

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO & ORGANIZAÇÃO
DO CONHECIMENTO

Viviane Lílian dos Santos Barrozo

CURADORIA E PRESERVAÇÃO DIGITAL: DIRETRIZES PARA A
GESTÃO DE DADOS CIENTÍFICOS

Belo Horizonte

2022

Viviane Lílian dos Santos Barrozo

**CURADORIA E PRESERVAÇÃO DIGITAL: DIRETRIZES PARA A
GESTÃO DE DADOS CIENTÍFICOS**

Dissertação apresentada à Universidade Federal de Minas Gerais, no Programa de Pós-Graduação em Gestão & Organização do Conhecimento da Escola de Ciência da Informação para obtenção do título de mestre em Gestão & Organização do Conhecimento.

Área de concentração: Ciência da Informação.
Linha de Pesquisa: Arquitetura & Organização do Conhecimento (AOC).

Orientador: Prof. Dr. Carlos Henrique Marcondes.

Belo Horizonte

2022

B278c

Barrozo, Viviane Lilian dos Santos.

Curadoria e preservação digital [recurso eletrônico] : diretrizes para a gestão de dados científicos / Viviane Lilian dos Santos Barrozo. – 2022.

1 recurso online (140 f. : il., color.) : pdf.

Orientador: Carlos Henrique Marcondes.

Dissertação (Mestrado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

Referências: f. 105-110.

Apêndices: f. 111-128.

Anexos: f. 129-140.

Exigência do sistema: Adobe Acrobat Reader.

1. Ciência da informação – Teses. 2. Ciência – Processamento de dados – Teses. 3. Preservação digital – Teses. 4. Gestão do conhecimento - Teses. I. Marcondes, Carlos H. II. Universidade Federal de Minas Gerais. Escola de Ciência da Informação. III. Título.

CDU 02:004

Ficha catalográfica: Rosimeire Silva Campos de Lima - CRB:6/3145

Biblioteca Prof^ª Etelvina Lima, Escola de Ciência da Informação da UFMG



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO - PPGOC

FOLHA DE APROVAÇÃO

CURADORIA E PRESERVAÇÃO DIGITAL: DIRETRIZES PARA A GESTÃO DE DADOS CIENTÍFICOS

VIVIANE LILIAN DOS SANTOS BARROZO

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, como requisito para obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, área de concentração CIÊNCIA DA INFORMAÇÃO, linha de pesquisa Arquitetura e Organização do Conhecimento.

Aprovada em 25 de julho de 2022, por videoconferência, pela banca constituída pelos membros:

Prof(a). Carlos Henrique Marcondes de Almeida (Orientador)
Professor Visitante ECI/UFMG

Prof(a). Cláudio José Silva Ribeiro
UNI RIO

Prof(a). Eli Lopes da Silva
Universidade do Estado de Santa Catarina

Belo Horizonte, 25 de julho de 2022.



Documento assinado eletronicamente por **Carlos Henrique Marcondes de Almeida, Professor do Magistério Superior - Visitante**, em 27/07/2022, às 15:01, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Claudio Jose Silva Ribeiro, Usuário Externo**, em 27/07/2022, às 16:31, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Eli Lopes da Silva, Usuário Externo**, em 29/07/2022, às 19:05, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador 1628114 e o código CRC F4CD4BFB.

Referência: Processo nº 23072.244167/2022-56

SEI nº 1628114



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO - PPGOC

ATA DA DEFESA DA DISSERTAÇÃO DA ALUNA

VIVIANE LILIAN DOS SANTOS BARROZO

Realizou-se, no dia 25 de julho de 2022, às 14:00 horas, por videoconferência, da Universidade Federal de Minas Gerais, a defesa de dissertação, intitulada *CURADORIA E PRESERVAÇÃO DIGITAL: DIRETRIZES PARA A GESTÃO DE DADOS CIENTÍFICOS*, apresentada por VIVIANE LILIAN DOS SANTOS BARROZO, número de registro 2020660878, graduada no curso de SISTEMAS DE INFORMAÇÃO, como requisito parcial para a obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Carlos Henrique Marcondes de Almeida - Professor Visitante ECI/UFMG (Orientador), Prof(a). Cláudio José Silva Ribeiro - UNIRIO, Prof(a). Eli Lopes da Silva - Universidade do Estado de Santa Catarina.

A Comissão considerou a dissertação:

Aprovada

Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.

Belo Horizonte, 25 de julho de 2022.

Assinatura dos membros da banca examinadora:



Documento assinado eletronicamente por Carlos Henrique Marcondes de Almeida, Professor do Magistério Superior - Visitante, em 27/07/2022, às 15:01, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por Claudio Jose Silva Ribeiro, Usuário Externo, em 27/07/2022, às 16:31, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por Eli Lopes da Silva, Usuário Externo, em 29/07/2022, às 19:07, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador 1628074 e o código CRC 2E139613.

Referência: Processo nº 23072.244167/2022-56

SEI nº 1628074

Este trabalho é dedicado aos meus amados pais, Vilma Aparecida dos Santos Barrozo (minha Flor) e José Maria Barrozo (vivo no meu coração e na memória), e a minha incomparável avó, Dona Jandira Frazão dos Santos (também viva no meu coração e na memória), por todo amor, ensinamentos, orações, dedicação e incentivo que sempre recebi de vocês desde o início da minha trajetória escolar e profissional.

Todas as minhas conquistas serão dedicadas a vocês, meus amores!

AGRADECIMENTOS

Agradeço a Deus a escrita de mais uma página vitoriosa no livro de minha vida. Toda ansiedade e toda angústia foram e serão convertidas em fé, porque em todos os meus propósitos, Ele sempre estará comigo.

À minha mãe, Flor, o amor, o companheirismo, a dedicação e o incentivo durante toda a minha trajetória acadêmica. Pelas vezes em que permaneceu acordada até altas horas, fazendo-me companhia, durante os registros da pesquisa.

Ao meu companheiro e noivo, Lindo, o amor e a alegria que suavizaram os dias mais tensos deste percurso acadêmico. Agradeço por me encorajar, recordando-me da minha capacidade.

Aos meus irmãos, Davidson, David e Diego, que sempre colaboraram e me auxiliaram em tarefas diversas. Se não fosse a ajuda deles, eu não conseguiria conciliar todas as atividades demandadas no meu cotidiano de vida, no decorrer do andamento da pesquisa.

Aos meus sobrinhos, Thalles, Arthur e Bruna que compreenderam a minha ausência em algumas das brincadeiras propostas por eles. Mesmo sendo crianças, entenderam que, às vezes, não podiam fazer tanto barulho para não me desconcentrar. Meus pequenos, obrigado por serem tão carinhos comigo, mesmo naqueles dias em que não fui tão atenciosa com vocês! Ah! Agradeço, também, por dizerem sempre que “a UFMG é da Viane”, e quando crescerem querem estudar e trabalhar lá (na UFMG), como eu. Ah, crianças! Como vocês são sábias!

Ao meu orientador, Professor Carlos Henrique Marcondes que, pela modéstia de um sábio, atuou com parceria, paciência e cuidado durante as orientações, possibilitando a conclusão deste estudo junto ao Programa de Pós-Graduação em Gestão e Organização do Conhecimento (PPGGOC).

Aos professores da Escola de Ciência da Informação que, por meio da docência, colaboraram com minha formação acadêmica, oportunizando a minha especialização requerida durante os estudos do Mestrado. Estendo essa menção aos professores: Elisângela Cristina Aganette (ECI/UFMG), Cláudio José Silva Ribeiro (UNIRIO) e Eli Lopes Silva (UDESC) pois, gentilmente, aceitaram o convite e deram valiosas contribuições durante o exame de qualificação.

Aos secretários e técnicos, Gildenara, Vinícius, Álvaro e Leandro, atuantes no PPGGOC, a presteza nos atendimentos.

Aos professores da Universidade Federal de Minas Gerais (UFMG), participantes da pesquisa, por serem tão gentis e solícitos ao receberem o meu convite.

Aos professores Wagner José Corradi Barbosa, Eliane Marina Palhares Guimarães, diretores de Educação a Distância da UFMG, e Graça Moreira, vice-diretora, por possibilitarem o meu afastamento do trabalho, para que eu ampliasse a minha dedicação ao término da pesquisa.

Aos meus amigos e colegas do Centro de Apoio à Educação a Distância, Desirré, Rosana e Vânia, sobretudo a equipe da Assessoria Pedagógica, Ana Carolina e Márcia, do Setor de Design Carolina, da Assessoria de Tecnologia da Informação, especialmente, Júnio, Leonardo e Thiago. Agradeço a essas pessoas o incentivo, a colaboração e a compreensão durante o caminhar, concomitante, das tarefas acadêmicas e profissionais. Estendo esse agradecimento aos amigos e colegas que não fazem mais parte do quadro de colaboradores e nem do convívio diário, mas que os guardo no coração: Mayara, Paloma, Greiziele, Cirdes e Pedro.

Agradeço ao colega do Comitê de Ética em Pesquisa (COEP), Anderson, as orientações quanto aos procedimentos éticos da pesquisa e a gentileza no atendimento.

Ao colega do PPGGOC, Jorge Santa Anna, a gentileza e a educação com que sempre me tratou, desde a disciplina eletiva. Agradeço a sua dedicação na revisão e na normalização deste trabalho acadêmico.

Enfim, aos colegas do PPGGOC, especialmente, Graça, Mayara, Jordânia e Fabíola, considerando a partilha, a parceira e o encorajamento durante todo o percurso acadêmico.

A todos vocês, os meus sinceros agradecimentos!

“[...] A preservação digital precisa ser uma responsabilidade compartilhada, principalmente pelo grande volume de informação digital que está sendo produzido e pela natureza da tecnologia digital [...]” (MÁRDERO ARELLANO, 2008, p. 58).

RESUMO

As transformações digitais ocorridas na segunda metade do século XX impactaram, sobremaneira, o fazer científico e suscitaram a reflexão acerca do uso da informação nos ambientes digitais. O aumento da produção e do consumo de dados em formato digital atingiu um patamar de destaque para a ciência e reforçou a importância da curadoria e da preservação digitais, seja para acesso imediato ou em longo prazo. Dessa forma, a preservação de objetos digitais tem sido um dos grandes desafios na contemporaneidade, visto que necessita recorrer a métodos e tecnologias que integrem o aspecto físico (armazenamento), lógico (formato) e intelectual desses objetos. A presente pesquisa objetiva recomendar o uso de princípios e diretrizes de preservação digital para subsidiar uma política de preservação digital na Universidade Federal de Minas Gerais. Metodologicamente, o estudo caracteriza-se como de natureza aplicada e de abordagem qualitativa. Quanto aos objetivos, a pesquisa é exploratória e explicativa e, quanto aos procedimentos técnicos, é bibliográfica, documental e de campo. A investigação ocorreu em quatro fases: a primeira consistiu no referencial teórico, acerca da curadoria e da preservação digitais; a segunda recorreu à aplicação de questionário semiestruturado para diagnosticar as práticas dos pesquisadores, caracterizados como produtores/consumidores de dados científicos, e o nível de preservação dos dados científicos; na terceira, aplicou-se a entrevista semiestruturada para conhecer as práticas dos entrevistados quanto à produção e/ou consumo de dados científicos; por fim, a quarta fase contemplou pesquisa documental, por meio da apreciação a documentos institucionais (planos, políticas, relatórios, manuais, projetos e resoluções) da instituição, a fim de identificar aspectos de governança digital. O estudo identificou as práticas dos pesquisadores referentes aos conjuntos de dados e das ações institucionais no que tange a esse tipo de governança. A análise mostrou que os pesquisadores lidam com dados heterogêneos e singulares, de tipologias e suportes variados e, em muitos casos, salvos em formatos inadequados à preservação e sem documentação e metadados apropriados aos domínios disciplinares. A maior parte dos dados hospeda-se em nuvem (serviço terceirizado), mediante o acesso restrito e custeado pelo próprio pesquisador, ou ainda, armazena-se em equipamentos pessoais ou em mídias frágeis, sem monitoramento, migração e devida atualização tecnológica. Identificou-se a necessidade de aprimorar as práticas desses pesquisadores quanto aos conjuntos de dados, de suprir a falta de orientações com a oferta de instrumentos, guias e planos e de apoiá-los quanto aos serviços de dados com o auxílio de equipe especializada composta por bibliotecário, além de incentivar o uso de recursos tecnológicos munidos de uma infraestrutura adequada de tecnologia da informação, já oferecida pela universidade. Assim, identificou-se a preservação de dados científicos e indicou-se um conjunto de princípios e diretrizes que possa servir de mecanismo para apoiar a universidade ao implementar a sua política de preservação digital. Espera-se que este estudo suscite uma reflexão sobre o aprimoramento das práticas de gestão de dados e, principalmente, ratifique a importância do comprometimento institucional com a pesquisa. Com isso, formaliza-se a política de preservação digital, haja vista salvaguardar os conjuntos de dados científicos e disponibilizá-los para reuso e, desse modo, contribuir para o avanço da ciência brasileira.

Palavras-chave: dados de pesquisa; governança digital; política de preservação; princípios FAIR; repositórios de dados.

ABSTRACT

The digital transformations that took place in the second half of the 20th century had a great impact on scientific practice and raised a reflection on the use of information in digital environments. The increase in the production and consumption of data in digital format has reached a level of prominence for science and reinforced the importance of digital curation and preservation, whether for immediate or long-term access. Thus, the preservation of digital objects has been one of the great challenges in contemporary times, since it needs to resort to methods and technologies that integrate the physical (storage), logical (format) and intellectual aspects of these objects. This research aims to recommend the use of digital preservation principles and guidelines to support a digital preservation policy at the Federal University of Minas Gerais. Methodologically, the study is characterized as being of an applied nature and of a qualitative approach. As for the objectives, the research is exploratory and explanatory and, as for the technical procedures, it is bibliographical, documentary and field. The investigation took place in four phases: the first consisted of the theoretical framework, about digital curation and preservation; the second resorted to the application of a semi-structured questionnaire to diagnose the practices of researchers, characterized as producers/consumers of scientific data, and the level of preservation of scientific data; in the third, the semi-structured interview was applied to know the practices of the interviewees regarding the production and/or consumption of scientific data; finally, the fourth phase included documental research, through the appreciation of institutional documents (plans, policies, reports, manuals, projects and resolutions) of the institution, in order to identify aspects of digital governance. The study identified the practices of researchers regarding data sets and institutional actions regarding this type of governance. The analysis showed that researchers deal with heterogeneous and singular data, of different types and supports, and, in many cases, saved in formats that are inappropriate for preservation and without documentation and metadata appropriate to the disciplinary domains. Most of the data is hosted in the cloud (outsourced service), through restricted access and paid for by the researcher, or stored on personal equipment or fragile media, without monitoring, migration and due technological updating. It was identified the need to improve the practices of these researchers in terms of data sets, to supply the lack of guidelines with the provision of instruments, guides and plans and to support them in terms of data services with the help of a specialized team composed of librarian, in addition to encouraging the use of technological resources equipped with an adequate information technology infrastructure, already offered by the university. Thus, the preservation of scientific data was identified and a set of principles and guidelines was indicated that can serve as a mechanism to support the university in implementing its digital preservation policy. It is hoped that this study will prompt a reflection on the improvement of data management practices and, above all, ratify the importance of institutional commitment to research. With this, the digital preservation policy is formalized, with a view to safeguarding scientific data sets and making them available for reuse and, in this way, contributing to the advancement of Brazilian science.

Keywords: research data; digital governance; preservation policy; FAIR principles; data repositories.

LISTA DE QUADROS

QUADRO 1 - Estrutura da dissertação	20
QUADRO 2 – Princípios FAIR e suas aplicações	33
QUADRO 3 – Levantamento de documentos institucionais da UFMG	47
QUADRO 4 – Estudos correlatos desenvolvidos no PPGGOC (2017 a 2021)	48
QUADRO 5 – Detalhamento dos estudos correlatos desenvolvidos no PPGGOC.....	49
QUADRO 6 – Artigos publicados no ENANCIB (2017 a 2021).....	52
QUADRO 7 - Teses e dissertações selecionadas da BDTD (2017 a 2021)	53
QUADRO 8 – Áreas consultadas e respostas do pré-teste - PCDC.....	55
QUADRO 9 – Setores consultados e respostas do pré-teste - FPI	55
QUADRO 10 – Extrato de validação dos questionários – pré-testes.....	56
QUADRO 11 – Cursos de Pós-Graduação <i>stricto sensu</i> da UFMG	57
QUADRO 12 - Perfil dos participantes da entrevista	73
QUADRO 13 - Seção produção/consumo de dados	74
QUADRO 14 – Seção armazenamento/cópias	75
QUADRO 15 - Seção integridade dos dados	77
QUADRO 16 - Seção metadados	78
QUADRO 17 - Seção controle e reuso	80
QUADRO 18 - Seção Plano de Gestão dos Dados.....	81
QUADRO 19 – Seção boas práticas	83
QUADRO 20 – Seleção dos documentos institucionais da UFMG	88
QUADRO 21 – Documentos institucionais selecionados	90
QUADRO 22 – Análise dos documentos institucionais da Reitoria	91
QUADRO 23 - Análise dos documentos da Diretoria de Arquivos.....	92
QUADRO 24 - Análise do documento da Diretoria de Governança Informacional	92
QUADRO 25 - Análise de documentos da Diretoria de Tecnologia da Informação	93
QUADRO 26 - Diagnóstico da curadoria e da preservação digital de dados científicos na UFMG	94
QUADRO 27 - Diagnóstico da governança digital de dados científicos na UFMG	96
QUADRO 28 - Recomendações e diretrizes para a política de preservação dos dados científicos.....	98

LISTA DE FIGURAS

FIGURA 1 – Fluxograma para a elaboração da política de preservação digital.....	26
FIGURA 2 - Modelo do Ciclo de Vida dos Dados - DCC	29
FIGURA 3 - Modelo do Ciclo de Vida dos Dados - <i>DataONE</i>	30
FIGURA 4 - Proposta para elaboração do Plano de Gestão de Dados	32
FIGURA 5 – Estrutura de uso de repositório de dados científicos.....	36
FIGURA 6 – Repositório de dados científicos do IBICT	37
FIGURA 7 - Percurso metodológico da pesquisa.....	40
FIGURA 8 - Organização do questionário - PCDC.....	54

LISTA DE GRÁFICOS

GRÁFICO 1 - Distribuição dos respondentes por área do conhecimento.....	58
GRÁFICO 2 – Distribuição dos respondentes por tempo de atuação na Pós-Graduação	59
GRÁFICO 3 – Produção e uso de dados em pesquisas da Pós-Graduação.....	59
GRÁFICO 4 - Instituições que fornecem dados	60
GRÁFICO 5 – Armazenamento e localização de cópias dos dados.....	61
GRÁFICO 6 - Dispositivos de armazenamento das cópias dos dados.....	62
GRÁFICO 7 – Formatos e extensões dos arquivos de dados.....	63
GRÁFICO 8 – Integridade dos arquivos de dados	65
GRÁFICO 9 - Metadados para arquivos dos dados	66
GRÁFICO 10 – Idiomas utilizados nas descrições dos campos.....	66
GRÁFICO 11 – Descrição de campos e atribuição de metadados.....	67
GRÁFICO 12 – Autorização para reuso e cópias de dados	68
GRÁFICO 13 – Licenças de uso atribuídas aos dados	68
GRÁFICO 14 - Uso de sites e/ou repositórios de dados	69
GRÁFICO 15 – Formatos de arquivos nos repositórios de dados.....	70
GRÁFICO 16 – Controle de versão dos dados	71
GRÁFICO 17 - Repositório de dados na UFMG.....	72

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
BDTD	Biblioteca Brasileira de Teses e Dissertações
CAED	Centro de Apoio à Educação a Distância
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CCSDS	Consultative Committee for Space Systems
CGU	Controladoria Geral da União
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
COEP	Comitê de Ética em Pesquisa
CONARQ	Conselho Nacional de Arquivos
DCC	Digital Curation Center
DCMI	Dublin Core Metadata Initiative
DOI	Digital Object Identifier
DPC	Digital Preservation Coalition
ECI	Escola de Ciência da Informação
E-ARQ	Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivísticas de Documentos
FAIR	Findable, Accessible, Interoperable and Reusable
Fala.BR	Plataforma Integrada de Ouvidoria e Acesso à Informação
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FORCE11	Community of Scholars, Librarians, Archivists, Publishers and Research Funders
FOSTER	Facilitate Open Science Training for European Research
GO FAIR	Global Open FAIR Initiative
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
IES	Instituição de Ensino Superior
IFLA	International Federation of Library Associations and Institutions
InterPARES	International Research on Permanent Authentic Records in Electronic Systems
ISO	International Organization for Standardization
LOP	Levels of Digital Preservation
MCTI	Ministério da Ciência, Tecnologia e Inovações
METS	Metadata Encoding and Transmission Standard
NDSA	National Digital Stewardship Alliance
OAI	Open Archive Initiative
OAIS	Open Archival Information System
OD	Objetos Digitais
OGP	Parceria para o Governo Aberto
OpenAIRE	Open Access Infrastructure for Research in Europe
ORCID	Open Researcher and Contributor ID
PGD	Plano de Gestão de Dados
PPD	Plano de Preservação Digital
PPGGOC	Programa de Pós-Graduação em Gestão & Organização do Conhecimento
PREMIS	Preservation Metadata: Implementation Strategies
RDP - Brasil	Redes de Dados de Pesquisa Brasileira
RI-UFGM	Repositório Institucional da Universidade Federal de Minas Gerais

RNP	Rede Nacional de Ensino e Pesquisa
SEDIC	Sociedade Espanhola de Documentação e Informação Científica
THE	Times Higher Education
UFMG	Universidade Federal de Minas Gerais
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Objetivos	18
1.1.1	Objetivo geral	18
1.1.2	Objetivos específicos	19
1.2	Estrutura da dissertação	19
2	REFERENCIAL TEÓRICO	21
2.1	Curadoria e Preservação Digital de Dados	21
2.2	Preservação Digital	22
2.2.1	Política e Plano de Preservação Digital	25
2.3	Gestão de Dados Científicos	27
2.3.1	Princípios FAIR e repositório de dados científicos	32
3	METODOLOGIA	39
3.1	Caracterização da pesquisa	39
3.1.1	Pesquisa bibliográfica – Primeira fase	41
3.1.2	Questionário – Segunda fase	41
3.1.3	Entrevista – Terceira fase	44
3.1.4	Pesquisa documental – Quarta fase	46
4	RESULTADOS E DISCUSSÕES	48
4.1	Análise dos trabalhos correlatos	48
4.2	Análise do questionário	54
4.3	Análise das entrevistas	72
4.4	Análise dos documentos institucionais	87
4.5	Diagnósticos	93
4.5.1	Diagnóstico da curadoria e preservação de dados	94
4.6	Diretrizes	97
4.6.1	Recomendações de diretrizes e princípios para preservar dados	97
5	CONSIDERAÇÕES FINAIS	101
	REFERÊNCIAS	105
	APÊNDICE A – ROTEIRO DO SURVEY - PCDC	111
	APÊNDICE B – ROTEIRO DO SURVEY - FPI	115

APÊNDICE C – TERMO CONSENTIMENTO - PCDC	117
APÊNDICE D – TERMO CONSENTIMENTO - FPI	120
APÊNDICE E – ROTEIRO COMPLETO DA ENTREVISTA - PCDC ..	123
APÊNDICE G - ENDEREÇOS - DOCUMENTOS INSTITUCIONAIS ..	126
APÊNDICE H - ENDEREÇOS - CURSOS DE PÓS-GRADUAÇÃO ...	127
ANEXO A – ARTICULAÇÃO METODOLÓGICA	129
ANEXO B – MATRIZ NÍVEIS DE PRESERVAÇÃO DIGITAL (LOP)..	130
ANEXO C – PARECER DO COMITÊ DE ÉTICA EM PESQUISA.....	131
ANEXO D – MANIFESTAÇÃO DE ACESSO À INFORMAÇÃO	137

1 INTRODUÇÃO

Na contemporaneidade, o intenso uso das tecnologias digitais tem potencializado o aumento da produção e o reuso de conteúdos digitais. Esse crescimento fez os dados atingirem um patamar de destaque para a ciência e, conseqüentemente, tornou relevante a curadoria e a preservação digital, com a finalidade de possibilitar o acesso imediato ou em longo prazo.

No âmbito da informação digital, a preservação dos objetos dessa natureza mostra-se desafiadora, devido ao crescimento vertiginoso da quantidade de dados produzidos, da heterogeneidade intrínseca e da singularidade desses dados. Em virtude disso, é possível recorrer às atividades de curadoria e de preservação digital como estratégia de garantia do acesso aos dados.

A curadoria digital envolve a manutenção, a preservação e a agregação de valor aos acervos digitais ao longo do ciclo de vida dos objetos que compõem esses acervos. No modelo de ciclo de vida proposto pelo *Digital Curation Center* (DCC), são fornecidas as etapas necessárias que viabilizam a realização adequada da curadoria e da preservação de dados. Nesse modelo, os dados (objetos digitais ou banco de dados) são elementos centrais, cujas ações de preservação são contínuas e permeiam todo o ciclo de vida dos dados, possibilitando o seu acesso, uso e reuso.

De acordo com a *Digital Preservation Coalition* (DPC, 2021), instituição situada no Reino Unido e responsável por fornecer boas práticas ao processo de preservação de dados, a preservação digital refere-se ao conjunto de atividades de gestão necessárias para garantir o acesso contínuo a materiais digitais pelo tempo que for necessário. Essas atividades englobam todas as ações de manutenção dos dados, além de envolverem práticas referentes aos limites de falha de mídia ou tecnologia e práticas associadas às mudanças organizacionais (DPC, 2021).

Assim como o DPC, no Brasil, existe a Rede Brasileira de Serviços de Preservação Digital (REDE CARINIANA). Segundo a Rede, para se preservar objetos digitais, é importante adotar métodos e tecnologias que integram a preservação, incluindo aspectos de natureza física (armazenamento), lógica (formato) e intelectual desses objetos. O objetivo principal desse processo é assegurar a autenticidade, a perenidade e o acesso aos conteúdos dos objetos em longo prazo (REDE CARINIANA, 2014).

No âmbito das Instituições de Ensino, encontram-se os dados científicos, originários de diversos domínios dessas instituições (pequenos laboratórios, equipes de pesquisa e pesquisadores individuais). Esses domínios, na visão de Sales e Sayão (2018), são denominados de pequena ciência, expressão oriunda do termo em inglês *small science*,

e constituem a “cauda longa” da ciência. Esses autores consideram a “cauda longa” da ciência como um grande número de pequenas equipes de pesquisadores e laboratórios independentes, os quais geram no seu dia a dia de pesquisa uma ampla variedade de coleções de dados de pesquisa.

Entende-se que os dados advindos da “cauda longa” “[...] representam a maior parcela de dados produzida pela ciência e constituem um território de constante criatividade e inovação que precisam ser revelados, integrados e compartilhados” (SALES; SAYÃO, 2018, p. 4184). Nesse sentido, Silva (2019, p. 21) considera que os dados científicos de pesquisa “[...] são informações registradas ou produzidas através de qualquer forma ou meio durante o decurso de uma pesquisa”. Assim, o planejamento, a gestão, a preservação e a reutilização dos dados produzidos nessas pequenas dimensões apresentam-se desafiadores devido à heterogeneidade e à singularidade, características peculiares dos dados científicos (SALES; SAYÃO, 2018).

Desse modo, é importante “[...] analisar a preservação a partir da necessidade de assegurar o acesso e a recuperação da informação científica como fundamento para a pesquisa acadêmica [...]” (REDE CARINIANA, 2014, não paginado). Assim, compete às Instituições de Ensino Superior (IES), gerenciar e salvaguardar os seus dados científicos, pois a preservação é a etapa antecessora à reutilização dos dados, atividades essas essenciais para fortalecer o progresso da ciência.

Souza e Aganette (2020) apontam que poucas IES possuem uma política de preservação digital publicada e nenhuma das instituições pesquisadas pelas autoras possui um plano oficial de preservação digital. A partir dessas constatações, o estudo citado concluiu que há necessidade de se amparar as IES brasileiras no processo de desenvolvimento da política e do plano de preservação relativo aos documentos digitais.

Nesse sentido, a ausência de instrumentos para a governança de dados, a obsolescência tecnológica e a fragilidade das mídias digitais acentuam o risco de perda e dificultam o gerenciamento dos dados científicos. Presume-se que a falta de governança dos conjuntos de dados produzidos no interior dos laboratórios acadêmicos faz com que os pesquisadores recorram aos repositórios de dados externos às instituições, hospedem esses dados em nuvem ou mantenham o acesso aos dados restritos aos grupos e parceiros de pesquisa. Além disso, grande parte desse estoque informacional não está disponível para acesso, compartilhamento e reuso. Isso porque, na maioria das vezes, os dados permanecem armazenados nos computadores pessoais de seus criadores, sem nenhuma gestão que os torne visíveis (ciência invisível), conforme mencionado no estudo de Sales e Sayão (2018).

Dias, Anjos e Araújo (2019) corroboram o entendimento de Sales e Sayão (2018) e concluem que as práticas de gestão de dados conduzidas pelos pesquisadores precisam

ser aprimoradas. Portanto, salienta-se que os produtores de dados “[...] possuem postura favorável com relação ao compartilhamento de dados, desde que exista algum controle formal sobre os mesmos” (DIAS; ANJOS; ARAÚJO, 2019, p. 5-6).

Diante dos apontamentos mencionados, esta pesquisa visa a endereçar as seguintes questões: 1 - Como ocorre a gestão e a preservação de dados científicos no interior dos laboratórios acadêmicos dos Programas de Pós-Graduação (PPG) da Universidade Federal de Minas Gerais (UFMG)? Quais normas, métodos e tipos de recursos são necessários para a criação de uma política de preservação digital?

Pretende-se com esta pesquisa recomendar o uso de diretrizes para subsidiar uma política de preservação digital na UFMG e diagnosticar as práticas dos pesquisadores quanto à gestão e à preservação de dados científicos. Tal proposta é necessária porque os conjuntos de dados de pesquisa necessitam de tratamento especializado para que possam ser acessados em longo prazo e reutilizados na produção de novos conhecimentos. Essa intenção vai ao encontro do que defende a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 2017, não paginado), ao relatar que os dados resultantes de financiamento público “[...] sejam gerenciados e compartilhados de forma a garantir o maior benefício possível para o avanço científico e tecnológico”.

Assim, considerando a importância da temática da preservação digital quanto à possibilidade de garantir o acesso, o uso e o reuso dos conjuntos de dados, esta pesquisa mostra-se relevante para a comunidade acadêmica da UFMG, assim como para a sociedade. Ademais, a pesquisa é relevante para as instituições de pesquisa, sobretudo para a UFMG e demais IES, pois fornecerá insumos para a formalização de diretrizes a serem utilizadas quando da implementação de uma política de preservação de dados.

1.1 Objetivos

Este estudo pretende atender ao objetivo geral e aos objetivos específicos, os quais constituirão a base que direciona os procedimentos para alcance dos resultados. Os objetivos da pesquisa são formalizados nas subseções seguintes.

1.1.1 Objetivo geral

Em função das questões levantadas, o objetivo geral foi formulado do seguinte modo:

- Recomendar diretrizes para subsidiar uma política de preservação digital de dados científicos a ser utilizada pela UFMG.

1.1.2 Objetivos específicos

A fim de atingir o objetivo geral do estudo, foram delineados os seguintes objetivos específicos:

- Compreender as normas, métodos e recursos necessários para garantir a preservação digital.
- Diagnosticar o estágio em que a UFMG encontra-se quanto à gestão de dados científicos e sua conformidade às atividades e processos de curadoria e de preservação digital, conforme padrões e práticas adotados pela comunidade internacional.
- Recomendar a constituição de diretrizes de preservação digital para dados científicos nas IES.

