impacto, colaboração e reconhecimento.

Conclui-se que na análise desta etapa, foi possível observar que os respondentes opinam que todos os seus dados devem ser preservados em PCs, informação retratada anteriormente no Gráfico 8, quando pontuam que todos os dados são armazenados em computadores pessoais. Essa percepção dos respondentes reitera, mais uma vez, que é fundamental uma intervenção por parte da biblioteca ou dos programas de pós que auxilie na mudança de comportamento dos pesquisadores em questão, porque, como bem salientaram Lima e Farias (2019, p.119), “dados guardados em gavetas e computadores são invisíveis e não causam impacto na comunidade científica”.

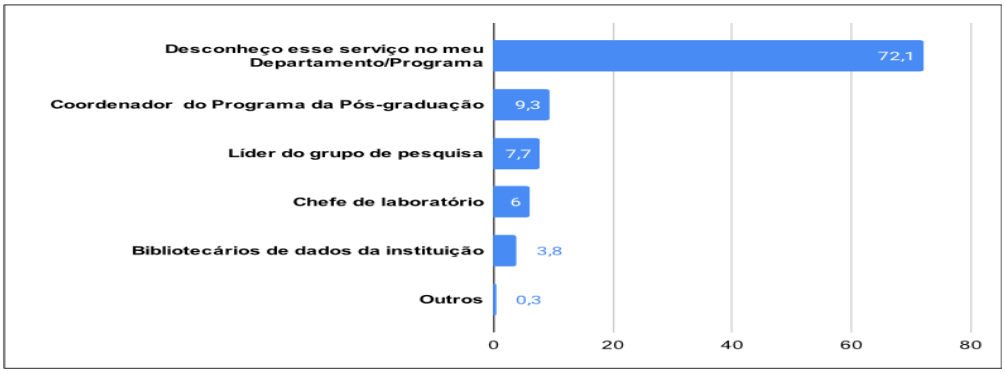
5.1.4 Informações organizacionais

A quarta parte do questionário versa sobre as atividades administrativas que os respondentes possam usufruir no tocante à gestão e à curadoria dos dados científicos dos laboratórios, departamentos, núcleos de pesquisa, entre outros ambientes, isto é, são questões para identificar os serviços e a assistência que os Centro/Departamento/Programa disponibilizam para que os dados científicos sejam processados durante suas pesquisas.

5.1.4.1 Dados recolhidos

Quando questionados de quem é a responsabilidade pelos serviços de gestão de dados científicos no Programa, cerca de 72,1% informaram desconhecer esses serviços, 9,3% afirmaram que fica a cargo do coordenador do Programa e 7,7% a cargo do líder de grupo de pesquisa, como demonstrado no Gráfico 18.

Gráfico 18 – Serviço de gestão de dados científicos no seu Departamento/Programa, quem é o responsável?

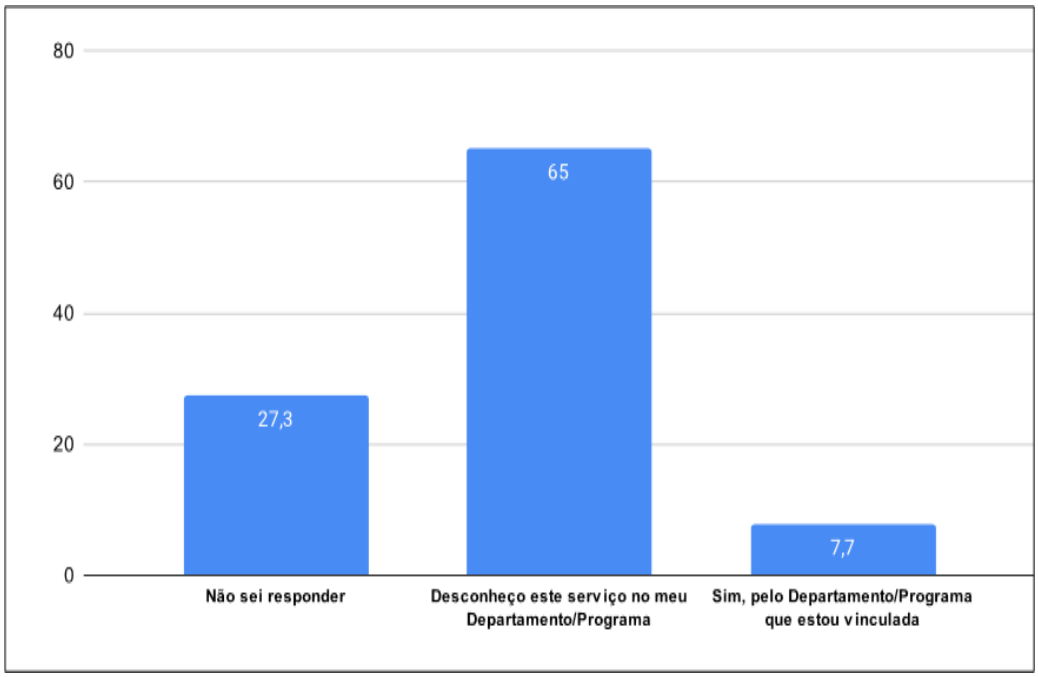


Fonte: Elaborado pela autora, 2021.

Em relação à questão 4.2 do formulário, cerca de 60,7% dos respondentes não souberam dizer se o Programa indica ou utiliza alguma coleção de dados e 33,9% afirmaram que não existe este tipo de serviço e quando respondem que sim, os participantes descrevem o que entendem como coleção de dados, ou seja, BR-CORTE e CQBAL, Periódicos Capes e NCBI, Microdados de pesquisas do IBGE, Coleções de teses e dissertações do repositório do programa e do institucional Locus, Coletas botânicas, Banco de dados gerido pelos orientadores.

Em relação à preservação dos dados produzidos, 65% dos respondentes indicam que desconhecem esse serviço no Programa a que estão vinculados, e cerca de 27,3% não souberam responder (Gráfico 19).

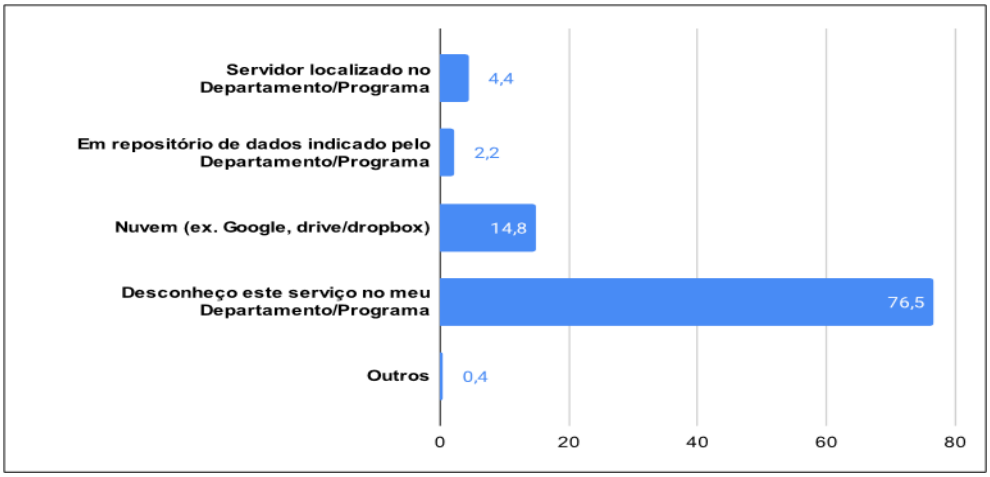
Gráfico 19 - Conhecimento se os dados produzidos pela sua pesquisa são de alguma forma preservados pelo Departamento/Programa



Fonte: Elaborado pela autora, 2021.

Sobre as opções de armazenamento de dados oferecidas pelo Departamento/Programa, o Gráfico 20 mostra que a maioria dos respondentes, em torno de 76,5%, desconhece esse perfil de assistência no Programa; alguns, cerca de 14,8%, responderam que o Programa oferece ou indica alguns armazenamentos em nuvem.

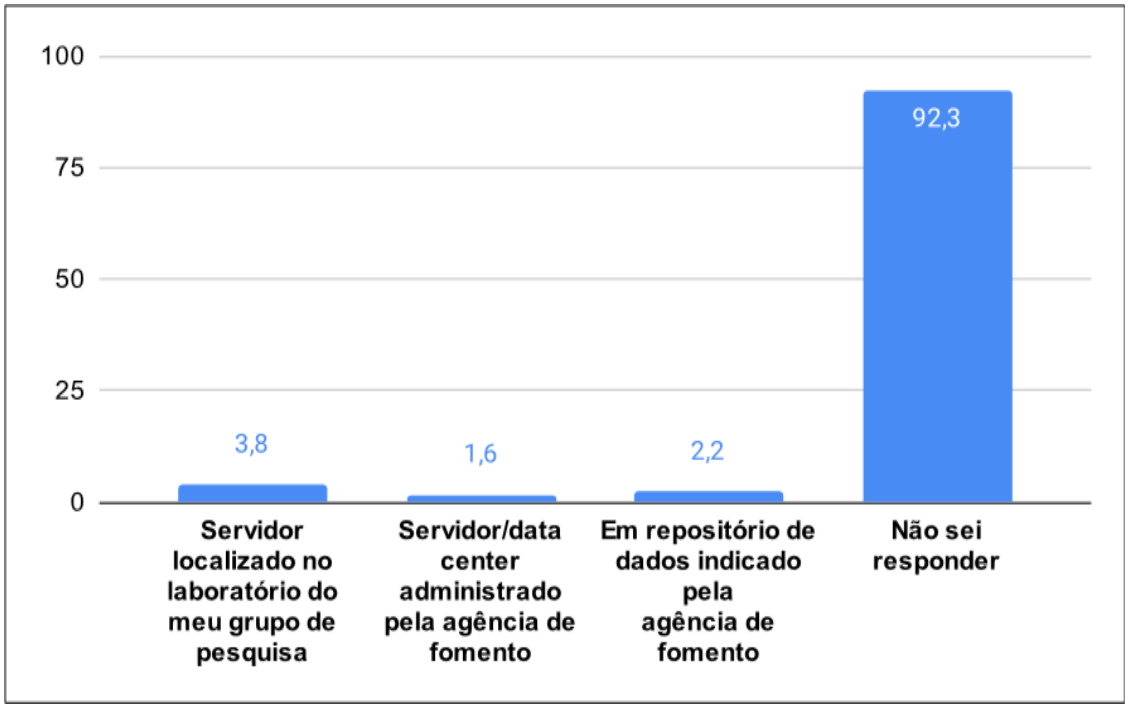
Gráfico 20 - Opções de armazenamento dos seus dados que o Departamento/Programa oferece ou indica



Fonte: Elaborado pela autora, 2021.

Quando questionados se as agências de fomento que atuam no Programa auxiliam na questão do armazenamento dos dados produzidos, 92,3% dos respondentes afirmam não saber informar.

Gráfico 21 – Opções de armazenamento que a agência de fomento oferece ou indica



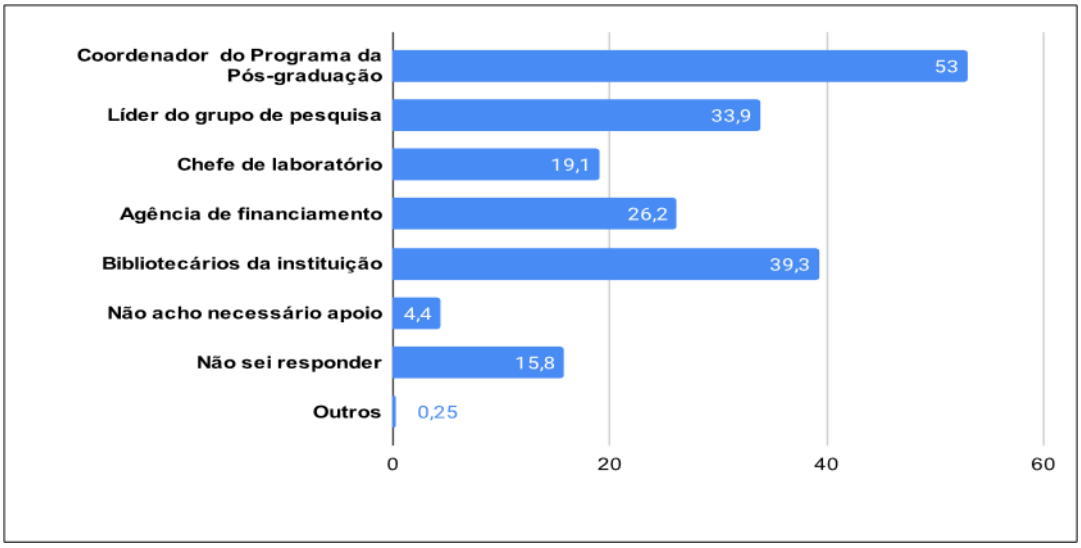
Fonte: Elaborado pela autora, 2021.

Sobre diretrizes fornecidas pelo Departamento/Programa para o uso e compartilhamento de dados científicos, em torno de 92,3% responderam que o Programa não fornece tais diretrizes e 7,7% afirmam que existem diretrizes traçando alguns comentários sobre tal documento e/ou procedimento.

> “Uso as recomendações do Comitê de Ética da UFV”
>
> “Sigo as diretrizes do Programa da Pós-Graduação”
>
> “Oferece parcialmente. Exemplo: após a utilização de um banco de dados no mestrado é obrigatório, para a defesa, que o aluno envie ao professorar os dados com novas variáveis criadas pelo estudante ao seu orientador e que isso seja documentado, assinado e entregue à secretaria do departamento. Além disso, as dissertações e teses ficam protegidas (sem acesso aberto) por um tempo, até que os artigos sejam publicados em revistas”
>
> “Não tenho conhecimento explícito sobre estas diretrizes, entretanto, rotineiramente, dados de experimentos e/ou dados de propriedade intelectual são mantidos sob sigilo até a publicação do artigo.”

Outro aspecto abordado nesta quarta parte do questionário é que os respondentes podiam opinar sobre quem deveria oferecer serviços de apoio à gestão dos dados científicos produzidos nos Programas, cerca de 53% indicam que esta função poderia ficar a cargo da coordenação dos Programas, e em torno de 39,3% afirmam que a biblioteca poderia oferecer esse apoio (Gráfico 22).

Gráfico 22 – Quem você acha que deve oferecer serviços de apoio à gestão dos dados de pesquisa no Departamento/Programa a que você está vinculado?



Fonte: Elaborado pela autora, 2021.

5.1.4.2 Análise

No que tange à assistência na preservação dos dados científicos, em torno de 65% dos respondentes afirmam que desconhecem esse perfil de serviço no departamento, pontuando também que o setor a que estão vinculados não oferece ou indica algum tipo de

opções de armazenamento dos seus dados. Apenas 14,8% informam que o Departamento/Programa indica o armazenamento em nuvem. Neste contexto, quase a maioria dos respondentes sinaliza que não sabem informar se as agências de fomento oferecem ou indicam opções de armazenamento.