1.2 Estrutura da dissertação

A presente dissertação está estruturada em cinco capítulos necessários ao desenvolvimento da pesquisa. Esses capítulos discorrem sobre a curadoria, a preservação digital e a gestão de dados, sendo esses assuntos considerados elementos essenciais para apoiar a investigação científica.

O primeiro capítulo, do qual faz parte esta subseção, é introdutório e contextualiza a curadoria e a preservação digital de dados científicos. Além disso, elucida as questões e a proposta do estudo (objetivos a serem atingidos).

Em seguida, apresenta-se o referencial teórico, explorando os fundamentos teórico-conceituais referentes à curadoria, à preservação digital e à gestão de dados. Esses assuntos constituem o “condutor” da pesquisa, ao nortear os demais procedimentos realizados.

O terceiro capítulo, reservado à metodologia, detalha o objeto da pesquisa, indicando os tipos de dados coletados, tanto os dados primários (pessoas) quanto os secundários (documentos). Assim, no decorrer do capítulo, o estudo é caracterizado, como também é mencionado os tipos de pesquisa, os métodos e as técnicas adotadas para garantir a aferição dos resultados.

O capítulo de número quatro expõe o relato e a análise dos dados, discutindo os resultados. Também apresenta o diagnóstico da UFMG no que tange aos dados científicos e recomenda a adoção de diretrizes que garantam a gestão e a preservação dos dados.

Para finalizar a dissertação, o quinto capítulo descreve as conclusões da pesquisa de acordo com os seus resultados. Com efeito, apresentam-se as recomendações de boas práticas, as intercorrências no decorrer do estudo e as perspectivas para realização de

estudos futuros. Arelado às conclusões e visando a garantir o fechamento do estudo, tem-se a exposição da lista de referências, dos apêndices e dos anexos.

Importante frisar que os capítulos foram desenvolvidos de modo sistematizado, ou seja, um complementando o outro. A fim de demonstrar essa correlação, apresenta-se o Quadro 1, o qual detalha a estrutura que dá forma a este estudo e vincula cada seção ao respectivo objeto que busca atender.

QUADRO 1 - Estrutura da dissertação

Ordem/numeração	Nome do capítulo	Resumo
1	Introdução	Delineamento do tema, do objeto, bem como dos objetivos, da justificativa e das contribuições da pesquisa.
2	Referencial teórico	Apresentação dos conceitos pertinentes ao objeto de pesquisa, qual seja, a preservação digital de dados científicos. O quadro teórico corresponde ao atendimento do primeiro objetivo específico: compreender as normas, métodos e recursos necessários para garantir a preservação digital.
3	Metodologia	Exposição dos procedimentos metodológicos, elucidando os métodos, a tipologia da pesquisa e os instrumentos adotados para a condução do estudo.
4	Análise e discussões	Relato do diagnóstico, da análise e da discussão dos resultados obtidos por meio de questionário e de entrevista semiestruturados. Esse capítulo está correlacionado ao segundo objetivo específico: diagnosticar o estágio em que a UFMG encontra-se quanto à gestão de dados científicos e sua conformidade às atividades e processos de curadoria e preservação digital, conforme padrões e práticas adotados pela comunidade internacional. O capítulo também apresenta o quadro de diretrizes para preservação digital, de modo a atender o terceiro objetivo específico da dissertação: recomendar a constituição de diretrizes de preservação digital para dados científicos nas IES.
5	Considerações finais	Descrição das conclusões de acordo com os resultados, indicando, também, as recomendações, as intercorrências e as perspectivas para a realização de estudos futuros.
Não numerado	Referências	Registro de todas as obras consultadas e citadas no estudo, constituindo, pois, a lista de referências.
Não numerado	Apêndices	Formulação do Roteiro do <i>Survey</i> - PCDC, do Roteiro do <i>Survey</i> – FPI, do Termo de Consentimento da Pesquisa - PCDC, do Termo de Consentimento da Pesquisa - FPI, do Roteiro completo da Entrevista - PCDC, do Roteiro simplificado da Entrevista – PCDC, dos Endereços Virtuais – URLs, dos Documentos Oficiais e Endereços Virtuais – URLs e dos Documentos Institucionais.
Não numerado	Anexos	Detalhamento dos modelos adotados para o delineamento do objeto da pesquisa, tais como: Articulação Metodológica, Metodologia Matriz dos Níveis de Preservação Digital, <i>Organization National Digital Stewardship Alliance</i> (NDSA) e o Parecer Consubstanciado do Comitê de Ética em Pesquisas.

Fonte: Elaborado pela autora (2022).

2 REFERENCIAL TEÓRICO

Neste capítulo, são apresentadas pesquisas que abordam a mesma temática deste estudo, a saber: a curadoria e a preservação digital de dados científicos. Além disso, o capítulo explora outros temas correlatos às temáticas, tais como: políticas e planos de preservação digital; Princípios FAIR (*Findable, Accessible, Interoperable and Reusable*) traduzidos como: localizáveis, acessíveis, interoperáveis e reutilizáveis; e os repositórios de dados, que amparam esta pesquisa.

2.1 Curadoria e preservação digital de dados

Na área da Ciência da Informação (CI), o termo curadoria digital surgiu em 2001. A produção internacional sobre esse tema data de 2004, mesmo ano de fundação do DDC. Porém, somente a partir de 2012, foram encontrados trabalhos publicados em bases de dados de língua portuguesa, conforme exposto por Siebra (2019). Assim, de acordo com a autora, tanto no cenário nacional quanto internacional, a partir de 2015, observou-se o crescimento da curadoria digital como campo de pesquisa, devido ao aumento substancial da quantidade de publicações. Desde então, a curadoria digital vem expandindo tanto no viés da prática quanto da pesquisa (KIM, 2015).

Em meados da década de 1990, a preservação digital entrou em vigor nos Estados Unidos e no Reino Unido, conforme relatado por Kim (2015). O autor afirma que o interesse inicial pela preservação digital começou com os bibliotecários e depois se fundiu com o interesse dos arquivistas. *A posteriori*, o trabalho com a preservação cresceu para além dessas comunidades e, desde o início dos anos 2000, ampliou os limites da preservação passiva adentrando-se à curadoria ativa (KIM, 2015).

Nesse sentido, na curadoria e na preservação digital, as áreas de atuação de ambas complementam a gestão da informação digital como um todo. Assim, há uma inter-relação entre a curadoria e a preservação digital, visto que a curadoria digital envolve a manutenção, a preservação e a agregação de valor aos acervos digitais ao longo do ciclo de vida desses acervos (SANTOS, 2016).

Uma pesquisa para ser consistente e gerar resultados satisfatórios precisa de dados legítimos e confiáveis. A curadoria ativa insere-se nesse âmbito, quando praticada nos repositórios, visto que reduz as ameaças ao seu valor de longo prazo e atenua o risco de obsolescência digital (DCC, 2020).

Nesse contexto, entram em cena os modelos de ciclo de vida dos dados. A finalidade desses modelos é “[...] proporcionar uma estrutura para a consideração das muitas

operações que devem ser realizadas em um registro de dados durante o uso” (SILVA, 2019, p. 49).

De acordo com Sayão e Sales (2015), há uma série de concepções de modelos de ciclo de vida de dados de pesquisa, cada um com particularidades e objetivos determinados, muitas vezes orientados para domínios de conhecimentos específicos. Os autores mencionam que alguns modelos tornaram-se referência para pesquisadores, bibliotecários e gestores de dados, a saber: *DCC Curation Lifecycle Model*¹, *DataONE Data Lifecycle*², *DDI Combined Lifecycle Model*³ e o *UK Data Archive Data Lifecycle*⁴.

Conforme sinalizado por esses autores, a escolha do modelo deve considerar os domínios do conhecimento e quais as etapas serão seguidas para atendimento às necessidades dos projetos. Justifica-se a verificação dos projetos, tendo em vista que “[...] alguns projetos podem não seguir de forma linear o caminho delineado pelo modelo” (SAYÃO; SALES, 2015, p. 13).

A ideia de se escolher as etapas também é mencionada por Santos (2016), ao relatar que os modelos de gestão de dados podem ser utilizados como guia para a planificação, o que não significa, necessariamente, que sejam um conjunto de regras que deve ser seguido passo a passo, mas podem ser úteis para a elaboração de relatórios e para a planificação dos serviços bibliotecários. Silva (2019) salienta que muitos pesquisadores têm focado no ciclo de vida para a gestão de ativos digitais, elaborando modelos de ciclo de vida específicos.

Com esse entendimento, a curadoria digital de dados implica em gestão, uma vez que atua no sentido de garantir a preservação e o armazenamento em longo prazo. Assim, uma gestão ativa de dados garante a qualidade desses dados e possibilita o acesso, o compartilhamento, o reuso e a abertura dos dados gerenciados (SILVA, 2019).

2.2 Preservação digital

A Carta para a Preservação do Patrimônio Digital, aprovada e adotada pela Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO, 2021), em 2003, manifesta a necessidade de os Estados-Membros, incluindo o Brasil, estabelecerem políticas e ações para proteger o patrimônio digital.

Assim como a Carta para a Preservação do Patrimônio Digital, no cenário internacional, há várias iniciativas voltadas para a preservação digital, sobretudo no que tange ao estabelecimento de boas práticas. Em alguns países, há projetos de pesquisa que

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

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

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

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

exploram questões de confiança e confidencialidade de registros e dados em ambientes *on-lines*. A título de exemplo, destaca-se um projeto multinacional e interdisciplinar, denominado *International Researchon Permanent Authentic Records in Electronic Systems* (INTERPARES). O principal objetivo do InterPARES é gerar os quadros teóricos e metodológicos para desenvolver políticas, procedimentos, regulamentos, normas e legislação locais, nacionais e internacionais, a fim de garantir a confiança pública baseada em evidências de boa governança, uma economia digital forte e uma memória digital persistente (INTERPARES, 2021).

Ainda no contexto internacional, reconhecida no domínio da Biblioteconomia, a *International Federation of Library Association* (IFLA) desenvolve vários padrões para a gestão dos dados bibliográficos e para o acesso à informação. Ademais, organizações de destaque contribuem para garantir a preservação digital, tal como o *Consultative Comitee for Space Systems* (CCSDS), que iniciou um esforço junto à *Intenational Organization for Standardization* (ISO), com o intuito de “[...] desenvolver um conjunto de normas capazes de regular o armazenamento a [sic] longo prazo das informações espaciais” (FERREIRA, 2006, p. 27). O autor relata que, desse esforço, nasceu o modelo de referência *Open Archival Information System* (OAIS).

No que tange aos modelos para preservação, o OAIS visa “[...] a identificar os componentes funcionais que deverão fazer parte de um sistema dedicado à preservação digital [...]”. Além disso, o modelo descreve as interfaces internas e externas do sistema e os objetos de informação que são manipulados no seu interior. Esse modelo foi aprovado como uma norma internacional, no ano de 2003 (FERREIRA, 2006, p. 27).

No cenário brasileiro, desde 2002, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) vem citando a preservação digital como um dos temas mais importantes relativos à missão do Instituto. Essa preocupação se deve à emergência dos dados em formato digital (REDE CARINIANA, 2014).

Historicamente, o Conselho Nacional de Arquivos (CONARQ, 2005, p. 2) vem alertando que a preservação das informações digitais, em longo prazo, tem sido seriamente ameaçada. Tal ameaça é resultado de dois diferentes fatores: “[...] a vida curta das mídias e o ciclo cada vez mais rápido de obsolescência dos equipamentos de informática, dos softwares e dos formatos”. No que tange à obsolescência, Boeres (2004) afirma que a informação digital é frágil, fugaz e necessita de gerenciamento cuidadoso desde o momento da sua criação. A autora acrescenta a importância da elaboração e da publicação de uma política estratégica de modo a garantir a preservação.

Portanto, considerando a perspectiva de Boeres (2004) e do Conarq (2005), cabe a atenção quanto ao modo com que tem sido realizada a gestão e a preservação dos acervos

digitais, tendo em vista preservar e proteger os dados e os documentos produzidos nas instituições. Nesse sentido, o IBICT empregou esforços e, em 2014, instituiu a Rede Cariniana. Desde então, a Cariniana apresenta-se como uma alternativa para organizações que desejam preservar seus acervos, oferecendo a preservação digital de documentos eletrônicos brasileiros, com o objetivo de garantir o acesso continuado em longo prazo (REDE CARINIANA, 2014).

A preservação digital, segundo Márdero Arellano (2008), precisa ser uma responsabilidade compartilhada, principalmente pelo grande volume de informação digital que está sendo produzido e pela natureza da tecnologia digital. Ela pode ser definida como um “conjunto de ações gerenciais e técnicas exigidas para superar as mudanças tecnológicas e a fragilidade dos suportes, garantindo o acesso e a interpretação dos documentos digitais pelo tempo que for necessário” (E-ARQ BRASIL, 2020, p. 220). Esse tipo de preservação pode ser caracterizado, também, como uma etapa de um longo processo, constituindo-se a parte mais longa e também a última do ciclo de gerenciamento de objetos digitais (REDE CARINIANA, 2014).

O objeto digital é “[...] todo e qualquer objeto de informação que possa ser representado através de uma sequência de dígitos binários [...]” (FERREIRA, 2006, p. 21). O autor relata que essa definição contempla a informação gerada em um contexto tecnológico digital (objeto nato-digital), como também a informação digital produzida com o uso de suportes analógicos (objetos digitalizados).

A fim de assegurar a preservação de objetos dessa natureza, recorre-se às tarefas de preservação digital, ou seja, o conjunto de procedimentos que engloba “[...] o manuseio e o armazenamento da mídia digital, a cópia da informação contida, a migração para novas mídias e a preservação da integridade da informação digital” (REDE CARINIANA, 2014, não paginado). No entendimento de Ferreira (2006), a estratégia de preservação digital é uma abordagem técnica que garante o acesso continuado à informação existente em formatos digitais.

Nesse sentido, Moore (2008, p. 63, tradução nossa) explica que “[...] as informações geradas no passado são enviadas para o futuro pelo ambiente de preservação atual [...]”. O autor também reforça a importância do tema, sobretudo os princípios que garantirão a comunicação da informação contida nos artefatos tecnológicos. Logo, “[...] a comprovação de que o ambiente de preservação preserva a autenticidade e a integridade ao realizar a comunicação constitui uma teoria da preservação digital [...]” (MOORE, 2008, p. 63, tradução nossa), teoria essa que precisa ser aprofundada, a partir das práticas e estratégias utilizadas pelas instituições no trabalho com a informação digital produzida. Nessa perspectiva,

reforça-se a importância de assegurar o acesso e a recuperação da informação científica como fundamento para a pesquisa acadêmica.

2.2.1 Política e plano de preservação digital

Uma política institucional é constituída por diretrizes e por princípios gerais que definem parâmetros a serem seguidos nas ações e estratégias da instituição. Essas ações e estratégias estão voltadas para o cumprimento da missão e o alcance da visão institucional, estando fundamentada em limites éticos estabelecidos por valores compartilhados (ARQUIVO NACIONAL, 2019).

Souza (2020), a partir da revisão de literatura realizada no decorrer da elaboração de sua tese, sinaliza dois elementos-chave quanto ao processo de oficialização das ações de preservação: a política e o plano de preservação digital. A autora relata que poucas IES possuem uma política de preservação de documentos e nenhuma desenvolveu um plano de preservação.

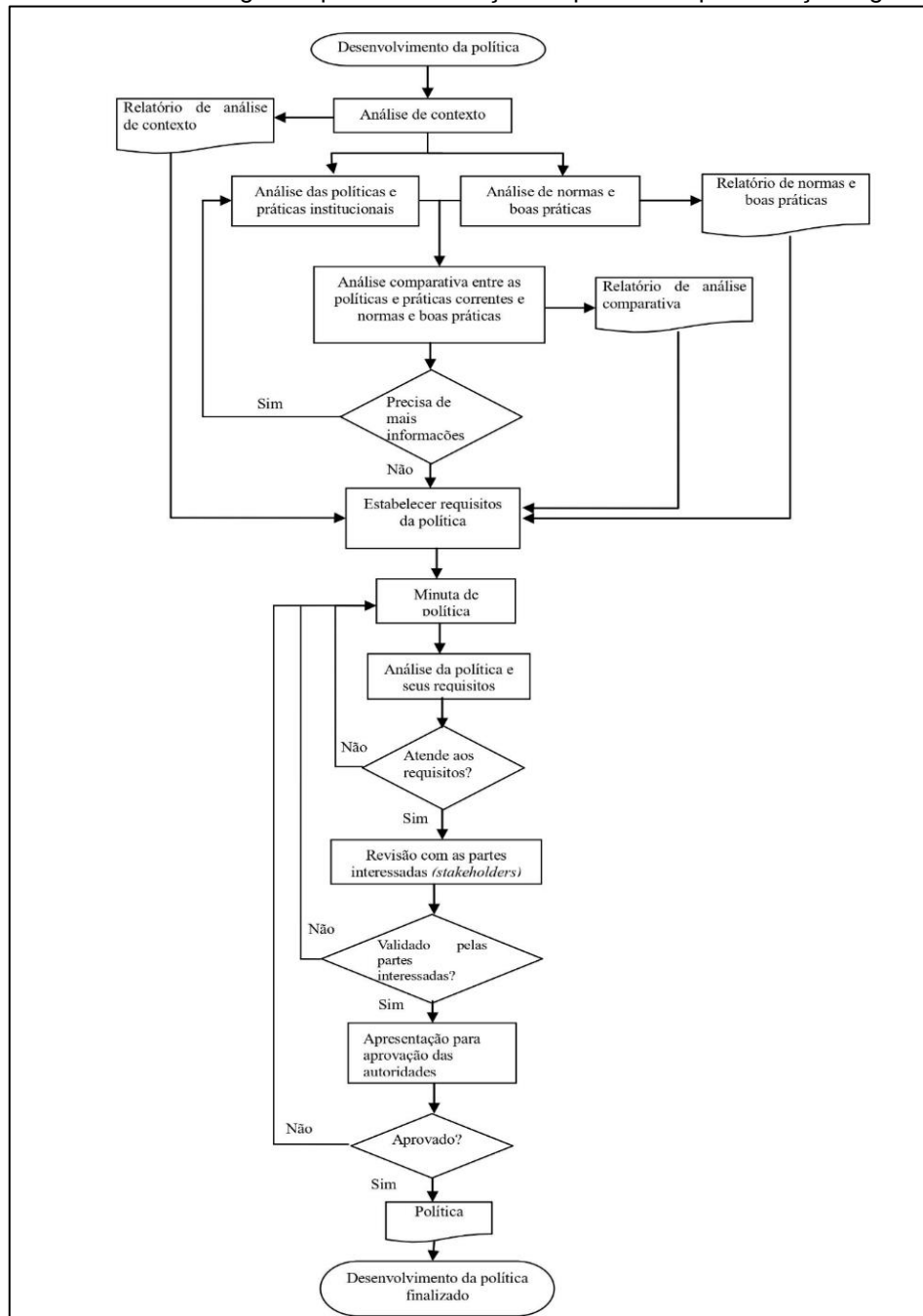
O plano de preservação digital é elaborado em conformidade com a política de preservação e apresenta o detalhamento das estratégias e das ações de preservação a serem adotadas (ARQUIVO NACIONAL, 2019). Portanto, as instituições precisam ter urgência no desenvolvimento de documentos que oficializam o compromisso institucional para com a preservação dos conteúdos digitais produzidos (SOUZA, 2020). Entende-se que essa preocupação é extensiva aos dados científicos, visto que eles constituem uma parte fundamental no desenvolvimento da ciência, conforme pontuado no estudo de Sayão e Sales (2016).

As infraestruturas de **gestão de coleções de dados de pesquisa precisam estar ancoradas em uma política formal** que defina todos os procedimentos do fluxo de curadoria: procedimentos práticos, formatos, metadados, arquivamento, preservação e segurança, licenças e, sobretudo, o tratamento que deve ser conferido ao compartilhamento de dados sensíveis, pessoais, confidenciais, de informações de interesse comercial ou que serão importantes nos processos de patenteamento, além das questões associadas à proteção de direitos de propriedade intelectual (SAYÃO; SALES, 2016, p. 69, grifo nosso).

Em virtude da importância da preservação digital e da formulação da política e do plano, como mencionado por Sayão e Sales (2016) e Souza (2020), constata-se que a problemática da preservação deve fazer parte das preocupações das entidades, com esforços direcionados à publicação desses documentos (AMORIM, 2008). Um programa de preservação, por exemplo, para ser consistente, “[...] deve garantir o acesso ao recurso para sempre, mas se o conteúdo estiver protegido por direitos de autor e restrições de outra natureza, isso deve ser assegurado, tal como no período ‘ativo’ do documento” (AMORIM, 2008, não paginado).

A partir dessa importância e da complexidade do tema, o Arquivo Nacional (2019) desenvolveu um roteiro que conduz a criação de uma política de preservação digital. Nesse roteiro, no formato de fluxograma, apresentam-se os passos a serem seguidos para o desenvolvimento da política de preservação de documentos, conforme apresentado na Figura 1.

FIGURA 1 – Fluxograma para a elaboração da política de preservação digital



Fonte: Arquivo Nacional (2019, p. 18).

Pelo que consta na Figura 1, o desenvolvimento da política é permeado por diversos processos integrados, com avaliação contínua, o que possibilita a validação pelas equipes, tendo em vista garantir consistência ao documento quando publicado. Cabe lembrar

que as recomendações oferecidas pelo Arquivo Nacional (2019) são fundamentais na constituição das diretrizes do projeto InterPARES.

2.3 Gestão de dados científicos

Os dados científicos são representados por todas as evidências que um investigador necessita para validar suas conclusões com o término da pesquisa (SILVA, 2019). O autor entende que os dados constituem os componentes ou insumos básicos a partir dos quais será evidenciada a informação e o conhecimento.

Os dados científicos podem ser classificados quanto à origem, natureza ou fase da pesquisa. Quanto à origem, Sayão e Sales (2015) propõem as seguintes classificações:

- **Observacionais:** são dados obtidos por meio de observações diretas, que podem ser associadas a lugares e tempos específicos. Por sua natureza instantânea, guardam uma importância crítica que os qualifica como registros históricos, pois não podem ser coletados uma segunda vez e, portanto, devem ser submetidos a processos de curadoria que os preservem para sempre;
- **Computacionais:** são resultados da execução de modelos computacionais ou de simulações. Para essa categoria de dados, a preservação por longo prazo pode não ser necessária, visto que os dados podem ser replicados ao longo do tempo;
- **Experimentais:** são provenientes de situações controladas em bancadas de laboratórios. Assim, dados experimentais provenientes de experimentos que podem ser precisamente reproduzidos não necessitam ser armazenados indefinidamente.

No que tange à natureza, os dados podem ser: números, imagens, vídeos, áudios, softwares, algoritmos, equações, animações ou modelos e simulações (SAYÃO; SALES, 2015). Por sua vez, no que concerne à fase da pesquisa, esses autores indicam a existência de três tipologias, a saber:

- **Dados brutos, crus ou preliminares (*raw data* em inglês):** são dados provenientes dos instrumentos científicos, de modo direto;
- **Dados derivados:** originam-se a partir dos resultados do processamento ou combinação de dados brutos ou de outros dados;
- **Dados canônicos ou dados referenciais:** são coleções de dados, consolidadas e arquivadas, na maioria das vezes, em grandes centros de dados, por exemplo, sequência genética, estrutura química, dentre outros locais. Os dados canônicos também podem ser representados por dados produzidos por órgãos do governo, pois “[...] se tornam dados de pesquisa uma vez que tenham sido modificados, processados ou expandidos” (SAYÃO; SALES, 2015, p. 9).

Com a preponderância do suporte digital no contexto científico, os dados deixam de ser um mero subproduto da atividade de pesquisa e se tornam protagonistas na geração de novos conhecimentos e descobertas (SAYÃO; SALES, 2016). Sobre isso, os autores esclarecem que a geração de dados científicos nos vários domínios das instituições (pequenos laboratórios, equipes e pesquisadores individuais), chamados de pequena ciência⁵, constituem a “cauda longa” da ciência. Assim, na *small science* (pequena ciência), “[...] o planejamento, a gestão, a preservação e a reutilização dos dados produzidos nessas pequenas dimensões apresentam-se desafiadores devido à sua heterogeneidade e singularidade [...]” (BARROZO; ALMEIDA, 2021, p. 3).

Corroborando as ideias de Sales e Sayão (2018), a Fundação Oswaldo Cruz (FIOCRUZ, 2021, não paginado) destaca que “[...] os investimentos para implementar a gestão ativa de dados priorizam as pesquisas da “cauda longa”, cujos dados não são cuidadosamente indexados e armazenados [...]”, o que possibilita um maior risco de se tornarem inacessíveis e subutilizados. A instituição salienta que a atividade começa no planejamento da coleta e/ou geração de dados, “[...] passando por processos de limpeza, curadoria, indexação e registro das transformações até o seu armazenamento e preservação de longo prazo” (FIOCRUZ, 2021, não paginado).

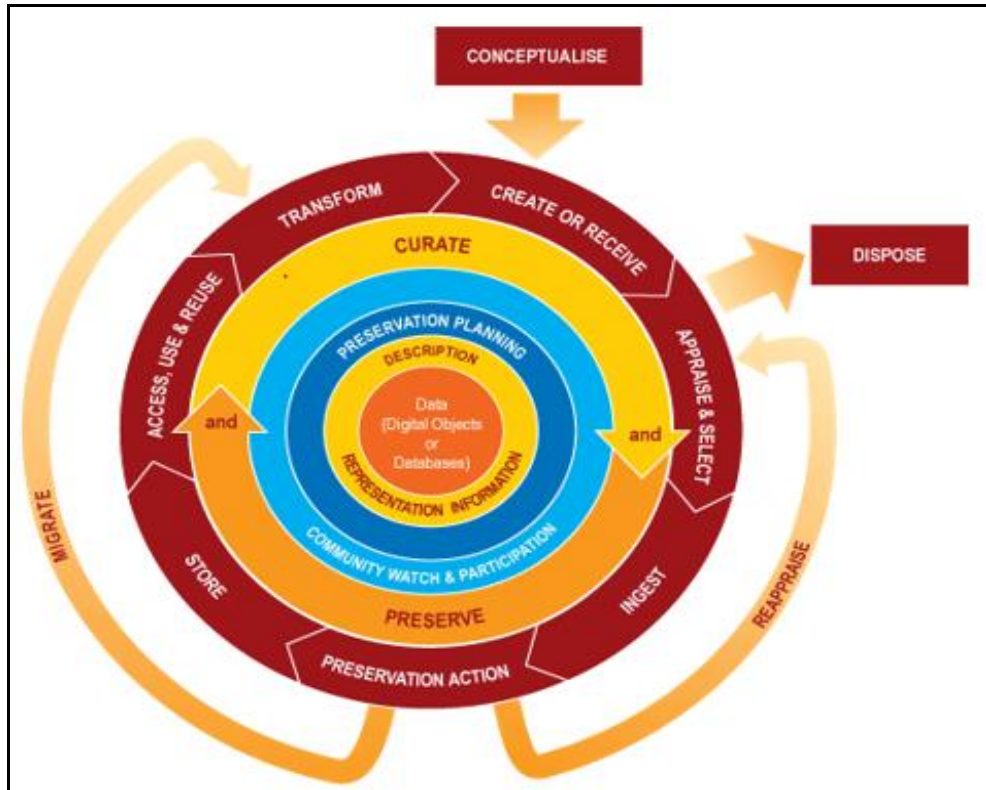
Silva (2019) reitera que a importância dos modelos de ciclo de vida é que proporciona uma estrutura para a consideração das muitas operações que devem ser realizadas durante o uso dos dados. Assim, cabe escolher o modelo de ciclo de vida dos dados de pesquisa, capaz de atender às particularidades e aos objetivos determinados, de acordo com os domínios de conhecimentos específicos (SILVA, 2019).

A curadoria e a preservação descrevem com detalhes a maioria das ações do modelo, “[...] mas também servem para representar a execução da gestão das ações planejadas e administrativas de apoio à curadoria” (SILVA, 2019, p. 51). O modelo proposto por Higgins (2008, 2011), intitulado *Curation Life Cycle Model* (Modelo de Ciclo de Vida de Curadoria), propõe um método de gerenciamento documental com o objetivo de alinhar as tarefas da curadoria às etapas do ciclo.

Semelhante ao modelo desenvolvido por Higgins (2008, 2011), o DCC elaborou modelo próprio. Nesse modelo, são elencadas oito ações necessárias que garantam a gestão ativa dos dados, conforme ilustrado na Figura 2.

⁵ Em inglês, *small science*.

FIGURA 2 - Modelo do Ciclo de Vida dos Dados – DCC



Fonte: Higgins (2008, 2011) e DCC (2022).

As oito ações que compõem o modelo sugerido pelo DCC são detalhadas pela FIOCRUZ (2021). Assim, cada ação é constituída por diversas atividades, tais como:

- **Conceituar e planejar:** conceitualizar e planejar a coleta de dados, incluindo os métodos de captura e as opções de armazenamento;
- **Produzir ou receber dados:** criar dados e seus respectivos metadados administrativos, descritivos, estruturais e técnicos. Se receber dados de terceiros, faça de acordo com políticas de coleta, de criadores de dados, outros arquivos, repositórios ou centros de dados;
- **Avaliar os dados e selecionar:** avaliar os dados e selecionar aqueles que passarão por curadoria e por preservação de longo prazo;
- **Alimentar/importar dados:** transferir os dados para arquivos, repositórios, centro de dados ou outros espaços de custódia, aderindo a guias, políticas e requerimentos legais;
- **Preservar:** realizar ações de preservação para garantir que os dados permaneçam autênticos, confiáveis, íntegros e utilizáveis;
- **Armazenar:** armazenar os dados de maneira segura, seguindo padrões relevantes;

- **Acessar, usar e reusar:** garantir que os dados estejam acessíveis para usuários designados pela instituição ou outros potenciais reutilizadores, registrando informação sobre eles e, quando necessário, aplicando controles e procedimentos de autenticação;
- **Transformar:** criar novos dados a partir do original mediante a migração para diferentes formatos, criação de subconjuntos, seleção ou *query* ou por meio da criação de novos resultados derivados (FIOCRUZ, 2021).

A partir das atividades requeridas no modelo do DCC, entende-se que a curadoria de dados corresponde ao conjunto de ações voltadas para o gerenciamento de dados científicos durante o ciclo de vida desses dados, no decorrer da pesquisa. Esse processo “[...] envolve manter, preservar e adicionar valor aos dados” (SAYÃO; SALES, 2015, p. 80). Os autores tomam como referência o modelo *DataONE*, específico para conduzir as atividades dos bibliotecários e dos pesquisadores no que tange à gestão dos dados científicos. O modelo *DataONE* está disposto na Figura 3.

FIGURA 3 - Modelo do Ciclo de Vida dos Dados - *DataONE*



Fonte: Sayão e Sales (2015).