Diante dessas informações, observa-se que, em questão de preservação e armazenamento dos dados, os setores analisados estão totalmente desprovidos de procedimentos e ferramentas que possam auxiliar os pesquisadores. Este levantamento vai ao encontro da teoria de Heidorn (2008), descrita na seção 3.4 (Cauda longa da ciência ou “long tail science”), em que ele observa que nos departamento e laboratórios das universidade são desenvolvidos inúmeros projetos de pesquisa, mas os registros dos processos dessas investigações científicas se perdem, porque não existe algum procedimento que faça qualquer tipo de preservação, ou seja, a grande maioria dos dados produzidos nestes ambientes ficam retidos em computadores pessoais.

De modo não muito diferente, foi identificado que cerca de 92,3% revelam que os departamentos/programas não fornecem diretrizes para uso e compartilhamento dos dados científicos. Evidenciando que, neste quesito, os programas também não conseguem dar o apoio necessário a seus pesquisadores, o que contribui de alguma forma com a invisibilidade das etapas realizadas durante as investigações científicas. Isto presume que os trabalhos científicos processados nestes locais não conseguem ser reconhecidos como objetos informacionais de alto valor, porque a invisibilidade dos procedimentos torna os dados científicos dispersos e com poucas oportunidades de serem reusados em outros estudos.

Perante o exposto, percebe-se que os respondentes esperam que a instituição crie formas de amparar os pesquisadores em relação à atividade no que concerne ao manuseio dos dados científicos, melhor dizendo, os respondentes assinalam que necessitam de oferta de orientação em temas como compartilhamento de dados, elaboração de PGDs, utilização de repositórios de dados, armazenamento e preservação dos dados e de outras ramificações que contenham o tema gestão de dados científicos. Nesta direção, Semeler e Pinto (2019), na retro Seção 3.7 (A biblioteca universitária no contexto dos dados científicos: desafios e perspectivas), reforçam que as práticas tradicionais podem ser retomadas com o objetivo de serem aplicadas a conjuntos de dados, por exemplo, a catalogação e a organização de materiais; a preservação e a curadoria; serviços de usuários e de referência; consultoria e treinamentos são repensados para serem aplicados junto às novas práticas, que envolvem o gerenciamento, a curadoria e o compartilhamento de dados. Pelos resultados desta etapa, conclui-se que uma infraestrutura que atenda essa demanda é crucial para o avanço do tema na UFV.

Com isso, a biblioteca tem que repensar sua relação com seus usuários, no caso presente, com os alunos da pós-graduação, pois o levantamento desta quarta parte do questionário sinaliza que a biblioteca tem um grande desafio a enfrentar. Pois é o momento em que as atitudes e iniciativas terão que partir da biblioteca, sendo dela a responsabilidade de reunir todos os atores que podem ser envolvidos na elaboração de atividades que supram a demanda apresentada, ou seja, fazer uma integração entre todos os envolvidos com a pesquisa científica, incluindo coordenadores de pós-graduação, líderes de grupos de pesquisa, professores da pós-graduação. Na Seção 3.7, já mencionada, Santa Anna, Dias e Maculam (2019), citando Costa e Cunha, apontam que o primeiro passo a ser realizado pelas bibliotecas é conhecer a forma de tratamento dos dados nos devidos departamentos e unidades acadêmicas que estão diretamente ligadas à produção das pesquisas. Além disso, as bibliotecas precisam reconhecer outras iniciativas adotadas pela instituição, de modo a consolidar um panorama ou diagnóstico sobre o tratamento de dados em nível institucional. A partir disso, planos emergentes de ação podem ser realizados para que as bibliotecas contribuam com a gestão dos dados.

Conclui-se esta etapa com o entendimento de que, nos requisitos necessários para iniciar atividades em gestão de dados na comunidade científica da UFV, será necessário reconhecer as especificidades de cada programa, reconhecer que é necessário agir em consonância com os anseios dos alunos de pós-graduação e, em paralelo, sempre dialogar com os dirigentes dos cursos.

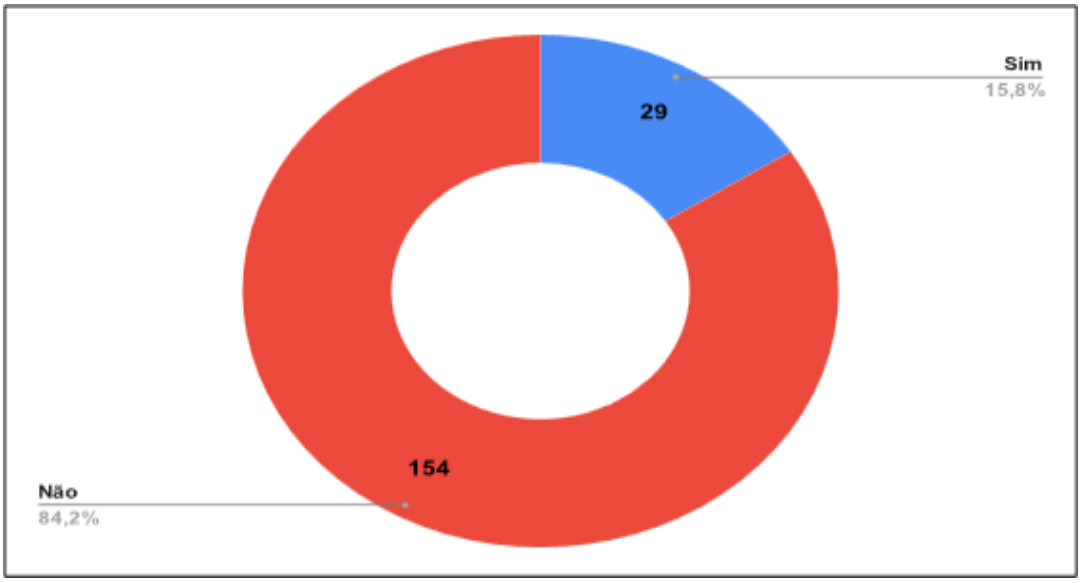
5.2 APRESENTAÇÃO DOS DADOS OBTIDOS ATRAVÉS DAS ENTREVISTAS

Os resultados desta etapa correspondem ao objetivo 3, que foi levantar informações a respeito de como é feita a gestão de dados nos programas de pós-graduação da UFV. Aqui serão analisados os dados recolhidos durante a entrevista, ressaltando que, nesta etapa, optamos por não utilizar apoio de *softwares*, geralmente utilizados para a geração de gráficos, análise de *cluster* e outros recursos de análises. Descartamos esses tipos de sistemas porque o número de consentimentos para participação na entrevista foi muito baixo. Entretanto, foram utilizados alguns gráficos e tabelas advindos do questionário *on-line*, tendo esses recursos sido usados para fundamentar os dados qualitativos recolhidos com a entrevista.

Pontua-se também que esta fase só foi possível em razão do formulário *on-line*, que, ao final do seu preenchimento, identificava o perfil do gestor e, logo após, fornecia opções de consentimento para a participação da entrevista estruturada, no sentido de identificar entre os

respondentes quem atua na UFV como técnico administrativo ou em alguma outra função acadêmica, para que pudesse ser direcionado o convite da entrevista. Assim, foi constatado que 15,8% dos respondentes detinham os perfis necessários, como mostra o Gráfico 23.

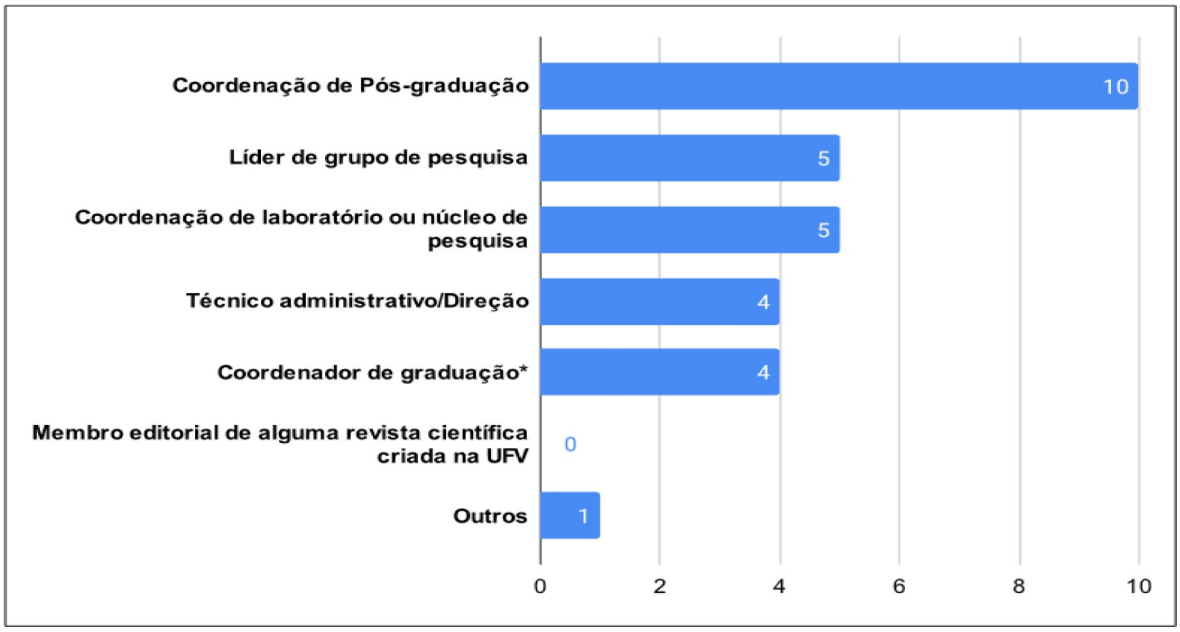
Gráfico 23 – Você exerce alguma função administrativa na UFV?



Fonte: Elaborado pela autora, 2021.

O Gráfico 24 apresenta o resultado do levantamento dos cargos exercidos pelos respondentes do questionário.

Gráfico 24 – Sua função



Fonte: Elaborado pela autora, 2021.

Destaca-se aqui também o baixo retorno de consentimentos para a participação na entrevista. Dos 183, somente 29 respondentes declararam ter alguma função administrativa ou acadêmica e, dentro deste conjunto, 16 funcionários aceitaram conceder a entrevista. No entanto, apenas 7 realizaram a entrevista. O Quadro 6 mostra os Centros e o perfil dos participantes que concederam a entrevista.

Quadro 6 - Perfil dos gestores participantes da entrevista

CENTROS	PERFIL DOS GESTORES	N° de participantes
Centro de Ciências Biológicas e da Saúde (CCB)	Professor e pesquisador	**2**
	Professor e integrante de Comissão de Ensino	
Centro de Ciências Exatas e Tecnológicas (CCE)	Técnico de laboratório	**2**
	Coordenador de programa de pós-graduação	
Centro de Ciências Humanas, Letras e Artes (CCH)	Professor e líder de grupo de pesquisa	**1**
Centro de Ciências Agrárias	Coordenador de programa de pós-graduação	**1**
Pró-reitoria de Extensão e Cultura (PEC)*[19]	Assistente em administração	**1**
	TOTAL	**7**

Fonte: Elaborado pela autora, 2021.

Dado esse levantamento, esta etapa se pautou em analisar e compreender o pequeno grupo que conseguimos reunir, assim, no intuito de entender e apreender o sentido e os significados das falas e dos diálogos dos participantes, buscou-se encontrar indicações que correspondessem aos objetivos desta fase da investigação. Para isso, pautamo-nos no método de análise de conteúdo, utilizando o procedimento de análise categorial, já expostos anteriormente na Seção 2.4 (Procedimentos das análises dos dados) da Metodologia. Segue abaixo a categorização das perguntas/questões criada para auxiliar neste processo.

Q1 Diretrizes de coleta, tratamento, armazenamento e preservação dos dados científicos
Q2 Existência de documentação formal para política de dados científicos
Q3 Indicação de repositórios de dados
Q4 Obrigações e incentivos ao compartilhamento de dados científicos
Q5 Ética, privacidade, propriedade intelectual e *copyright*
Q6 Atuações das agências de fomento: planos de gestão de dados
Q7 Atuações das agências de fomento: depósito dos dados em repositórios
Q8 Atuações das agências de fomento: responsabilidade da gestão dos dados
Q9 Política de dados para a instituição

[19] É o órgão responsável por coordenar, estimular e compatibilizar as atividades de extensão e cultura desenvolvidas pelas diversas unidades da UFV

No primeiro momento da entrevista, foi identificado o perfil do respondente e seu local de atuação, como já apresentado no Quadro 6. Através da primeira pergunta, o participante podia se apresentar e, ao mesmo tempo, a pesquisadora contextualizava o objetivo da entrevista e como seriam os procedimentos da interlocução.

A análise de coleta de dados deste instrumento consistiu em identificar nas mensagens dos respondentes detalhes que pudessem fornecer informações sobre os procedimentos e práticas exercidos nos programas de pós-graduação, no que refere à gestão de dados científicos. Desta forma, nos quadros a seguir, são descritos os trechos mais relevantes da entrevista de cada um dos participantes.

Q1 Diretrizes de coleta, tratamento, armazenamento e preservação dos dados científicos

No Quadro 7, nota-se que não foram identificados documentos, ferramentas ou procedimentos que pudessem ser citados pelos respondentes ou que indicassem alguns resquícios que atendessem ao objetivo da pergunta. As informações contidas neste quadro coadunam com os dados do Gráfico 20, em que 65% declaram desconhecer esse serviço no Programa e 27% não sabem responder. O Gráfico 8 também informa que 77,6% não utilizam servidor ou *data center* do Centro ou programa para o armazenamento dos dados produzidos.

Quadro 7 - Diretrizes de coleta, tratamento, armazenamento, e preservação dos dados científicos

Centros	Respostas
CCB	“Cada vez mais as instituições de financiamento estão exigindo e nos projetos que eu tô coordenando elas estão exigindo uma publicação e disponibilização dos dados.”
	“Não, que eu saiba não tem nenhuma diretriz para todos. Tem a diretriz do programa de pós-graduação que envolve o armazenamento de tese e está mais focado em cada pesquisa específica dos alunos no nível de pós-graduação, mas no departamento, entre os docentes, não conheço nenhuma diretriz que oriente."
CCH	“Olha, na área do Direito, o armazenamento de dados é uma realidade muito rara… Mas agora, com relação aos meus alunos que eu oriento em pesquisas, a gente sempre tem a ideia de manter pastas compartilhadas em nuvem para o armazenamento de dados das pesquisas, dos protocolos de pesquisa”
CCA	“Infelizmente não! … nunca vi uma política no departamento que se preocupasse com o armazenamento dos dados, entendo cada professor é responsável pelos dados gerados em suas pesquisas.”
CCE	“Nós não temos nenhuma coisa estruturada para armazenar isso, para a memória do programa, exceto nas próprias teses e dissertações que vão ter os apêndices cheios dos resultados brutos… cada orientador fica responsável por isso”.
	“Se tiver, eu desconheço.”
PEC	“Eu lido mais com sistema de registro de atividades de extensão, né? Que é o RAEX (Registro de Atividade de Extensão), um sistema que registra as atividades de extensão.”