No modelo do ciclo de vida proposto pelo *DataONE*, também são apresentadas oito etapas. Sayão e Sales (2015) detalham cada uma das etapas, sendo elas:

- **Planejar:** descrição dos dados que serão compilados e como eles serão gerenciados e disponibilizados para acesso durante o seu tempo de vida;

- **Coletar:** as observações são realizadas, de modo manual, ou por sensores ou outros instrumentos, e os dados são colocados em formato digital;
- **Assegurar:** a qualidade dos dados é assegurada por meio de controles e inspeção;
- **Descrever:** a descrição é realizada de modo acurado usando padrões de metadados apropriados;
- **Preservar:** processo de submissão a um arquivo apropriado (centro de dados), tendo em vista garantir a preservação de longo prazo;
- **Descobrir:** dados potencialmente úteis são descobertos e acessados em conjunto com informações relevantes sobre os dados (metadados);
- **Integrar:** diversas fontes são combinadas para formar um conjunto de dados homogêneo que pode ser prontamente analisado;
- **Analisar:** dados são analisados (SAYÃO; SALES, 2015).

Importante frisar que as etapas a serem cumpridas dependerão do tipo de projeto. Assim, apenas partes do ciclo de vida podem ser executadas (SAYÃO; SALES, 2015). No entanto, as etapas do ciclo vão sendo executadas, de modo integrado, o que confirma que a gestão dos dados científicos também é uma maneira de manter a integridade da pesquisa (SILVA, 2019). Logo, para uma efetiva gestão de dados de pesquisa, é fundamental haver planejamento, primeira etapa do ciclo de vida a ser desenvolvida durante o delineamento da pesquisa, nas ações iniciais. Para formalização das ações e dos compromissos referente aos dados, pode-se adotar um Plano de Gestão de Dados (PGD). (SAYÃO, SALES, 2015).

O PGD caracteriza-se como “[...] um documento formal que descreve todo o ciclo de vida dos dados, desde sua coleta até a documentação completa do processo de pesquisa” (SILVA, 2019, p. 55). Esse instrumento é considerado por Sayão e Sales (2015) como um documento burocrático e, como tal, sua elaboração não deve ser pensada como uma mera tarefa administrativa. Esses últimos autores reforçam que, no PGD, deve ser explicitado o que realmente é necessário à preservação, compartilhamento e reuso dos dados.

Um PGD pode ser criado a partir de um template elaborado pela própria instituição onde a pesquisa será desenvolvida ou pelas agências de financiamento do projeto. Podem ser desenvolvidos modelos, cujos elementos sejam recomendados pelo *DataONE* ou por outras organizações internacionais que apontam os requisitos necessários à criação de um plano de gestão (SAYÃO; SALES, 2015). Os autores apresentaram um modelo a ser seguido, quando da criação do PGD, conforme descrito na Figura 4.

FIGURA 4 - Proposta para elaboração do Plano de Gestão de Dados

RESUMO DAS INFORMAÇÕES RECOMENDADAS PARA COMPOR O PLANO DE GESTÃO DE DADOS	
DESCRIÇÃO DOS DADOS	<ul style="list-style-type: none"> • Tipo de dados produzidos pela pesquisa • Quantidade de dados que será coletada • Como os dados serão coletados • Como os dados serão processados • Formatos de arquivo que serão usados • Como os arquivos serão nomeados • Medidas para garantir a qualidade dos dados • Coleções de dados disponíveis • Dados existentes que serão usados • Preservação de curto prazo • Responsáveis pela gestão de curto prazo
METADADOS	<ul style="list-style-type: none"> • Metadados necessários • Como os metadados serão criados • Esquema que será usado
POLÍTICA DE ACESSO, COMPARTILHAMENTO E REUSO	<ul style="list-style-type: none"> • Obrigações de compartilhamento • Como os dados serão compartilhados • Questões éticas e de privacidade • Propriedade intelectual e copyright • Usos futuros e usuários potenciais • Citação dos dados
GESTÃO DO ARQUIVAMENTO DE LONGO PRAZO: PRESERVAÇÃO DIGITAL DOS DADOS DE PESQUISA	<ul style="list-style-type: none"> • Que dados serão preservados • Onde os dados serão arquivados • Necessidade de formatação dos dados • Responsável pelo contato com o centro de dados
ORÇAMENTO: CUSTOS ENVOLVIDOS NA GESTÃO DE DADOS	<ul style="list-style-type: none"> • Custos previstos • Como os custos serão cobertos

Fonte: Sayão e Sales (2015, p. 26).

Como se percebe, para elaboração de um PGD, a instituição precisa, *a priori*, considerar cinco elementos essenciais: modo de descrição dos dados, metadados, política, processo de gestão e orçamento. Para constituição de cada um desses elementos, um conjunto de ações e requisitos deve ser detalhado no plano (SAYÃO; SALES, 2015).

2.3.1 Princípios FAIR e repositório de dados científicos

Os Princípios FAIR foram definidos pelo grupo multidisciplinar *Community of Scholars, Librarians, Archivists, Publishers and Research Funders* (FORCE11), responsável por discussões sobre o futuro e a modernização da comunicação científica (WILKINSON *et al.*, 2016). Publicados em 2016, na *Scientific Data*, esses princípios são direcionados à fundamentação da prática da gestão de dados científicos nas instituições de pesquisa. Assim, a comunidade FORCE11 descreve o FAIR como um conjunto de princípios orientadores para tornar os dados localizáveis, acessíveis, interoperáveis e reutilizáveis, favorecendo a boa gestão de dados de pesquisa (FORCE11, 2021).

Um dos grandes desafios da ciência com o uso intensivo de dados é facilitar a ampliação do conhecimento. Nesse sentido, a reutilização e o acesso aos dados auxiliarão os pesquisadores em suas descobertas, possibilitando “[...] o acesso, a integração e a análise dos dados científicos apropriados para tarefas e seus algoritmos e fluxos de trabalho associados” (FORCE11, 2021, não paginado, tradução nossa).

Nesse contexto, a iniciativa FAIR objetiva que “[...] dados fragmentados e desconectados sejam encontráveis, acessíveis, interoperáveis e, portanto, reutilizáveis por máquinas e pessoas” (GLOBAL OPEN FAIR INITIATIVE BRASIL, 2021, não paginado). Tais princípios incluem padrões, protocolos, políticas, diretrizes e boas práticas, cujo propósito é tornar os dados disponíveis para todos acessarem, usarem e compartilharem (GLOBAL OPEN FAIR INITIATIVE BRASIL, 2021). No Quadro 2, são apresentados os princípios e algumas aplicações.

QUADRO 2 – Princípios FAIR e suas aplicações

	Princípio	Aplicações
ENCONTRÁVEL	E.1 (meta): dados devem ter identificadores globais, únicos e persistentes.	Adotar identificador único persistente tanto para o conjunto de dados quanto para os metadados (ex.: DOI, ARK, RRID e PID).
	E.2: dados devem ser descritos utilizando metadados ricos (impacta diretamente R1)	O conjunto de dados deve ser descrito por metadados ricos o suficiente para que, uma vez indexados em um mecanismo de busca, possam ser encontrados mesmo sem o seu identificador único persistente.
	E.3: metadados devem incluir, de modo claro e explicitamente, os identificadores dos dados que descrevem.	Como não podemos prever que os dados e seus metadados estejam sempre juntos, a associação entre eles deve ocorrer pela inclusão do identificador persistente dos dados nos metadados.
	E.4 (meta): dados devem ser registrados ou indexados em mecanismos de busca.	Para que os dados sejam encontrados, seus metadados devem ser indexados em mecanismos de busca (<i>searchengine</i>), que possibilitem aos computadores e aos usuários encontrá-los com facilidade.
ACESSÍVEL	A.1 (meta): dados devem ser recuperáveis pelos seus identificadores usando protocolo de comunicação padronizado.	Com o identificador persistente do conjunto de dados e/ou de seus metadados, o usuário deverá recuperá-los mais facilmente por meio de protocolos de comunicação padronizados (ex.: HTTP ou FTP).
	A.1.1: o protocolo deve ser aberto, gratuito e universalmente implementável.	Independente de licenciamento dos dados e dos metadados, o protocolo de comunicação usado para dar acesso a eles deve ser aberto, gratuito e passível de ser implementado por qualquer interessado (ex.: HTTP ou Ftp).
	A.1.2: o protocolo deve permitir procedimentos de autenticação e autorização, quando necessário.	Dependendo das restrições de acesso aos dados e/ou metadados, um mecanismo de autenticação e autorização para o acesso deve ser liberado pelo protocolo de comunicação (ex.: os repositórios confiáveis oferecem essa opção).

	Princípio	Aplicações
	A.2: metadados devem ser acessíveis, mesmo quando os dados não estiverem mais disponíveis.	É preciso existir um conjunto de estratégias de preservação para dados e metadados. Os metadados devem ser sempre acessíveis, possibilitando a criação de índices para o conjunto de dados atuais vigentes e aqueles não mais disponíveis.
INTEROPERÁVEL	I.1 (meta): dados devem ser representados por meio de uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento.	Para que se possa representar dados e metadados, devem ser adotadas linguagens de representação do conhecimento que sejam padronizadas, acessíveis e amplamente aplicáveis (ex.: RDF, XML, DICOM, dentre outras).
	I.2 (meta): dados devem usar vocabulários de acordo com os princípios FAIR.	Dados e metadados devem possuir referências a vocabulários e/ou ontologias que os descrevem. Devemos garantir que esses também sigam os Princípios FAIR.
	I.3 (meta): dados devem incluir referências qualificadas para outros (meta) dados.	É necessário referenciar o conjunto de dados, possibilitando que aqueles gerados a partir de outros conjuntos sejam interligados, assegurando a ligação semântica entre eles.
REUTILIZÁVEL	R.1 (meta): dados são descritos com uma pluralidade de atributos precisos e relevantes.	Prover metadados descritos com alto nível de detalhes que permitam ao pesquisador avaliar a possibilidade do seu reuso, bem como a adequação às suas necessidades.
	R.1.1 (meta): dados devem ser disponibilizados com licenças de uso claras e acessíveis.	É fundamental que o responsável pelos dados e metadados defina, explicitamente, quem pode ter acesso a eles, com que finalidade e sob quais condições. Essas informações são definidas por meio de licenças de uso.
	R.1.2 (meta): dados devem estar associados à sua proveniência.	Especificar a proveniência (linhagem) dos dados é importante não só para que o pesquisador possa avaliar a utilidade dos dados ou metadados, mas também para que possa atribuir o devido crédito a quem produziu, manteve ou editou esses dados.
	R.1.3 (meta): dados devem estar alinhados com padrões relevantes do seu domínio.	Além de atender aos padrões específicos da área de cada comunidade, deve-se dar atenção às boas práticas de arquivamento e de compartilhamento específicos da área de pesquisa.

Fonte: Elaborado pela autora (2021).

No processo científico, a descrição de procedimentos e métodos utilizados nas pesquisas sempre foi primordial e atende ao princípio inerente à publicação científica. Com base nos Princípios FAIR, isso quer dizer que “[...] qualquer pesquisador deve ser capaz de verificar, replicar ou reproduzir a pesquisa realizada, até mesmo utilizar os resultados de pesquisas para promover novos avanços no conhecimento científico e tecnológico” (DUDZIAK, 2016, não paginado). A partir desse argumento, entende-se que

[...] o **conhecimento científico como um bem da humanidade** demandam [sic] que metodologias, instrumentos, softwares e **dados estejam abertos** para **garantir o princípio da reprodutibilidade e de autocorreção da**

ciência e a transparência de seus fluxos e de sua trajetória de erros e acertos (SAYÃO; SALES, 2016, p. 67, grifo nosso).

Curty (2019) salienta acerca da importância da reutilização dos dados. Pondera que “[...] o efetivo reuso requer não somente a possibilidade de identificar e acessar dados preexistentes, alinhados aos interesses dos pesquisadores; [sic] mas principalmente que os dados atendam a condições mínimas de reusabilidade” (CURTY, 2019, p. 178). Salienta-se que a reusabilidade é determinada por características técnicas de conformidade, preconizadas pelo FAIR, “[...] e considerando os propósitos de sua pesquisa e com base em seu melhor julgamento sobre os atributos e condições de qualidade dos dados” (CURTY, 2019, p. 179).

O compartilhamento dos dados de pesquisa mantém relações diretas com o reuso desses dados, tanto no sentido de promover a validação, quando no sentido de oportunizar novas interpretações, gerando novos conhecimentos (CAREGNATO; ROCHA; GABRIEL JUNIOR, 2021). Todavia, mesmo que a disponibilidade dos dados seja condição “[...] para que o reuso possa materializar-se, o reuso não pode ser simplesmente presumido como consequência certa do compartilhamento” (CURTY, 2019, p. 178).

Vidotti, Torino e Coneglian (2021) afirmam que a reutilização fica facilitada na medida em que quão bem estão descritos os dados e os metadados, incluindo informações sobre os direitos de uso, a proveniência e o contexto dos dados, torna-se possível combinar e reutilizar os dados, incluindo outras instâncias. Corroborando as ideias de Vidotti, Torino e Coneglian (2021), Sayão e Sales (2021) afirmam que o grau de reusabilidade coloca em destaque a necessidade de licenças⁶ apropriadas às condições de reuso requerido.

Esses últimos autores reafirmam que a gestão de dados de pesquisa se manifesta, atualmente, com alto foco de interesse e constitui um dos maiores desafios para as organizações de pesquisa. Portanto, no contexto das universidades, cuja gestão de dados é decorrente das funções de ensino, pesquisa e extensão,

são muitas formas de planejar e executar as tarefas de gestão de dados que variam de acordo com referências objetivas como graus de investimento, sistemas técnicos disponíveis, volume e tipo de dados e de como a gestão de dados está integrada aos seus fluxos de trabalhos e processos; e com percepções mais subjetivas como cultura disciplinar e prestígio acadêmico (SAYÃO; SALES, 2021, p. 231).

⁶ O desenvolvimento de estudos científicos depende de acesso a outros trabalhos acadêmicos. O licenciamento aberto dos escritos é a forma ideal de conceder-lhes a flexibilidade, liberdade de uso e acessibilidade de que necessitam sem prejudicar os autores moral ou patrimonialmente (BRANCO; BRITTO. 2013, p.151).

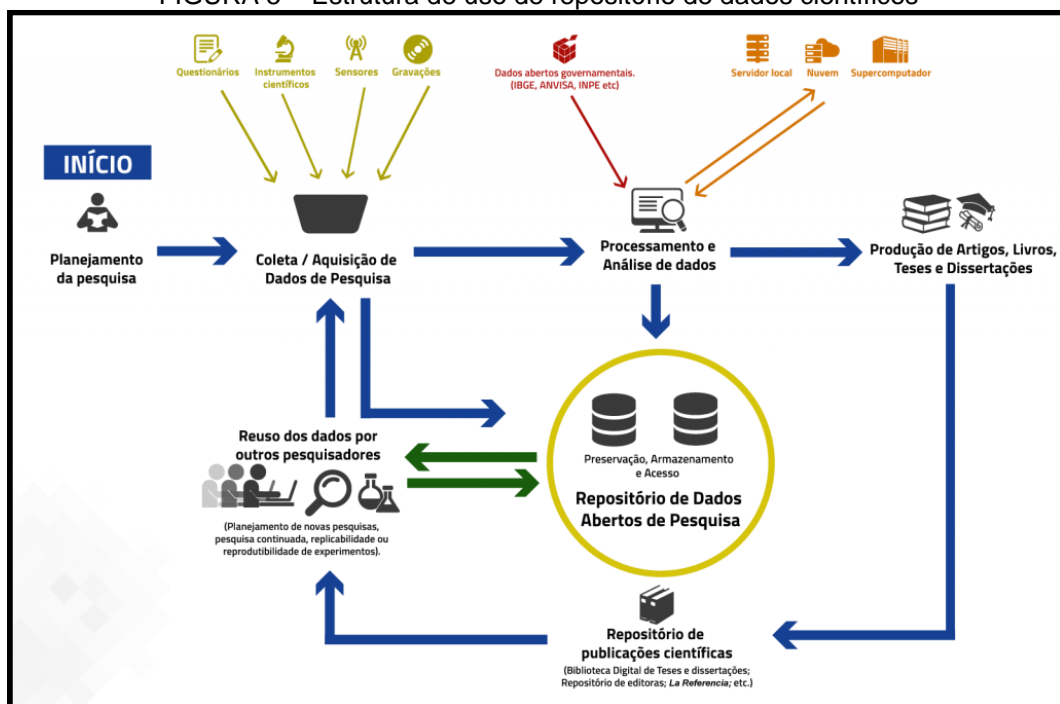
O discurso sobre a gestão de dados nas universidades implica no papel dessas instituições quanto à garantia da preservação. Sobre isso, os autores supracitados reforçam que a governança e a preservação dos dados requerem um compromisso institucional e

[...] a consolidação dos serviços de gestão de dados reflete o nível de aceitação organizacional incorporada a eles e o grau de planejamento das várias ações necessárias: orçamento sustentável em vigor, política de dados apropriada, conexão orgânica com as comunidades-alvo, conformidade com os códigos éticos e legais, alinhamento com os objetivos estratégicos institucionais e uma estratégia de desenvolvimento que considere os percursos possíveis para cada instituição [...] (SAYÃO; SALES, 2021, p. 231).

Esse discurso acerca do compromisso com a preservação é complementado quando se promove a função dos repositórios institucionais. Sayão e Sales (2016) apontam os repositórios de dados como um importante recurso informacional, constituindo parte essencial das infraestruturas de pesquisa em escala global. Os repositórios, discorrem Sayão e Sales (2016), tornam as práticas das pesquisas visíveis e abertas, cujo intuito inicial é que toda a sociedade acesse uma parcela importante da atividade de pesquisa, sobretudo quando esses ambientes são acessados em vários momentos da pesquisa, de modo que os interessados acompanhem o percurso do estudo.

Semelhante a Sayão e Sales (2016), Paganine *et al.* (2020) acreditam no potencial dos repositórios de dados científicos, destacando que esses ambientes têm sido sistematicamente propostos pela literatura, como uma ferramenta adequada para o compartilhamento, armazenamento, organização e divulgação dos dados científicos. A Figura 5 ilustra a estrutura funcional de um repositório dessa natureza.

FIGURA 5 – Estrutura de uso de repositório de dados científicos



Fonte: Rede Nacional de Ensino e Pesquisa (RNP) (2019).

Constata-se que os repositórios de dados científicos passam por diversos desafios, sobretudo no Brasil. A realidade brasileira aponta um baixo número de repositórios, sem a devida padronização e sem conformidade com os padrões internacionais, por conseguinte, sem as certificações conferidas pelas diretrizes dos órgãos responsáveis (PAGANINE; AMARO, 2020).

Paganine *et al.* (2021) mencionam o esforço do IBICT e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), instituições vinculadas ao Ministério da Ciência, Tecnologia e Inovações (MCTI), quanto à implementação dos repositórios de dados científicos dessas instituições, denominados, respectivamente, de Aleia e de LattesData. Os autores relatam que as criações do Aleia e do LattesData também foram motivadas pelo compromisso 3, do 4º Plano de Ação Nacional da Parceria para Governo Aberto (OGP) do Brasil, por meio de um Acordo de Cooperação Técnica (ACT) firmado entre CNPq e IBICT, em dezembro de 2019. A Figura 6 apresenta a interface inicial do Aleia.

FIGURA 6 – Repositório de dados científicos do IBICT



Fonte: IBICT (2021).

O repositório Aleia tem por objetivo “[...] prover uma ferramenta com as funcionalidades de registrar, reunir, organizar, disseminar, compartilhar e preservar os dados científicos de pesquisas [...]”. Essas pesquisas podem ser “[...] executadas por colaboradores do IBICT e conjuntos de dados científicos externos ao órgão, mas de comunidades científicas específicas” (PAGANINE *et al.*, 2021, p. 173). Já o Lattes-Data tem por objetivo principal servir como ferramenta oficial que possibilite a seus pesquisadores financiados realizarem depósitos

de conjuntos de dados que surgiram como resultado dos projetos desenvolvidos com recursos do CNPq (LATTESDATA, 2021).

Paganine *et al.* (2021) reiteram que os dois repositórios pretendem abranger conjuntos de dados de pesquisadores de diversas instituições, incluindo diversas áreas e realidades. Nesse sentido, as instituições demonstram preocupação em gerenciar os conjuntos de dados da pesquisa brasileira, atendendo aos requisitos, normas e padrões estabelecidos em âmbito internacional e, assim, contribuir para o desenvolvimento da pesquisa brasileira.

Cabe salientar que o conteúdo produzido no âmbito das IES, de acordo com o artigo 6º da Lei de Direitos Autorais (LDA) não serão de domínio da União, dos Estados e demais entes federados as obras por eles simplesmente subvencionadas (BRASIL, 1998). Assim, a disponibilização de trabalhos acadêmicos e seus conjuntos de dados, ainda que seja plausível não se trata de uma obrigação, salvo os casos previstos em editais (agências de fomento) ou expresso em termos de compromisso (acordos formais).

3 METODOLOGIA

Esta seção apresenta o procedimento metodológico adotado neste estudo, caracteriza o ambiente delimitado e explicita as técnicas e os métodos utilizados. A seção também se reserva à apresentação da tipologia da pesquisa, dos instrumentos de coleta e de análise de dados adotados para construção da dissertação, assim como do procedimento ético para a coleta de dados.

3.1 Caracterização da pesquisa

A presente pesquisa possui caráter teórico-metodológico, efetivado pelo referencial teórico e pelo levantamento de dados com a aplicação de uma pesquisa de campo. Para tanto, utiliza uma abordagem indutiva, generalizando suas conclusões a partir dos dados levantados.

De natureza aplicada e de abordagem qualitativa, o estudo tenta compreender as práticas dos pesquisadores atuantes nos PPG da UFMG concernentes à curadoria e à preservação dos dados científicos. De acordo com Creswell (2010, p. 92), “[...] distinto da orientação teórica, estão os estudos qualitativos em que a teoria (ou alguma outra explicação ampla) torna-se o ponto final”. O autor explica que esse é um processo indutivo da construção do conhecimento a partir dos dados para temas amplos e para modelos generalizados da teoria.

Quanto aos objetivos, a investigação é exploratória e explicativa e, quanto aos procedimentos técnicos, bibliográfica, documental e de campo. Nessa última modalidade, utiliza-se como instrumento de coleta de dados o questionário semiestruturado e como técnica a entrevista semiestruturada, tendo em vista fortalecer o aprofundamento da pesquisa.

Assim, convencionou-se organizar o percurso metodológico em quatro fases, a saber:

- **Primeira fase:** consiste na pesquisa bibliográfica (fontes secundárias), conduzida por meio da consulta às fontes de informação como livros impressos e eletrônicos, revistas eletrônicas, arquivos oficiais, teses, dissertações e artigos;
- **Segunda e terceira fases:** aplica a pesquisa de campo (coleta de dados);
- **Quarta fase:** recorre à pesquisa documental.

A organização das fases aplicadas para o desenvolvimento do estudo é apresentada na Figura 7.

FIGURA 7 - Percurso metodológico da pesquisa



Fonte: Elaborada pela autora (2021).

O desenvolvimento dessas fases dependeu da aprovação do projeto junto ao Comitê de Ética em Pesquisa (COEP) da UFMG. O referido projeto foi submetido por meio da Plataforma Brasil, cujas coletas de dados foram iniciadas junto aos participantes somente após a tramitação legal e aprovação do projeto, conforme consta no Termo de Aprovação apresentado no Anexo C. A quarta fase, reservada à pesquisa documental (fontes primárias), foi realizada por meio da consulta aos planos, relatórios, manuais, políticas e resoluções para complementar informações não obtidas na coleta de dados. A partir desses procedimentos, foi possível desenvolver a análise dos dados.

Para análise de dados, recorreu-se à metodologia proposta pela organização *National Digital Stewardship Alliance* (NDSA), sendo uma matriz aspiracional para avaliação comparativa dos aspectos técnicos do programa de preservação digital, traduzida para o português por Laura Vilela Rodrigues Rezende, pesquisadora membra do grupo Estudos e Práticas de Preservação Digital (DRÍADE), grupo de pesquisa ligado à Rede CARINIANA e ao IBICT. Esta metodologia usa a Matriz Níveis de Preservação Digital⁷ (LOP) como instrumento para diagnosticar o grau de maturidade das práticas de preservação digital.

A LOP (2019) consiste de cinco áreas funcionais: armazenamento, integridade, controle, metadados e conteúdo. Ademais, o instrumento baseia-se em quatro níveis de preservação, do mais baixo ao mais alto nível, a saber: nível 1 (conhecer), nível 2 (proteger), nível 3 (supervisionar) e nível 4 (sustentar). O detalhamento da matriz utilizada consta registrado no Anexo B.

⁷ Conjunto hierárquico de recomendações sobre como as organizações devem começar a construir ou melhorar suas atividades de preservação digital (PHILLIPS et al. 2013).

Após a apresentação geral das fases da pesquisa, descrevem-se os detalhes de cada uma. A subseção 3.1.1 menciona as práticas da primeira fase da pesquisa.

3.1.1 Pesquisa bibliográfica – Primeira fase

A primeira fase da pesquisa consistiu na consulta à literatura para construção do referencial teórico, de modo a atender o primeiro objetivo: “Compreender as normas, os métodos e os recursos necessários para a preservação digital”. Os resultados dessa fase são apresentados na seção 4.6.

Nesse contexto, os procedimentos utilizados para a obtenção dos dados foram as pesquisas bibliográfica, documental e contatos diretos, escolha essa alinhada às recomendações de Marconi e Lakatos (2016). Entende-se que a pesquisa bibliográfica corresponde ao levantamento dos principais estudos até então realizados e que são capazes de fornecer indícios atuais e relevantes associados à temática investigada (MARCONI; LAKATOS, 2016). As autoras enfatizam que a investigação preliminar constitui os estudos exploratórios e deve ser realizada por intermédio de dois recursos: documentos (fontes primárias e/ou secundárias) e contatos diretos (pessoas), podendo esses recursos serem executados ao mesmo tempo.

Nesse sentido, as fontes de informação consultadas para esta pesquisa foram: livros impressos e eletrônicos, revistas eletrônicas, arquivos oficiais, teses, dissertações e artigos. Os dados coletados desses estudos possibilitaram a formulação das bases teóricas, conforme descrito no Capítulo 2, reservado ao Referencial Teórico. Além disso, a análise apurada e contextual desses estudos é detalhada na Subseção 4.1, destinada à análise dos trabalhos correlatos.

3.1.2 Questionário – Segunda fase

No desenvolvimento da articulação metodológica proposta por Pereira (2019), definiu-se o questionário como sendo um dos instrumentos para a coleta de dados. Os resultados quanto aos dados levantados por esse instrumento são apresentados na Subseção 4.2.

Vieira (2009) considera o questionário como um instrumento de coleta bastante utilizado em várias áreas do conhecimento. Ele se caracteriza como um formulário, elaborado de modo prévio e “[...] constituído por uma série ordenada de perguntas em campos fechados e abertos, que devem ser respondidas por escrito e sem a presença do entrevistador” (MARCONI; LAKATOS, 2017, p. 219).

Desse modo, para consultar os pesquisadores atuantes nos 92 cursos de Pós-Graduação (amostra desta pesquisa), foi idealizado o grupo Produtores/Consumidores de Dados Científicos (PCDC). A consulta ocorreu por meio do envio do questionário contendo 25 questões, as quais abordavam as práticas relacionadas à preservação digital de dados científicos na UFMG (Apêndice A). Esse questionário foi organizado em oito seções, distribuídas na seguinte ordem:

1. Aceite ao Termo de Consentimento Livre e Esclarecido (TCLE);
2. Atuação na Pós-Graduação;
3. Produção/Consumo de dados;
4. Armazenamento/Cópias;
5. Integridade;
6. Metadados;
7. Controle/Reuso;
8. Entrevista.

Na estrutura do questionário, a seção 1 refere-se aos aspectos éticos da pesquisa e a seção 2 à identificação da área de conhecimento e o tempo de atuação do pesquisador na Pós-Graduação. Da seção 3 a 7, apresentam-se as questões baseadas nas cinco áreas funcionais da LOP, que estão detalhadas no Anexo B. Por fim, a seção 8 convida os participantes à concessão de entrevista.

Na consulta aos dirigentes da UFMG, quanto à governança de dados científicos (políticas e planos de curadoria e de preservação digital), foi idealizado o grupo Formuladores de Políticas Institucionais (FPI). Tal grupo foi composto por assessores, diretores, pró-reitores e tecnologistas, atuantes nos órgãos e setores ligados à Administração Central da UFMG.

A coleta de dados junto ao grupo FPI também ocorreu por meio da aplicação de questionário. Esse instrumento foi composto por 14 questões e organizado em duas seções, assim distribuídas:

1. Aceite ao TCLE;
2. Aspectos institucionais da gestão e preservação digital de dados científicos.

Para esse questionário, a seção 1 refere-se aos aspectos éticos da pesquisa e a seção 2 aborda a temática da governança de dados. A governança de dados teve-se aos seguintes tópicos temáticos: aspectos institucionais da gestão e da preservação digital de dados científicos, políticas e as estratégias de preservação adotadas pela instituição.

A escolha da amostra que compõe o grupo FPI foi embasada na composição do Comitê Gestor do Repositório Institucional (COGERI)⁸. Esse comitê foi instituído em 2019,

cuja finalidade foi a de organizar, fortalecer e aprimorar as ações da UFMG referentes ao Repositório Institucional. Assim, os setores/participantes do grupo FPI foram:

1. Biblioteca Universitária/Sistema de Bibliotecas da UFMG (BU/SiB);
2. Diretoria de Tecnologia da Informação (DTI);
3. Pró-Reitoria de Pesquisa (PRPq);
4. Pró-Reitoria de Extensão (PROEX);
5. Pró-Reitoria de Graduação (PROGRAD);
6. Pró-Reitoria de Pós-Graduação (PRPG).

Devido à proximidade com o tema e à participação efetiva na gestão da universidade, considerou-se importante incluir quatro outros órgãos ou setores da instituição. São eles:

1. Diretoria de Arquivos (DIARQ);
2. Diretoria de Governança Informacional (DGI);
3. Pró-Reitoria de Planejamento e Desenvolvimento (PROPLAN);
4. Assessoria de Saúde (representação da UFMG no Comitê de Enfrentamento ao Novo Coronavírus).

Considerando que o conjunto de pessoas com características semelhantes formam o universo de uma pesquisa (VERGARA, 2014), foram definidas as amostras de interesse desta pesquisa. Para amostragem, foram selecionados dez pesquisadores atuantes na Pós-Graduação da UFMG, como representantes do PCDC, e três membros do COGERI para representarem o grupo FPI.

Assim, ao enviar o questionário convidando os participantes, foi explicitada a abordagem da preservação digital e dos dados científicos no contexto da pesquisa e que a participação seria voluntária. Também foram detalhados e explicados os procedimentos éticos adotados, como é de praxe nas pesquisas envolvendo seres humanos.

A princípio, foi aplicado o pré-teste, cujo objetivo desse procedimento é “verificar se [o] questionário é compreensível para todos os respondentes e se [a] pesquisa está bem planejada” (VIEIRA, 2009, p. 103). A autora afirma que os questionários devem ser submetidos ao pré-teste na forma em que serão aplicados durante a coleta definitiva. Nesse percurso, “[...] os respondentes do pré-teste precisam ser adequados” (VIEIRA, 2009, p. 104). Então, foram submetidos dois modelos de questionários para atendimento às especificidades de cada grupo: questionário aplicado aos PCDC (Apêndice A) e questionário aplicado aos FPI (Apêndice B). Como mencionado, o convite aos docentes foi encaminhado às secretarias e

⁸ Resolução n. 12, de 10 de outubro de 2019. Esse documento institui o Comitê Gestor do Repositório Institucional (COGERI) da Universidade Federal de Minas Gerais (UFMG) e estabelece a sua estrutura (UFMG, 2019).

colegiados dos PPG, de acordo com os endereços listados na referida manifestação, para serem repassados aos docentes (Apêndice H). Alguns PPG responderam que haviam encaminhado a pesquisa aos docentes. Outros não o fizeram, impossibilitando aferir se as mensagens chegaram aos docentes.