Fonte: Elaborado pela autora, 2021.

Q2 Existência de documentação formal para política de dados científicos

No Quadro 8, percebe-se que os participantes da entrevista pontuam algo em torno da autonomia dos professores e gestores em relação a este tema. Foi observado também que não existe alguma regra estabelecida, indicando, assim, a inexistência de algo formal.

Quadro 8 - Existência de documentação formal para política de dados científicos

Centros	Respostas
CCB	"Existem compromissos com coleta de material biológico e eu preciso fazer relatórios regulares do que foi coletado… mas não há, não estou muito ciente do que há de formal nesse sentido, acho que é pouca coisa."
	"Segue as normas do curso, que eu saiba são omissas quanto às questões de uso de dados, não tem uma regra explícita, muito clara … Cada um faz de maneira autônoma o registro dos seus dados."
CCH	"Não, não há. Inclusive eu aceitei em participar da pesquisa para me informar mais a respeito do assunto, pois isso me chamou bastante atenção."
CCA	"Não existe, é cada professor no seu quadrado. Tô lembrando aqui que no seu questionário, há pergunta sobre armazenamento. Eu não faço armazenamento na nuvem porque não confio. As minhas formas de armazenamento são em HD externo, backup, meus dados são armazenados nestes sistemas."
CCE	"Não, isso não existe. Cada orientador, cada pesquisador, cuida dos seus próprios resultados de dados."
	"Para mim, nunca foi apresentado nada não. Eu desconheço."
PEC	"Eu desconheço algum documento neste sentido."

Fonte: Elaborado pela autora, 2021.

Q3 Indicação de repositórios de dados

Nota-se no Quadro 9 que este perfil de serviço não é oferecido aos discentes, entretanto, no trecho "alguns projetos estão usando o repositório e não vou lembrar qual, né? Por exemplo, um promovido pela Fapesp que está mais avançada nisso … Que é um repositório de dados sobre espécies biológicas, que eu não lembro o nome", há algum indício de que existe algo neste sentido, mas nada que indique com mais firmeza a informação concedida. E os dados obtidos através do questionário afirmam que 60,7% dos Programas não indicam nem ofertam algum tipo de coleção de dados ou repositórios de dados.

Quadro 9 – Indicação de repositórios de dados

Centros	Respostas
CCB	"Não, existe uma tendência, por exemplo, alguns projetos estão usando o repositório e não vou lembrar qual, né? Por exemplo, um promovido pela Fapesp que está mais avançada nisso …

	Que é um repositório de dados sobre espécies biológicas, que eu não lembro o nome”
	“Não, o meu trabalho específico é muito com coleção, então todo o dado que está relacionado à coleção fica sendo documentado no livro de tombo e isso é atualizado no trabalho rotineiro com a coleção e sempre está disponível e não fica aberto na internet, porque não vejo que há uma vontade institucional, não [estimulam] isso, não têm esse desejo aparente, mas também não impedem que a gente faça se quiser. Mas não tem uma orientação geral quanto a isso.”
CCH	“Não, não tenho conhecimento de indicação de repositórios de dados.”
CCA	“Cada professor cuida do seu. Então, assim, em termos de programa, em termos de departamento, em termos de instituição não existe uma política, quer dizer, eu desconheço, no nosso departamento não tem. O nosso departamento é diferente, um exemplo é o departamento de Fitotecnia, que mexe com patente, que mexe com outras gerações e conjuntos de dados de pesquisa, eu não sei como é feita essa questão lá. O nosso caso é diferente, a gente trabalha muito com dados econométricos, o pessoal da economia principalmente. Já o pessoal da extensão gera muitos dados a partir de entrevista, a partir de pesquisa, ou seja, cada professor que gerencia seus dados”
CCE	“Nosso departamento não tem isso não… O departamento não tem nenhuma provisão para isso.“
	“Nunca me foi apresentado nada neste sentido, eu sempre armazenei meus dados no meu computador, uso do *google drive* também, faço *backup* em HD. Não conheço nada institucional neste sentido.”
PEC	“Não temos instrução neste sentido. Nem um documento que eu conheça. Desconheço esse serviço, infelizmente, fiquei até curiosa.”

Fonte: Elaborado pela autora (2021).

Q4 Obrigações e incentivos a compartilhamento de dados científicos

Os respondentes colocam que este incentivo ou obrigação para que haja um compartilhamento de dados não existe nos programas. Entretanto, alguns entrevistados falam que a PPG já indicou alguma ferramenta *open access*, falam também de algumas agências de fomento que pedem o compartilhamento dos dados de seus projetos em algum órgão, como o Ministério do Meio Ambiente. Pontuam também que os setores podem ter alguma dificuldade em oferecer este serviço em razão da diversidade de dados que existe, até mesmo dentro de um mesmo departamento. Mas não há elemento que indique que em algum programa haja algo estabelecido nesta direção. Um indicativo que evidencia esta questão é que os respondentes do questionário também afirmam que 92,3% dos programas não têm diretrizes que auxiliem em questões de uso e compartilhamento de dados (Quadro 10).

Quadro 10 - Obrigações e incentivos a compartilhamento de dados científicos

Centros	Respostas
CCB	“Dos órgãos financiadores e das revistas, do departamento não. Aliás, justamente o departamento e a instituição têm, antes de mais nada, respeito, me parece, pela independência e pela liberdade

	de atuação do cientista, que é fundamental para o progresso da ciência.”
	“Não existe nenhum incentivo para isso pra ter um banco comum. Até mesmo porque são também pesquisas muito diferentes, tem uma abrangência muito grande no departamento de pesquisas. A minha pesquisa, por exemplo, não tem muita relação com a de outros docentes e isso é muito comum, é um departamento múltiplo, tem várias frentes de trabalho com biologia animal e alguns nem diretamente com biologia animal.”
CCH	“Nada, não há incentivo nem obrigação”.
CCA	“Não existe nenhuma obrigação e não existe nenhuma política nesse sentido, isso não existe. O que existe é quando um professor consegue captar recursos de uma agência [para um projeto] … e esses dados que foram gerados e disponibilizados para o órgão que contratou o departamento. Foi algo que não foi disponibilizado para a comunidade científica, foi disponibilizado para o Ministério do Meio Ambiente.”
CCE	“Tem sim, a própria PPG já falou sobre esse *open access* de como é interessante ter realmente os nossos dados lá para o mundo… Ele é [o dado] liberado na hora que você publica o artigo, né? Então isso eu acho que entre as publicações do departamento, teses e dissertações, melhor dizendo, do programa e de artigos publicados, aí você tem os dados suplementares que vão com os artigos. Essas são as formas como a gente dissemina nossos dados.”
	“Quando a gente vai publicar, a gente segue as normas da revista, tem revistas que até pedem uma base de dados complementar. É o relacionamento meu com a revista, agora em relação a alguma demanda institucional, nunca chegou até a mim.”
PEC	“Não é feita essa exigência ao público que a gente atende não, que são estudantes que trabalham com extensão … A única coisa que a gente exige, que os editais exigem, é a apresentação dos trabalhos no SIA, Eles têm que apresentar os resumos no SIA, que podem ser final ou pode ser informações sobre o andamento do projeto, podendo apresentar os dados parciais que ele tiver, a princípio, é isso que é exigido.”

Fonte: Elaborado pela autora, 2021.

Q5 Ética, privacidade, propriedade intelectual e *copyright*

Diante das respostas transcritas no Quadro 11, verifica-se que o CEP é um órgão que rege toda a estrutura das pesquisas dos programas de pós-graduação, como também direciona e estabelece algumas regras que delineiam pesquisas em seres humanos. Já em relação à Comissão Permanente de Propriedade Intelectual, que é o órgão que promove a disseminação da cultura de proteção à propriedade intelectual, o assunto não foi muito abordado.

Quadro 11 - Ética, privacidade, propriedade intelectual e *copyright*

Centros	Respostas
CCB	“Se eu tenho pesquisa com dados com seres humanos, ela tem que ser aprovada pelo comitê de ética em pesquisa com seres humanos … Eu não trabalhei com registro de patente, sempre há essa possibilidade no futuro, distante ou próximo. Mas a minha área é mais de ciência básica, mas também aí a cobrança da ética é muito maior, de dar isso, é uma coisa que está aprendendo dar um retorno à comunidade do que está sendo feito, quais os resultados e quais as implicações para a vida cotidiana.”
	“A gente segue orientações da PPG que não força, a prerrogativa é nossa de liberar imediatamente os dados ou não. Eu imagino que é mais comum não liberar os dados, porque

	são muitos trabalhos que rendem resultados originais e não é imediata a publicação. Então é comum fazer uso de ferramenta de manter sigilo por um tempo … Não tem um coisa única no departamento, mas isso envolve muito a parte de publicação em geral. A gente cede esse *copyright*, a revista quando publica, partilha esse *copyright*. Não conheço muitos registros de patente dentro do nosso departamento, é mais copyright mesmo de dados de pesquisas originais que são publicados em uma revista e esse *copyright* fica partilhado com as revistas para viabilizar as publicações. E não tem também, não conheço nenhuma orientação geral do departamento para seguir o mesmo. Acho que é sempre resolvido mais caso a caso, cada docente resolve dentro da natureza de sua pesquisa.”
CCH	“O que eu sei é que há uma preocupação a mais, eu acho que é mais exacerbada entre os pesquisadores em direito, porque a gente conhece a legislação a respeito, né? Da proteção de dados, proteção da propriedade intelectual e tal. Então hoje tem esse conhecimento, a gente tem até mais receio, né, dessas bases e desses repositórios de dados. Então funciona como que a gente sabe que é perigoso, a gente não utiliza.”
CCA	“Quando os dados são gerados a partir de entrevistas, como você está fazendo, ele tem que seguir um ritual, ele tem que seguir um ritual que o comitê estabelece. E quem é o guardião desses dados? É o professor responsável por aquela pesquisa. Não é o departamento, não é o programa de pós-graduação, é o professor responsável pela submissão do projeto ao Comitê de ética, ele que é o guardião do dado... Dependendo [da] defesa da tese ou da dissertação aqui na UFV, tem uma pergunta lá, é para verificar se a dissertação é sigilosa ou se ela é aberta … Nesses 11 anos que eu tô aí no programa e tal, eu só participei de uma defesa cujos dados das informações geradas foram sigilosas, elas não tiveram acesso ao público … o estudante exigiu que isso fosse sigiloso com direitos autorais, entendeu?”
CCE	“Sim, tem gente que trabalha com questionários. Pesquisas com pessoas têm que ter o consentimento. Tem que ter o comitê de ética, né? E com uso de animais, fizemos alguns trabalhos com peixes, essas coisas aí temos que ter realmente o comitê de ética ou com pessoas ou com usos de animais. Isso faz parte dos registros dos projetos, então, nesses casos, tem que manter sigilo. O *copyright* não seria tão aplicado, dependendo, porque tem alguns projetos de gerar mapas ou gerar coisas que podem ser sim. Com certeza, tem áreas que trabalham muito mais em cooperação com empresas ou com diferentes órgãos do governo. E aí tem que fazer quando realmente eles estão buscando essa propriedade intelectual entre a universidade e essas outras instituições, com certeza, tem essas questões”
	“Na época do meu mestrado, o meu projeto teve que passar pelo comitê de ética, teve que ser avaliado. Isso era uma preocupação que o meu antigo orientador tinha. Porque quando a gente foi fazer um estudo no caso, a gente aplicou um questionário, entrei em contato previamente com o comitê de ética para submeter a pesquisa. No meu caso, diante da natureza das perguntas, eu tinha que criar um TCLE, tive que submeter o projeto na plataforma Brasil. Essa preocupação, a instituição tem. …Agora, em relação à propriedade intelectual eu não sei informar, mas eu imagino que tenha, porque às vezes tem desenvolvimento de produtos, de processos, mas eu não sei dar maiores detalhes.”
PEC	“A gente procura respeitar a questão da propriedade intelectual … através do portal transparência, chegam pedidos de acesso a documentos, e aí o que é público no Raex a gente libera os documentos. O que não é e que vai depender de propriedade intelectual, a gente precisa da autorização do professor que escreveu, para perguntar se pode mandar os dados daquele projeto para o usuário externo. Esses casos que acontecem, a gente não libera sem autorização do autor do projeto. Dados da escrita do projeto, mesmo os relatórios públicos, a gente libera.”

Fonte: Elaborado pela autora, 2021.

Q6 Atuações das agências de fomento: planos de gestão de dados

Não foi possível identificar, nas falas do Quadro 12, traços que correspondessem ao objetivo da pergunta. Alguns entrevistados assinalaram que trabalham diretamente com agências de fomentos, entretanto, esses órgãos nunca solicitaram algum documento semelhante ao abordado. Quando outros participantes mencionam, não é possível reconhecer se realmente se está falando sobre o PGD ou outros tipos de relatórios. Assessorando nesta direção, o Gráfico 14 indica que 85% nunca elaboraram um PGD, fortalecendo que os programas desconhecem tal documento.

Quadro 12 - Atuações das agências de fomento: planos de gestão de dados

Centros	Respostas
CCB	“Cada vez mais as instituições de financiamento estão exigindo e nos projetos que eu tô coordenando elas estão exigindo uma publicação e disponibilização dos dados. Como? Onde? Isso ainda não existe, tá? Então é um dos desafios que estão sendo cobrados, inclusive, mas até agora não foi solucionado.”
	“Depende do edital e do projeto específico ... os dados foram disponibilizados com a quantidade de informações que eles queriam, o que não era muita informação. Eram alguns dados básicos do acervo como abrangência de espécies ... esse tipo de coisa. Isso foi disponibilizado na época por meio de relatórios para essas agências.”
CCH	“Não, a gente, na verdade, trabalha em parceria mais com as agências de fomento da UFV, como a Funarbe, tem uma outra fundação de fomento, só que esqueci o nome. Que eu saiba, eu nunca vi essa exigência, assim, das pesquisas que eu já conduzi, tanto CNPQ com bolsa, com a Fapemig ou mesmo na Funarbe, nunca vi essa exigência, nem nos editais, nunca vi.”
CCA	“Eu desconheço, para mim esse tema que você traz é novo. E eu posso falar com uma certa segurança porque fui coordenador do programa por quatro anos e meio. Estou à frente do programa e nunca vi nenhuma agência, o que as agências fazem é exigir crédito, né? Uma pesquisa financiada pela Fapemig na hora de publicar um artigo, que dê crédito de pesquisa financiada pela Fapemig ou pela Capes, mas nenhuma delas, nunca tocou nesse assunto de gerenciamento de dados.”
CCE	“Por enquanto eu nunca tive exigência ainda, pode ser que esteja por vir. Eu participo das reuniões com a Capes, com os coordenadores, quando se está solicitando bolsa da Fapemig ou do CNPq, projetos que têm que submeter nunca têm esse quesito para abordar.”
	“Olha, eu não sei informar, assim, o contato com as agências de fomento geralmente é o orientador que faz. Eu particularmente não me lembro de ter preenchido algum documentação quanto a isso. Projetos, essas coisas é o pessoal da secretaria ou o pessoal do orientador.”
PEC	“Olha, vou te falar que tivemos uma discussão recente sobre isso. Um dos servidores do setor, ele está com esta parte. Eu não estou muito por dentro. Parece que é uma exigência de fora, né? Tem uma pessoa que já está olhando isso, adaptando uns documentos internos da divisão para atender a essa instrução. Não sei em que pé está, mas tem alguém fazendo, sim.”

Fonte: Elaborado pela autora, 2021.

Q7 Atuações das agências de fomento: depósito dos dados em repositórios

Sobre esta questão, é possível identificar, de acordo com as afirmações dos respondentes no Quadro 13, que existe depósito de dados científicos em algum projeto de pesquisa, enquanto em outros, só é exigido o trabalho final como a tese ou dissertação, ou alguns relatórios administrativos que são utilizados na prestação de contas.

Quadro 13 - Atuações das agências de fomento: depósito dos dados em repositórios

Centros	Respostas
CCB	"Às vezes o *Github*, por exemplo, eu já vi indicarem. Mas geralmente a própria revista disponibiliza um link para os dados, como que eles põem não sei, mas geralmente está associado à publicação."
	"O que eu lidei diretamente como responsável com esse tipo de informação foi em atividade de campo docente, que eram disciplinas de levar alunos pro campo. A ideia da disciplina era ensinar técnicas de coleta de vertebrados com os alunos, a gente capturava o organismo e soltava. E isso tinha uma licença específica associada e ela, tanto no âmbito federal como estadual e direto com o Parque Estadual do Rio Doce, a gente, ao final dessas atividades, eu sempre enviava um relatório dando todos os dados de todos os animais capturados e observados e partilhava com essas instâncias que controlam, esses dados foram disponibilizados diretamente para eles por meio de relatórios … então eles pediam que desse os dados para a base de dados deles, em geral não tem essa exigência."
CCH	"Não, assim, não na área do Direito, né? Eu só conduzi pesquisas na área do direito, mesmo não sei de outras áreas."
CCA	"Não, a Capes tem por exigência que toda tese concluída eu tenho que colocar no Sucupira, a tese e a dissertação, mas os dados em si, brutos, não."
CCE	"Nem sei se eles exigem, só nessas questões quando se você vai usar o programa de identificação de bioinformática. Aí tem que depositar em tal lugar para você depois rodar programa, identificar similaridades com tal espécie, com tal gênero, essa é a única que eu sei que se transfere os dados. Mas, mais uma vez, eu não tenho certeza na área, por exemplo, de informações espaciais, eles utilizam muito fotos e mapas que eu não sei como que isso funciona, se eles têm que transferir as imagens pra ter essas imagens de satélite. Eu não sei como isso funciona. Eu acho que se pode fazer essa doação ou essa transferência de dados. Mas eu não sei se é uma exigência também não."
	"Olha, os contatos que eu tive com as agências de fomento, foi na época do mestrado, em questão de bolsa e no começo do doutorado também, mas quando passei no concurso, cortou. Eu não sei informar, para mim nunca pediram. Ou o orientador também não quis passar, eu não sei se tem demanda disso."
PEC	"Olha só, a princípio, as bolsas do meu setor de extensão são financiadas com recurso interno e a gente tinha também financiado pela Funarbe e ela pede relatórios, todos eles fazem relatórios, mas os relatórios vão ser usados internamente só. Atualmente, nós estamos com dois editais, um com o Ministério Público de Trabalho de Ouro Preto, que tem uma verba da Vale pelo rompimento da barragem, ela aplicou uma multa, e o dinheiro da multa está sendo usado para financiar projetos de extensão para recuperar aquela área ou minimizar os impactos dos problemas da mineração daquela área. Então, esses recursos têm projetos nossos da UFV e aí a gente precisa mandar um relatório com todos os esquemas que eles pedem. A gente manda para PPO e ela encaminha para o Ministério do Trabalho, aí, no caso, para a Vara de Ouro Preto. E em 2016 foi encerrado um programa do MEC que chamava ProExT, é um programa de extensão do MEC, mas a última edição foi em 2016, nele também ele tinha que fazer um relatório no final dos projetos para o

	MEC, para o pessoal do ProExT. Esses são os órgãos financiadores com que a gente trabalha. Esses relatórios têm um padrão exigido pelos financiadores, como nome, equipe, o público atendido pelo projeto, prestação de contas, as atividades que foram desenvolvidas. Essas informações variam um pouco de exigências de um financiador por outro. Mas, basicamente é isso. Perfil da comunidade atendida, por aí, quais metodologias utilizadas, se foi curso, ou algum treinamento.”

Fonte: Elaborado pela autora, 2021.

Q8 Atuações das agências de fomento: responsabilidade da gestão dos dados

No Quadro 14, as opiniões divergem entre si, porém alguns consideram que seria vantajoso que houvesse uma gestão colaborativa entre a universidade e as agências de fomento.

Quadro 14 - Atuações das agências de fomento: responsabilidade pela gestão dos dados

Centros	Respostas
CCB	“Neste sentido, elas já estão colaborando, a partir dessa chamada CNPQ, Ana, não sei o que era chamada 10, de 2016, eu acho. Começou a estabelecer seminários integrados de acompanhamento dos projetos. Então, foi uma chamada direcionada de várias linhas de pesquisa com um foco, que era a bacia do Rio Doce, e o que acontece, cada projeto, cada rede de pesquisa expõe os seus resultados e o andamento para todo mundo. Isso dá uma transparência muito grande.”
	“Eu acho que seria interessante ter na própria universidade, imediatamente me disponibilizei para esta entrevista que achei muito relevante essa pesquisa. Finalmente, eu estou na UFV desde 2011, nunca tinha ouvido nenhum interesse institucional claro nesse sentido. Eu acho que você tem uma esfera institucional dentro da Universidade que quem retém o fomento vem das agências, mas a infraestrutura é da Universidade, então é um fomentador também de pesquisa e já que ela é a instituição-mãe onde está sendo gerado aquele conhecimento, ela tem um canal específico para isso.”
CCH	“Eu não sei a quem caberia, sabe, essa questão da gestão, mas que deveria ser uma exigência sim, né? Eu acho que deveria, mas para falar a verdade eu não sei quem seria melhor ou mais habilitado para fazer esse tipo de prestação de serviço, sabe? Não sei se as agências de fomento ou mesmo a biblioteca da UFV, não sei, sinceramente não tenho base para te dar essa informação, essa opinião.”
CCA	“Eu acho que coordenação não tem que fazer gestão de dados, agências, também eu não sei. Eu acho que quem tem que fazer a gestão dos dados tem que ser o responsável pela geração dos dados. Eu acho que as agências e a universidade, por exemplo, ela pode pautar uma política normativa de como os dados deveriam ser armazenados, guardados. Mas quem tem que ser responsável tem que ser um indivíduo … não vai oferecer um funcionário por departamento para ficar responsável por isso. Então eu acho que é a instituição, ela deveria criar uma normativa e cada professor responsável pela geração desses dados que abastece o sistema, entendeu?”
CCE	“Eu acho que tem que ser descentralizado ao máximo. Porque as agências de fomento não têm mão de obra, nem estrutura para esse tipo de coisa. Eu acho que é muito melhor você saber quem tem esses dados que nós pagamos para gerar. E aí pode ter acesso sim, quando precisar. Mas eu

	acho que quem teve a ideia ou o projeto, quem realmente gerou os dados deve realmente ser o responsável por guardar e, lógico, para disponibilizar a quem realmente financiou. Mas ter um repositório geral disso, eu não vejo sentido.”
PEC	“A gente encaminha [para as agências de fomento] os dados individuais que os professores e os bolsistas dos projetos mandam. E o que ela faz com isso? a gente não tem acesso a isso depois que vai para lá. A gente não tem esse contato. Eu imagino que seja para comprovar a prestação de conta.”

Fonte: Elaborado pela autora, 2021.

Q9 Política de dados para a instituição

Na última questão, as falas são quase unânimes sobre a relevância de se elaborar uma política de dados para a instituição, mesmo que a concepção de cada participante divirja quanto aos elementos necessários para formulá-la (Quadro 15).

Quadro 15 - Política de dados para a instituição

Centros	Respostas
CCB	“Seria muito bom, seria muito útil, se nós tivéssemos uma assessoria competente e experiente para gestão de dados, para preparação de material de popularização ou popularização da ciência, tudo isso seria maravilhoso! … Se for possível, seria interessante, eu acho, mas tem que tomar muita cautela, a construção dessa política tem que ser de baixo para cima, né? E isso não é trivial, mas não é impossível. Então essa política se for, tem que ser construída com muita cautela e princípios éticos, provavelmente ela tem que ser construída a partir de experiências e não de experimentos.”
	“Acho que sim, é importante, principalmente com o pessoal da sua área. Tem treinamento próprio para isso, para gerenciar os metadados, tem que ter um suporte de tecnologia para que esses dados sejam seguros também. Controlar o acesso dentro do que a política institucional decide que deve ser feito. Acho que não deve ser uma decisão de um pesquisador específico, mas um comitê decide as coisas e norteia o que a gente disponibiliza. Mas eu acho que dado disponível, quanto mais dados disponíveis, melhor, não sei por que esconder dados de pesquisa, não tem sentido isso.”
CCH	“Eu acho que sim, eu acho que é uma condição necessária para o próprio desenvolvimento da ciência. Se não há isso, vai acontecer o que acontece nas pesquisas em Direito, né? Os pesquisadores não sabem o que os pesquisadores antes dele fizeram, então, às vezes, pesquisam as mesmas coisas e chegam aos mesmos resultados. Você não vê desenvolvimento, isso não é desenvolvimento da ciência. E é uma realidade muito presente nas pesquisas em Direito, hoje.”
CCA	“Eu acho que sim … Eu acho que o que precisa é ter uma política talvez na instituição de como armazenar esses dados, como esses dados estão em concordância com a legislação nacional, entendeu? Eu acho que talvez uma diretriz nesse sentido é interessante, mas agora tô falando que a responsabilidade da gestão tem que ser compartilhada, não dá para ser centralizada … A universidade pode até falar: “Olha, tem aqui dentro da PPG, um funcionário responsável pelo armazenamento de dados, mas quem vai ficar responsável em abastecer isso, não é o programa, não é o departamento, é o pesquisador.”
CCE	“Eu acho que ela poderia pelo menos discutir isso e dar tipo orientações, diretrizes em quais áreas é importante e como que isso deveria ser feito. Mas, mais uma vez, é muito diverso. Se dentro de um programa de pós-graduação se tem essa diversidade enorme, imagina entre os

	diferentes centros da ciência! Eu entendo que tem coisa, quando se está trabalhando com ser humano ou animal aí pode ter que ter um tratamento diferenciado do que quando você tá fazendo uma reação química na bancada, sabe? O tipo de resultado que vai ser gerado e que tem que ser feito com isso. Mas todos merecem atenção, isso eu acho que eu e meus colegas também somos muito crus nisso, nem fazem ideia do que poderia ser feito, porque com certeza seria interessante, até para mostrar o que nós estamos fazendo, isso que eu entendo que quando você lança seus dados em um repositório, várias pessoas podem visualizar. É uma coisa bem valiosa, podem enriquecer a pesquisa. Através da disponibilização dos dados, é possível criar outras discussões científicas.”
	“Eu acho que sim, deveria ter um cenário mais macro, a gente consegue saber do que foi publicado, de trabalhos que foram feitos. A gente não sabe do que está em andamento, do que tem de dados. Uma dificuldade que eu tive, talvez seria facilitado, não só no âmbito da UFV, mas nacional, talvez até internacional. Eu estou fazendo uma metanálise, uma revisão sistemática com metanálise. Então para eu poder utilizar os dados, eu tive que tirar dos artigos, só que em muitos artigos que tinham dados, eu não conseguia processar. Eu entrei em contato com vários autores e eles não responderam, sei lá, porque não quiseram ou não tiveram interesse. Então de certa forma essa limitação restringiu um pouco o meu estudo. Eu podia ter uma robustez estatística maior de representatividade e de significância do meu estudo, abranger mais o estudo, mais dados. Mas o fato de eu ser limitado ao acesso aos dados, eu consegui fazer, então essa disponibilização de dados seria interessante. Mas por outro lado, tem a questão de sigilo da pesquisa, teria que encontrar o meio termo nisso.”
PEC	“Eu penso que sim, eu penso que isso é primordial. A gente lida com propriedade intelectual, né? Em quase todos os setores da universidade e a gente sabe hoje aí que para pegar um dado desse e fazer um plágio, né? Então esse repositório tem que estar bem, tem informação que precisa estar bem assegurada, mas também, por outro lado, o que pode ser feito, divulgado para outras pessoas utilizarem também como parâmetro de pesquisa, eu também acho muito importante. Os trabalhos que dão para outras universidades, outros pesquisadores ou até mesmo da universidade, aproveitando esses dados em alguma pesquisa. Eu acho que é um reaproveitamento de recursos, inclusive...Olha outra coisa, eu tô fazendo um projeto em que tenho que entrevistar os operários da Esav e da Uremg, então tem pessoas que não são vivas mais, a maioria já faleceu. E aí alguns casos entrevistei os herdeiros, os descendentes, mas os que eu consegui, dois ou três, eu consegui. Daqui a um tempo, eles não estarão mais para entrevista. Se a minha entrevista não for depositada em algum lugar e algum outro pesquisador quiser saber a história da vida daquele indivíduo, não terá jeito de entrevistá-lo.”

Fonte: Elaborado pela autora, 2021.

5.2.1 Análise dos dados referentes à entrevista

Nesta etapa das entrevistas individuais, conseguimos a participação de funcionários de todos os centros de ensino da UFV e nesse processo, foi possível identificar que é um tema emergente na instituição, como bem colocou esse respondente:

> Seria interessante, pelo menos com os professores que como eu tem pouco conhecimento sobre essa área, de repente, sei lá, conversar mais em termos de conceitos e definições para a gente entender também mais, né? O que é esse universo todo de gestão de dados? Muitos não sabem do que se trata, entendeu? Às vezes, esse cuidado na definição,

> no tratamento dos conceitos mesmo é interessante. Não sei, se para o nível da entrevista em si. Mas eu acho que é interessante até para chamar a atenção dos pesquisadores de que existe e que isso é necessário. Eu não conhecia muito o tema, conhecia pouco, mas até por conhecer pouco, eu me interessei em participar da entrevista. (Participante do CCH).

Outro participante, neste sentido, também pontua:

> Esse assunto aí é novo. Inclusive para a maior parte dos pesquisadores que eu conheço, não vou falar de todos. Esse assunto é novo. É um assunto, que eu diria, que ele não rola na roda de conversa. (Participante do CCA).