Assim, submeteram-se os questionários por meio da conta institucional da UFMG, garantindo formalidade e sigilo aos dados pessoais dos participantes.

Para o PCDC, o convite foi encaminhado por e-mail institucional às secretarias e aos colegiados dos PPG, para então, ser repassado aos docentes, repetindo o mesmo fluxo quando da aplicação do pré-teste. A mensagem eletrônica direcionada aos professores convidava-os a participarem da pesquisa e apresentava o link do questionário (Apêndice A). Por sua vez, o convite ao grupo FPI, diferente do pré-teste, foi encaminhado diretamente no e-mail institucional dos órgãos ou setores envolvidos e apresentava o link do questionário (Apêndice B).

Em ambos os casos, os questionários foram disponibilizados pelo *Forms Microsoft Office 365* (disponível para membros da comunidade MinhaUFMG), contendo restrições por período e acessível, apenas, aos e-mails institucionais (domínio @ufmg). Essas restrições foram necessárias, na garantia de assegurar a participação, apenas, dos sujeitos selecionados para a pesquisa. Junto ao questionário, foi disponibilizado o Termo de Consentimento Livre e Esclarecido (TCLE) (Apêndice C). A seguir, na subseção 3.1.3, será apresentada a terceira fase da pesquisa, ou seja, a entrevista submetida ao grupo PCDC.

3.1.3 Entrevista – Terceira fase

Na articulação metodológica proposta por Pereira (2019), definiu-se a entrevista como o segundo instrumento para auxiliar na coleta de dados. A entrevista pode ser caracterizada como “[...] um procedimento utilizado na investigação social, para a coleta de dados, ou para ajudar no diagnóstico ou no tratamento de um problema social” (MARCONI; LAKATOS, 2017, p. 213). Os resultados das entrevistas são apresentados no capítulo 4, em específico, na seção 4.3.

Para atender ao segundo objetivo, que é “Diagnosticar o estágio em que a UFMG encontra-se quanto à gestão de dados científicos e sua conformidade às atividades e processos de curadoria e de preservação digital, conforme padrões e práticas adotados pela comunidade internacional”, adotou-se a entrevista semiestruturada. Na visão de Marconi e Lakatos (2016), a entrevista configura-se como uma estratégia utilizada na tentativa de explorar com mais profundidade uma questão.

Entre as vantagens na aplicação da entrevista, esse instrumento fornece melhor amostragem, é mais flexível e precisa, além de revelar os esclarecimentos e as confirmações (MICHEL, 2015). Além disso, essa técnica dá maior oportunidade para a obtenção de dados relevantes, muitas vezes não constantes em fontes documentárias, e o seu objetivo principal é estimular a obtenção de informações do entrevistado sobre determinado assunto ou problema (MICHEL, 2015).

Nesse sentido, com o intuito de aprofundar sobre as práticas dos pesquisadores da UFMG quanto à gestão de dados científicos, os 11 sujeitos escolhidos para conceder a entrevista foram contatados. Entre eles, nove tiveram disponibilidade e participaram da entrevista semiestruturada, cujo roteiro completo com as questões abordadas encontra-se estruturado no Apêndice E.

Conforme mencionado anteriormente, adotou-se a técnica da entrevista para auxiliar no diagnóstico, de modo a atender ao segundo objetivo desta pesquisa. Para isso, os professores foram contatados por e-mail institucional (@ufmg) e/ou mediante o aplicativo Whatsapp, recursos esses indicados pelos participantes quando sinalizaram a concessão da entrevista. As entrevistas foram realizadas entre os meses de março e abril de 2022 pela plataforma de Conferência Web da Rede Nacional de Ensino e Pesquisa (RNP).

Seguiu-se o roteiro da entrevista passando pelas sete seções. Todos os participantes responderam às 14 questões, acerca de suas práticas no que tange à gestão e à preservação de dados científicos na UFMG. Além das questões constantes do roteiro, por se tratar de uma entrevista semiestruturada, foi perguntado a respeito de apoio por parte dos profissionais da informação, levando em conta a possibilidade do uso dos repositórios de dados e das políticas institucionais.

Essas entrevistas foram gravadas na plataforma da RNP e, posteriormente, foram salvas no computador da pesquisadora nos formatos de áudio e vídeo, assim como foi realizado o *upload* de uma cópia de cada entrevista para o *Google Drive* e outra para a área restrita da pesquisadora no canal do YouTube. Nesse canal, foram aplicadas as devidas restrições de acesso para atendimento às exigências do COEP. Essas exigências dispõem que todo o material (questionários e entrevistas) obtido na coleta de dados será utilizado, apenas, para fins de pesquisa e será mantido em arquivo digital, sob guarda e responsabilidade da pesquisadora, por um período de cinco anos após o término da pesquisa.

Quanto à transcrição, a princípio, as entrevistas foram convertidas em texto escrito com o auxílio da ferramenta *Microsoft Office 365*. Foi necessário revisar as transcrições, pois o software apresentou falhas na captura de determinadas palavras. Após a transcrição, foi utilizado um editor de planilha para facilitar a organização das seções constantes do roteiro da entrevista (Apêndice E). Os textos das respostas fornecidas pelos participantes, ao serem

inseridos na planilha, passaram por revisão gramatical (sintática, ortográfica e de estilo), considerando as normas da Gramática da Língua Portuguesa e o linguajar acadêmico. No entanto, esse processo foi realizado de modo superficial, tendo em vista preservar a integridade e a originalidade das respostas fornecidas pelos participantes. Na seção 3.1.4, a seguir, será apresentada a quarta fase da pesquisa.

3.1.4 Pesquisa documental – Quarta fase

No âmbito desta pesquisa, para consultar os dirigentes da UFMG quanto à governança de dados científicos, políticas e planos de curadoria e preservação digital, foi idealizado o grupo FPI, conforme já mencionado. Contudo, recorreu-se à análise de documentos para suprir a lacuna deixada pela ausência de respostas durante a aplicação do instrumento de coleta direcionado ao referido grupo. Os resultados da pesquisa documental serão apresentados na seção 4.4.

Mesmo tendo validado o questionário, o grupo FPI hesitou em responder às questões submetidas à coleta de dados. Então, devido à ausência de respostas ao questionário destinado ao FPI, optou-se pela pesquisa documental, cuja fonte de coleta de dados “[...] está restrita a documentos, escritos ou não, constituindo o que se denomina de fontes primárias” (MARCONI; LAKATOS, 2016, p. 48).

Para concretização da pesquisa documental, realizou-se um levantamento dos órgãos e dos setores partícipes das tomadas de decisões quando da formulação de políticas na instituição. Os órgãos ou setores da UFMG escolhidos para essa coleta foram:

1. Administração Central;
2. Biblioteca Universitária/Sistema de Bibliotecas da UFMG (BU/SiB);
3. Diretoria de Tecnologia da Informação (DTI);
4. Pró-Reitoria de Pesquisa (PRPq);
5. Pró-Reitoria de Extensão (PROEX);
6. Pró-Reitoria de Graduação (PROGRAD);
7. Pró-Reitoria de Pós-Graduação (PRPG);
8. Diretoria de Arquivos (DIARQ);
9. Diretoria de Governança Informacional (DGI);
10. Pró-Reitoria de Planejamento e Desenvolvimento (PROPLAN);
11. Assessoria de Saúde (representação da universidade no Comitê de Enfrentamento ao Novo Coronavírus).

Posteriormente, visitaram-se os 11 sites dos órgãos e setores da UFMG, sendo levantados 13 documentos, entre eles: seis planos, três relatórios, uma política, duas resoluções e um manual. A documentação encontrada foi relacionada no Quadro 3, que indica

os tipos de documentos, os órgãos ou setores responsáveis pela emissão deles, assim como a gestão ou período de vigência desses documentos.

QUADRO 3 – Levantamento de documentos institucionais da UFMG

Tipo de documento	Nome do documento	Órgão ou setor	Gestão
Plano	Plano de Desenvolvimento Institucional (PDI)	Administração Central - Reitoria	2018-2023
Plano	Plano de Dados Abertos (PDA)	Administração Central – Reitoria	2021-2022
Relatório	Relatório Gerencial da BU	Biblioteca Universitária (BU)	2017
Política	Política de Desenvolvimento do Acervo da BU	Biblioteca Universitária (BU)	2015
Relatório	Relatório Anual de Gestão da DIARQ	Diretoria de Arquivos (DIARQ)	2017
Manual	Manual de Gestão de Documentos Arquivísticos da UFMG - DIARQ	Diretoria de Arquivos (DIARQ)	2019
Relatório	Relatório Gestão da DGI	Diretoria de Governança Informacional (DGI)	2018-2021
Plano	Plano Diretor de Tecnologia da Informação e Comunicação da UFMG (PDTIC)	Diretoria de Tecnologia da Informação (DTI)	2021-2024
Plano	Planejamento Estratégico de TI da UFMG	Diretoria de Tecnologia da Informação (DTI)	2013-2015 2015-2020
Plano	Plano de Gestão da PROEX	Pró-Reitoria de Extensão (PROEX)	2018-2022
Nenhum documento encontrado	—	Pró-Reitoria de Graduação (PROGRAD)	—
Nenhum documento encontrado	—	Pró-Reitoria de Pós-Graduação (PRPG)	—
Resolução	Resolução n. 01, de 28 de janeiro de 2021 - Infraestruturas de Apoio Institucional à Pesquisa (AIPq)	Pró-Reitoria de Pesquisa (PRPq)	2021
Resolução	Resolução n. 01, de 06 de fevereiro de 2018 - Infraestruturas sobre pesquisa na UFMG.	Pró-Reitoria de Pesquisa (PRPq)	2018
Plano	Plano Estratégico Institucional da PROPLAN	Pró-Reitoria de Planejamento e Desenvolvimento (PROPLAN)	Atualização semestral

Fonte: Elaborado pela autora (2022).

Para selecionar os documentos a serem analisados, definiu-se como critério de inclusão a ocorrência de uma ou mais palavras-chave ou termos relacionados à temática da pesquisa. Não havendo ocorrência de palavras-chave ou de termos, o documento foi descartado. Os termos procurados nos documentos foram: Governança Digital, Preservação, Preservação Digital, Plano de Preservação, Curadoria, Curadoria de Dados, Curadoria Digital, Dados Científicos, Dados de Pesquisa e Conjunto de Dados. A apresentação e a discussão quanto ao uso desses termos e as evidências encontradas são discorridas na terceira seção do quarto capítulo.

4 RESULTADOS E DISCUSSÕES

Neste capítulo, são apresentadas as análises dos dados coletados junto aos pesquisadores dos PPG da UFMG, tendo como base a aplicação do questionário e da entrevista. O capítulo também analisa os dados do levantamento bibliográfico, referentes aos trabalhos correlatos, como também analisa e discute os dados da pesquisa documental, realizada no site da universidade.

4.1 Análise dos trabalhos correlatos

Com o intuito de compreender como o tema da preservação digital de dados científicos tem sido tratado no campo da Ciência da Informação, buscaram-se, a princípio, os estudos realizados no âmbito do PPG em Gestão e Organização do Conhecimento (PPGGOC). Partiu-se do pressuposto de que alguns estudos poderiam ter recorrido à UFMG como campo de análise, como é o caso desta dissertação. Na sequência, foi consultada a base de dados da Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) do IBICT e a base do Google Acadêmico, além da consulta a outros materiais bibliográficos, como livros e revistas.

Para identificar os estudos correlatos defendidos no PPGGOC, nos últimos cinco anos (2017-2021), foi consultado o Repositório Institucional da UFMG (RI-UFMG), ambiente que armazena a produção intelectual da comunidade universitária em formato digital. A busca ocorreu com uso das seguintes palavras-chave ou termos: Preservação Digital, Plano de Preservação, Curadoria de Dados, Curadoria Digital, Dados Científicos e Dados de Pesquisa. O resultado dessa consulta inicial é apresentado no Quadro 4.

QUADRO 4 – Estudos correlatos desenvolvidos no PPGGOC (2017 a 2021)

Base	Termos		Teses	Dissertações
			(46)	(59)
Repositório Institucional da UFMG	Curadoria Digital	Curadoria de Dados	01	02
	Dados de Pesquisa	Dados Científicos	01	-
	Preservação Digital	Plano de Preservação	01	-
	Total recuperado:		03	02

Fonte: Dados da pesquisa (2021).

Entre as coleções de teses (46) e dissertações (59), foram recuperados apenas cinco documentos, sendo três teses e duas dissertações. Esse resultado inicial evidencia que esta investigação pode complementar tais estudos, ao pesquisar o tema da preservação digital de dados científicos e contribuir para melhoria das práticas de gestão desses dados na UFMG.

Considerando a relevância da temática para esta pesquisa, os estudos acerca da preservação digital no âmbito das IES são bastante significativos, assim como outros estudos que abordam a curadoria de dados científicos. O Quadro 5 apresenta o foco temático ou abordagem de cada um dos cinco estudos desenvolvidos no PPGGOC, com menção ao ano de defesa, título, autor da pesquisa e também a tipologia documental.

QUADRO 5 – Detalhamento dos estudos correlatos desenvolvidos no PPGGOC

Ano	Título	Autor(a)	Tipo
2019	A curadoria de dados científicos na Ciência da Informação: levantamento do cenário nacional	Liliane Chaves de Resende	Dissertação
2019	Expansão das fronteiras da educação na Ciência da Informação: a formação de profissionais para atuação no contexto dos dados de pesquisa	Gracielle Mendonça Rodrigues Gomes	Tese
2019	Suporte à gestão de dados de pesquisa: uma ampliação dos serviços oferecidos pelas bibliotecas	Fernanda Gomes Almeida	Tese
2020	Preservação digital em Instituições de Ensino Superior: instrumentos para a oficialização e operacionalização da preservação de documentos arquivísticos digitais	Luciana Gonçalves Silva Souza	Tese
2020	Repositórios de dados científicos na América do Sul: uma análise da conformidade com os Princípios FAIR	Marcello Mundim Rodrigues	Dissertação

Fonte: Dados da pesquisa (2021).

Entre os estudos correlatos desenvolvidos no PPGGOC, a dissertação de Liliane Chaves de Resende (2019) teve como objetivo investigar a importância que as atividades de curadoria digital de dados científicos têm para a área da Ciência da Informação, no Brasil, frente à tendência internacional. Resende (2019) relatou que o desconhecimento atual da curadoria digital de dados e do grau de adesão da Ciência da Informação a esse novo cenário do fazer científico traz crescentes desafios para o profissional da informação. Esse profissional busca se adequar e contribuir para solucionar os novos problemas que surgem no âmbito da curadoria digital.

Para obter um panorama da curadoria digital no âmbito internacional, Resende (2019) realizou uma revisão sistemática da literatura, complementada com uma pesquisa *survey* que objetivou coletar informações sobre a opinião dos pesquisadores brasileiros da Ciência da Informação. A autora concluiu que esse campo do conhecimento, no Brasil, está buscando se adaptar. Assim, investigar as tendências do cenário mundial das atividades da curadoria digital e como essas atividades são realizadas por pesquisadores brasileiros fornecerá subsídios para o desenvolvimento desse campo científico, no Brasil (RESENDE, 2019).

A tese de Gomes (2019) objetivou propor orientações curriculares que apoiem projetos de formação de profissionais no âmbito da gestão e da curadoria de dados de pesquisa pertinentes ao campo da Ciência da Informação.

Gomes (2019) constatou a existência de cursos de dados com vários focos e objetivos, sendo que esses cursos recebem influências dos currículos de diversas áreas como Ciência da Computação, Ciência da Informação, Estatística, Matemática, Administração, Comunicação e Ciências Políticas. A autora mencionou que há escassez de cursos de dados direcionados ao contexto dos dados de pesquisa, e que as formações são direcionadas aos estudantes de Pós-Graduação e profissionais. Além disso, o estudo evidenciou que há ênfases ou concentrações em cursos de Graduação e Pós-Graduação *stricto sensu* na área de dados como estratégia recorrente de inserir a temática na formação oferecida pelas Escolas de Ciência da Informação.

Portanto, é necessário que as Escolas de Ciência da Informação disponibilizem uma formação específica que habilite os profissionais às atividades e às formas de manipulação dos dados de pesquisa (GOMES, 2019). Há necessidade, também, de abertura dos docentes e dos profissionais da Ciência da Informação no que concerne à aprendizagem de novos conhecimentos no ambiente de dados, envolvendo áreas correlatas, como a Estatística e a Computação. Nesse sentido, sugere-se que os profissionais da Ciência da Informação busquem as habilidades e as competências que lhes conferirão um papel crucial nas instituições, tendo em vista oferecer suporte à gestão e à curadoria dos dados de pesquisa (GOMES, 2019).

Em sentido semelhante, a tese de Almeida (2019) objetivou a proposição de uma taxonomia de serviços voltada à gestão dos dados de pesquisa gerados nas universidades. A pesquisa buscou sintetizar as dificuldades, os desafios e as ações necessárias para viabilização e manutenção desses serviços nas instituições.

Almeida (2019) apresentou os temas que fundamentaram sua tese: os Dados de Pesquisa, o Paradigma da *eScience*, *Big Science*, *Small science*, Geração e Estruturação de Dados, Gestão de Dados de Pesquisa, Ciclo de Vida dos Dados, Compartilhamento e Reuso de Dados, Publicação e Citação, Infraestruturas, Padrões e Políticas para Dados e as Bibliotecas e a Gestão de Dados de Pesquisa. Ao discorrer sobre esses assuntos, foram discutidos os desafios enfrentados para a entrega de serviços de apoio à gestão de dados pelas bibliotecas, às práticas e às necessidades dos pesquisadores, assim como os Serviços de Suporte à Gestão de Dados de Pesquisa (SSGDPs).

O estudo supracitado concluiu que os objetivos propostos foram atendidos no que tange a identificar os serviços, as ações, as dificuldades e os desafios; a contextualizar o oferecimento de serviços e infraestruturas nas universidades entrevistadas; a sistematizar e a propor uma taxonomia de serviços; a sintetizar as dificuldades e os desafios; e a sintetizar as ações para viabilização e manutenção dos serviços nas instituições. Nota-se que alguns assuntos que fundamentaram a tese de Almeida (2019) são correlatos à temática desta

pesquisa, tais como: a Gestão de Dados Científicos, o Ciclo de Vida e o Reuso; os Padrões e as Políticas para Dados, assim como as Práticas e Necessidades dos Pesquisadores.

Assim como Almeida (2019), a tese de Souza (2020) delimitou como ambiente de investigação as universidades, cujo enfoque se concentrou na preservação digital. Tal pesquisa buscou enfatizar a importância acerca da preservação dos documentos arquivísticos digitais e da concretização de ações, tendo em vista os instrumentos de oficialização e de operacionalização da preservação digital. Assim, Souza (2020) propôs referências para o desenvolvimento de políticas e de planos de preservação de documentos arquivísticos digitais no âmbito das IES brasileiras.

Souza (2020), a partir de uma revisão de literatura, sinalizou dois elementos-chave do processo de oficialização das ações de preservação: a política e o plano de preservação digital. Em virtude da importância dos dois instrumentos, a autora mapeou, entre as IES, nas modalidades públicas, estaduais e federais brasileiras, a presença de políticas e de planos de preservação digital.

Percebeu-se que poucas IES publicaram a política de preservação de documentos arquivísticos digitais e nenhuma delas desenvolveu o plano de preservação (SOUZA, 2020). Como conclusão desse estudo, foi apontada a urgência no desenvolvimento de documentos que oficializem o compromisso institucional para com a preservação dos conteúdos digitais produzidos no cotidiano das instituições.

Por fim, o último estudo identificado no PPGGOC é a dissertação de Rodrigues (2020) que teve como finalidade estudar o fenômeno dos dados gerados por meio do processo científico e o desenvolvimento de serviços relacionados aos crescentes desafios que permeiam os processos de gestão e de curadoria. Nesse contexto, foi mencionado o crescente volume de recursos digitais em constante expansão, nos dias atuais.

O autor esclareceu que os objetos do estudo dessa investigação foram os dados, enquanto fenômeno; os conjuntos de dados, que são unidades informacionais; os Princípios FAIR, caracterizados como diretrizes à gestão e à curadoria de dados científicos; e os repositórios digitais institucionais de dados científicos, manifestados como ambientes virtuais de organização informacional. Assim, nesse estudo, investigaram-se os conjuntos de dados científicos e os respectivos repositórios digitais institucionais sul-americanos à luz dos Princípios FAIR.

Rodrigues (2020) também salientou que lidar com o fenômeno dos dados é assistir a comunidade acadêmica em seus diversos campos, assim como a sociedade civil em suas dificuldades tecnológicas e digitais, disponibilizando meios de acesso mais ágeis e assertivos em um ambiente digital cada vez mais nebuloso e caótico. O autor concluiu que profissionais da informação devem buscar sua capacitação em dados, a começar pelo planejamento de

projetos e de políticas institucionais voltado à implementação de repositórios de dados científicos. Para instituir esses ambientes, Rodrigues (2020) mencionou que o processo tramita pelo entendimento das divergentes necessidades entre as comunidades, pelo conhecimento técnico computacional exigido a tais práticas e, idealmente, pela busca da padronização e da manutenção desses serviços.

Além dos estudos desenvolvidos no PPGOC, a fim de subsidiar o referencial teórico desta investigação, buscou-se os artigos apresentados no Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ENANCIB). Semelhante aos estudos anteriores, foi delimitado o período de cinco anos (2017-2021), com a intenção de perceber como os assuntos vêm sendo tratado no interior dos PPG, de modo a reconhecer as pesquisas em andamento.

A busca a esses estudos ocorreu na plataforma do Google Acadêmico, ambiente digital que consolida boa parte da literatura acadêmica. Nesse contexto, a plataforma possibilitou recuperar os artigos publicados nas cinco edições do referido evento. Foram utilizadas as mesmas palavras-chave das buscas realizadas no PPGOC, a saber: Curadoria de Dados, Curadoria Digital, Dados Científicos, Dados de Pesquisa, Preservação Digital e Plano de Preservação. O quantitativo de artigos publicados no ENANCIB e recuperados na busca e as respectivas afiliações institucionais de seus autores são expostos no Quadro 6.

QUADRO 6 – Artigos publicados no ENANCIB (2017 a 2021)

Base	Termos	Artigos recuperados	Artigos selecionados	Afiliação (IES)
Google acadêmico	Curadoria de Dados	02	01	PPGCI/UFPE
	Curadoria Digital	09	07	Câmara do FCRB/UNIRIO IBICT/CNEN RECIFE/UFPE UFBA UFRGS UNIRIO/JBRJ
	Dados Científicos	09	04	FIOCRUZ/UFF UFPB UFRGS UNESP/ Universidad Carlos III
	Dados de Pesquisa	14	02***	UNESP/USP/UFG UNIRIO
	Preservação Digital	17	03*****	UFPB

				UFPB/UFPE
				UFSCar
	Plano de Preservação	01	01	Câmara do RECIFE/UFPE
	Total	52	18	XXXXXXXXXXXXXXXXXXXX

Fonte: Dados da pesquisa (2021).

Nota: O asterisco indica a quantidade de artigos repetidos, ou seja, já selecionados nas buscas com os termos anteriores.

A partir das buscas, ao ler os títulos e os resumos, houve redundância no retorno dos artigos (artigos repetidos), pois indexavam as mesmas palavras-chave. Desse modo, o número de artigos selecionados com o termo “Dados de Pesquisa” foi de cinco, porém três artigos haviam sido selecionados quando do resultado da busca por “Dados Científicos”. O mesmo fato ocorreu na busca por “Preservação Digital”, quando cinco estudos dos oito selecionados já haviam sido recuperados na busca por “Curadoria Digital”.

Com o propósito de finalizar o levantamento e a análise dos estudos correlatos ao tema desta dissertação, recorreu-se à última consulta idealizada na base de dados da BDTD. Durante essa busca, no recurso “pesquisa avançada”, aplicaram-se os seguintes filtros: termo, campo de busca, grupo de busca, tipo de documento, idioma e período, sendo esses ordenados por relevância. Quanto ao uso das palavras-chave, foi definida a expressão “Política de Preservação Digital”, por considerar a relevância dessa expressão no contexto deste estudo.

Com efeito, a busca por teses e dissertações na BDTD resultou em um número mais expressivo, visto que essa base integra e dissemina os textos completos das teses e dissertações defendidas nas instituições brasileiras de ensino e de pesquisa. Os registros encontrados e selecionados na BDTD estão descritos no Quadro 7.

QUADRO 7 - Teses e dissertações selecionadas da BDTD (2017 a 2021)

Base	Expressão de busca	Tipo de documento	Documentos recuperados	Documentos selecionados
BDTD	Política de preservação digital	Teses	08	03
		Dissertações	20	02
Total:			28	05

Fonte: Dados da pesquisa (2021).

Após a leitura dos resumos e das teses e dissertações recuperadas, foram selecionadas aquelas que abordam a temática no contexto da Ciência da Informação. Assim, foi obtido um total de três teses e duas dissertações. Esses documentos, em conjunto com 18 artigos (publicados no ENANCIB e artigos recuperados do Google Acadêmico), foram utilizados no referencial teórico desta pesquisa.

4.2 Análise do questionário

As perguntas do questionário foram elaboradas e analisadas à luz da matriz LOP (2019), tendo em vista as cinco categorias e níveis de preservação estabelecidos nesse instrumento. Destaca-se, também, que atributos evidenciados na literatura foram utilizados para embasar as análises. A Figura 8 apresenta a estrutura do questionário aplicado ao grupo PCDC. Tal estrutura representou a “espinha dorsal” quanto ao desenvolvimento das análises para aferição dos resultados.

FIGURA 8 - Organização do questionário - PCDC



Fonte: Elaborada pela autora (2022).

Como mencionado na seção da metodologia, para validação do questionário, foi aplicado o pré-teste. O envio do questionário do pré-teste concretizou-se com o uso do e-mail, sendo que foi enviada uma mensagem formal às secretarias dos cursos de Pós-Graduação, solicitando o envio aos coordenadores de cada curso. Assim, não foi possível identificar, em alguns casos, a quais cursos as respostas se referiam. O Quadro 8 elucida o quantitativo de consultas por áreas de conhecimento e as respostas referentes ao grupo PCDC.

QUADRO 8 – Áreas consultadas e respostas do pré-teste - PCDC

Áreas do conhecimento	Subáreas consultadas (professores)	Respostas
Ciências Agrárias	-	-
Ciências Biológicas	Biologia	0
Ciências da Saúde	Enfermagem	01
Ciências Exatas e da Terra	Química e Física	01
Engenharias	Engenharia	01
Ciências Humanas	Educação, Educação Física e Psicologia	03
Ciências Sociais Aplicadas	Administração	01
Linguística, Letras e Artes	-	-
Total geral		
08 áreas	10 subáreas	07 respostas válidas

Fonte: Dados da pesquisa (2021).

Conforme mostra o Quadro 8, na validação do questionário, destinado ao grupo PCDC, dos dez pesquisadores consultados, sete respostas foram obtidas. Os pesquisadores consultados estavam vinculados às seguintes áreas: Ciências da Saúde (01), Ciências Exatas e da Terra (01), Engenharias (01), Ciências Humanas (03) e Ciências Sociais Aplicadas (01).

Na consulta ao grupo FPI, durante o pré-teste, entre os dez setores que compõem a amostra do estudo, foram selecionados três, a saber: Biblioteca Universitária (BU), Diretoria de Tecnologia da Informação (DTI) e Pró-Reitoria de Graduação (PROGRAD). Assim como se procedeu com o grupo PCDC, o instrumento de coleta para o pré-teste foi enviado por e-mail à secretaria da Administração Central, com pedido de envio aos respectivos setores/órgãos solicitados. O Quadro 9 discrimina o retorno das respostas do pré-teste com esse grupo de participantes.

QUADRO 9 – Setores consultados e respostas do pré-teste - FPI

Setores da amostra	Setores consultados (membros)	Respostas
Biblioteca Universitária/Sistema de Bibliotecas da UFMG (BU/SiB)	Biblioteca Universitária/Sistema de Bibliotecas da UFMG (BU/SiB)	01
Diretoria de Tecnologia da Informação (DTI)	Diretoria de Tecnologia da Informação (DTI)	
Pró-Reitoria de Pesquisa (PRPq)	-	
Pró-Reitoria de Extensão (PROEX)	-	
Pró-Reitoria de Graduação (PROGRAD)	Pró-Reitoria de Graduação (PROGRAD)	
Pró-Reitoria de Pós-Graduação (PRPG)	-	
Diretoria de Arquivos (DIARQ)	-	
Diretoria de Governança Informacional (DGI)	-	
Pró-Reitoria de Planejamento e Desenvolvimento	-	
Assessoria de Saúde	-	
Total geral		
10 setores	03 participantes	01 resposta válida

Fonte: Dados da pesquisa (2021).

Conforme descrito no Quadro 9, entre os membros consultados, obteve-se, apenas, uma resposta, validando o questionário. Em suma, apresenta-se o extrato das validações ocorridas entre outubro e novembro de 2021 para ambos os questionários, de acordo com os registros do Quadro 10.

QUADRO 10 – Extrato de validação dos questionários – pré-testes

Pré-teste (PCDC)		Pré-teste (FPI)	
Enviados	Recebidos	Enviados	Recebidos
10	07	03	01

Fonte: Dados da pesquisa (2021).

Mediante a validação dos instrumentos de pesquisa submetidos no período de outubro a novembro de 2021, e após tramitação legal e aprovação do projeto no COEP, a coleta de dados junto aos grupos PCDC e FPI foi concretizada. Essas atividades foram realizadas em um prazo de três meses, de março a maio de 2022.

Posto isso, cabe lembrar que o questionário é um instrumento de valor significativo de coleta de dados, pois “[...] obtém-se grande número de dados, atinge maior número de pessoas e possibilita o anonimato (que permite respostas mais verdadeiras e seguras)” (MICHEL, 2015, p. 92). Em contrapartida, “pesquisadores costumeiramente enfrentam uma grande dificuldade: as pessoas hesitam – ou, até mesmo, resistem – em responder às muitas perguntas que lhes são feitas” (VIEIRA, 2009, p. 16). Ainda que haja risco quanto à utilização do questionário, considera-se uma possível colaboração por parte dos pesquisadores, dirigentes e membros da UFMG.

Na coleta junto ao grupo PCDC, obtiveram-se 42 respostas ao questionário proposto, abrangendo cinco das oito áreas do conhecimento⁹. Entre os respondentes, onze demonstraram interesse em conceder a entrevista para aprofundamento sobre as práticas de preservação de dados (Apêndice E). Por outro lado, durante a coleta junto ao grupo FPI, os dirigentes hesitaram em responder ao questionário. Essa intercorrência alterou a metodologia, incluindo a análise de documentos institucionais no que tange à governança de dados científicos (pesquisa documental), conforme resultados apresentados no tópico 4.4.

No grupo PCDC, foram consultados os pesquisadores atuantes nos 92 cursos de Pós-Graduação da UFMG. Os contatos dos colegiados dos PPG foram requisitados pela Plataforma Integrada de Ouvidoria e Acesso à Informação Fala.BR¹⁰. A manifestação foi registrada por meio do protocolo de número 23546.012771/2021-11 e atendida pela Pró-Reitoria de Pós-Graduação (PRPG) da UFMG (Anexo D).

⁹ Áreas do conhecimento ou especialidades do conhecimento são: Ciências Agrárias; Ciências Biológicas; Ciências da Saúde; Ciências Exatas e da Terra; Engenharias; Ciências Humanas; Ciências Sociais Aplicadas; e Linguística, Letras e Artes.