Dialogando com as demandas expressas nos diálogos dos participantes acima, Silva *et al,* (2019) na retro Seção 3.7 (A biblioteca universitária no contexto dos dados científicos: desafios e perspectivas) chamam a atenção dos profissionais bibliotecários sobre disseminar a importância da gestão de dados, como também, criar estratégias para a capacitação dos pesquisadores que querem adquirir informação e conhecimento sobre o tema, mas não sabem por onde começar. Compactuando com essa ideia, na mesma Seção, Santos (2019) também reconhece que fica a cargo do bibliotecário promover consultoria sobre dados científicos, incentivar e planejar a gestão de dados de pesquisa, incluindo *advocacy* e orientação aos pesquisadores.

Em outros pontos de algumas entrevista, podemos notar que algumas falas se confundem em relação aos termos ou às questões formuladas, principalmente entre a 2° e a 3° questão. O responsável pela “gestão” dos dados processados durante as pesquisas é o professor, ele é o “guardião” dos dados, é o docente que, de alguma forma, administra os dados gerados no andamento das pesquisas. Este processo de “guarda” dos dados científicos na UFV não difere do de outras instituições de pesquisas de nosso país, Resende (2019), na Seção 3.2 (Dados científicos e dados de pesquisa) deste estudo, aponta que são poucas as iniciativas de curadoria de dados científicos implantadas nas comunidades científicas brasileiras, a autora assinala que existe interesse das instituições em desenvolver a curadoria dos dados científicos, porém, ainda não se tem conhecimento de uma política sólida, com definições claras, para tornar uma atividade obrigatória nas instituições.

Foi possível identificar também que o Comitê de Ética é quem dita as regras em relação ao que é ou não permitido nos processos das pesquisas científicas com seres humanos. Em determinados pontos, não foi possível identificar se os relatórios criados durante ou no

final dos projetos de pesquisa atendiam a questões burocráticas das agências de fomento, ou se realmente eram documentos que remetiam aos dados científicos.

Outro ponto identificado é que pesquisadores, que têm alguma função administrativa e que no momento estão com pesquisas em curso, sentem falta de uma estrutura que dê apoio e acesso a dados científicos em etapas cruciais de seus estudos.

> … em alguns casos, entrevistei os herdeiros, os descendentes, mas os que eu consegui, dois ou três, eu consegui. Daqui a um tempo, eles não estarão mais para entrevista. Se a minha entrevista não for depositada em algum lugar e algum outro pesquisador quiser saber a história da vida daquele indivíduo, não terá jeito de entrevistá-lo. (Participante do CCH-PEC)

Através da fala do participante do CCH-PEC, podemos perceber que a gestão e a curadoria dos dados científicos em uma instituição de pesquisa é crucial para a preservação, acesso e uso futuro. Nesta direção, na Seção 3.3 (Curadoria digital) Siebra (2019) afirma que a gestão e curadoria desses dados possibilita a preservação da memória individual e coletiva, como também, preserva a herança cultural da produção acadêmica e bibliográfica, resultando na preservação digital, o que ocasiona o acesso, o uso e o reuso a longo prazo.

Outro ponto a considerar também é a questão das agências de fomento neste contexto. De acordo com os participantes, até o presente momento não foi exposto nada a respeito.

> Não existe uma pergunta nesse sentido, se existe uma política do programa de guardião ou repositório de dados gerados. Eu posso te dizer o seguinte: os programas e os coordenadores também vão muito na banda do que as agências cobram. (Participante do CCA)

A fala do participante do CCA coaduna com a declaração de Costa (2017) na Seção 3.5 (Política institucional de dados científicos), onde a autora aponta que as agências de fomento no Brasil estão extremamente atrasadas se comparadas com as de países como o Reino Unido e os EUA, quanto ao entendimento sobre o tema em questão, ou seja, são poucas as agências brasileiras que têm se envolvido em formulação de políticas de gestão de dados científicos.

Em contrapartida, teve outro entrevistado que declarou que as agências de fomento, neste contexto, podem ajudar em um projeto de gestão de dados.

> “As agências de fomento, neste sentido, podem ser uma coisa muito mais concreta, as agências de fomento como ferramentas de mudança de pesquisa … as agências de fomentos são, talvez, as formas mais eficientes de fomentar isso, inclusive porque elas precisam compartilhar, o ideal de um banco de dados é que eles sejam compatíveis com diferentes pesquisas ou que eles possam ser disponibilizados para outros pesquisadores.” (Participante do CCB).

Nota-se pela fala dos participantes, que eles reconhecem que uma política de dados pode viabilizar uma disponibilização dos dados científicos e que essa ação poderá gerar ganhos significativos para a comunidade científica da UFV, como, por exemplo, transparência, progresso e desenvolvimento das pesquisas. Ao mesmo tempo, os entrevistados percebem a dificuldade de gerenciar a tipologia diversificada dos dados científicos, como também conseguem discernir que é vital estabelecer um meio termo quando se trata de questões como sigilo, confidencialidade, propriedade intelectual e outros termos afins. Essa postura dos entrevistados, percebendo como é importante formular uma política de dados, vai ao encontro do entendimento dos autores Briney, Goben e Zilinski (2017), na Seção 3.5 (Política institucional de dados científicos), onde eles defendem que a política de dados institucionais deve cobrir a propriedade de dados de pesquisa, administração e expectativas, bem como fornecer definições claras, identificar reivindicações de acesso e propriedade aos dados, especificar os períodos de retenção e definir responsabilidades de todas as partes interessadas nos dados. Neste contexto, é notório que os participantes não conseguiram expressar quais seriam os atores que poderiam compor esta política, mas conseguiram vislumbrar sua relevância em uma instituição de pesquisa.

Por fim, vale lembrar que esta etapa do trabalho buscou, dentro do que foi possível, analisar e identificar as formas de procedimentos de gestão dos dados científicos nos programas de pós-graduação, ou seja, objetivou criar um diálogo com os funcionários que estão na linha de frente com as agências de fomentos, nas coordenações dos programas, nos departamentos, entre outros. Entretanto, o estudo não conseguiu abordar todos os cenários existentes, havendo ainda inúmeras questões que precisam ser identificadas e algumas outras mais aprofundadas, resultando, assim, numa fonte abundante de vários estudos e pesquisas.

6 CONSIDERAÇÕES FINAIS

Esta pesquisa investigou e mapeou as práticas e percepções em relação aos dados científicos criados na UFV e, paralelamente, mapeou o tratamento desses dados que as unidades de pós-graduação executam e dispõem aos pesquisadores. Para a realização desse trabalho, foram utilizados instrumentos de coleta de dados, como a aplicação de um questionário *on-line*. Através deste instrumento, os participantes poderiam conceder ou não uma entrevista. As etapas executadas acima foram cruciais para a compreensão e entendimento do tema “Gestão de dados e seus assuntos correlatos” são pouco discutidos na referida universidade. Essa constatação se deu através da entrevista, porque o número de interessados em participar desse processo esteve muito abaixo do esperado, isso também foi comprovado nas falas das entrevistas, tendo em vista que alguns participantes desconheciam alguns termos ou informações relacionados ao tema trabalhado.

Nesse contexto, também desconheciam o apoio ou a assistência em relação aos dados científicos processados em suas unidades de atuação, ou seja, nenhuma diretriz que fosse direcionada aos discentes das pós-graduações. Alguns participantes também pontuaram o papel das agências de fomento, opinando que se elas não pedem ou exigem algo direcionado à gestão dos dados, o tema não será relevante entre os setores que coordenam os projetos de pesquisa, enquanto as agências não posicionarem seu papel neste cenário. Em relação ao CEP, alguns reconheceram o papel estratégico do órgão, como também suas diretrizes fundamentais para o andamento dos projetos de pesquisa na instituição.

Pontua-se que as informações e dados coletadas durante a entrevista refletem pouco a realidade de todos os programas de pós-graduação, pois, como já mencionado, a adesão para a participação na entrevista foi baixa. O trabalho realizado não conseguiu abranger todas as unidades que lidam diretamente com os trâmites da pesquisa científica, como, por exemplo, membros do CEP, membros de revistas científicas editadas pela UFV, membros do CPPI, membros dos institutos de pesquisa que existem na UFV, entre outros, e isso deixa muitas questões a serem levantadas e discutidas, assim, nota-se que é um cenário que precisa ser mais investigado e explorado.

Outro ponto a ressaltar é que a pesquisa não se preocupou em fazer cruzamentos de dados entre departamentos ou programas de pós-graduação, pois julga-se que este não é o momento adequado, os esforços direcionados da pesquisa foram para obter um cenário de como é tratado e discutido o assunto abordado.

Vale pontuar, que a pesquisa em sua fase inicial projetava, após todos os processos efetuados, elaborar subsídios para a formulação da política de dados científicos da UFV. Entretanto, repensando na complexidade de tal tarefa e no pouco tempo disponível, optou-se por realizar algumas tarefas que, de alguma forma, contribuíssem com tal empreitada. Assim, preferimos realizar esta pesquisa, na qual foram mapeados dados e informações relevantes das características e perfis da comunidade científica, pois considera-se que a pesquisa executada pode contribuir futuramente, quando a instituição for elaborar uma política institucional de dados científicos.

Com a aplicação das ferramentas de coleta de dados, obtivemos uma informação muito pertinente, que, apesar da baixa participação nas entrevistas, em torno de 89,9% dos respondentes do questionário *on-line*, houve interesse em participar de cursos ou eventos sobre a temática, sendo essa informação muito importante para fundamentar projetos de letramentos sobre dados científicos na UFV. Esse interesse dos respondentes, mais uma vez, fortalece que o tema entre os pesquisadores é um assunto em evolução e isso dá subsídios para a elaboração de estratégias para ampliar a discussão sobre dados científicos na instituição.

O objetivo geral e os específicos foram alcançados, pois com a metodologia aplicada conseguimos identificar e mapear a tipologia, o acesso, o compartilhamento, reuso e a gestão dos dados científicos gerados nos ambientes de pesquisa da UFV. Sendo assim, com o resultado da identificação e mapeamento desses ativos informacionais foi adquirido um panorama de como a comunidade científica da instituição pratica e percebe o tema abordado.

Contudo, a pesquisa forneceu subsídios: a UFV necessita de um letramento em dados, principalmente após a entrevista com os membros de unidades de pós-graduação. Verificou-se que o tema necessita ser desenvolvido amplamente, carecendo de planejamentos sólidos sobre o assunto. Agora é o momento de inserir o tema nas comunidades de pesquisas, dessa forma, a pesquisa sugere algumas estratégias de maneira ampla, mas que, futuramente, podem ser esmiuçadas e mais pontuais em seu modo de planejamento e execução. Assim, seguem algumas estratégias:

- criar vínculos com departamentos e programas de pós-graduação em prol de fomentar a discussão “Gestão de dados de pesquisa e suas vertentes” entre a comunidade científica;
- no planejamento do letramento em dados, é imprescindível que se divida em programas de pós-graduação, ou seja, respeitando cada domínio existente;

- planejar e desenvolver projetos com professores e coordenadores para realizar levantamentos de repositórios de dados científicos que podem ser utilizados pelos programas de pós-graduação;
- realizar levantamentos para identificar quais professores e pesquisadores têm interesse na temática e, após a identificação, criar grupos de estudos sobre o tema; e
- em todas as atividades desenvolvidas sobre a temática, sempre fomentar discussões sobre proteção e privacidade dos dados.

REFERÊNCIAS

ALBAGLI, Sarita. Ciência aberta em questão. In: ALBAGLI, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud (Orgs.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: Unirio, 2015. p. 9-25. Disponível em: http://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20%285%29.pdf. Acesso em: 10 fev. 2020.

ALBAGLI, Sarita; CLINIO, Anne; RAYCHTOCK, Sabryna. Ciência aberta: correntes interpretativas e tipos de ação open science: interpretive trends and types of action **Liinc em revista**, Rio de Janeiro, v. 10, n. 2, 2014. DOI:10.18617/liinc.v10i2.749. Disponível em: http://revista.ibict.br/liinc/article/view/3593. Acesso em: 24 nov. 2020.

ALMEIDA, Fernanda Gomes. **Suporte à gestão de dados de pesquisa: uma ampliação dos serviços oferecidos pelas bibliotecas**. Orientadora: Beatriz Valadares Cendón. 2019. 249 f. Tese (Doutorado) - Programa de Pós- Graduação em Gestão & Organização do Conhecimento, Escola da Ciência da Informação, Universidade Federal de Minas Gerais. Belo Horizonte, 2019. Disponível em: https://repositorio.ufmg.br/bitstream/1843/31596/4/Tese_PPGGOC_Fernanda%20Gomes%20Almeida_11out.pdf. Acesso em: 10 fev. 2020.

ASSIS, Maria Cristina de. Metodologia do trabalho científico. *In*: FARIA, Evangelina Maria B. de; ALDRIGUE, Ana Cristina S. (Orgs.). **Linguagens:** usos e reflexões. João Pessoa: Ed. Universitária/UFPB, 2008.

ATKINS, Daniel Ewell *et al*. **Revolutionizing science and engineering through cyberinfrastructure:** report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. Washington, DC, Jan. 2003. Disponível em: https://www.nsf.gov/cise/sci/reports/atkins.pdf. Acesso em: 15 dez. 2020.

BORGMAN, Christine L. Research Data: Who Will Share What, with Whom, When, and Why? **RatSWD Working Paper**, n. 161. Oct. 2010. Disponível em: https://ssrn.com/abstract=1714427. Acesso em: 04 fev. 2020

BORGMAN, Christine L. *et al*. Data Management in the Long Tail: science, software, and service. **International Journal of Digital Curation.** v. 11, n.1, p. 128-149, 2016. ISSN: 1746-8256. Disponível em: http://www.ijdc.net/article/view/11.1.128. Acesso em: 04 fev. 2020.

BRINEY, Kristin ; GOBEN, Abigail; ZILINSKI, Lisa. Institutional, funder, and journal data policies. *In:* Johnston, Lisa R. (ed.) **Curating Research Data**: practical strategies for your digital. Chicago, Illinois: Association of College and Research Libraries Repository, 2017. v.1, cap. 2, p. 60-78.

BUENO, Wilson Costa. Comunicação científica e divulgação científica: aproximações e rupturas conceituais. **Informação & Informação**, Londrina, v. 15, n. 1, p. 1-12, dez. 2010. ISSN 1981-8920. Disponível em: http://www.uel.br/revistas/uel/index.php/informacao/article/view/6585. Acesso em: 03 fev. 2020.

CARIBÉ, Rita de Cássia do Vale. Comunicação científica: reflexões sobre o conceito. **Informação & Sociedade:** Estudos, João Pessoa, v. 25, n. 3, p. 89-104, 2015. Disponível em: https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/23109. Acesso em: 03 fev. 2020.

CARLSON, Jake R. **Demystifying the data interview**: developing a foundation for reference librarians to talk with researchers about their data. 2011. 18 p. Disponível em: https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1186&context=lib_research. Acesso em:

CARVALHO, Isabel Cristina Louzada. **A socialização do conhecimento no espaço das bibliotecas universitárias**. Niterói: Interciência, 2004.

CHAN, Leslie; OKUNE, Angela; SAMBULI, Nanjira. O que é ciência aberta e colaborativa, e que papéis ela poderia desempenhar no desenvolvimento? *In*: ALBAGLI, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud (orgs.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: Unirio, 2015. p. 91-119. Disponível em: http://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes %20abertas_PORTUGUES_DIGITAL%20%285%29.pdf. Acesso em: 10 fev. 2020.

CHIZZOTTI, Antonio. **Pesquisa em ciências humanas e sociais**. 8. ed. São Paulo: Cortez, 2000. 164 p.

CHIZZOTTI, Antonio. **Pesquisa qualitativa em ciências humanas e sociais**. 5 ed. Petrópolis: Vozes, 2013. 144 p.

CHOUDHURY, Sayeed *et al*. **Research data curation**: a framework for an Institution-WideServices Approach. Louisville, CO: ECAR, May 2018. Disponível em: https://library.educause.edu/-/media/files/library/2018/5/ewg1803pdf.pdf~. Acesso em: 04 ago. 2020.

CÓRDULA, Flavio Ribeiro; ARAÚJO, Wagner Junqueira de. O compartilhamento de dados científicos na era do e-science. *In*: DIAS, Guilherme Ataíde; OLIVEIRA, Bernardina Maria Juvenal Freire da (Orgs.) **Dados científicos**: perspectivas e desafios. João Pessoa: UFPB, 2019. cap. 9, p. 189-207.

CORRÊA, Pedro Luiz Pizzigatti. O modelo organizacional para gestão integrada de dados da biodiversidade brasileira. *In*: DIAS, Guilherme Ataíde; OLIVEIRA, Bernardina Maria Juvenal Freire da (orgs.) **Dados científicos**: perspectivas e desafios. João Pessoa: UFPB, 2019. cap. 7, p. 153-175.

COSTA, Maíra Murrieta. **Diretrizes para uma política de gestão de dados científicos no Brasil.** 2017. Orientador: Murilo Bastos da Cunha. 2017. 288 f. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017.

COSTA, Maíra Murrieta; CUNHA, Murilo Bastos da. A necessidade de uma política nacional para a gestão de dados de pesquisa no Brasil. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, dez. 2019. Disponível em: http://revista.ibict.br/liinc/article/view/4763/4330. Acesso em: 28 ago. 2020.

COSTA, Maíra Murrieta; CUNHA, Murilo Bastos da. O bibliotecário no tratamento de dados oriundos da *e-science*: considerações iniciais. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 19, n. 3, p. 189-206, set. 2014. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362014000300010&lng=en&nrm=iso. Acesso em 07 jan. 2021.

COSTA, Michelli Pereira da; LEITE, Fernando César Lima. Fatores que exercem influência na comunicação dos dados de pesquisa: uma revisão sistematizada da literatura no campo da Ciência da Informação. *In:* ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19., 2018, Londrina. **Anais eletrônicos**... Londrina: Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina (PPGCI/UEL), 2018. Disponível em: http://enancib.marilia.unesp.br/index.php/XIXENANCIB/xixenancib/paper/viewFile/1265/1869. Acesso em: 26 nov. 2018.

ENCONTRO DE GESTÃO E PRESERVAÇÃO DE DADOS DE PESQUISA EM HUMANIDADES, 1., 2019, Rio de Janeiro. [Manifesto]. Rio de Janeiro: GO Fair Brasil, 2019.

EUROPEAN COMMISSION. **Scientific information in the digital age**: Ensuring current and future access for research and innovation. Brussels, Feb. 2007. Disponível em: https://ec.europa.eu/digital-single-market/en/news/scientific-information-digital-age-ensuring-current-and-future-access-research-and-innovation. Acesso em: 23 dez. 2020.

FUNDAÇÃO DE AMPARO A PESQUISA DO ESTADO DE SÃO PAULO. Programa FAPESP de pesquisa em eScience. [2019]. Folder eletrônico do programa. Disponível em https://fapesp.br/publicacoes/2015/folder_escience.pdf. Acesso em: 20 dez. 2020.

FUNDAÇÃO OSWALDO CRUZ. Grupo de Trabalho em Ciência Aberta. **Termo de referência:** gestão e abertura de dados para pesquisa na Fiocruz. Rio de Janeiro: FIOCRUZ, 2018. 15 p.

GABRIEL JÚNIOR, Rene Faustino *et al.* Acesso aberto a dados de pesquisa no brasil: mapeamento de repositórios, práticas e percepções dos pesquisadores e tecnologias. **Ciência da Informação**, Brasília, v. 48, n. 3, p. 87-101, 2019. Disponível em: http://hdl.handle.net/20.500.11959/brapci/136453. Acesso em: 22 jan. 2021.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. São Paulo. Atlas, 2008.

GOFAIR BRASIL SAÚDE. Declaração Aberta da Rede de Implementação do Campo da Saúde. [2019]. Disponível em: https://www.arca.fiocruz.br/bitstream/icict/42368/2/declaracao_implementacao_gofairbrasilsaude_portugu%C3%AAs.pdf. Acesso em: 28 ago. 2020.

GOMES, Gracielle Mendonça Rodrigues. **Expansão das fronteiras da educação na Ciência da Informação**: a formação de profissionais para atuação no contexto dos dados de pesquisa. Orientadora: Beatriz Valadares Cendón. 2019. 244 f. Tese (Doutorado) - Programa de Pós-Graduação em Gestão & Organização do Conhecimento, Escola de Ciência da Informação, Universidade Federal de Minas Gerais. Belo Horizonte, 2019. Disponível em:

https://repositorio.ufmg.br/bitstream/1843/31349/1/Tese%20-%20Vers%c3%a3o%20Final.pdf. Acesso em: 20 dez. 2020.

GROGAN, Denis. **A prática do serviço de referência**. Brasília: Briquet de Lemos, 1995.

HEIDORN, Patrick Bryan. Shedding light on the dark data in the long tail. **Library Trends**, v. 57, n. 2, p. 280-299, 2008. Disponível em: https://www.ideals.illinois.edu/bitstream/handle/2142/9127/Heidorn_LongTail_PreprintwEdits.doc.pdf?sequence=7. Acesso em: 22 set. 2020.

HENNING, Patricia Corrêa *et al.* Go fair e os princípios fair: o que representam para a expansão dos dados de pesquisa no âmbito da ciência aberta. **Em Questão**, Porto Alegre, v. 25, n. 2, p. 389-412, 2019. DOI:10.19132/1808-5245252.389-412. Disponível em: https://seer.ufrgs.br/EmQuestao/article/view/84753/52667. Acesso em: 23 nov. 2020.

HEY, Tony; TREFETHEN, Anne. Cyberinfrastructure for e-Science. **Science,** Washington, v. 308, n. 5723, p. 817-821, May 2005. Disponível em: https://science.sciencemag.org/content/308/5723/817/tab-article-info. Acesso em: 05 de jan. 2021.

HUANG, Yingshen; COX, Andrew; SBAFFI, Laura. Research data management policy and practice in China. **International Journal of Digital Curation**, v. 15. n. 1, 2020. Disponível em: http://www.ijdc.net/article/view/718. Acesso em: 20 dez. 2020.

HODSON, Simon. FAIR data, FAIR services and the FAIR data action plan. *In*: INTERNATIONAL CONFERENCE “DATA ANALYTICS AND MANAGEMENT IN DATA INTENSIVE DOMAINS” (DAMDID/ RCDL’2018), 20., 2018, Moscow, Russia.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Manifesto de acesso aberto a dados da pesquisa brasileira para ciência cidadã.** 2016. Disponível em: ibict.br/sala-de-imprensa/noticias/item/478-ibict-lanca-manifesto-de-acesso-aberto-a-dados-da-pesquisa-brasileira-para-ciencia-cidada. Acesso em: 23 dez. 2020.

LAKATOS, Eva Maria MARCONI, Marina de Andrade. **Técnicas de pesquisa**. São Paulo. Atlas, 2002.

LIMA, Juliana Soares; FARIAS, Maria Giovanna Guedes. Ciência aberta e gestão de dados científicos: competências necessárias para a atuação do bibliotecário. *In*: FARIAS, Gabriela Belmont de; FARIAS, Maria Giovanna Guedes (org.). **Competência e mediação da informação**: percepções dialógicas entre ambientes abertos e científicos. São Paulo: ABECIN Editora, 2019. p. 185-204.

LUCE, Richard E. Grand challenges and new roles for the twenty-first-century research library in an era of e-science. *In:* MARCUM, Deanna B.; GEORGE, Gerald (Ed.). **The data deluge**: can libraries cope with e-science? Santa Barbara, California: Libraries Unlimited, 2010. cap. 1.

MARÍN-ARRAIZA Paloma; VIDOTTI Silvana. Implementação de serviços institucionais de administração de dados. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, 11 dez. 2019. Disponível em: http://revista.ibict.br/liinc/article/view/4819. Acesso em: 20 dez. 2020.

MARQUES, Fabrício. Ciência transparente. **Revista Fapesp,** São Paulo, n. 218, p. 54-58, abr. 2014.

MARTÍNEZ-URIBE, Luis; MACDONALD, Stuart. Un nuevo cometido para los bibliotecarios académicos: data curation. **El Profesional de la Información**, v. 17, n. 3, p. 273–280, 2008. Disponível em: http://eprints.rclis.org/19304/. Acesso em: 05 de jan. 2021.

MINAYO, Maria Cecília de Souza (org.). **Pesquisa Social:** Teoria, método e criatividade. 18. ed. Petrópolis: Vozes, 2001.

MONTEIRO, Elizabete Cristina de Souza de Aguiar; SANTANA, Ricardo César Gonçalves. Repositórios de dados científicos na infraestrutura de pesquisa: adoção dos princípios *fair*. **Ciência da Informação**, Brasília,v. 48, n. 3, 2019. Disponível em: http://hdl.handle.net/20.500.11959/brapci/136407. Acesso em: 15 jul. 2021.

MORIN, Edgar. **Os setes saberes necessários à educação do futuro**. São Paulo: Cortez, 2000.

MUELLER, Suzana Pinheiro Machado. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, Brasília, v. 35, n. 2, p. 27-38, ago. 2006. Disponível em: http://www.scielo.br/scielo.phpscript=sci_arttext&pid=S010019652006000200004&lng=en&nrm=iso. Acesso em: 03 fev. 2020.

MUELLER, Suzana Pinheiro Machado; PASSOS, Edilenice J. L. As questões da comunicação científica e a ciência da informação. *In:* MUELLER, Suzana Pinheiro Machado; PASSOS, Edilenice J. L. (org.). **Comunicação científica**. Brasília: Departamento de Ciência da Informação da Universidade de Brasília, 2000. p. 13-22. Disponível em: https://repositorio.unb.br/bitstream/10482/1444/1/CAPITULO_QuestaoComunicacao.pdf. Acesso em: 10 fev. 2020.

OECD. **Declaration on Access to Research Data From Public Funding**. Paris, Jan. 2004. Disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0321?_ga=2.53862248.1976794085.1642513271-143398942.1623405705. Acesso em: 23 dez. 2020.

OECD. **Principles and Guidelines for Access to Research Data from Public Funding**. Paris, 2007. Disponível em: https://www.oecd.org/sti/inno/38500813.pdf. Acesso em: 23 dez. 2020.

OPEN KNOWLEDGE FOUNDATION. ***A fair, free and open future***. *2020*. Disponível em: https://okfn.org/. Acesso em: 23 dez. 2020.

PAGANINE, Lucas Nóbrega; AMARO, Bianca. Características dos repositórios de dados científicos no brasil. **BIBLOS** - Revista do Instituto de Ciências Humanas e da Informação, Rio Grande, v. 34, n. 1, p. 176-188, 2020. DOI:10.14295/biblos.v34i1.11132. Acesso em: 05 set. 2020.

PAGANINE, Lucas Nóbrega *et al*. **Princípios e diretrizes orientadoras para políticas institucionais de dados científicos abertos em apoio à Ciência Aberta**. Brasília, DF: IBICT; Embrapa; Fiocruz, 2020. 21 p.

PINHEIRO, Lena Vania Ribeiro. Do acesso livre à ciência aberta: conceitos e implicações na comunicação científica. **RECIIS - Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, Rio de Janeiro, v. 8, n. 2, p. 153-165, 2014. Disponível em: https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/629/1269. Acesso em: 24 nov. 2020.

PRÍNCIPE, Pedro *et al*. Estudo sobre os dados científicos gerados no âmbito da investigação produzida na Universidade do Minho. **Cadernos Bad**, 2014. Disponível em: http://repositorium.sdum.uminho.pt/handle/1822/30497. Acesso em: 02 ago. 2019.

RAJASEKAR Arcot *et al*. Sociometric methods for relevancy analysis of long tail science data. *In:* INTERNATIONAL CONFERENCE ON SOCIAL COMPUTING, 2013, Alexandria, VA, 2013, p. 1-6.

RANS, J.; JONES, S. **RDM strategy:** moving from plans to action. *DCC RDM Services case studies*. Edinburgh: Digital Curation Centre, 2013. Disponível em: https://www.dcc.ac.uk/resources/developing-rdm-services/rdm-strategy-moving-plans-action. Acesso em: 19 dez. 2020.

REILLY, Susan *et al*. **ODE Opportunities for Data Exchange: r**eport on integration of data and publications. 2011. Disponível em: https://www.stm-assoc.org/2011_12_5_ODE_Report_On_Integration_of_Data_and_Publications.pdf. Acesso em: 23 dez. 2020.

RESEARCH INFORMATION NETWORK. **Research Funders Policies for the Management of Information Outputs:** a report commissioned by the Research Information Network. 2007. Disponível em: http://www.ijdc.net/article/view/36. Acesso em: 23 dez. 2020.

RESENDE, Liliane Chaves de. **A curadoria de dados científicos na Ciência da Informação**: levantamento do cenário nacional. Orientador: Marcello Peixoto Bax. 2019. 134 f. Dissertação (Mestrado) - Programa de Pós- Graduação em Gestão & Organização do Conhecimento, Escola da Ciência da Informação, Universidade Federal de Minas Gerais. Belo Horizonte, 2019. Disponível em: https://repositorio.ufmg.br/bitstream/1843/32413/1/Disserta%c3%a7%c3%a3o%20CD-ECI-UFMG%20-%20Liliane%20CResende.pdf. Acesso em: 24 nov. 2020.

SALES, Luana Farias. Gestão de Dados de Pesquisa e o Papel dos Bibliotecários. *In:* SEMINÁRIO DO SISTEMA INTEGRADO DE BIBLIOTECAS DA USP, 1. 2018. **Gestão de Dados de Pesquisa e Boas Práticas para o Desenvolvimento da Ciência**. São Paulo:

Sibiusp, 2018. Disponível em: https://www.slideshare.net/sibiusp/gesto-de-dados-de-pesquisa-e-o-papel-dos-bibliotecrios. Acesso em: 25 jul. 2018.

SALES, Luana Farias; SAYÃO, Luis Fernando. A ciência invisível: revelando os dados da cauda longa da pesquisa. *In:* ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19. 2018. Disponível em: http://hdl.handle.net/20.500.11959/brapci/103678. Acesso em: 23 nov. 2020.