¹⁰ Constitui um canal integrado para encaminhamento de manifestações (acesso à informação, denúncias, reclamações, solicitações, sugestões, elogios e simplificações) a órgãos e entidades do poder público. Disponível no endereço: <<https://falabr.cgu.gov.br>>.

No Quadro 11 estão relacionados os cursos de Pós-Graduação (Doutorado e Mestrado) existentes na UFMG.

QUADRO 11 – Cursos de Pós-Graduação *stricto sensu* da UFMG

Area (Pós-Graduação)	Respondidos
Ciências Agrárias	0
- Ciência Animal; Ciência de Alimentos; Ciências Florestais; Produção Animal; Produção Vegetal; e Zootecnia.	
Ciências Biológicas	09
- Bioinformática; Biologia Celular; Biologia Vegetal; Bioquímica e Imunologia; Ciências Biológicas: Fisiologia e Farmacologia; Ecologia, Conservação e Manejo da Vida Silvestre; Ensino em Biologia (Mestrado Profissional) – Coordenação Local; e Ensino em Biologia (Mestrado Profissional) – Coordenação Nacional.	
Ciências da Saúde	11
- Análises Clínicas e Toxicológicas; Ciências Aplicadas à Cirurgia e à Oftalmologia; Ciências Aplicadas à Saúde do Adulto; Ciências da Reabilitação; Ciências da Saúde: Saúde da Criança e do Adolescente; Ciências do Esporte; Ciências Farmacêuticas; Ciências Fonoaudiológicas; Educação Física (Mestrado Profissional); Enfermagem; Estudos da Ocupação; Gestão de Serviços de Saúde (Mestrado Profissional); Infectologia e Medicina Tropical; Medicamentos e Assistência Farmacêutica; Medicina Molecular; Nutrição e Saúde; Odontologia; Odontologia em Saúde Pública (Mestrado Profissional); Patologia; Promoção da Saúde e Prevenção da Violência (Mestrado Profissional); Saúde da Mulher; e Saúde Pública.	
Ciências Exatas e da Terra	02
- Ciência da Computação; Estatística; Física; Geologia; Inovação Tecnológica e Biofarmacêutica; Matemática; e Química.	
Ciências Humanas	02
- Antropologia; Ciência Política; Educação e Docência (Mestrado Profissional); Educação: Conhecimento e Inclusão Social; Filosofia; Geografia; História; Psicologia; Psicologia: Cognição e Comportamento; e Sociologia.	
Ciências Sociais Aplicadas	12
- Administração; Arquitetura e Urbanismo; Ciência da Informação; Comunicação Social; Controladoria e Contabilidade; Demografia; Direito; Economia; Gestão e Organização do Conhecimento; e Sociedade, Ambiente e Território.	
Engenharias	06
- Ciências e Técnicas Nucleares; Construção Civil; Engenharia de Estruturas; Engenharia de Produção; Engenharia Elétrica; Engenharia Mecânica; Engenharia Metalúrgica, Materiais e de Minas; Engenharia Metalúrgica, Materiais e de Minas (Mestrado Profissional); Engenharia Química; Geotecnia e Transportes; e Saneamento, Meio Ambiente e Recursos Hídricos.	
Interdisciplinar	0
Alimentos e Saúde; Ambiente Construído e Patrimônio Sustentável; Análise e Modelagem de Sistemas Ambientais; Estudos do Lazer; e Neurociências.	
Linguística, Letras e Arte	0
Artes; Artes (Mestrado Profissional); Educação e Docência (Mestrado Profissional); Estudos Linguísticos; Letras (Mestrado Profissional); Letras: Estudos Literários; e Música.	0
Total de cursos de Pós-Graduação <i>stricto sensu</i>:	92
Total de respostas obtidas:	42

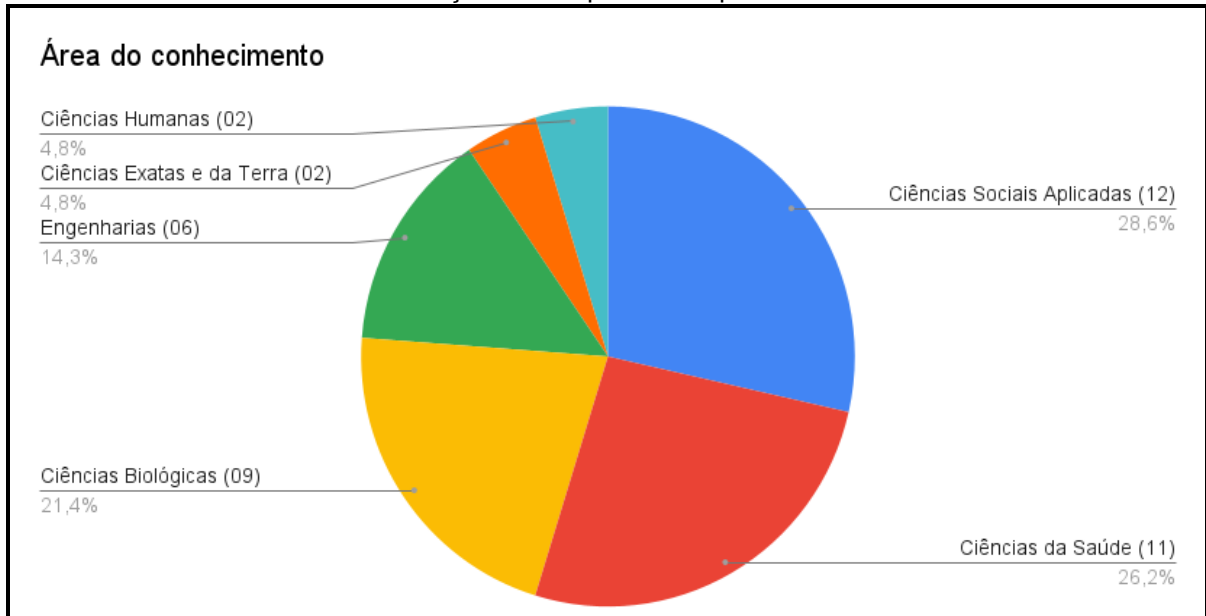
Fonte: Dados da pesquisa (2022).

O questionário foi organizado em oito seções e conteve 25 questões, entre elas, 18 abordavam as práticas relacionadas à gestão e à preservação de dados científicos na UFMG. A primeira questão apresentou o TCLE, aprovado pelo COEP. Os resultados

indicaram que todos os sujeitos aceitaram participar da pesquisa, ou seja, estavam de acordo com os termos e com demais informações constantes nesse documento.

A segunda questão do questionário, vinculada à seção 2, buscou identificar em qual área do conhecimento os respondentes atuavam. Os resultados podem ser observados no Gráfico 1.

GRÁFICO 1 - Distribuição dos respondentes por área do conhecimento

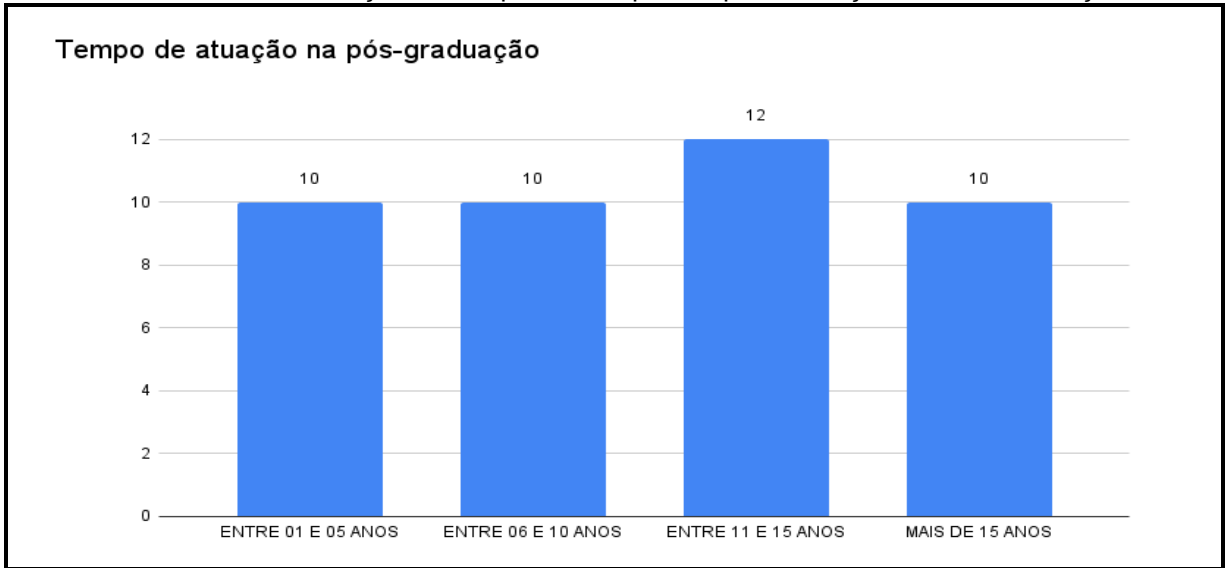


Fonte: Dados da pesquisa (2022).

Dos 42 pesquisadores consultados, as áreas de conhecimento com maior representatividade foram: Ciências Sociais Aplicadas, Ciências da Saúde e Ciências Biológicas, seguidas das Engenharias, Ciências Humanas e Ciências Exatas e da Terra. Não houve representação nas áreas das Ciências Agrárias e Linguística, Letras e Artes.

Ainda na seção 2, a terceira questão teve a finalidade de identificar o tempo de atuação do pesquisador na Pós-Graduação. Os resultados para essa pergunta estão ilustrados no Gráfico 2.

GRÁFICO 2 – Distribuição dos respondentes por tempo de atuação na Pós-Graduação

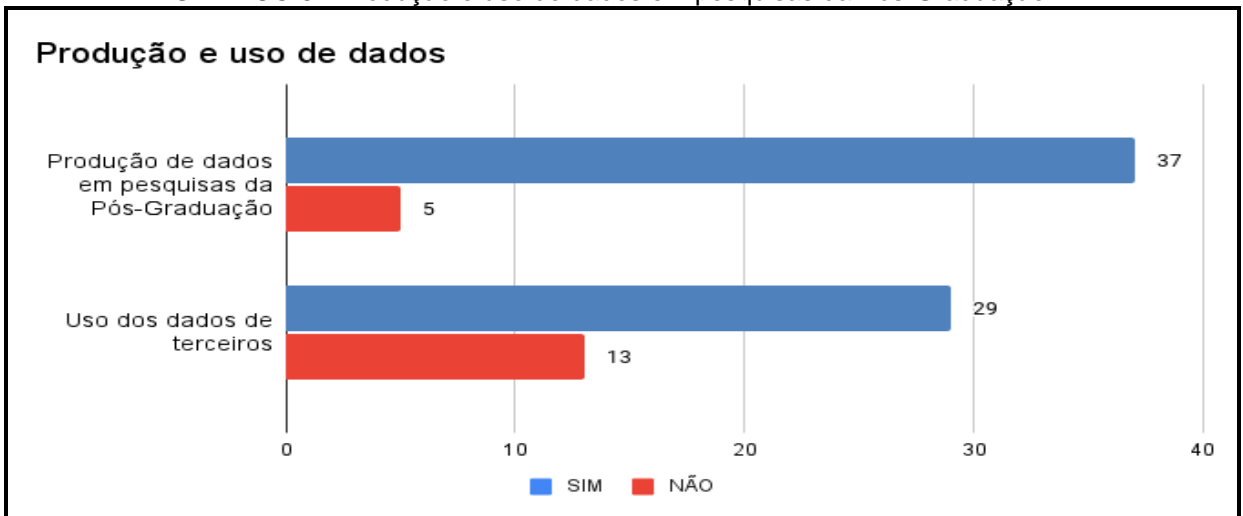


Fonte: Dados da pesquisa (2022).

Consultados sobre o tempo de atuação na Pós-Graduação, as respostas dos sujeitos revelaram os seguintes intervalos: entre 11 e 15 anos (28 %), entre 15 anos ou mais (24%), de 6 a 10 anos (24%) e de 1 a 5 anos (24%). Esses dados possibilitam atestar que os pesquisadores (52%) atuam na Pós-Graduação há mais de 11 anos. Isso mostra que são pesquisadores experientes em suas áreas de atuação e que têm práticas consolidadas, devido à expertise adquirida ao longo dos anos, atuando em pesquisa.

Na seção 3, em específico na quarta pergunta, os pesquisadores foram consultados sobre os dados gerados como resultados de suas pesquisas. Os dados oriundos com a análise das respostas constam expostos no Gráfico 3.

GRÁFICO 3 - Produção e uso de dados em pesquisas da Pós-Graduação



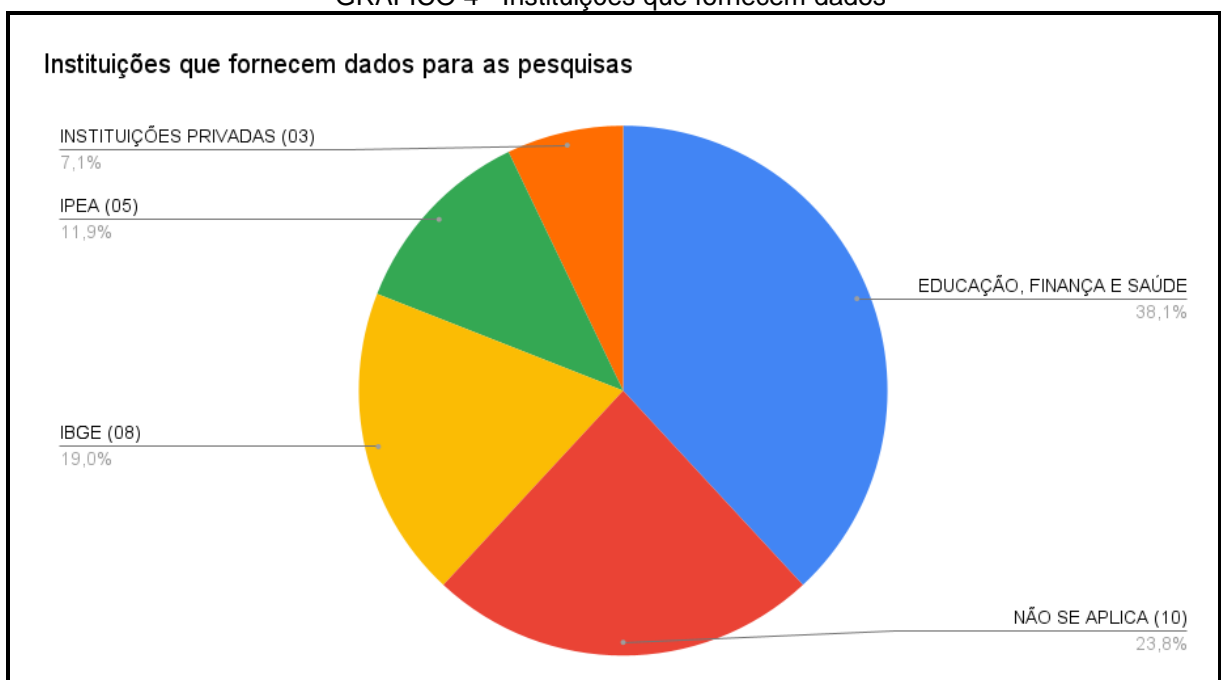
Fonte: Dados da pesquisa (2022).

Os dados revelam que a grande maioria dos pesquisadores (88%) lida com dados científicos em suas pesquisas. Responderam não gerar dados em suas pesquisas apenas cinco dos consultados (12%). Esse resultado confirma as afirmações da literatura, de que a quantidade de dados gerados nas pesquisas científicas tem atingindo proporções elevadas, sobretudo com o uso das tecnologias digitais, como apontado nos estudos de Sales e Sayão (2018), Silva (2019) e Rodrigues (2020).

Na quinta questão, sobre o uso de dados de terceiros no desenvolvimento das pesquisas, a maioria dos respondentes (69%) mencionou que usa esses tipos de dados. Todavia, chama a atenção o fato de que ainda há uma parcela que não utiliza os dados de terceiros (31%), o que pode evidenciar certa resistência quanto à possibilidade do reuso de dados de outras pesquisas.

Ainda na terceira seção, o sexto questionamento abordou as instituições que fornecem os dados usados pelos pesquisadores, sendo permitido selecionar mais de uma instituição. As instituições públicas da área da Educação, Finanças e da Saúde, as quais, em conjunto, foram apontadas pela maioria, totalizaram (38%) das menções. Em segundo lugar, o Instituto Brasileiro de Geografia e Estatística (IBGE), considerado o principal provedor de informações geográficas e estatísticas do Brasil, apresentou-se com (19%) das menções. O Instituto de Pesquisa Econômica Aplicada (IPEA), que disponibiliza dados por meio do *IpeaData*, está entre os institutos mais apontados pelos respondentes (11%). Esses resultados estão apresentados no Gráfico 4.

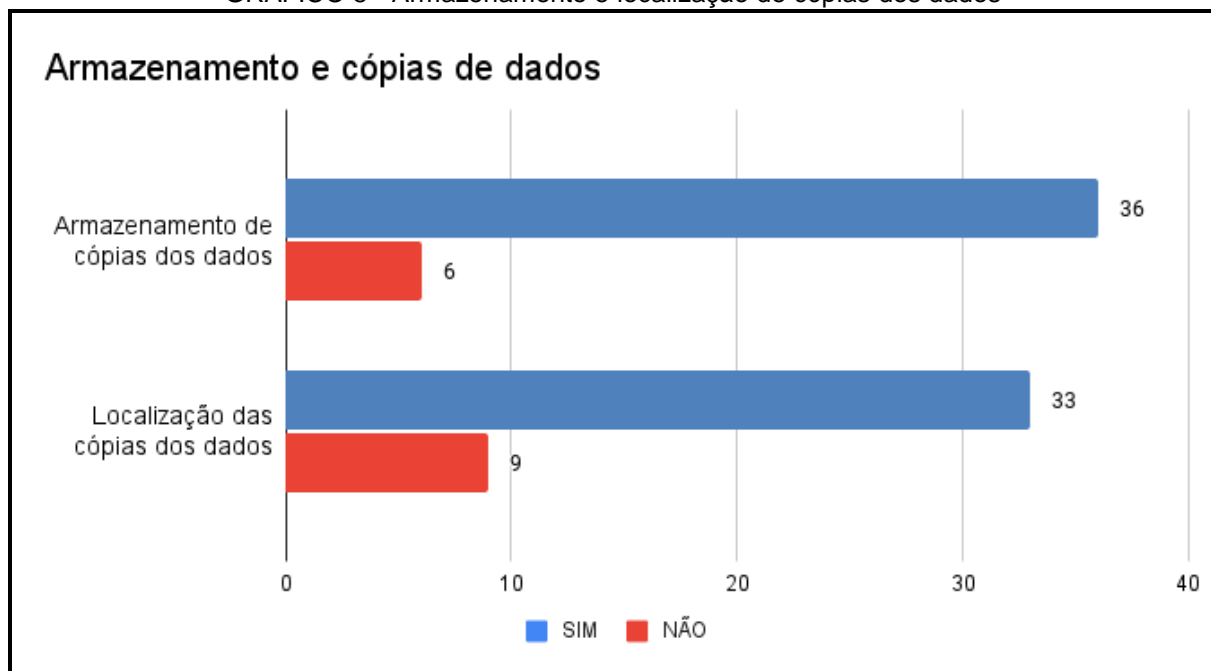
GRÁFICO 4 - Instituições que fornecem dados



Fonte: Dados da pesquisa (2022).

A quarta seção do questionário contemplou aspectos referentes ao armazenamento, primeira área funcional da matriz de preservação. Assim, por meio da sétima questão, foi consultada a existência de mais de uma cópia dos arquivos de dados. Pelos resultados, notou-se que a grande maioria mantém mais de uma cópia dos dados (86%) e apenas (14%) não possuem cópias dos dados (Gráfico 5).

GRÁFICO 5 - Armazenamento e localização de cópias dos dados



Fonte: Dados da pesquisa (2022).

A maioria das respostas vai ao encontro do que recomenda a LOP (2019), ao declarar que é importante manter cópias dos dados em diferentes suportes, formatos e ambientes. Essa ação correspondente ao nível 1 de armazenamento dos dados, tendo em vista o conhecimento sobre o conteúdo dos dados (LOP, 2019).

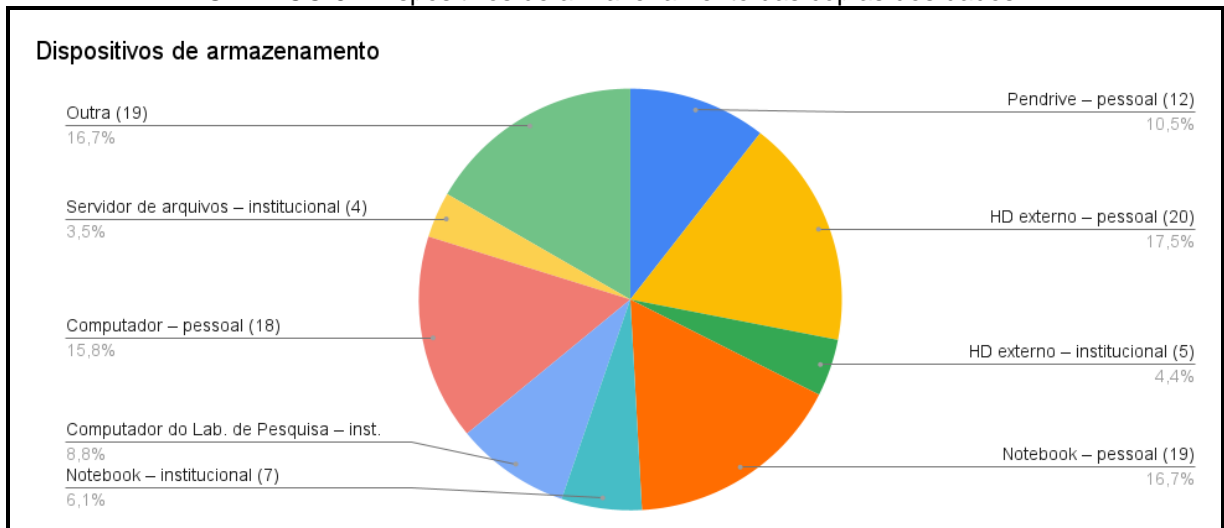
Ainda sobre o armazenamento de dados, na oitava questão, foram consultados se as cópias encontram-se em lugares distintos. A maioria dos respondentes informou que mantém cópias em outra localização (79%), ao passo que 21% registraram que não possuem essas cópias (Gráfico 5).

De acordo com a matriz LOP (2019), ter mais de uma cópia assegurada em localização geográfica distinta possibilita atingir o nível 2 (proteja o seu conteúdo). Portanto, os resultados com essa pergunta revelam que muitos pesquisadores estão atentos às estratégias de manter os dados assegurados, tendo em vista garantir a proteção e a integridade dos dados produzidos.

A nona questão analisou o tipo de dispositivo onde as cópias dos dados encontram-se armazenadas, cujo respondente poderia selecionar mais de uma opção. O Gráfico 6 demonstra que os pesquisadores preferem utilizar os recursos de tecnologia da

informação próprios do que recursos oferecidos pela instituição ou adquiridos pelos PPG para manterem os dados armazenados durante a realização das pesquisas.

GRÁFICO 6 - Dispositivos de armazenamento das cópias dos dados



Fonte: Dados da pesquisa (2022).

Na categoria “dispositivos pessoais”, os HD externos (20), computadores portáteis (19) e *desktops* (18) foram as opções mais escolhidas pelos participantes, seguidas dos *pendrives* (12). Na categoria “dispositivos institucionais e de laboratório de pesquisa”, foram apontados o uso de computadores *desktops* (10), de computadores portáteis (07), de HD externos (05) e de servidores de arquivos (04).

Concomitante ao uso dos dispositivos físicos supracitados, os pesquisadores fizeram menção ao uso de serviços em nuvem. O termo “on-line” ocorreu uma vez, sem ser vinculado à ferramenta. Contudo, o termo “nuvem” ocorreu nove vezes, sendo que, em dois momentos, foi especificado o uso das ferramentas *Google Drive* e *OneDrive*. Portanto, entre as ferramentas que os pesquisadores utilizam para armazenar os dados das pesquisas, estão o *Google Drive* (05), o *OneDrive* (02), o *Dropbox* (01), o controlador de versões *GitHub* (01), os sites e os e-mails pessoais.

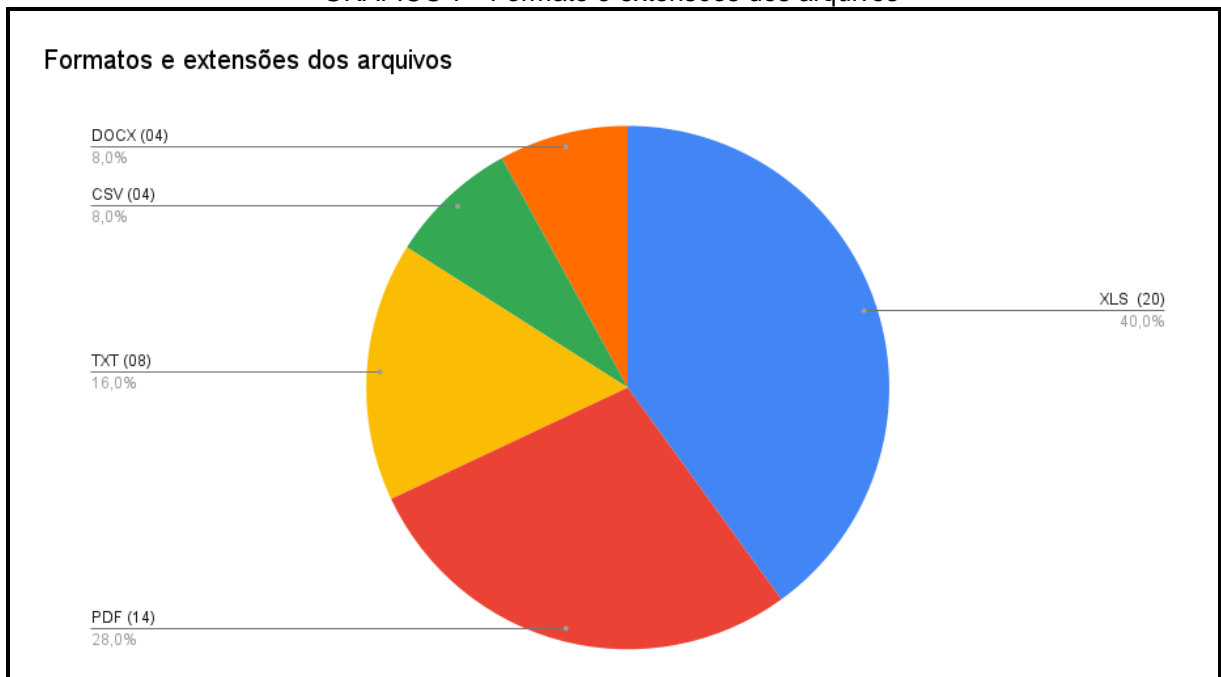
A computação em nuvem permite acessar arquivos, programas e recursos por meio da rede conectada à internet. Este tipo de serviço vem se popularizando devido à economia de custos e por possibilitar melhorias na segurança e na gestão dos dados (*STATE OF THE CLOUD, 2021*). Segundo o relatório *State of the Cloud (2021)*, a computação em nuvem, que demonstrava tendência de crescimento para os próximos anos, foi impulsionada pela pandemia da COVID-19, fazendo com que as estratégias de computação em nuvem ganhassem relevância entre as organizações e os indivíduos.

Nesse sentido, os três serviços de armazenamento em nuvem mais citados pelos pesquisadores (*Google Drive, OneDrive e Dropbox*) estão entre os dez serviços mais bem avaliados nos sites de tecnologia da informação. Importante reforçar que, ter no mínimo uma

cópia em um tipo de suporte de armazenamento distinto, armazenar a cópia em diferentes localizações geográficas e monitorar a obsolescência do armazenamento são condições que possibilitam atender o nível 3 da preservação digital relacionado à supervisão do conteúdo. Além disso, se houver iniciativas de maximizar a diversidade de armazenamento, visando a evitar pontos concretos de falha, atinge-se o nível 4 da preservação (sustente o seu conteúdo) (LOP, 2019).

A décima questão do questionário, também localizada na seção referente ao armazenamento, investigou em quais formatos as cópias dos arquivos encontram-se armazenadas, cujo participante poderia selecionar mais de uma opção. Pelos resultados, constatou-se que os formatos com maior ocorrência foram: XLS com 48% e PDF com 33%, seguidos do txt (19%) e csv e docx com 10% cada. Em menor número, com 5% de ocorrência nas respostas, foram mencionados os formatos JPGE, ppt e dat. Além desses, com ocorrência de 2% nas respostas, foram apontados os formatos ASCII, mp4, mat, TIFF e WAV. Por fim, com menor ocorrência (de 1%) estão os formatos zip, sav, dta, pzf, spw, dwg, dxf, shp, ecw, grid, img, pzm, adicht, e.g e Py. Os resultados da décima pergunta são explanados no Gráfico 7.

GRÁFICO 7 - Formato e extensões dos arquivos



Fonte: Dados da pesquisa (2022).

Entre os formatos apontados pelos pesquisadores, há formatos de arquivos aceitáveis para a preservação. Contudo, o formato XLS, mais usado pelos pesquisadores, não é adequado à preservação. Trata-se de uma extensão de software proprietário e a recomendação é que sejam utilizados softwares não proprietários (*open source*) (REDE CARINIANA, 2017). De qualquer modo, se os dados estiverem disponíveis de forma tabulada,

é possível salvar os arquivos na extensão csv, formato aceito para a preservação (ISOTANI; BITTENCOURT, 2015). O segundo formato mais indicado pelos pesquisadores, o PDF, constitui um formato aceito para a preservação, desde que esteja sob uma licença de uso aberta. Quanto aos formatos txt e csv, eles são aceitáveis e favorecem a preservação, mas é preciso observar os tipos de dados (REDE CARINIANA, 2017). Já o formato docx não é recomendado para a preservação de dados, por se tratar de uma extensão de software proprietário. Nesses casos, a literatura indica o uso de softwares de código aberto com salvamento na extensão txt ou odt (ISOTANI; BITTENCOURT, 2015).