SALES, Luana Farias; SAYÃO, Luis Fernando. A grande e a pequena Ciência: análise das diferenças na gestão de dados de pesquisa. **Informação & Sociedade:** Estudos, João Pessoa, v. 29, n. 3, 30 set. 2019b. Disponível em: https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/47615. Acesso em: 10 out. 2020.

SALES, Luana Farias; SAYÃO, Luis Fernando. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 21, n. 2, p. 90-115, dez. 2016. ISSN 1981-8920. Disponível em: http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939. Acesso em: 04 fev. 2020.

SALES, Luana Farias; SAYÃO, Luis Fernando. Avaliação de infraestrutura para gestão de dados de pesquisa. *In:* ENCONTROS PRÉ-CONFOA: DADOS CIENTÍFICOS EM SAÚDE E REPOSITÓRIOS, 2., 2017, Rio de Janeiro. **[Encontros...]** Rio de Janeiro: Fiocruz/ICICT, 2017. 48 p. Inclui folder do evento.

SALES, Luana Farias; SAYÃO, Luis Fernando. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade:** Estudos, João Pessoa, v. 22, n. 3, p. 179-191, 2012b. Disponível em: http://hdl.handle.net/20.500.11959/brapci/92680. Acesso em: 16 mar. 2020.

SALES, Luana Farias; SAYÃO, Luis Fernando. O impacto da curadoria digital dos dados de pesquisa na Comunicação Científica. **Encontros Bibli:** revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, p. 118-135, dez. 2012a. ISSN 1518-2924. Disponível em: https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17nesp2p118. Acesso em: 03 fev. 2020.

SALES, Luana Farias; SAYÃO, Luis Fernando. Uma proposta de taxonomia para dados de pesquisa. **Revista Conhecimento em Ação**, v. 4, n. 1, p. 31-48, 2019a. Disponível em: http://hdl.handle.net/20.500.11959/brapci/118944. Acesso em: 21 fev. 2020.

SANTA ANNA, Jorge; DIAS, Célia da Conceição; MACULAN, Benildes Coura. A gestão dos dados de pesquisa nas universidades e o papel dos serviços informacionais oferecidos nas bibliotecas: Uma revisão narrativa. **Múltiplos Olhares em Ciência da Informação**, Belo Horizonte, v. 9, n. 2, 2019. Disponível em: https://periodicos.ufmg.br/index.php/moci/article/view/19169. Acesso em: 28 jan. 2021.

SANTOS, Marcelle Costal de Castro dos. Competência em Informação para a gestão de dados de pesquisa. *In:* ENCONTRO DA REDE SUDESTE DE REPOSITÓRIOS INSTITUCIONAIS, 1., 2019, Rio de Janeiro. [**Anais...]** Rio de Janeiro: Fiocruz/Icict/UFRJ, 2019. 38 p.

SANTOS, Thayse Natália Cantanhede. Curadoria digital e preservação digital: cruzamentos conceituais. **Revista Digital de Biblioteconomia & Ciência da Informação**, v. 14, n. 3, p. 450-464, 2016. DOI: 10.20396/rdbci.v14i3.8646336. Disponível em: https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8646336. Acesso em: 16 maio 2020.

SAYÃO, Luis Fernando. **Gestão de dados de pesquisa: o que todo pesquisador espera da sua biblioteca**. Fortaleza: Plurissabares BCH UFC, 14 jul. 2020. Disponível em: https://www.youtube.com/watch?v=_HT_nrmwjUE&list=PLOhWAljyF7rreYsaQuwb6LdUZ7fd_cTN8&index=16&t=5423s. Acesso em: 14 jul. 2020.

SAYÃO, Luís Fernando; SALES, Luana Farias. A ciência invisível: os dados da cauda longa da pesquisa científica. *In*: DIAS, Guilherme Ataíde; OLIVEIRA, Bernardina Maria Juvenal Freire de. (orgs.) **Dados científicos:** perspectivas e desafios. João Pessoa: UFPB, 2019, p. 33-52. Disponível em: http://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/view/359/508/2949-1. Acesso em: 23 nov. 2020.

SAYÃO, Luís Fernando; SALES, Luana Farias. Afinal, o que é dados de pesquisa? **BIBLOS** - Revista do Instituto de Ciências Humanas e da Informação, Rio Grande, v. 34, n. 2, 2020. Disponível em: https://periodicos.furg.br/biblos/article/view/11875. Acesso em: 23 fev. 2021.

SAYÃO, Luís Fernando; SALES, Luana Farias. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 6, n. 1, 2013. Disponível em: http://hdl.handle.net/20.500.11959/brapci/119469. Acesso em: 19 jan. 2021.

SAYÃO, Luís Fernando; SALES, Luana Farias. **Guia de Gestão de Dados de Pesquisa para Bibliotecários e Pesquisadores.** Rio de Janeiro: CNEN, 2015. 90 p. Disponível em: http://www.icb.usp.br/~sbibicb/images/guia%20gestaoPDF/Guia%20de%20gestao%20dados%20de%20pesquisa.pdf. Acesso em: 23 fev. 2021.

SEMELER, Alexandre Ribas; PINTO, Adilson Luiz. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. **Ciência da Informação**, Brasília, v. 48, n. 1, 3 maio 2019. Disponível em: http://revista.ibict.br/ciinf/article/view/4461. Acesso em: 05 de jan. 2021.

SHEARER, Kathleen. **Comprehensive brief on research data management policies**. 2015. Disponível em: https://portagenetwork.ca/wp-content/uploads/2016/03/Comprehensive-Brief-on-Research-Data-Management-Policies-2015.pdf. Acesso em: 18. dez.2020.

SIEBRA, Sandra de Albuquerque. Curadoria digital: uma área em expansão. **Archeion Online**, João Pessoa, v. 6, n. 2, 2019. DOI: 10.22478/ufpb.2318-6186.2019v6n2.47089. Disponível em: https://periodicos.ufpb.br/ojs/index.php/archeion/article/view/47089/27666. Acesso em: 16 mar. 2020.

SIEBRA, Sandra de Albuquerque; BORBA, Vildeane da Rocha; MIRANDA, Májory Karoline Fernandes de Oliveira. Curadoria digital: um termo interdisciplinar. **Informação &**

Tecnologia, Marilia, João Pessoa, v. 3, n. 2, p. 21-38, 2016. Disponível em: http://hdl.handle.net/20.500.11959/brapci/41848. Acesso em: 16 mar. 2020.

SILVA, Fabiano Couto Corrêa da. **Gestão de dados científicos.** Rio de Janeiro: Interciência, 2019. 130 p.

SILVA, Fabiano Couto Corrêa da. O papel dos bibliotecários na gestão de dados científicos. **RDBCI:** Revista Digital de Biblioteconomia e Ciência da Informação, Campinas, SP, v. 14, n. 3, p. 387–406, 2016. DOI: 10.20396/rdbci.v14i3.8646333. Disponível em: https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8646333. Acesso em: 14 jul. 2020.

SILVA, Maria Helena Ferreira Xavier da. *et al*. Competências dos bibliotecários na gestão dos dados de pesquisa. **Ciência da Informação**, Brasília, v. 48, n. 3, 2019. Disponível em: http://hdl.handle.net/20.500.11959/brapci/136517. Acesso em: 29 jan. 2021.

SILVA SEGUNDO, Sanderli José da; ARAÚJO, Wagner Junqueira de. Curadoria e ciclo de vida dos dados. *In*: DIAS, Guilherme Ataíde; OLIVEIRA, Bernardina Maria Juvenal Freire da (orgs.). **Dados científicos**: perspectivas e desafios. João Pessoa, PB: UFPB, 2019. cap. 6, p. 113-152.

TARGINO, Maria das Graças. Comunicação científica: uma revisão de seus elementos básicos. **Informação & Sociedade:** Estudos, João Pessoa**,** v. 10 n.2 2000, n. 2, 2000. Disponível em: http://hdl.handle.net/20.500.11959/brapci/92195. Acesso em: 03 fev. 2020.

TARTAROTTI, Roberta Cristina D.; DAL'EVEDOVE, Paula Regina; FUJITA, Mariângela Spotti Lopes. Biblioteconomia de dados em repositórios de pesquisa: perspectivas para a atuação bibliotecária. **Informação & Informação**, Londrina, v. 24, n. 3, p. 207-226, 2019. DOI:10.5433/1981-8920.2019v24n3p207. Disponível em: https://www.uel.br/revistas/uel/index.php/informacao/article/view/38732/pdf. Acesso em: 24 nov. 2020.

TENOPIR, Carol *et al*. Data Sharing by scientists: practices and perceptions. **PLoS ONE**, v.6, n.6, e21101, 2011. DOI:10.1371/journal.pone.0021101. Disponível em: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101. Acesso em: 02 ago. 2019.

TENOPIR, Carol *et al*. Research data services in European academic research libraries. **LIBER Quarterly,** v. 27, n. 1, p. 1–22, 2017. DOI: http://doi.org/10.18352/lq.10180. Disponível em: https://www.liberquarterly.eu/article/10.18352/lq.10180/. Acesso em 07 jan. 2021.

THE ROYAL SOCIETY. **Science as an Open Enterprise.** The Real Society Science Policy Centre Report. n. 2, 2012. Disponível em: https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/. Acesso em: 23 dez. 2020.

UNIVERSIDADE FEDERAL DE VIÇOSA. **Relatório de atividade de 2018**: ano base 2017. Viçosa: UFV, 2018.

VANZ, Samile Andrea de Souza *et al*. Acesso aberto a dados de pesquisa no Brasil: práticas e percepções dos pesquisadores: relatório 2018. Porto Alegre: UFRGS, 2018. Disponível em: https://lume.ufrgs.br/handle/10183/185195. Acesso em: 02 ago. 2019.

VEIGA, Viviane Santos de Oliveira. **Percepção dos pesquisadores portugueses e brasileiros da área de neurociências quanto ao compartilhamento de artigos científicos e dados de pesquisa no acesso aberto verde**: custos, benefícios e fatores contextuais. Orientadora: Cícera Henrique da Silva. 2017. 294 f. Tese (Doutorado) - Programa de Pós-Graduação em Informação, Comunicação e Saúde, Instituto de Comunicação e Informação em Saúde, Fundação Oswaldo Cruz. Rio de Janeiro, 2017.

VEIGA, Viviane Santos de Oliveira *et al.* Compartilhamento de dados de pesquisa na fiocruz: diagnóstico e percepção do pesquisador. **Ciência da Informação**, Brasília, v. 48, n. 3, 2019. Disponível em: http://hdl.handle.net/20.500.11959/brapci/136386. Acesso em: 22 jan. 2021.

ZEELAND, Hilde van; RINGERSMA, Jacquelijn. The development of a research data policy at Wageningen University & Research: best practices as a framework. **The Journal of the Association of European Rechearch Libraries**. v. 27, n. 1, p. 153-170, 2017. Disponível em: https://www.liberquarterly.eu/articles/10.18352/lq.10215/. Acesso em: 19 dez. 2020.

WALLIS, Jillian C.; ROLANDO, Elizabeth; BORGMAN, Christine L. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. **PLoS ONE**, v. 8, n. 7, 2013. Disponível em: https://doi.org/10.1371/journal.pone.0067332. Acesso em: 28 out. 2020.

APÊNDICE 1: QUESTIONÁRIO

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Prezado(a) pesquisador(a),

Eu sou Andréia Cipriano Maciel, bibliotecária e documentalista da Biblioteca Central da UFV e convido você a participar da pesquisa intitulada: "Práticas e percepções sobre dados científicos na Universidade Federal de Viçosa", que é parte integrante da minha dissertação de mestrado no Programa de Pós-Graduação em Biblioteconomia Universidade Federal do Estado do Rio de Janeiro, sob a orientação do Prof. Dr. Luís Fernando Sayão.

Para iniciar o preenchimento, é preciso que você concorde com o Termo de Consentimento Livre e Esclarecido (TCLE). O tempo estimado para o preenchimento do questionário é de 10 minutos.

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Você está sendo convidado para participar voluntariamente da pesquisa "Práticas e percepções sobre dados científicos na Universidade Federal de Viçosa". Especificamente, o questionário tem como objetivo coletar informações sobre os dados científicos gerados na UFV, identificando as práticas, percepções dos pesquisadores, como também, mapear as características de gestão desses elementos científicos.

Informo que ao responder este questionário você está concordando em participar desta pesquisa. Assim, você não terá nenhum custo, nem receberá qualquer vantagem financeira. Para a participação, você contará com o esclarecimento sobre o estudo em qualquer aspecto que desejar, em qualquer momento.

A sua participação é voluntária e de caráter estritamente confidencial e estará livre para participar ou recusar-se a participar a qualquer tempo e sem quaisquer prejuízos. As informações obtidas serão totalmente confidenciais e suas respostas serão registradas anonimamente.

Se você tiver alguma dúvida sobre este estudo ou sobre sua participação, entre em contato com a pesquisadora Andréia Cipriano Maciel pelos emails: andreia.maciel@ufv.br ou andreiacmaciel@gmail.com que se compromete a sanar dúvidas ou buscar soluções para resolver o problema apresentado pelo participante.

DECLARO que li as informações contidas neste Termo de consentimento, fui devidamente informado(a) pela pesquisadora Andréia Cipriano Maciel dos procedimentos que serão utilizados, e da confidencialidade das informações, concordando em participar da pesquisa. Foi-me garantido que posso retirar meu consentimento a qualquer momento e que poderei pedir esclarecimentos a qualquer tempo. *

- ( ) Sim, concordo em responder este questionário.

1. Identificando seu perfil

Descrição (opcional)

1.1 Qual seu nível de escolaridade? *

- Pós-Doutorado
- Doutorado
- Mestrado
- Graduação

1.2 Em qual perfil profissional você se encaixa? *

- Professor
- Pesquisador
- Professor visitante
- Professor pesquisador
- Pesquisador colaborador
- Pesquisador externo
- Estagiário
- Bolsista
- Outros...

1.2.1 Caso você tenha respondido "Outros" especifique:

Texto de resposta curta

1.3 Em qual Centro a sua pós-graduação ou sua pesquisa está vinculada? *

- Centro de Ciências Agrárias
- Centro de Ciências Biológicas e da Saúde
- Centro de Cências Exatas e Tecnológicas
- Centro de Ciências Humanas, Letras e Artes

1.4 Quais das alternativas melhor descrevem o principal financiador da sua pesquisa? *

- ( ) Agência internacional
- ( ) Agência nacional
- ( ) Agência regional
- ( ) A própria instituição
- ( ) Empresa privada
- ( ) Não possuo financiador
- ( ) Outros...

1.4.1 Caso você tenha respondido "Outros" especifique:

Texto de resposta curta

2 As questões desta seção referem-se às práticas adotadas por você em relação aos seus dados científicos.

Para melhor elucidar alguns termos desta seção, seguem os significados:

DADOS CIENTÍFICOS: todo e qualquer tipo de registro coletado, observado, gerado ou utilizado no âmbito da pesquisa científica, que pode ser interpretado, tratado e aceito como evidência pela comunidade científica e necessário para analisar, validar e produzir resultados de pesquisa.

CIÊNCIA ABERTA ou ACESSO ABERTO: propõe que conhecimentos científicos de todos os tipos deveriam ser compartilhados abertamente tão logo quando possível no processo de descoberta.

GESTÃO DE DADOS CIENTÍFICOS: conjunto de práticas voltadas para o tratamento de dados científicos durante o seu ciclo de vida, o que inclui todos os aspectos de manutenção, compartilhamento, segurança e preservação.

PLANO DE GESTÃO DE DADOS: documento formal que descreve todo o ciclo de vida dos dados, desde sua coleta até a documentação completa do processo de pesquisa, e registra as decisões tomadas em relação aos padrões de metadados, formatos, bases de dados, métodos, segurança e períodos de armazenamento, assim como os custos associados com a gestão dos dados

2.1 Que termos melhor descrevem os tipos de dados de pesquisa que você produz? (marque quantas alternativas julgue necessário) *

- [ ] Dados automaticamente gerados por programas de computador
- [ ] Dados coletados de sensores e equipamentos
- [ ] dados de simulação
- [ ] dados documentais
- [ ] dados experimentais
- [ ] dados observacionais
- [ ] entrevistas
- [ ] questionários
- [ ] Outros...

2.2 Quais tipos de dados científicos você produz? (marque quantas alternativas julgue necessário) *

- [ ] Algoritmos
- [ ] Animações, modelos ou simulações
- [ ] Áudios
- [ ] Banco de dados
- [ ] Imagens
- [ ] Planilhas
- [ ] Slides
- [ ] Software
- [ ] Textos
- [ ] Vídeos
- [ ] Websites
- [ ] Outros...

2.3 Qual o volume estimado dos dados científicos que você produz em um ano? *

- ◯ <1 GB
- ◯ 1-50 GB
- ◯ 100 - 500 GB
- ◯ 500 GB - 1 TB
- ◯ 1 - 50 TB's
- ◯ 50 - 100 TB's
- ◯ >100 TB's
- ◯ Não sei responder

2.4 Em quais locais você armazena seus dados científicos? *

	Todos os meus da...	Parte dos meus da...	Nenhum dos meus...	Não sei responder
Meu computador p...	○	○	○	○
Servidor localizado...	○	○	○	○
Servidor/ data cent...	○	○	○	○
Nuvem (ex. Google ...	○	○	○	○
Arquivos em papel	○	○	○	○
Mídia externa (ex. ...	○	○	○	○

2.5 Você já compartilhou seus dados científicos em acesso aberto? (marque quantas alternativas julgue nece *

- [ ] Sim, por solicitação do meu principal financiador
- [ ] Sim, por solicitação do Programa onde estudo
- [ ] Sim, para publicação em periódico ou em anais de evento
- [ ] Sim, por iniciativa própria
- [ ] Nunca compartilhei (se você escolheu esta opção, passe para a questão 2.7)

2.6 Se você respondeu que compartilha seus dados científicos em acesso aberto, onde você compartilha?

	Todos os meus da...	Parte dos meus da...	Nenhum dos meus...	Não sei responder
Repositório temáti...	○	○	○	○
Repositório de revi...	○	○	○	○
Repositório da min...	○	○	○	○
Repositório nacion...	○	○	○	○

2.7 Se você NÃO compartilha, marque as possíveis razões. (marque quantas alternativas julgue necessário)

- [ ] Falta de financiamento
- [ ] Falta de padronização dos dados
- [ ] As pessoas não precisam deles
- [ ] Não tenho tempo para compartilhá-los
- [ ] Não há infraestrutura para compartilhá-los
- [ ] Eles não podem ser compartilhados
- [ ] Não me foi exigido
- [ ] Eu não tenho permissão para torná-los públicos
- [ ] Eu perderia o controle dos dados
- [ ] Eu preciso publicar os resultados da pesquisa antes de compartilhá-los
- [ ] Eu não sei como compartilhar, pois não possuo conhecimentos e ferramentas de compartilhamento
- [ ] Os dados não estão totalmente documentados
- [ ] Os dados estão em um formato que não permite esse compartilhamento
- [ ] Outros...

2.8 Quão satisfeito você está com a organização dos seus dados nas seguintes situações *

	Satisfeito	Parcialmente satis...	Insatisfeito	Não sei responder
Coleta de dados	○	○	○	○
Descrição dos dad...	○	○	○	○
Armazenamento d...	○	○	○	○
Preparação dos rel...	○	○	○	○
Recuperação dos ...	○	○	○	○
Preservação dos d...	○	○	○	○

2.9 Qual sua familiaridade com a gestão de dados científicos? *

- ○ Conheço bastante
- ○ Conheço pouco
- ○ Desconheço

2.10 Você já elaborou um Plano de Gestão de Dados? (marque quantas alternativas julgue necessário). *

- [ ] Sim, por solicitação do meu principal financiador
- [ ] Sim, por solicitação do Programa Pós-graduação que estou vinculado
- [ ] Sim, para publicação em periódico
- [ ] Sim, por outras razões
- [ ] Nunca elaborei

2.11 Você utiliza algum repositório de dados para acessar dados científicos de outros pesquisadores? Não se aplica nessa questão infraestrutura para a gestão de documentos bibliográficos (software de bibliotecas, biblioteca de teses e dissertações) *

- ( ) Sim
- ( ) Não

2.12 Se SIM, se for possível, especifique o repositório que você usa.

Texto de resposta curta

3 Nesta seção, são apresentadas questões referentes às suas percepções em relação ao uso de dados científicos e ao seu compartilhamento.

Descrição (opcional)

3.1 Especifique a sua concordância com cada uma das afirmativas a seguir *

	Concordo	Concordo parcialm...	Discordo	Não sei responder
Eu utilizaria dados ...	○	○	○	○
Eu compartilharia p...	○	○	○	○
Eu compartilharia t...	○	○	○	○
Eu estaria mais incl...	○	○	○	○
O compartilhament...	○	○	○	○
Eu poderia perder o...	○	○	○	○
Minhas ideias de p...	○	○	○	○
Meus dados poderi...	○	○	○	○

3.2 Para outros pesquisadores usarem meus dados de pesquisa, eu espero como retorno: *

	Sempre	As vezes	Não faço questão	Não sei responder
Um acordo mútuo ...	○	○	○	○
Co-autoria em publ...	○	○	○	○
Agradecimento em ...	○	○	○	○
Citação em publica...	○	○	○	○
A oportunidade de ...	○	○	○	○
A solicitação de u...	○	○	○	○
Reembolso de part...	○	○	○	○
Ter acesso a uma c...	○	○	○	○
Ter acesso a todos...	○	○	○	○

3.3 Onde você acha que seus dados científicos devem ser preservados? *

	Todos os meus da...	Parte dos meus da...	Nenhum dos meus...	Não sei responder
Em servidor/datac...	○	○	○	○
Em servidor localiz...	○	○	○	○
Em meu computad...	○	○	○	○
Em um repositório ...	○	○	○	○
Em um repositório ...	○	○	○	○
Em um repositório ...	○	○	○	○
Em outros tipos de ...	○	○	○	○
Na nuvem	○	○	○	○
Em mídia externa (...	○	○	○	○
Não devem ser pre...	○	○	○	○

4 Nesta seção, são apresentadas questões para identificar os serviços e assistência que o seu Centro/Departamento/Programa disponibiliza para que seus dados científicos sejam processados durante suas pesquisas.

Descrição (opcional)

4.1 Se houver serviço de gestão de dados científicos no seu Departamento/Programa, quem é o responsável? *

- ( ) Coordenador do Programa da Pós-graduação
- ( ) Líder do grupo de pesquisa
- ( ) Chefe de laboratório
- ( ) Agência de financiamento
- ( ) Bibliotecários de dados da instituição
- ( ) Desconheço esse serviço no meu Departamento/Programa
- ( ) Outros...

4.1.1 Caso você tenha respondido "Outros" especifique:

Texto de resposta curta

4.2 Existe alguma coleção de dados que o Departamento/Programa indica ou utiliza que pode ser usado em sua pesquisa? *

- Sim
- Não
- Não sei responder

4.2.1 Se for SIM a sua resposta, se for possível, especifique o nome a Coleção de Dados

Texto de resposta curta

4.3 Você tem conhecimento se os dados produzidos pela sua pesquisa são de alguma forma preservados pelo Departamento/Programa? (Não se aplica nessa questão, a preservação digital de documentos bibliográficos resultantes da pesquisa, como por exemplo: tese, dissertação, relatório de pesquisa, artigos.) *

- Sim, pelo Departamento/Programa que estou vinculada
- Desconheço este serviço no meu Departamento/Programa
- Não sei responder

4.4 Quais as opções de armazenamento dos seus dados que o Departamento/Programa oferece ou indica? *

- ( ) Servidor localizado no Departamento/Programa
- ( ) Em repositório de dados indicado pelo Departamento/Programa
- ( ) Nuvem (ex. Google, drive/dropbox)
- ( ) Desconheço este serviço no meu Departamento/Programa
- ( ) Outros...

4.4.1 Caso você tenha respondido "Outros" especifique:

Texto de resposta curta

4.5 Quais as opções de armazenamento que a agência de fomento oferece ou indica? *

- ( ) Servidor localizado no laboratório do meu grupo de pesquisa
- ( ) Servidor/data center administrado pela agência de fomento
- ( ) Em repositório de dados indicado pela agência de fomento
- ( ) Não sei responder
- ( ) Outros...

4.5.1 Caso você tenha respondido "Outros" especifique:

Texto de resposta curta

4.6 Seu Departamento/Programa fornece diretrizes para o uso e compartilhamento de dados científicos? *

- Sim
- Não

4.7 Se o Departamento/Programa fornece diretrizes, se possível, comente sobre elas.

Texto de resposta longa

4.8 Quem você acha que deve oferecer serviços de apoio a gestão dos dados de pesquisa no Departamento/Programa que você está vinculado? (marque quantas alternativas julgue necessário) *

- [ ] Coordenador do Programa da Pós-graduação
- [ ] Líder do grupo de pesquisa
- [ ] Chefe de laboratório
- [ ] Agência de financiamento
- [ ] Bibliotecários da instituição
- [ ] Não acho necessário apoio
- [ ] Não sei responder
- [ ] Outros...

Caso você tenha respondido "Outros", especifique:

Texto de resposta curta

4.9 Você participaria de eventos criados pela UFV sobre o tema abordado neste questionário? *

- Sim
- Não tenho interesse por esse tema

4.10 Você exerce alguma função administrativa na UFV? *

- Sim
- Não

5 Esta seção é para identificar sua função administrativa e verificar seu interesse em participar de uma entrevista, que objetiva mapear a estrutura organizacional no que tange à curadoria e gestão de dados científicos criados na UFV.

Descrição (opcional)

5.1 Qual é sua função? *

- Coordenação de Pós-graduação
- Líder de grupo de pesquisa
- Coordenação de laboratório ou núcleo de pesquisa
- Membro editorial de alguma revista científica criada na UFV
- Técnico administrativo/Direção
- Outros...

5.2 Sendo você colaborador das funções administrativas da organização e se tiver disponibilidade de tempo e interesse, você aceita participar de uma curta entrevista que devida as circunstâncias da pandemia do Covid-19 será realizada remotamente? *

- ( ) Sim
- ( ) Não

Após a seção 6 Continuar para a próxima seção

Seção 7 de 8

6 Agendamento da entrevista

Descrição (opcional)

6.1 Se você respondeu afirmativo na questão anterior, por gentileza, forneça seu email ou telefone do seu setor para o agendamento da entrevista. *

Texto de resposta curta

Minhas cordiais saudações e meus sinceros agradecimentos por participar desta pesquisa!

APÊNDICE 2: ENTREVISTA

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Você está sendo convidado a participar voluntariamente da pesquisa “Práticas e percepções sobre dados científicos na Universidade Federal de Viçosa”. A **entrevista** tem como objetivo coletar informações sobre os dados científicos gerados na UFV, a pesquisa busca identificar as práticas, percepções e características de gestão desses elementos resultantes de investigações científicas. Informo que, ao concordar em realizar a entrevista, você não terá nenhum custo, nem receberá qualquer vantagem financeira.

Sua participação é voluntária e de caráter estritamente confidencial e estará livre para participar ou se recusar a participar a qualquer tempo e sem quaisquer prejuízos. As informações obtidas serão totalmente confidenciais e os dados que identificam os participantes serão mantidos e guardados em sigilo. Se os resultados forem publicados, seu nome será sempre mantido em sigilo e nunca aparecerá.

O tempo de duração da será de 15 a 25 minutos e será gravado em áudio.

Se você tiver alguma dúvida sobre este estudo ou sobre sua participação, entre em contato com a pesquisadora Andréia Cipriano Maciel, pelos e-mails andreia.maciel@ufv.br ou andreiacmaciel@gmail.com ou pelo telefone 31 988448599, que se compromete a sanar dúvidas ou buscar soluções para resolver o problema apresentado pelo participante.

DECLARO que li as informações contidas neste Termo de Consentimento Livre e Esclarecido, fui devidamente informado(a) pela pesquisadora Andréia Cipriano Maciel dos procedimentos que serão utilizados, e da confidencialidade das informações, concordando em participar da pesquisa. Foi-me garantido que posso retirar meu consentimento a qualquer momento e que poderei pedir esclarecimentos a qualquer tempo.

( ) Sim, concordo em participar da entrevista.

ENTREVISTA COM OS COORDENADORES E DIRIGENTES DOS AMBIENTES DE PESQUISA DA UNIVERSIDADE FEDERAL DE VIÇOSA - MG

1) Em qual Centro e Departamento você atua? Qual sua função administrativa?

2) No seu Centro/Departamento/Programa existe alguma diretriz de coleta, tratamento, armazenamento, preservação dos dados científicos gerados pelas pesquisas?

3) O Centro/Departamento/Programa tem algum documento formal para a política de dados científicos processados por seus pesquisadores?

4) Em seu Centro/Departamento/Programa, é oferecido ou indicado aos pesquisadores algum repositório de dados? Se SIM, por favor, cite-o.

5) Em seu Centro/Departamento/Programa existe alguma obrigação ou incentivo de compartilhamento dos dados científicos processados por seus pesquisadores?

6) Há questões éticas e de privacidade associadas aos dados de pesquisa gerados em seu Centro/Departamento/Programa?

7) Há questões associadas à propriedade intelectual e de copyright aos dados científicos gerados em seu Centro/Departamento/Programa?

8) As agências de fomento que atuam em seu Centro/Departamento/Programa têm exigido Plano de Gestão de Dados dos projetos financiados por elas?

9) As agências de fomento têm pedido/exigido o depósito dos dados em algum repositório? Se SIM, por favor, cite-os.

10) Na sua opinião, quem deve ser o responsável pela gestão de dados científicos gerados por pesquisas financiadas pelas agências de fomento?

11) Em sua opinião, a UFV precisa de uma política para a gestão de dados científicos?

12) Quer acrescentar algo sobre a entrevista? Alguma crítica ou sugestão?