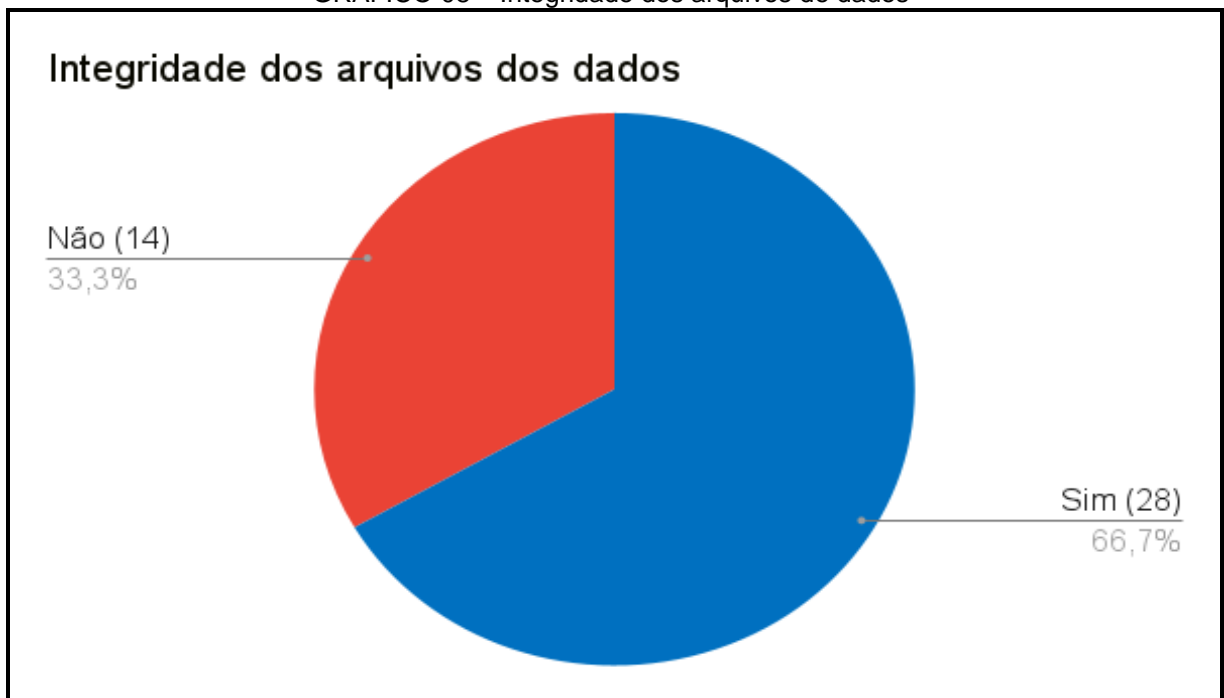
A respeito dos formatos com menor ocorrência, o conjunto de respostas apresentou uma diversidade de formatos. No entanto, houve maior abrangência no que tange aos tipos de arquivos, como: áudio, vídeo, imagens, desenhos, gráficos, linguagens de programação, ferramentas de estatística, análise e apresentação de dados científicos, entre outros.

Posto isso, é importante enfatizar que a variedade de formatos é necessária, tendo em vista atender às especificidades das áreas do conhecimento. Nesse sentido, ressalta-se a importância do uso de padrões de metadados adequados aos objetos digitais, além do constante monitoramento quanto à obsolescência tecnológica. Essa discussão está em sintonia com as recomendações da matriz LOP (2019), ao dispor que a existência de um plano e a execução das ações para tratar a obsolescência de hardware, de software e de suportes de armazenamento são condições indispensáveis para se atingir o nível 4 (sustente o conteúdo do objeto).

Cabe lembrar que, na matriz LOP (2019), os formatos de arquivos são tratados na quinta área funcional, dedicada ao conteúdo. Nessa categoria, recomendam-se duas ações básicas: 1 - documentar e verificar os formatos de arquivos e de outras características essenciais do conteúdo; e 2 – monitorar a obsolescência e as mudanças das tecnologias das quais os documentos sejam dependentes (LOP, 2019). Nesse contexto, de acordo com a matriz LOP (2019), realizar migrações, normalizações, emulações e atividades similares que garantam o acesso ao conteúdo são estratégias que possibilitarão o alcance do nível 4 da preservação digital (sustente o seu conteúdo).

Na seção 5 do questionário, foram exploradas as informações a respeito da integridade, segunda área funcional da matriz de preservação. Seguindo a ordem das perguntas, na estrutura do questionário, a décima primeira questão consultou os pesquisadores acerca da verificação dos arquivos de dados por meio de antivírus, cujos resultados podem ser visualizados no Gráfico 08.

GRÁFICO 08 – Integridade dos arquivos de dados

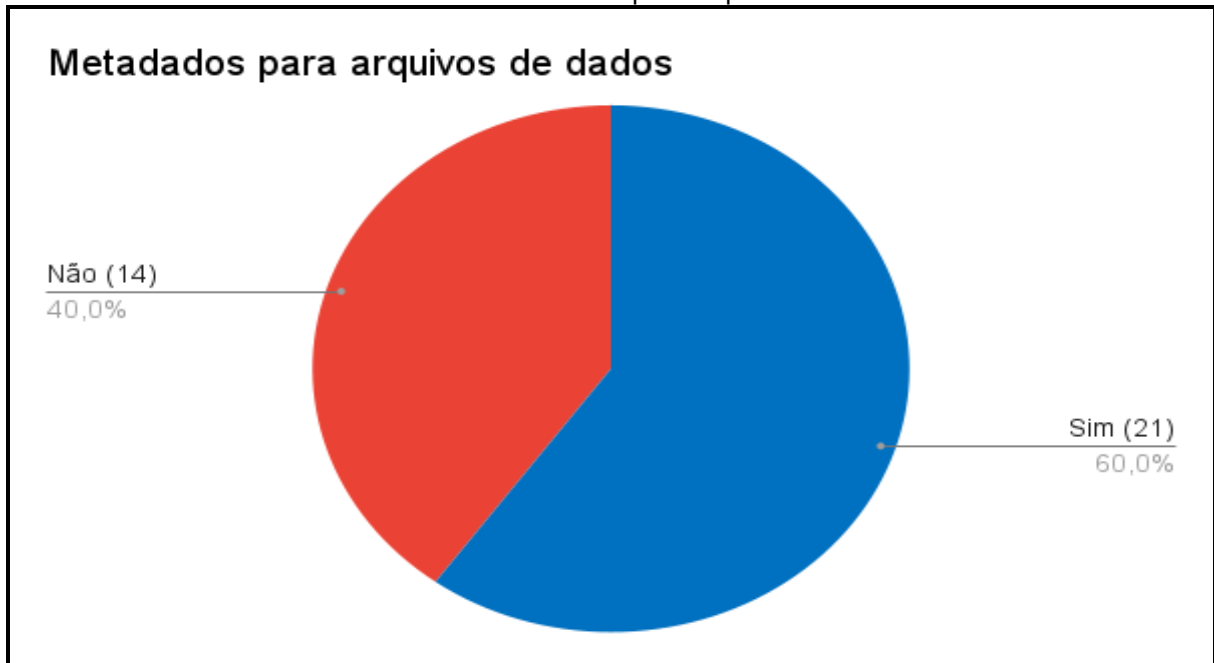


Fonte: Dados da pesquisa (2022).

Constata-se que a maioria dos pesquisadores verifica a integridade dos arquivos antes de armazená-los (67%). Uma parcela considerável (33%) não tem essa preocupação ao armazenar os dados. De acordo com a matriz LOP (2019), escanear vírus em todo o conteúdo; isolar o conteúdo em quarentena, caso seja necessário; fazer *backup* da informação de integridade; e armazenar uma cópia em localização distinta são requisitos para se atingir o nível 2 da preservação (proteja o seu conteúdo).

Na sexta seção, foram abordadas questões referentes aos metadados, quarta área funcional da matriz de preservação. Por meio da décima segunda questão, os pesquisadores foram consultados acerca da descrição dos campos quando do salvamento e da organização dos arquivos. Pelos dados coletados, nota-se que a maioria dos respondentes descrevem os campos de cada arquivo de dados (60%), ao passo que (40%) não faz essa descrição, conforme apontado no Gráfico 09.

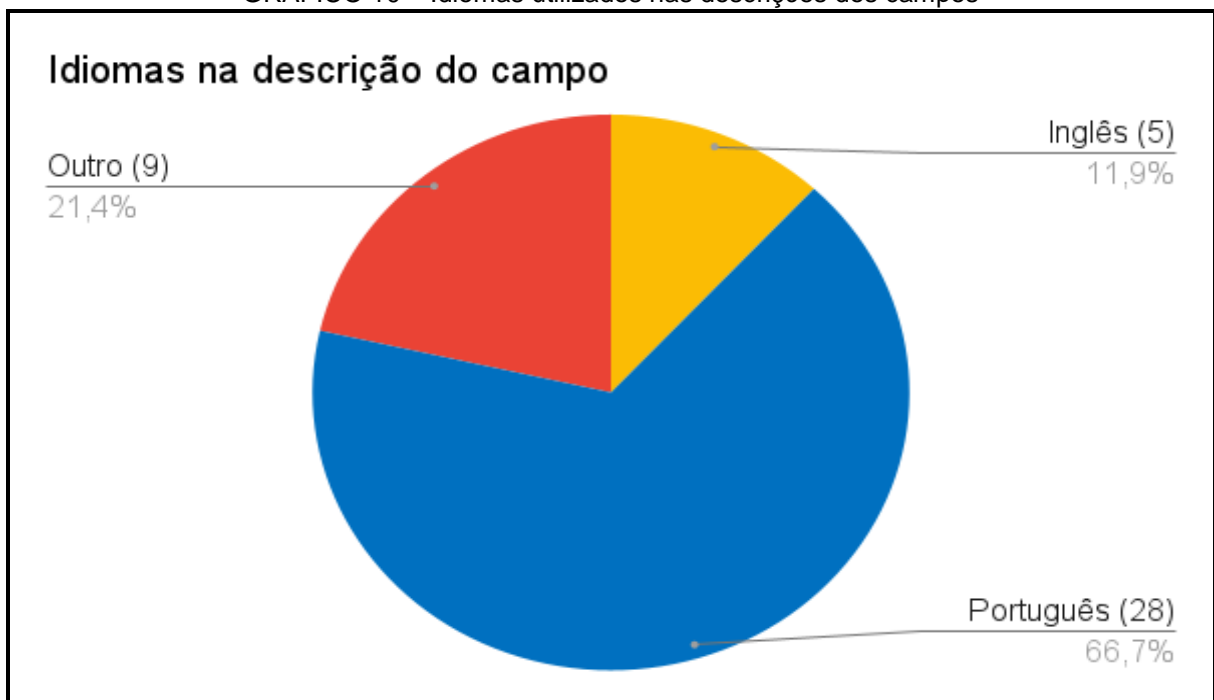
GRÁFICO 09 - Metadados para arquivos de dados



Fonte: Dados da pesquisa (2022).

Na décima terceira questão, objetivou-se saber em qual idioma os campos são descritos. Pelas respostas obtidas, verificou-se que a maioria dos pesquisadores descreve os campos dos arquivos em português (66,7%), em outros idiomas não especificados (21,4%) e em inglês (11,9%), nos idiomas Alemão, Espanhol e Francês não houve ocorrências. Tais resultados podem ser visualizados no Gráfico 10.

GRÁFICO 10 – Idiomas utilizados nas descrições dos campos

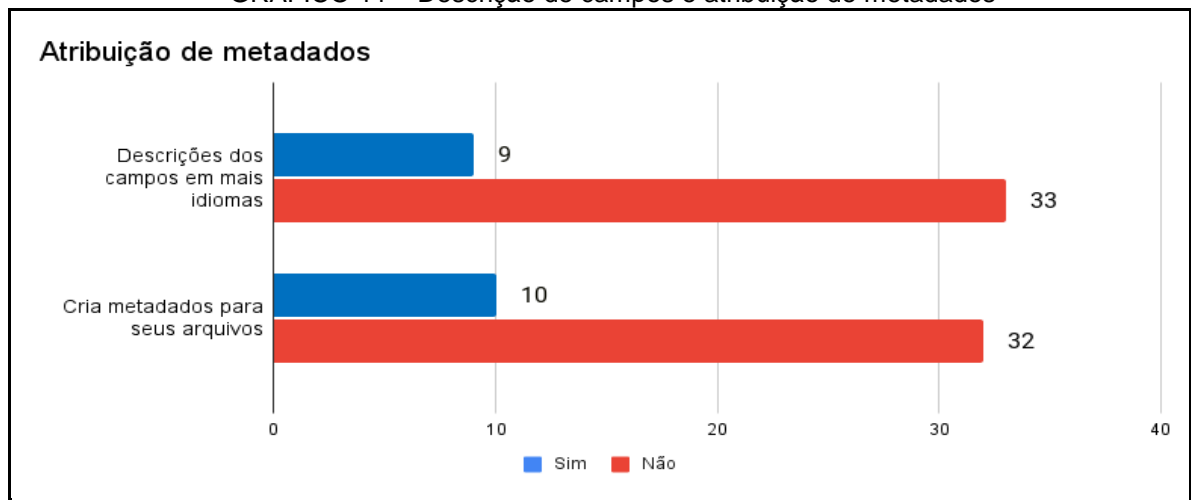


Fonte: Dados da pesquisa (2022).

A questão de número 14 consultou se as descrições estão em mais de um idioma. Responderam que não descrevem em mais de um idioma 79%, enquanto 21% informaram que descrevem os campos dos arquivos em mais de um idioma (Gráfico 11).

A questão 15, ainda na seção de metadados, indagou acerca da criação de metadados para os arquivos de dados científicos. No enunciado dessa questão, foi explicado que a criação de metadados “referia-se aos arquivos de dados como um todo: identificação do projeto de pesquisa, descrição dos dados, fonte de financiamento, data ou período de coleta, ferramenta ou instrumento de coleta”. Processando as respostas, aferiu-se que 76% responderam que não criam metadados e 24% mencionaram que criam metadados para seus arquivos de dados (Gráfico 11).

GRÁFICO 11 – Descrição de campos e atribuição de metadados



Fonte: Dados da pesquisa (2022).

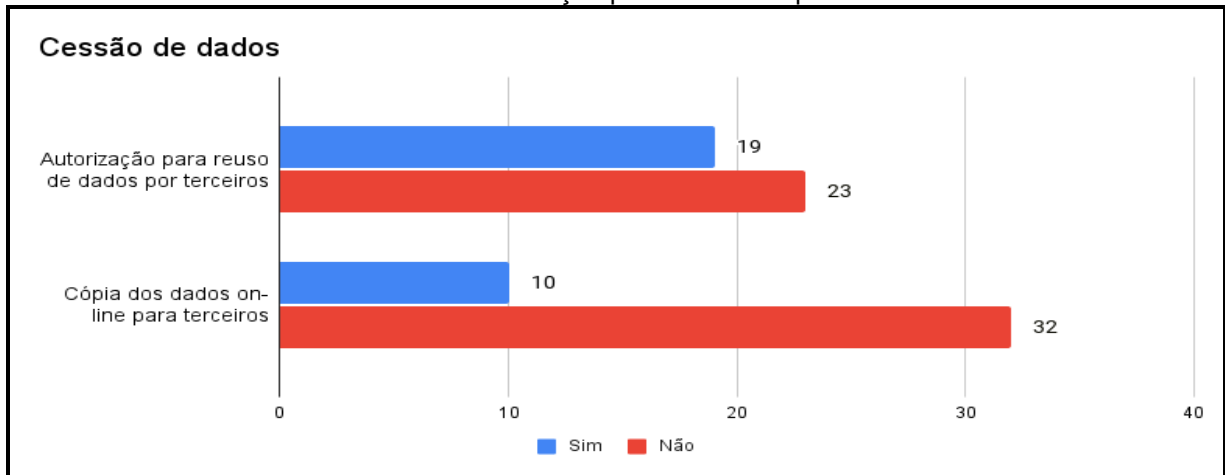
De acordo com a matriz LOP (2019), ressalta-se a importância de determinar quais padrões de metadados serão aplicados (nível 3). Ademais, é recomendável registrar as ações de preservação associadas ao conteúdo, incluindo quando essas ações ocorreram. Com efeito, o registro e a inclusão das ações possibilitam adentrar-se ao nível 4 da preservação.

A sétima seção do questionário explorou aspectos relativos à terceira área funcional da matriz LOP e ao quarto pilar dos princípios FAIR: o controle e o reuso dos dados. Mediante a décima sexta questão, foi perguntado sobre a autorização em conceder a reutilização dos dados por terceiros. Após análise dos dados coletados, constatou-se que a maioria dos participantes não cede seus dados a terceiros (55%). Os resultados mostram que há uma falta de consenso entre a comunidade científica, tendo em vista que muitos optam pela autorização do reuso de seus dados e outros, em quantidade aproximada, não autorizam a reutilização. A divisão entre as respostas é apresentada no Gráfico 12.

A oferta de uma cópia on-line dos arquivos de dados para que terceiros possam utilizá-los constituiu o assunto da questão 17 do questionário. Foi constatado, após a análise

das respostas, que a maioria dos respondentes não disponibiliza cópia dos dados on-line (76%). Ao contrário da questão anterior, nota-se uma grande diferença entre os que disponibilizam as cópias e os que não realizam essa atividade. A diferença no quantitativo das respostas é exposta no Gráfico 12.

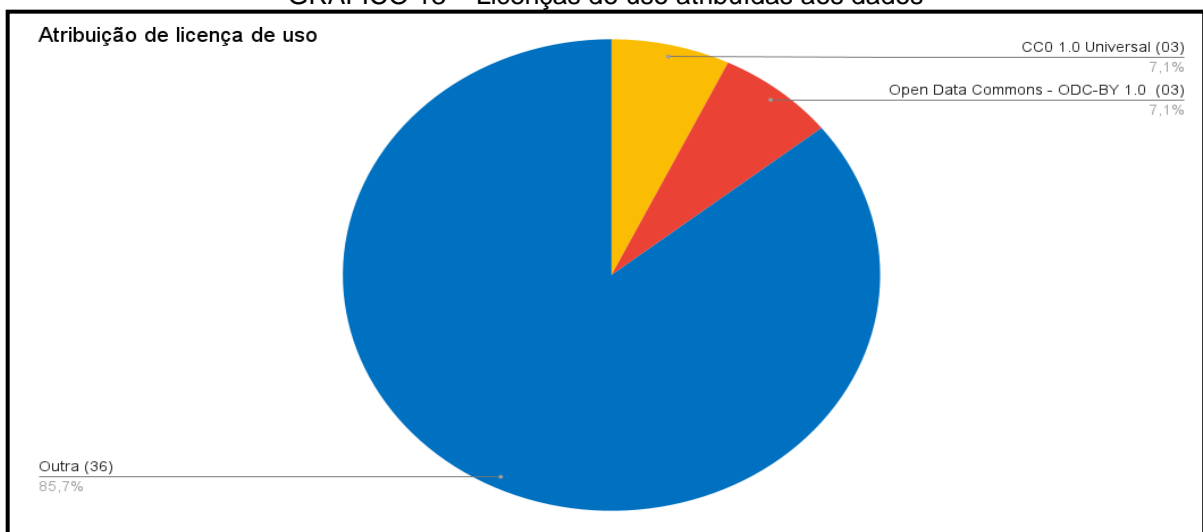
GRÁFICO 12 – Autorização para reuso e cópia de dados



Fonte: Dados da pesquisa (2022).

O tipo de licença em que os arquivos de dados são fornecidos foi explorado na questão de número 18. Pelo processamento das respostas, revelou-se que a maioria dos pesquisadores (85,7%) não aplica ou ignora o uso das licenças. Por outro lado, foi identificado o uso de algumas licenças, com a seguinte distribuição: 7,1% atribuem a licença CC0 1.0 Universal (domínio público) e outros 7,1% atribuem a licença *Open Data Commons* (ODC-BY 1.0), que permite aos usuários compartilhar, modificar e usarem, de modo livre, os dados de um determinado banco de dados. O quantitativo das respostas referente à atribuição das licenças é ilustrado no Gráfico 13.

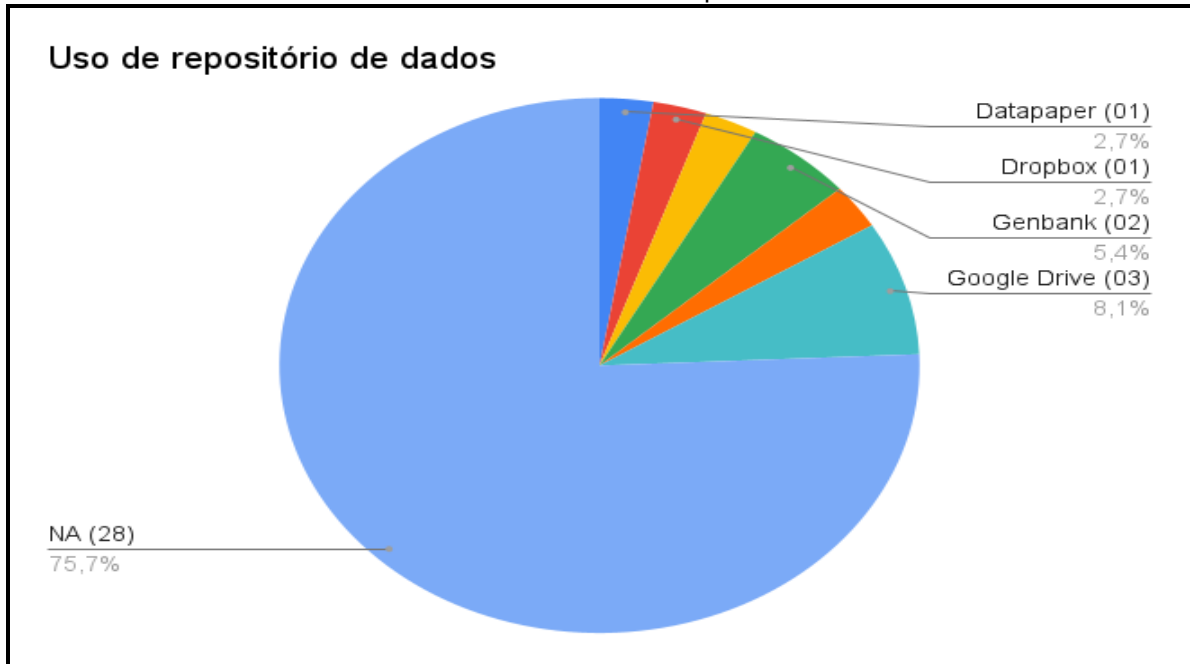
GRÁFICO 13 – Licenças de uso atribuídas aos dados



Fonte: Dados da pesquisa (2022).

O uso dos sites ou dos repositórios de dados foi consultado por meio da décima nona questão. A maioria das respostas considerou não utilizar repositórios e/ou sites para disponibilizar dados de pesquisa (75,7%). O *Google Drive* foi indicado por 8,1% dos respondentes e o *Genbank* por 5,4%, seguidos do *Dropbox*, *Dataper*, *Figshare* e *GitHub* que obtiveram, cada um, 2,7% de respostas (Gráfico 14).

GRÁFICO 14 - Uso de sites e/ou repositórios de dados



Fonte: Dados da pesquisa (2022).

A vigésima questão reforçou quanto ao principal formato dos arquivos de dados publicados nesses repositórios. Nessa questão, foi permitida a marcação de uma resposta, apenas, entre as opções. Pelos resultados, 38,1% indicaram outros formatos não relacionados entre as opções; 26,2% responderam que armazenam no formato XLS; 21,4% optaram pelo formato PDF; por fim, 14,3% declararam a escolha pelo formato csv. Não houve seleção para os formatos RDF e LOD, conforme dados expostos no Gráfico 15.

GRÁFICO 15 – Formatos de arquivos nos repositórios de dados



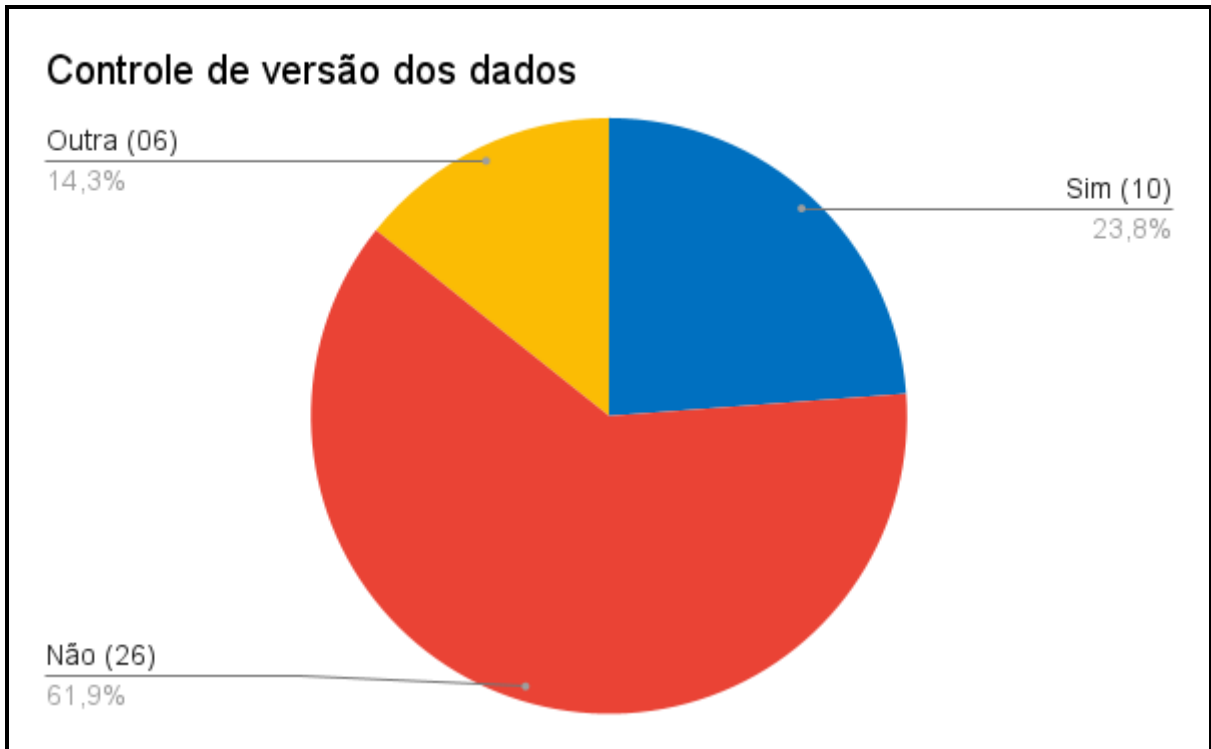
Fonte: Dados da pesquisa (2022).

A questão de número 20 muito se assemelhou à questão de número 10. Nessa questão, os formatos mais indicados foram: XLS com 48% das respostas e PDF com 33%, seguidos do txt, com 19% e csv e ocx, que obtiveram 10% cada, conforme ilustrado no Gráfico 7. Na questão 20, os consultados poderiam inserir, de modo livre, os formatos que lidam no decorrer de suas pesquisas. Assim, foi disponibilizada uma lista prévia para a seleção. Contudo, no caso de indicar outros formatos, os consultados poderiam preencher o campo especificado.

A análise dos resultados indicou que houve repetição nos formatos exibidos e analisados na décima questão e, na questão 20, os formatos XLS e PDF também foram os mais indicados pelos pesquisadores. Ressalta-se não ter havido seleção para os formatos *Resource Description Framework* (RDF) e nem para o formato *Linked Open Data* (LOD) para agregar um conjunto de dados.

Na vigésima primeira questão, ainda na seção 7, buscaram-se informações a respeito das versões e dos controles de versionamento dos arquivos. As respostas indicaram que 62% não controlam as versões, 24% controlam as versões e 14% realizam o processo, às vezes, ou não o realizam. Esses resultados estão apresentados no Gráfico 16.

GRÁFICO 16 – Controle de versão dos dados

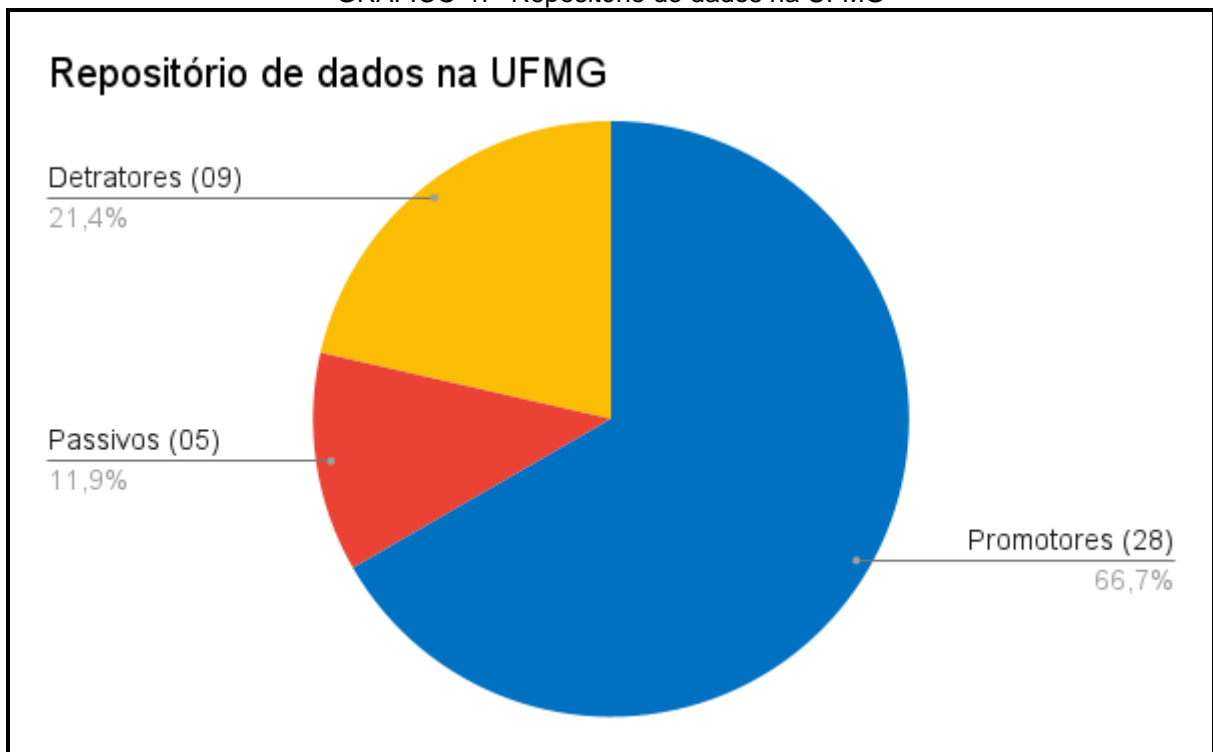


Fonte: Dados da pesquisa (2022).

De acordo com a matriz LOP (2019), determinar os agentes humanos e softwares que devem ter autorização para ler, gravar, mover e eliminar conteúdo é um fator básico para a efetivação do nível 1 da preservação; documentar e pôr em prática as definições do primeiro nível consolidará o nível 2; manter registros das ações (*log*) e identificar quais os agentes humanos e os softwares que realizam ações junto ao conteúdo oportunizará o nível 3; por fim, realizar revisões periódicas de ações/registros de acesso é uma condição específica para adentrar-se ao nível 4 da preservação.

Na vigésima segunda questão, foi realizada uma pesquisa de opinião acerca da disponibilização, pela UMFG, de um repositório de dados. Indagou-se aos participantes se esse repositório facilitaria o compartilhamento e o reuso dos dados de pesquisas. Para 66,7% dos participantes, o repositório facilitaria a gestão dos dados. Ao contrário, 21,4% acreditam que o repositório não promoveria a gestão. Por sua vez, 11,9% entendem que não haveria promoção definitiva da gestão, mas o repositório seria um facilitador para o compartilhamento e para o reuso dos dados. Os dados obtidos com as respostas à questão 22 aparecem no Gráfico 17.

GRÁFICO 47- Repositório de dados na UFMG



Fonte: Dados da pesquisa (2022).

Paganine e Amaro (2020, p. 177) salientam que “os repositórios de dados de pesquisa têm sido sistematicamente propostos pela literatura, como a ferramenta adequada para o compartilhamento desses dados de pesquisa [...]”. Os autores alertam que há desafios a serem superados quanto à padronização dos sistemas e à efetivação da curadoria dos dados. As atuais demandas sobre dados advêm da Ciência Aberta, que reforça a importância do uso desse tipo de ferramenta, haja vista facilitar o armazenamento, a organização, o compartilhamento e a divulgação dos dados científicos (PAGANINE; AMARO, 2020).

A seção 8 do questionário foi composta por três perguntas, cujos pesquisadores foram convidados a conceder uma breve entrevista, com a intenção de aprofundar os resultados da pesquisa. As perguntas foram pontuais, solicitando a aceitação ou não do convite em participar da entrevista, como também indicando um dia e um horário para a concretização dessa prática.

4.3 Análise das entrevistas

No intuito de aprofundar às práticas de gestão, de curadoria e de preservação dos dados científicos, os entrevistados foram submetidos a 17 perguntas. As perguntas foram distribuídas em sete seções ou conjuntos temáticos, considerando o que determina a LOP (2019). O roteiro simplificado da entrevista com o grupo PCDC, está relacionado no Apêndice F.

Nas áreas das Ciências Humanas e Sociais Aplicadas, a técnica da análise de texto é utilizada com frequência. Isso porque essa análise possibilita sintetizar e comparar o conteúdo qualitativo do material escrito ou falado, por exemplo, as entrevistas de inquérito, de recrutamento, de psicoterapia, dentre outras, constituindo, pois, “[...] um material verbal rico e complexo” (BARDIN, 2016, p. 93).

Quanto à organização do conteúdo a ser analisado, optou-se pela categorização. A categorização “é uma operação de classificação de elementos constitutivo de um conjunto por diferenciação, e em seguida, por reagrupamento segundo o gênero (analogia), com os critérios previamente definidos” (BARDIN, 2016, p. 147). A autora esclarece que o critério de categorização pode ser semântico, sintático, léxico e expressivo.

A categorização semântica é constituída pelas temáticas que têm significado e compõem a pesquisa (BARDIN, 2016). Assim, após a apresentação do perfil dos participantes, foram descritos extratos de algumas das afirmativas dos entrevistados, em cada categoria temática, considerando que o roteiro foi organizado em seções de interesse da pesquisa.

Como já mencionado, os entrevistados atuam como professores da Pós-Graduação, condição para participar da pesquisa. De imediato, 11 sinalizaram que poderiam conceder a entrevista. Porém, quando do agendamento, dois não estavam com agenda livre no período destinado à coleta dos dados. Esses justificaram suas ausências, não participando da pesquisa, devido ao início do semestre letivo na UFMG (28 de março de 2022), ao retorno das aulas presenciais, entre outros compromissos assumidos por eles.

Por ser facultativa a participação, nove professores se voluntariaram a conceder a entrevista, por meio da plataforma da RNP. Em atendimento aos termos aprovados pelo COEP, a identidade dos participantes foi mantida em sigilo, de modo a garantir o anonimato deles, como também atender à Lei n. 13.709/2018, nomeada como Lei Geral de Proteção de Dados Pessoais (LGDP). Desse modo, os nomes foram substituídos por identificação codificada, ou seja, foram representados pela letra E, que significada “entrevistado”, e seguidos por um número sequencial correspondente, assim: E1, E2, E3, E4, E5, E6, E7, E8 e E9.

As entrevistas foram representadas por cinco áreas do conhecimento. A área com maior representatividade foi a de Ciências Biológicas, com três participantes, seguida das áreas da Saúde (02), Humanas (02), Engenharias (01) e Ciências Econômicas (01). Os dados demográficos referentes a áreas do conhecimento, ao PPG em que o participante atua e ao tempo de seu vínculo junto à instituição são mencionados no Quadro 12.

QUADRO 12 - Perfil dos participantes da entrevista

ID	Área do conhecimento, conforme o CNPQ	Atuação no PPG	Tempo de vínculo na UFMG (em anos)
E1	Engenharias	Engenharia Eletrônica; Neurociências	Mais de 15

E2	Ciências da Saúde	Educação Física	6 a 10
E3	Ciências Humanas	Educação	11 a 15
E4	Ciências Biológicas	Biofísica	Mais de 15
E5	Ciências Humanas	Educação	Mais de 15
E6	Ciências Econômicas	Economia	Mais de 15
E7	Ciências da Saúde	Farmácia	Mais de 15
E8	Ciências Biológicas	Ecologia	11 a 15
E9	Ciências Biológicas	Evolução	11 a 15

Fonte: Dados da pesquisa (2022).

A maioria dos entrevistados atua na Pós-Graduação da UFMG há mais de 15 anos. Por esse resultado, evidencia-se que os pesquisadores são experientes em suas áreas de atuação e que têm práticas consolidadas, devido à expertise adquirida ao longo dos anos realizando pesquisas.

A seguir, apresenta-se a tabulação das respostas dos entrevistados em cada seção do roteiro da entrevista. Em específico, o Quadro 13 apresenta a tabulação da primeira seção, destinada a coletar informações sobre a produção/consumo de dados pelos pesquisadores. No referido quadro, são descritos alguns trechos extraídos do conteúdo das entrevistas.

QUADRO 13 - Seção produção/consumo de dados

ID	Produção/consumo de dados	
1PC	- Qual(ais) tipo(s) de dados científicos são produzidos em suas pesquisas?	
2PC	- Quais tecnologias são utilizadas?	
	Tipos de dados	Tecnologias utilizadas
E1	Dados de medições biológicas ou ligados ao ser humano; áudio (voz e fala), vídeo (pessoas e regiões de preservação ambiental); imagens e imagens hiperespectrais - preservação ambiental – fogo ou fumaça. Dados sobre encefalografias para potenciais evocados (relacionados a eventos).	Câmeras hiperespectral e Encefalografia.
E2	Dados experimentais (animais e seres humanos); Pesquisa observacional com seres humanos. Dados numéricos e observacionais. Figuras, gráficos, tabelas e textos.	Editores de planilhas e de textos e questionários.
E3	Dados de registro de observação, gravação de áudio e gravação de vídeo. Esses dados são registrados em formatos variados (dados qualitativos) e notas de leitura.	Editor de texto e transcrição dos áudios.
E4	Dados experimentais. Sinais elétricos em tecido. Gráficos e formatos do software específico.	Amplificadores especializados para gravar registro de sinais elétricos em tecido.
E5	Dados observacionais. Observação nos espaços escolares. Áudio e vídeo.	Suporte em fitas VHF (desde o início das pesquisas). Fitas K7. As fitas foram

		passadas para o digital. Fichas em papel foram digitalizadas. Uso de SPSS para preservar.
E6	Fotografias de satélites, filmes, relatórios e questionários. Dados computacionais.	Software de georreferenciamento.
E7	Dados experimentais. Vídeos.	Bioinformática. Ferramentas que buscam dados da literatura. Software para facilitar a busca por Toxicogenômicas.
E8	Dados observacionais (biológicos).	Tablet, celular, computador, base de dados. As fichas de campo são transcritas.
E9	Dados experimentais e observacionais. Imagens e gráficos.	Microscópio e computadores.

Fonte: Dados da pesquisa (2022).

O Quadro 14 mostra a heterogeneidade dos dados produzidos pelos entrevistados. Isso está em concordância com o que revelou o estudo de Silva (2019), destacando que os dados científicos ou de pesquisa são informações registradas ou produzidas por meio de qualquer recurso e em diferentes formatos durante o decurso de uma pesquisa (SILVA, 2019). Em virtude dessa variedade de possibilidades de registros, as amostras coletadas são heterogêneas, conforme constatado no quadro anterior e proclamado por Silva (2019).

De acordo com os consultados, os tipos de dados gerados e as tecnologias utilizadas pelas áreas são, na maioria das vezes, dados observacionais, experimentais e computacionais. Dentre as tecnologias utilizadas, predominam as simulações, vídeos, áudios, imagens, gráficos, textos e planilhas. Esse resultado evidencia a peculiaridade e a preferência nas práticas de pesquisa de cada área de conhecimento. Sobre isso, Sales e Sayão (2018) afirmaram que alguns aspectos da atividade de curadoria variam de acordo com o tipo de dados, com as tecnologias subjacentes aos dados e, sobretudo, com o domínio disciplinar específico.

Em seguida, foi analisada a segunda seção do roteiro de entrevista. Assim, no Quadro 14, são registrados os dados respondidos pelos entrevistados sobre o armazenamento e a perda de dados científicos durante as práticas de pesquisa no PPG.

QUADRO 14 – Seção armazenamento/cópias

ID	Armazenamento/Cópias	
1AC	- Você já perdeu arquivos de dados importantes para suas pesquisas? Se sim, comente o fato.	
	Perda de dados integral/parcial	Nunca perdeu dados
E1		Não ocorreu. Temos infraestrutura de HDs Raid5, com redundância para reconstrução de conteúdo em caso da queima de algum HD. Os backups são realizados diariamente. Os equipamentos ficam nos laboratórios do Centro de Estudos da Fala, Acústica, Linguagem e Música (CEFALA) e no Departamento da Engenharia Elétrica.
E2	Sim.	

E3		Não. Mas, os manuscritos em cadernos foram se apagando. Foi preciso digitar os dados.
E4	Sim. Um aluno de Doutorado havia guardado uma única cópia dos dados (parcial) no laptop que foi furtado.	O aluno precisou refazer os experimentos.
E5	Sim. A perda ocorreu em pesquisas longas. Houve perdas por causa de mudança de tecnologias.	
E6		Não perdeu.
E7		Não perdeu.
E8		Não perdeu. Salvamos os dados em nuvem.
E9		Não. Salvamos os dados em nuvem. Cada um cuida dos seus dados.

Fonte: Dados da pesquisa (2022).

Na consulta sobre perda de arquivos de dados, o E1 respondeu que nunca havia perdido dados. No entanto, relatou a perda de uma unidade de armazenamento de um equipamento importante. Contudo, devido à infraestrutura tecnológica que foi montada no laboratório, havia cópias de segurança, conforme explica o participante:

Teve uma vez que a gente perdeu todos os nossos SSDs de um servidor importante. Mas os dados estavam guardados no HD de 10 terabytes que ficam dentro do mesmo computador para armazenamento de dados (Comentário do E1).

Caso parecido aconteceu com o E2. Esse entrevistado informou que o impacto da perda foi mitigado devido ao uso do serviço em nuvem, conforme consta no registro de seus comentários:

Recentemente, eu tive um HD externo que parou de funcionar. E, a princípio, cerca de 80% das informações estavam salvas também numa nuvem. Mas, cerca de 20% foi perdido e não consegui restaurar, facilmente. Ainda tenho o HD externo, mas talvez precise de um serviço mais profissional. E, nesse HD externo, eu perdi alguns dados mais antigos da minha tese de Doutorado e da dissertação (Comentário do E2).

Nota-se que há ocorrências de perda e preocupações dos pesquisadores quanto ao armazenamento seguro dos dados. Ao relatar sobre a montagem do laboratório de pesquisa, o E6 mencionou a preocupação no que tange à manutenção dos dados coletados.

A área do laboratório é fechada e as coletas ocorrem e, diariamente, são deslocadas para o laboratório [...]. No laboratório, tem backups; tem tudo lá dentro. Mas, isso era muito importante, porque desde o início a nossa preocupação foi manter os dados (Comentário do E6).

Estes relatos corroboram as preocupações mencionadas na literatura e o alerta de que a informação digital pode ser perdida com facilidade. Boeres (2004, p. 35) reforçou que “[...] a informação digital é frágil e fugaz, necessita de gerenciamento cuidadoso desde o momento da sua criação e uma abordagem política e estratégica de modo a garantir a preservação”.

O próximo item analisado no contexto das entrevistas foi a garantia da integridade dos dados. No Quadro 15, os entrevistados responderam sobre essa integridade.

QUADRO 15 - Seção integridade dos dados

ID	Integridade				
1IN 2IN	- Algum arquivo se corrompeu ou foi recuperado parcialmente? De qual tipo ou formato? - Foi necessário coletar os dados, novamente? Quais as dificuldades encontradas durante a coleta?				
Arquivo corrompido integral/parcial	Nunca perdeu dados	Tipo/formato	Foi necessário coletar	Dificuldades encontradas	
E1	Sim. Porém, não foram dados de pesquisa.		Conteúdo do site dos professores da Engenharia Eletrônica e Elétrica.	Sim. Os professores tiveram que reconstruir suas páginas baseadas nos backups.	Ataque <i>ransomware</i> (sequestro de dados do site) e perda do conteúdo de um HD que armazenava os dados das páginas dos professores.
E2	Sim. Dados publicados.		XLS	Não precisou. São dados muito antigos. Dados publicados.	Trata-se de uma prática pessoal (uso do <i>Google Drive</i>). Não é aberto como gostaria que fosse.
E3		Nunca perdeu.			
E4		Nunca perdeu.			
E5	Sim.			Não. Invalida a pesquisa.	Mudanças de tecnologias (CDs e disquetes).
E6		Nunca perdeu.			
E7		Nunca perdeu.			
E8	Sim.			Não. Havia outras cópias.	
E9		Nunca perdeu.			

Fonte: Dados da pesquisa (2022).

Quando consultado sobre a integridade dos dados e as dificuldades encontradas, o E1 explicou acerca da existência de dois fatos:

A gente teve dois grandes problemas: a máquina virtual que cuidava das *homepages* de todo mundo sofreu um ataque *ransomware* e todos os dados foram seqüestrados [...].

Em 2010, isso há muito tempo, apesar da estrutura *Raid5* que a gente usava, eu fiz uma besteira e a gente acabou perdendo o conteúdo dos HDs. Mas, eram também as páginas dos professores. Dados mesmo de pesquisa estavam armazenados em lugar separado (Comentário do E1).

Os relatos indicam evidências sobre a importância de um trabalho colaborativo para garantir o armazenamento seguro dos dados de pesquisa nas instituições. Portanto, ficou evidenciado que, para lidar com os dados, é necessária a existência de uma equipe multidisciplinar para a realização de um trabalho conjunto entre pesquisadores, profissionais da informação, profissionais da computação, estatísticos, entre outros.

A seção seguinte do roteiro foi destinada à análise dos metadados. No quadro 16, estão registradas as respostas dos entrevistados quanto ao trabalho de representação dos dados.

QUADRO 16 - Seção metadados

ID	Metadados		
1ME	- Você mantém algum tipo de organização que facilita a localização dos seus dados como uma lista ou a disposição em pastas por tipo de dados?		
2ME	- Há algum responsável pelas coleções de dados de pesquisa? Que cargo ele ocupa?		
	Organização dos dados	Responsável pelas coleções	Cargo ocupado
E1	Sim. Organizados em bancos de dados (sistemas de e-mail, entre outros).	Especialista em banco de dados.	Pesquisador.
E2	Sim. Organizamos a partir dos nomes dos alunos, classificando-os em Mestrado ou Doutorado. Também utilizamos subpastas para dados mais brutos da coleta de dados.	Os responsáveis são o professor e o aluno envolvido na pesquisa.	Professor e aluno.
E3	Sim. Organização da documentação individual e digitada. Documentação salva em HD externo.		Professor e aluno.
E4	Sim. Fazemos uso do caderno de experimento. A estrutura dos dados junto ao protocolo é considerada. Uso de planilha no <i>Google Drive</i> .	Professor e o aluno.	Professor e o aluno.
E5	Sim. Uso de códigos de metadados para clareza dos dados quando das pesquisas multicêntricas.	Professor.	Professor.
E6	Sim. Organização interna ao laboratório. Os dados são organizados pela codificação definida pelo Tribunal da Justiça de Minas Gerais (TJMG).	Professor.	Professor da computação e outro da Arquitetura.
E7	Sim. É utilizado o Procedimento Operacional Padrão (POPs).	Pós-doutoranda.	Pós-doutoranda.
E8	Sim. Organizamos por nome e por data (sem padrão). Uso do Padrão <i>Darwin Core</i> .	Professor e aluno.	Professor e aluno.
E9	Sim. Dados organizados por planilhas com siglas criadas para identificar.	Professor e aluno.	Professor e aluno.

Fonte: Dados da pesquisa (2022).

Quando consultado sobre a organização e a descrição dos dados, o E1 explicou como os dados são organizados. Segue relato do participante:

[...] No CEFALA, a gente tem um dos pesquisadores que fez Doutorado com a gente e agora está na Universidade Federal dos Vales do Jequitinhonha e

Mucuri (UFVJN). Ele é um super especialista em banco de dados. Então, a gente tem tudo muito bem organizado em banco de dados, como os próprios sistemas nossos, os sistemas de e-mail, dentre outros. Está tudo organizado em banco de dados (Comentário do E1).

Sobre esse assunto, o E4 explicou que há esforços no sentido de organizar os dados em planilhas, conforme relato registrado:

Já houve situações em que tentamos várias coisas que podiam ser importantes. Criei uma planilha no Excel e coloquei à disposição de cada chave e alguns metadados. Eu já criei isso, mas a minha experiência é que nunca chegou na versão final porque cada vez que eu comecei a usar, percebi que faltava mais alguma coisa (Comentário do E4).

O E9 informou que a organização é realizada em tabelas, tal como demonstra o relato: “[...] Há uma tabela onde tem que colocar tudo que chega, e que é analisado e que todo mundo pode consultar [...]” (Comentário do E9). Segundo o entrevistado, essa iniciativa é super importante, pois garantirá que os dados não sejam perdidos ao longo dos anos.

As respostas e os relatos sinalizam que a organização dos dados é feita por pesquisadores e por seus orientandos. Em alguns casos, quando há especialistas ou algum egresso do curso participando do grupo de pesquisa, esses se encarregam das tarefas de organização para facilitar que o dado seja localizado.

Quando os participantes foram consultados sobre o uso dos serviços da biblioteca na oferta de apoio e de orientação, as respostas foram as seguintes:

A gente gostaria do apoio. O Departamento de Ciência da Informação, Biblioteconomia que chamava antigamente, é uma das poucas unidades da universidade que tenho tido pouco contato. Na época que eu coordenei a incubadora de empresas, eu tive contato com a gente de lá. Só nessa época (Comentário do E1).

Não procurei. E acho que é necessário. Às vezes, para que a gente busque alguma coisa, é preciso um empurrão [...]. Nós tendemos a acomodar no status atual. E, até por desconhecimento meu, eu não sabia que esse cuidado com os dados faz parte da profissão de um bibliotecário (Comentário do E2).

Não. Muito raramente, embora o pessoal aqui da biblioteca [...] sejam excelentes. Quando eu precisava [...], há muito tempo que eu não preciso. Mas, quando eu precisava, o pessoal atendia na maior boa vontade e competência. Só tenho elogios ao pessoal de lá (Comentário do E7).

O desconhecimento sobre a atuação dos bibliotecários e dos Sistemas de Bibliotecas na gestão de dados tem sido retratado na literatura. Do mesmo modo, a literatura tem revelado o potencial do bibliotecário para oferecer suporte aos serviços voltados para a curadoria e para a gestão dos dados, como salientou a tese de Almeida (2019). Silva (2019) afirmou que a maioria dos serviços de gestão de dados está relacionada com o apoio às necessidades dos investigadores em todo o processo de pesquisa, cujos bibliotecários podem desempenhar o seu papel na maioria das etapas desse processo.

No Quadro 17, os entrevistados responderam sobre controle e reuso de dados científicos. Assim, foi concretizada a quinta seção do roteiro de entrevista.

QUADRO 17 - Seção controle e reuso

ID	Controle/reuso			
1CR	- Você compartilha dados coletados em suas pesquisas com terceiros?			
2CR	- Se sim, com quem? Com seus colegas de laboratório, do Programa de Pós-Graduação ou com seus orientandos ou pesquisadores externos à UFMG?			
3CR	- Neste caso, existe algum termo de cessão ou obrigação de citar você como o produtor dos dados?			
4CR	- Se não há nenhum termo ou concessão, você cederia seus dados se fosse solicitado? Por que não cederia?			
Compartilha dados		Com quem	Termo de cessão/ Citação	Cessão de dados
E1	Sim. O Banco de dados do CEFALA (banco não é aberto, mas o interessado pode requerer acesso).	Qualquer pessoa. Gente do mundo inteiro. Mas, a maioria são pessoas que estão ligadas ao grupo de pesquisa.	Termo de cessão, não. Citar o artigo que descreve o banco de dados do CEFALA, sim.	Sim. Solicitar por e-mail.
E2	Sim. Com as revistas.	Por meio das revistas que exigem o envio dos conjuntos de dados, indiretamente. Compartilho com terceiros e com pesquisadores fora da UFMG.	Não há termo. Fica registrado no e-mail.	Solicitação por e-mail.
E3	Não.	Pesquisa nacional desenvolvida com 33 pesquisadores; compartilham-se os dados com eles.	Não. Tratou-se de uma pesquisa em parceria.	Depende. Com mestrando e doutorando, não faz sentido. É um trabalho que ele tem que fazer. Para um colega, não me aconteceu nunca. Mas, havendo autorização da fonte de dados, sim.
E4	Não.	Pesquisa entre professor e aluno.	Quando publicados, há uma declaração de que os dados serão disponibilizados mediante requisição.	É difícil alguém do laboratório ou do programa querer os dados, porque a pesquisa é individual.
E5	Não.	Alunos e colegas parceiros na pesquisa.	Não.	Sim. Desde que utilize o termo de cessão.
E6	Não.	Não é público. Somente o TJMG.	Sim. Proposto pelo TJMG.	Não.
E7	Sim. Nas revistas.			Sim.
E8	Sim.	Sim. Com os grupos que lideram as pesquisas.	Sim. Termo de parceria na instituição.	Sim.
E9	Sim.	Sim. Com os parceiros da pesquisa.	Não.	Se já estiver publicado, sim.

Fonte: Dados da pesquisa (2022).

Sobre o controle e o reuso dos dados, o E2 relatou como ocorre em sua área de pesquisa, conforme comentário:

Dados são cedidos sem nenhum problema. É comum, na minha área, nós recebermos contatos de outros pesquisadores, que solicitam dados para inclusão como metanálises, revisões sistemáticas e metanálise. Se o contato é gentil e, claro, o que ocorre, basicamente, em todas às vezes, nós cedemos os dados sem nenhum problema. Até porque nós vemos como benefício o fato de o nosso trabalho estar sendo citado. E, dentro do laboratório, esse compartilhamento é mais tranquilo e mais comum (Comentário do E2).

Contudo, o E2 externalizou a preocupação em compartilhar dados com terceiros e não receber os créditos pela coleta.

Há preocupação com o compartilhamento de dados, pois tenho receio de que terceiros peguem esses dados e construam o conhecimento sem dar crédito a quem coletou os dados. Eu lembro que, na época em que surgiu a preocupação com o vírus Zica, um grupo de pesquisadores brasileiros codificou o vírus, determinou a estrutura genética dele e compartilhou essa estrutura. Com isso, um grupo, se não me engano do Reino Unido, utilizou esses dados e produziu um trabalho na Nature. Então, eu vejo essa preocupação de muitos colegas disponibilizarem os dados e os grupos mais estruturados, com mais recursos, utilizem esses dados, produzam conhecimento e não deem crédito ao original a quem coletou aqueles dados (Comentário do E2).

Nesse sentido, o E8 demonstrou ser favorável ao compartilhamento e ponderou: “[...] Eu defendo o compartilhamento geral e público da informação, desde que ela não prejudique o andamento do trabalho” (Comentário do E8). O E4, por sua vez, mencionou que conhece iniciativas referentes à produção de novos conhecimentos a partir de dados compartilhados:

Eu já conheço algumas iniciativas nesse sentido em outros lugares. Eu conheço pesquisadores no Brasil que conseguiram ter acesso a dados de pesquisadores do mundo, com permissão, e conseguiram produzir resultados novos a partir desses dados (Comentário do E4)

O relato do E4 reforça a ideia do reuso dos dados, como foi apresentado pela FIOCRUZ:

No que tange à gestão de dados de pesquisa, destaca-se o princípio “Reusável”, pois considera-se a linhagem dos dados (gerado ou coletado); suas particularidades ou limitações; sua data da geração do conjunto de dados, condições de laboratório, a versão dos dados arquivados e/ou reutilizados deve ser claramente especificada e documentada, entre outras (FIOCRUZ, 2021, não paginado).

Nesse contexto, a gestão de dados adequada promove a investigação e assegura a reprodutibilidade da pesquisa, além de cumprir com os requisitos estabelecidos pelo financiador da pesquisa. De qualquer modo, a dinâmica da gestão de dados funcionará com qualidade se estiver atendendo às normativas registradas em um plano de dados. No Quadro 18, os entrevistados responderam sobre planos de gestão dos dados.

QUADRO 18 - Seção Plano de Gestão dos Dados

ID	Plano de Gestão dos Dados			
1PGD	- Você utiliza algum template fornecido pela UFMG, agências de fomento ou ferramenta on-line para registrar as etapas (ciclo de vida) de suas pesquisas?			
2PGD	- Se sim, qual template ou ferramenta on-line? De qual instituição?			
3PGD	- Se não utiliza templates, como registra o processo de pesquisa?			
Utiliza template		Qual template	Fornecido pela instituição	Registro do processo de pesquisa
E1	Não. Mas, tenho interesse.			Não.
E2	Não.			Não. O formato que utilizamos é o modelo sugerido pelo Comitê de Ética.
E3	Template do edital para participar de algum projeto.	Específico do edital.	Não.	Usam-se roteiros prontos para as pesquisas (registro de observação).
E4	Não. A gente não tem esse costume.			
E5	Não.			Pacotes de trabalho (relatórios, coleta documental e ficha de organização).
E6	Sim.	Modelo do TJMG.	Modelo do TJMG.	Relatórios para o jurídico analisar.
E7	Sim. Utilizamos o Procedimento Operacional Padrão (POP).	POP criado para o laboratório.	Contratou consultoria.	Somente com o POP.
E8	Sim.	Padrão <i>Darwin Core</i> .		Modelo <i>Darwin Core</i> .
E9	Não.			Protocolo de laboratório.

Fonte: Dados da pesquisa (2022).

Os PGD evidenciam e validam os resultados das pesquisas científicas. Segundo Silva (2019), esse plano caracteriza-se como um documento que descreve o ciclo de vida dos dados desde a fase da coleta até a documentação completa do processo de pesquisa.

Consultados sobre o uso do PGD e dos templates para registrar o ciclo da pesquisa, os entrevistados citaram o uso de protocolos dos laboratórios e de registro em cadernos de anotações (cadernos de laboratórios). Além desses instrumentos, alguns usam modelos dos editais e do Comitê de Ética quando submetem algum projeto, como também há utilização do POP.

Sobre o PGD, os entrevistados E1 e E2 afirmaram desconhecimento:

Eu não sabia que existia, tá? Se eu souber, até pode ser uma coisa interessante (Comentário do E1).

Quando li esse roteiro, realmente, nesta seção, foi a seção que eu pensei assim: é zero. Não sei nem por onde começar a te responder. A única coisa que nós temos bem estabelecido é que todo o projeto é enviado ao Comitê de Ética. E, para o Comitê de Ética, você, geralmente, apresenta um cronograma e, ao final da pesquisa, você tem que apresentar um relatório. Então, além disso, é nada (Comentário do E2).

Recentemente, eu tive a oportunidade de ler um livro. Trata da pesquisadora que foi prêmio Nobel na área de Química, em 2020. E, em dado momento do livro, existe uma disputa entre grupos diferentes de pesquisa, sobre qual grupo foi responsável primeiro de uma determinada descoberta. O processo se tornou jurídico e, entre as coisas que foram levadas ao júri, incluíram-se os cadernos de protocolos dos laboratórios, em que as pessoas registravam os objetivos e os resultados dos experimentos que estavam sendo feitos. Então, como não estamos numa pesquisa [...] competitiva, isso fora da nossa realidade (Comentário do E2).

No tocante à gestão de dados de pesquisa na Pós-Graduação, as respostas dos entrevistados vão ao encontro dos resultados do estudo de Dias, Anjos e Araújo (2019). Nesse estudo, foi concluído que as práticas de gestão de dados conduzidas pelos pesquisadores precisam ser aprimoradas. Ademais, muitos pesquisadores concordam com a prática de compartilhar os dados, considerando o uso de instrumentos legais de padronização e controle, de modo que a prática seja respaldada (DIAS; ANJOS; ARAÚJO, 2019).

A última seção do roteiro abordou as boas práticas. O Quadro 19 apresenta os dados oriundos das entrevistas com os participantes.

QUADRO 19 – Seção boas práticas

ID	Boas práticas			
1BP	- Os dados produzidos durante suas pesquisas são publicados ou ficam restritos aos dispositivos de armazenamento? Em qual plataforma? Essa plataforma é nacional ou internacional?			
2BP	- Relate sua experiência na UFMG referente à curadoria e à preservação de dados científicos e indique dificuldades, boas práticas e/ou sugestões.			
3BP	- Relate a experiência de seu campo de atuação ou área do conhecimento referente à curadoria e à preservação dos dados científicos, seja no âmbito nacional ou internacional.			
	Dados são publicados	Dados ficam restritos	Indicar plataforma	Plataforma nacional/internacional
E1	Sim.	Servidor do próprio laboratório.	Acesso pelo site.	Nacional. Site do CEFALA: <www.cefala.org>.
E2	Sim.		Revista <i>Jornal of Biology</i> . Os conjuntos de dados são enviados via <i>Mendeley</i> .	Internacional. Revista <i>Jornal of Biology</i> .
E3	Sim.		Blog e on-line.	Nacional.
E4	Sim (parcialmente).	Sim. Figuras já ficaram de fora.	Somente em periódicos da área.	
E5	Sim.	Sim. Suporte papel. Guarda-se em caixa arquivo.	Periódicos.	
E6	Sim.	Sim. Restritos ao TJMG.	Sistema interno do Projeto Brumadinho e site do TJMG (área restrita).	Nacional. Site do Projeto Brumadinho.
E7	Sim.	Sim. Restrito ao grupo.	<i>Google Drive</i> .	Internacional. Periódicos
E8	Sim.	Sim. Níveis de acesso restrito.	Sistema de Autorização e Informação em Biodiversidade	Nacional. SISBIO, ICMBio e IEF.

			(SISBIO) e Instituto Chico Mendes de Conservação da Biodiversidade (IMBio). Dados sobre mineração do Instituto de Florestas (IEF).	
E9	Não.	Sim.	<i>Google Drive.</i>	

Fonte: Dados da pesquisa (2022).

A FAPESP (2017) evidenciou que a gestão adequada dos dados de pesquisa é parte essencial das boas práticas de pesquisa. A instituição também declarou que dados resultantes de financiamento público devem ser gerenciados e compartilhados de modo a garantir o maior benefício possível para o avanço científico e tecnológico.

Ao ser consultado sobre as boas práticas em pesquisa, o E2 informou acerca da existência de um manual fornecido pela universidade, conforme o registro de seus comentários:

Há uns cinco anos, a universidade disponibilizou um manual de boas práticas em pesquisa. E nós temos esse manual impresso no laboratório [...], pois na internet é difícil encontrar esse manual. E, informalmente, faz parte da orientação trabalhar a partir desse manual (Comentário do E2).

O E4 sugeriu que o assunto desta entrevista seja tratado pelos bibliotecários, quando da disciplina de Metodologia de Pesquisa:

Essencial que, durante a disciplina Metodologia de Pesquisa, esse assunto seja tratado muito claramente com os doutorandos e pós-doutorandos. Acho que o trabalho das bibliotecas, nesse sentido, e do Centro de Documentação é essencial. Não há nada que impeça que esses profissionais façam contato direto com os alunos da Pós-Graduação. Em detalhes, os bibliotecários têm muito mais elementos informativos para os estudantes que a gente muitas vezes não tem (Comentário do E4).

Em virtude da relevância do tema e do interesse dos entrevistados, após a seção de boas práticas, os participantes foram consultados sobre uma possível adesão a um repositório de dados e sobre a importância de políticas de preservação digital para o campo da pesquisa. Sendo assim, indagou-se a opinião dos entrevistados quanto à possibilidade de uso de um repositório de dados fornecido pela UFMG em parceria com o IBICT. Pelos resultados obtidos com as respostas, a maioria compreende que a institucionalização do repositório é importante, visto que garantirá a preservação e ampliará a visibilidade da pesquisa. Nas palavras dos participantes:

A ideia de um repositório grande é uma ideia legal. Eu nem sei se existe ou não, mas é uma ideia legal (Comentário do E1).

Eu acredito que isso seja fundamental. As coisas são pouco institucionais e recaem muito nos indivíduos. E quando recaem nos muitos indivíduos, você rouba muita energia dessas pessoas. Se a decisão e as soluções estivessem mais a nível institucional, o gasto de energia seria menor. Então, eu, realmente, acredito que seria fundamental se a universidade tivesse um

repositório para disponibilizar para seus docentes e estudantes (Comentário do E2).

Absolutamente, válido e importante. No Promestre [Mestrado em Educação e Docência], além da dissertação, há produção de um recurso didático. Esse recurso didático demora muito para se tornar público. A gente tem algumas ideias que a pandemia dificultou, por exemplo, de fazer uma feira com um dia de apresentação desses dados. Mas, no próprio sistema e no repositório da universidade, é muito demorada a finalização e a divulgação. Os recursos não se dirigem a acadêmicos apenas; dirigem-se, especialmente, para profissionais da Educação Social, escolas, museus, teatros, jardins zoológicos, parques [...]. Então, ele [o recurso] estando no repositório público, no sentido de atingir mais pessoas, ele daria mais visibilidade e mais utilidade para essa produção (Comentário do E3).

A afirmativa do E3 está de acordo com Sayão e Sales (2016), quando caracterizaram os repositórios como recursos informacionais de armazenamento e gestão. Os repositórios são sistemas de armazenamento de dados e objetos decorrentes da pesquisa e possuem diversos recursos capazes de democratizar as práticas científicas, tornando a ciência aberta, dinâmica e em constante uso em todos os contextos sociais (SAYÃO; SALES, 2016).

Ainda sobre o uso do repositório de dados na UFMG, o E5 e o E7 afirmaram que usariam, tendo em vista a grande importância desses ambientes para a continuidade e ampliação das pesquisas. Nas palavras dos participantes:

Eu usaria com prazer. Eu acho que é necessário, pois facilita, inclusive, as boas práticas de metadados [...]. O custo individual da pesquisa acaba sendo alto para você manter a integridade dessas nuvens e desses espaços compartilhados e o volume é muito grande. [...] Arquivos brutos são grandes (Comentário do E5).

Com certeza. Nem precisa terminar a pergunta (Comentário do E7).

Para o E4, não faz sentido a universidade investir em repositório de dados. Para justificar sua opinião, o participante relatou a experiência de outra universidade com o uso de serviço em nuvem, porém de forma corporativa. O E4 acredita que:

O único problema de ser no *Google Drive* é se um dia o Google parar. Mas, eu acho difícil. Honestamente, eu não sei se hoje faz sentido a universidade querer investir nisso (repositório de dados). Na USP [Universidade de São Paulo], eles já usam o gmail para professores e alunos. Eles têm um gmail empresário. Às vezes, faz mais sentido terceirizar esse tipo de coisa. Pensa na responsabilidade da universidade ao investir nisso e dar algum problema (Comentário do E4).

Diferentemente do E4, o E6 mencionou os pilares que sustentam a universidade e, como tal, afirmando que a instituição deve promover recursos para dar visibilidade às pesquisas e salvaguardar os conjuntos de dados. O repositório seria uma iniciativa de garantir os pilares institucionais relacionados à visibilidade do que é produzido em favor da própria sociedade. Na opinião do E6:

Uma universidade não faz somente educação. Faz pesquisa, e isso eu não sei onde está. Na Educação, já notei que tem uma padronização, pois encontro no repositório as teses, dissertações e monografias. Agora as

pesquisas, eu não sei onde estão. Há uma quantidade imensa de pesquisas na FUNDEP [Fundação de Desenvolvimento da Pesquisa] e uma quantidade imensa que estão nas outras três fundações da UFMG. Existe uma quantidade imensa que está na mão de professores por aí afora que ninguém vê (Comentário do E6).

A preocupação do E6 é relatada nos estudos de Sales e Sayão (2018), ao afirmarem sobre o sigilo dos dados, mantidos sobre posse única dos seus criadores. Na maioria dos casos, os dados constituem uma literatura cinza, ou seja, que não é disseminada, estando armazenada em recursos escolhidos de modo individual pelos produtores e estando sobre responsabilidade exclusiva dessas pessoas (SALES; SAYÃO, 2018).

Paganine e Amaro (2020) também acreditam no valor dos repositórios e informam que esse recurso informacional é uma tendência originada em torno das discussões sobre o crescimento da Ciência Aberta. A Ciência Aberta, segundo os autores, por meio dos repositórios, torna-se mais acessível, ampliando a capacidade de se armazenar, organizar, compartilhar e disseminar dados e resultados de pesquisa.

Em diversos momentos das entrevistas, observou-se a menção à institucionalização dos processos de gestão e da preservação de dados. A esse respeito, Souza e Aganette (2020) perceberam que um planejamento adequado das atividades de preservação digital inclui o desenvolvimento e a implementação de uma política de preservação digital.

O E1 mostrou-se preocupado com a preservação, mas revelou a existência de falhas pela falta de orientações, durante o trabalho nas instituições:

A gente faz isso de uma maneira extremamente amadora. A gente até se preocupa em ter os dados preservados, mas nós não temos um guia, um conjunto de protocolos e de *guidelines*; a gente não tem um guia para a gente seguir, não. A gente faz; cada um faz do jeito que acha que é melhor. A gente se preocupa com isso, mas a gente não tem um padrão de procedimento. Definitivamente, não há. Todos se preocupam e, eventualmente, têm competências para fazer, mas não existe um padrão a ser seguido (Comentário do E1).

O E3 destacou um ponto importante, a salvaguarda do material de pesquisa no Setor de Documentação anexo à Biblioteca Universitária. Para esse entrevistado:

Os dados, propriamente, quando a pesquisa é minha, eu guardo por cinco anos. A gente tinha fitas e eu entreguei as fitas e a maior parte no lixo com coração partido [...]. Tinha entrevistas bacanas, entrevista com pessoas diversas e professores. Sempre guardamos por cinco anos. No caso da faculdade, tem o Centro de Documentação, o CEDOC, que guarda o relatório de pesquisa e cadernos com dados. Ele guarda materiais de pesquisa diversos. Eu mesma já entreguei alguns dados lá; fitas com entrevistas (Comentário do E3).

O E5 expôs sua preocupação referente à preservação para acesso em longo prazo. Discursou que:

A preservação de dados brutos exige muito custo e espaço digital. Preservação do presente para pensar o futuro [...], a minha preocupação é

muito dessa ordem. Hoje, nós produzimos, no nosso cotidiano, uma quantidade de dados digitais muito grande e nós não temos nenhuma política de preservação (Comentário do E5).

Essa constatação quanto à falta de política de preservação formal nas instituições foi evidenciada na pesquisa de Souza e Aganette (2020), cujos resultados dessa política não se restringem à padronização dos processos de trabalho, mas representam uma estratégia institucional para salvaguardar a produção intelectual das instituições. Além disso, conforme concluído pelas autoras, a política, como documento formal, fornecerá respaldo e segurança nas decisões e nos procedimentos realizados pelos pesquisadores, no momento do armazenamento dos dados de pesquisa.

Ao finalizar a análise dos dados das entrevistas, partiu-se para a análise da pesquisa documental. Assim, na subseção seguinte, são apresentados os resultados referentes à busca de evidências nos documentos institucionais consultados.

4.4 Análise dos documentos institucionais

Esta parte do quarto capítulo apresenta os trechos dos documentos selecionados e a análise acerca da preservação digital e da governança de dados. Como já mencionado, foram verificados os documentos institucionais da UFMG, constituintes do corpus da pesquisa documental.

A UFMG contém em seu brasão a frase “*Incipit Vita Nova*”, simbolizando a possibilidade do “início de uma nova vida”. O significado dessa descrição pode ser atribuído às descobertas e às renovações, tanto da comunidade acadêmica, quanto da própria instituição, que completa seus 95 anos, neste ano de 2022. Apesar de sofrer com os extremos cortes orçamentários dos últimos anos, a universidade vem se consolidando como uma das melhores universidades federais do Brasil e da América Latina, segundo o *Times Higher Education* (2022). Acredita-se que esse resultado seja devido ao esforço institucional, sobretudo pela atuação dos pesquisadores.

No âmbito das universidades públicas brasileiras, o tripé formado por ensino-pesquisa-extensão é indissociável e obrigatório, conforme assegura o artigo 207 da Constituição Federal de 1988 (BRASIL, 1988). Com efeito, a universidade tem autonomia didático-científica, administrativa e de gestão financeira e patrimonial, o que permite a adoção de estratégias para uma governança efetiva. Nesse contexto, salienta-se a importância da pesquisa científica para a instituição e para a sociedade. Portanto, considera-se “[...] que a gestão de dados é um **ponto-chave no apoio à pesquisa**” (SOCIEDADE ESPANHOLA DE DOCUMENTAÇÃO E INFORMAÇÃO CIENTÍFICA - SEDIC, 2022, não paginado, grifo nosso, tradução nossa).

Para a análise dos documentos, buscaram-se os termos (categorias) constantes no Quadro 20.

QUADRO 20 – Seleção dos documentos institucionais da UFMG

Documento	Gestão	Termo buscado	Termo encontrado
Plano de Desenvolvimento Institucional (PDI)	2018-2023	Governança Digital; Preservação; Preservação Digital; Plano de Preservação; Curadoria; Curadoria de Dados; Curadoria Digital; Dados Científicos; Dados de Pesquisa; Conjuntos de Dados.	Governança Digital (p. 9, 227 e 230); Preservação (p. 106, 203, 211, 230, 242, 253, 255, 261, 268, 271, 274, 276 e 277); Preservação Digital (p. 106); Curadoria (p. 276 e 279); Conjuntos de Dados (p. 230).
Plano de Dados Abertos (PDA)	2021-2022	Governança Digital; Preservação; Preservação Digital; Plano de Preservação; Curadoria; Curadoria de Dados; Curadoria Digital; Dados Científicos; Dados de Pesquisa; Conjuntos de Dados.	Governança Digital (p. 13, 36 e 39); Curadoria (p. 24); Conjuntos de Dados (p. 13, 22, 23, 24 e 25).
Relatório Gerencial da BU	2017	Governança Digital; Preservação; Preservação Digital; Plano de Preservação; Curadoria; Curadoria de Dados; Curadoria Digital; Dados Científicos; Dados de Pesquisa; Conjuntos de Dados.	Preservação (p. 26).
Política de Desenvolvimento do Acervo da BU	2015	Governança Digital; Preservação; Preservação Digital; Plano de Preservação; Curadoria; Curadoria de Dados; Curadoria Digital; Dados Científicos; Dados de Pesquisa; Conjuntos de Dados.	Preservação (p. 8, 9, 13, 15 e 16); Curadoria (p. 14).
Relatório Anual de Gestão da DIARQ	2017	Governança Digital; Preservação; Preservação Digital; Plano de Preservação; Curadoria; Curadoria de Dados; Curadoria Digital; Dados Científicos; Dados de Pesquisa; Conjuntos de Dados.	Preservação (p. 3, 5 e 6); Preservação Digital (p. 5).
Manual de Gestão de Documentos Arquivísticos da UFMG	2019	Governança Digital; Preservação; Preservação Digital; Plano de Preservação;	Preservação (p. 3, 4, 6, 8, 15, 25, 26, 27, 29, 37, 40, 41 e 42); Preservação Digital (p. 29).

Documento	Gestão	Termo buscado	Termo encontrado
		Curadoria; Curadoria de Dados; Curadoria Digital; Dados Científicos; Dados de Pesquisa; Conjuntos de Dados.	
Relatório de Gestão da DGI	2018-2021	Governança Digital; Preservação; Preservação Digital; Plano de Preservação; Curadoria; Curadoria de Dados; Curadoria Digital; Dados Científicos; Dados de Pesquisa; Conjuntos de Dados.	Conjuntos de Dados (p. 16).
Plano Diretor de Tecnologia da Informação e Comunicação da UFMG (PDTIC)	2021-2024	Governança Digital; Preservação; Preservação Digital; Plano de Preservação; Curadoria; Curadoria de Dados; Curadoria Digital; Dados Científicos; Dados de Pesquisa; Conjuntos de Dados.	Governança Digital (p. 1, 2, 9, 16, 23 e 91); Conjuntos de Dados (p. 76).
Planejamento Estratégico de Tecnologia da Informação da UFMG	2013-2015 2015-2020	Governança Digital; Preservação; Preservação Digital; Plano de Preservação; Curadoria; Curadoria de Dados; Curadoria Digital; Dados Científicos; Dados de Pesquisa; Conjuntos de Dados.	Governança Digital (p. 3, 16, 20 e 23).
Plano de Gestão - PROEX	2018-2022	Governança Digital; Preservação; Preservação Digital; Plano de Preservação; Curadoria; Curadoria de Dados; Curadoria Digital; Dados Científicos; Dados de Pesquisa; Conjuntos de Dados.	Nenhuma ocorrência.
Resolução 01/2021, de 28 de Janeiro de 2021 - Infraestruturas de Apoio Institucional à Pesquisa (AIPq)	2021	Governança Digital; Preservação; Preservação Digital; Plano de Preservação; Curadoria; Curadoria de Dados; Curadoria Digital; Dados Científicos; Dados de Pesquisa; Conjuntos de Dados.	Nenhuma ocorrência.
Resolução n01/2018, de 06 de fevereiro de 2018 -	2018	Governança Digital; Preservação;	Nenhuma ocorrência.

Documento	Gestão	Termo buscado	Termo encontrado
Infraestruturas de sobre pesquisa Universidade Federal de Minas Gerais. Pró-Reitoria de Pesquisa - PRPq		Preservação Digital; Plano de Preservação; Curadoria; Curadoria de Dados; Curadoria Digital; Dados Científicos; Dados de Pesquisa; Conjuntos de Dados.	
Plano Estratégico Institucional - PROPLAN	Atualização semestral	Governança Digital; Preservação; Preservação Digital; Plano de Preservação; Curadoria; Curadoria de Dados; Curadoria Digital; Dados Científicos; Dados de Pesquisa; Conjuntos de Dados.	Nenhuma ocorrência.

Fonte: Dados da pesquisa (2022).

Aplicado o critério de inclusão, foram pré-selecionados nove documentos, dois quais dois não continham informações sobre diretrizes ou normas para a gestão de dados científicos. Assim, foram selecionados sete documentos para análise, conforme listagem descrita no Quadro 21.

QUADRO 21 – Documentos institucionais selecionados

Documentos pré-selecionados	Documentos selecionados
Plano de Desenvolvimento Institucional (PDI) (2018-2023)	Plano de Desenvolvimento Institucional (PDI) (2018-2023)
Plano de Dados Abertos (PDA) (2021-2022)	Plano de Dados Abertos (PDA) (2021-2022)
Relatório Gerencial da BU (2017)	Os termos encontrados estão fora do escopo desta pesquisa. Documento excluído.
Política de Desenvolvimento do Acervo da BU (2015)	Os termos encontrados estão fora do escopo desta pesquisa. Documento excluído.
Relatório Anual de Gestão da DIARQ 2017	Relatório Anual de Gestão da DIARQ (2017)
Manual de Gestão de Documentos Arquivísticos da UFMG (2019)	Manual de Gestão de Documentos Arquivísticos da UFMG (2019)
Relatório de Gestão da DGI (2018-2021)	Relatório de Gestão da DGI (2018-2021)
Plano Diretor de Tecnologia da Informação e Comunicação da UFMG (PDTIC) (2021-2024)	Plano Diretor de Tecnologia da Informação e Comunicação da UFMG (PDTIC) (2021-2024)
Planejamento Estratégico de Tecnologia da Informação da UFMG (2013-2015 e 2015-2020)	Planejamento Estratégico de Tecnologia da Informação da UFMG (2013-2015 e 2015-2020)

Fonte: Dados da pesquisa (2022).

Após a seleção dos sete documentos, os trechos que continham informações relacionadas ao que estava sendo investigado, ou seja, estratégias ou quaisquer ações direcionadas à preservação e à gestão dos dados de pesquisa foram destacadas em quadros. A exposição dos trechos e dos termos de busca¹¹, objetivou facilitar a síntese das informações,

¹¹ Os trechos descritos no interior dos quadros foram retirados dos documentos, mediante descrição literal, ou seja, não foi realizada nenhuma mudança nas estruturas sintáticas e ortográficas da escrita, de modo a preservar a originalidade dos documentos. As palavras ou expressões negritadas dizem respeito aos termos que foram utilizados no momento das buscas.

possibilitando condensar os dados e gerar resultados. Os quadros seguintes, de números 22, 23, 24 e 25, apresentam o resumo do que consta informado nos documentos de cada um dos quatro setores analisados: Reitoria, DIARQ, DGI e DTI.

QUADRO 22 – Análise dos documentos institucionais da Reitoria

Trechos analisados
Administração Central - Reitoria
Plano de Desenvolvimento Institucional (2018-2023)
<p>- O termo “governança digital”, apresentado no subtópico “Governança de Tecnologia da Informação”, refere-se à formalização do Comitê de Governança Digital (CGD) que será responsável pela elaboração do Plano Diretor de TI (PDTI). Plano essencial para as atividades da Diretoria de Tecnologia da Informação (DTI), pois definirá as políticas, estratégias, procedimentos, estrutura e recursos necessários com vistas à otimização do uso de TI da UFMG. A partir de definições de onde a Instituição quer chegar, serão definidos quais os recursos da TI que serão necessários para suportar tais decisões, além de dar transparência às atividades de TI e à sua aplicação.</p> <p>- O termo “governança digital”, apresentado no subtópico “Comunicação e Informação”, refere-se à Tecnologia da Informação - item 24. Apoiar as ações do Comitê de Governança Digital na priorização e acompanhamento de projetos e na elaboração e implantação do PDTI da Universidade. Sendo uma ação prevista para os anos 2022-2023.</p> <p>- O termo “preservação digital”, apresentado no eixo “Pesquisa”, cujo objetivo geral é: incrementar a atividade de pesquisa na UFMG, visando a que esta se torne referência nacional e alcance reconhecimento internacional em todas as áreas do conhecimento, refere-se ao 8º objetivo específico: Reestruturar a Política de Periódicos Científicos da UFMG de forma a institucionalizá-la garantindo recursos e infraestrutura para promover a melhoria continuada, a visibilidade, o armazenamento e a segurança dos dados, bem como a preservação digital e recuperação da informação relacionada aos periódicos científicos editados e/ou produzidos pela comunidade UFMG.</p> <p>- O termo “conjuntos de dados”, apresentado no subtópico “Comunicação e Informação”, refere-se ao item 16. Organizar e avaliar conjuntos de dados de interesse público produzidos pela UFMG e que possam ser disseminados como dados abertos em portal institucional. Sendo uma ação prevista para os anos 2020/2021.</p>
Plano de Dados Abertos (2021-2022)
<p>- O termo “governança digital”, constante da “Apresentação”, do Plano de Dados Abertos (PDA), refere-se à elaboração do referido documento baseado na legislação vigente, assim como menciona Estratégia de Governança Digital (EGD) por meio do Decreto nº 8.638 que institui a Política de Governança Digital; Lei Complementar nº 101, de 4 de maio de 2000; entre outros normativos que abordam o tema de transparência da informação;</p> <p>- O Termo “governança digital”, apresentado no Glossário do referido plano, define o termo como: Política instituída pelo Decreto 8.638, de 15 de janeiro de 2016 e que estabelece a Estratégia de Governança Digital (EGD). Baseia-se no conceito de Governo Digital da OCDE e refere-se ao uso de tecnologias digitais, como parte integrada das estratégias de modernização governamentais, para gerar benefícios para a sociedade.</p> <p>- O termo “curadoria”, apresentado no subtópico “Modelo de sustentação”, refere-se à abertura de dados como sendo uma iniciativa que requer uma contínua e permanente curadoria sobre a publicação dos dados, encarregada de manter as ações de abertura, catalogação e publicação dos dados da UFMG.</p> <p>- O termo “conjuntos de dados”, constante da “Apresentação”, do Plano de Dados Abertos (PDA), refere-se que para priorizar a disponibilização dos conjuntos de dados, foram considerados os princípios da publicidade e da transparência da administração pública como cerne para todas as decisões e determinações aqui estabelecidas, assim como o grau de relevância das informações para o cidadão, o alinhamento com os instrumentos de planejamento estratégico da Instituição e a disponibilidade de dados já gerenciados pelos Sistemas Integrados de Gestão da UFMG; No subtópico “3.5 Definição dos dados a serem abertos”, consta a definição dos primeiros conjuntos de dados a serem abertos pelo UFMG foi realizada de forma a atender às solicitações dos Acórdãos</p>

TCU 3022/2015 e 1007/2016, considerando, sobretudo, a relevância das informações para a sociedade, a capacidade de operacionalização das ações necessárias dentro do prazo previsto e o alinhamento com os instrumentos de planejamento estratégico, de forma a garantir os princípios da publicidade e da transparência da administração pública; No subtópico “3.6.1 Estratégia definida para abertura dos dados” contempla ações: definição dos **conjuntos de dados** a serem abertos; distinção dos setores responsáveis pelos **conjuntos de dados**; identificar a necessidade de publicação de novos **conjuntos de dados** para abertura; analisar as solicitações de informações mais recorrentes no e-SIC e na Ouvidoria e sugerir novos conjuntos de dados candidatos à abertura de dados.

Fonte: Dados da pesquisa (2022).

QUADRO 23 - Análise dos documentos da Diretoria de Arquivos

Trechos analisados
Diretoria de Arquivos (DIARQ)
Relatório Anual de Gestão da DIARQ (2017)
- O termo “ preservação digital ”, apresentado no tópico “ Atividades de Divulgação da DIARQ (eventos, campanhas desensibilização) ”, refere-se ao IV Ciclo de Palestras da Universidade Federal de Minas Gerais, tendo como tema “O Sistema Eletrônico de Informações (SEI) e a Preservação Digital: O impacto desse sistema eletrônico na administração pública – mudanças e desafios”.
Manual de Gestão de Documentos Arquivísticos da UFMG (2019)
- O termo “ preservação digital ”, apresentado no tópico “ 5.7.2 Preservação Digital ”, traz a definição de preservação digital; afirma que “o desafio da preservação digital é enfrentar a obsolescência tecnológica, a fragilidade das mídias, a variedade dos documentos e de formatos, além de assegurar a estabilidade e a fixidez, e garantir o acesso a longo prazo. Documento digital não é virtual, está fixado em um suporte”. Esclarece que “a UFMG, por ser um órgão integrante do Sistema Nacional de Arquivos – SINAR, deve seguir a legislação e as orientações vigentes sobre o assunto. Uma delas estabelece diretrizes para a implementação de repositórios arquivísticos digitais confiáveis para o arquivamento e a manutenção de documentos arquivísticos digitais em suas fases corrente, intermediária e permanente. A gestão dos documentos produzidos digitalmente é mais um dos grandes desafios a serem enfrentados na implantação da gestão de documentos e das políticas arquivísticas na UFMG.
- O termo “ preservação ”, constante da “ Apresentação ”, menciona que “a ausência de normas, métodos e procedimentos de trabalho gera o acúmulo desordenado de documentos, tanto em suporte tradicional quanto em meio digital, dificultando o acesso e a recuperação de informações necessárias para a tomada de decisão no âmbito de toda a instituição”. Aponta a política de gestão documental da UFMG para preservação de documentos dos documentos indispensáveis à administração institucional, à memória da UFMG e à garantia dos direitos individuais e coletivos; tendo como diretrizes a eficiência gerencial e a preservação da memória, a Política de Gestão Documental apresenta-se como um passo fundamental para a melhoria dos procedimentos administrativos e de salvaguarda do patrimônio documental histórico da UFMG. Quando da digitalização, recomenda que os documentos digitalizados sejam salvos, preferencialmente, em formato PDF/A, aconselhável com relação à garantia de acesso e preservação por longos períodos.

Fonte: Dados da pesquisa (2022).

QUADRO 24 - Análise do documento da Diretoria de Governança Informacional

Trechos analisados
Diretoria de Governança Informacional (DGI)
Relatório de Gestão da DGI (2018-2021)
- O termo “ conjuntos de dados ”, apresentado no tópico “ Dados Abertos ”, refere-se ao Plano de Dados Abertos (PDA) é o documento orientador para as ações de implementação e promoção da publicação de conjuntos de dados pela Universidade. Informa que o PDA da UFMG foi publicado em janeiro de 2021, após trabalho conjunto de diversas áreas da instituição, especialmente do Centro de Comunicação da UFMG, da Diretoria de Tecnologia da Informação e da DGI. E, que em abril de 2021, foram publicados os primeiros conjuntos de dados que a Universidade havia se comprometido a abrir, sem atrasos no cronograma. A DGI, ambiciona ir além do previsto no PDA, implementando a abertura

de novos conjuntos de dados antes mesmo da próxima atualização do documento, prevista para janeiro de 2023.

Fonte: Dados da pesquisa (2022).

QUADRO 25 - Análise de documentos da Diretoria de Tecnologia da Informação

Trechos analisados
Diretoria de Tecnologia da informação (DTI)
Plano Diretor de Tecnologia da Informação e Comunicação da UFMG – PDTIC (2021-2024)
<ul style="list-style-type: none"> - O termo “governança digital”, apresentado no tópico “Organização da TI”, refere-se ao Comitê de Governança Digital (CGD), instituído em 2018, de caráter deliberativo e de natureza estratégica, tem a finalidade de deliberar sobre os assuntos relativos à Governança Digital na UFMG, auxiliando a instituição na tomada de decisões relacionadas à Tecnologia da Informação e Comunicação; apresenta o CGD como representação de um ponto forte na Análise SWOT. - O termo “conjuntos de dados”, apresentado no subtópico “Plano de Ações”, refere-se à Meta 45 – Ação 179: auxiliar os responsáveis pelos dados abertos na UFMG na organização e disponibilização de conjuntos de dados de interesse público a serem disseminados em portal institucional para ser cumprida de 2021-2024.
Planejamento Estratégico de Tecnologia da Informação da UFMG (2013-2015 e 2015-2020)
<ul style="list-style-type: none"> - O termo “governança digital”, apresentado no tópico “2. Introdução: Governança, PDI, PDTI e Comitê de TI”, refere-se à elaboração do PDI e instituição do CGD. Menciona “o realinhamento de algumas prioridades, ao longo do período, desviou servidores dos projetos estratégicos para a execução de projetos táticos e operacionais. Acredita-se que este problema pode ser mitigado com o novo Comitê de Governança Digital, que poderá atuar garantido maior alinhamento e priorização entre as ações estratégicas, táticas e operacionais.

Fonte: Dados da pesquisa (2022).

A partir dos trechos apresentados, extraídos dos documentos constituintes da pesquisa documental, infere-se que a instituição tem se preocupado com a governança digital ao instituir um Comitê de Governança Digital (CGD), além de buscar atendimento às exigências legais nos prazos estabelecidos pelos instrumentos legais. No que tange ao eixo “Pesquisa”, objetiva reestruturar a Política de Periódicos Científicos da UFMG para promover melhorias e garantir a preservação digital. Quanto à abertura dos dados, também em atendimento às exigências da legislação, disponibiliza os conjuntos de dados produzidos pela UFMG que sejam de interesse público. Por fim, no que se refere à preservação digital e à gestão de dados científicos, os documentos mencionam que as ações institucionais ainda são incipientes, porém, quando colocadas em prática e ampliadas, mediante esforços e trabalho coletivo das equipes gestoras, sinalizam compreender a relevância da temática para a comunidade acadêmica e para a sociedade.

4.5 Diagnósticos

O diagnóstico apresenta uma situação ou problema que deve ser tratado por uma determinada instituição. Considera-se que a tomada de decisão está atrelada à identificação e a possíveis causas de um problema. Após análise dos dados, temos o cenário da

preservação digital de dados científicos na UFMG, atendendo ao segundo objetivo deste estudo e apresentado na primeira parte deste capítulo.

4.5.1 Diagnóstico da curadoria e preservação de dados

Apresenta-se o diagnóstico das práticas dos pesquisadores atuantes nos PPG da UFMG, a partir das análises às respostas do questionário, assim como aos relatos da entrevista. As evidências que identificam o panorama atual da instituição foram sistematizadas, tendo como base os itens da matriz LOP e são apresentadas no Quadro 26.

QUADRO 26 - Diagnóstico da curadoria e da preservação digital de dados científicos na UFMG

Diagnóstico das práticas dos pesquisadores	
Temática	Situação (abril de 2022)
Diagnóstico	
Perfil/Atuação	- Pesquisadores experientes em suas áreas de atuação com práticas bem consolidadas.
Produção/Consumo	- Dados heterogêneos e singulares; - Tipologia diversificada; - Formatos proprietários e suportes variados.
Armazenamento/Cópias	- Uso frequente de equipamentos e dispositivos portáteis de aquisição pessoal; - Aquisição e uso de serviços em nuvem vinculados às contas pessoais (armazenamento em nuvem); - Cópias de segurança restritas aos equipamentos e contas pessoais; - Infraestruturas física e lógica montadas com recursos de projetos nos laboratórios dos PPG, sem a devida orientação dos tecnologistas da instituição.
Integridade	- Perdas de dados e arquivos, ocasionadas pela inexperiência quanto ao uso e ao gerenciamento de serviços computacionais; - Ausência de apoio multidisciplinar para lidar com dados científicos (profissionais da informação, da computação, estatísticos, entre outros).
Metadados	- Ausência de padronização na descrição dos dados científicos (desconhecimento de padrões de metadados).
Controle/Reuso	- Cessão de dados sob demanda; - Falta de instrumento para formalizar a cessão dos dados; - Acesso restrito aos dados (parte dos dados liberada apenas para os pesquisadores envolvidos e mediante o compartilhamento de acesso às contas privadas); - Registros manuscritos.
Plano de Gestão de Dados	- Ausência de Plano de Gestão de Dados (formalização dos processos das pesquisas); - Ausência de outros instrumentos norteadores para os pesquisadores quanto aos serviços de gestão de dados; - Desconhecimento do auxílio de bibliotecários no que concerne aos serviços de dados (curadoria, preservação e gestão de dados).
Repositório de Dados	- Promoção quanto ao uso de repositórios de dados de modo institucionalizado; Sinalização de adesão no caso da UFMG participar do projeto proposto pelo IBICT.
Boas Práticas	- Concordância quanto à inclusão de orientações sobre a curadoria e a preservação dos dados nas aulas de Metodologia

	de Pesquisa, nos Cursos de Pós-Graduação (seminários com a participação de bibliotecários).
Política de Preservação	- Falta de plano e de política de preservação digital na universidade.

Fonte: Dados da pesquisa (2022).

Este diagnóstico revela a situação dos PPG consultados quanto à gestão e à preservação digital dos dados científicos na UFMG. Foi possível identificar como os pesquisadores lidam com os dados originários das pesquisas que lideram em suas áreas de conhecimento e suas opiniões quanto às estratégias para fortalecer a prática da gestão e da preservação dos dados científicas nas instituições.

Os dados científicos são heterogêneos e singulares, apresentam tipologias e suportes variados e, em muitos casos, estão em formatos não convencionais à preservação. Além disso, não adotam metadados apropriados aos domínios disciplinares, cujo armazenamento está, na maioria dos casos, sob responsabilidade dos fornecedores terceirizados, ou armazenados em mídias frágeis, sem monitoração, sem migração e sem atualização tecnológica.

Posto isso, confirma-se o que é apresentado na literatura de que o planejamento, a gestão, a preservação e a reutilização dos dados produzidos nessas pequenas dimensões constituem processos permeados por desafios ainda não superados (SALES; SAYÃO, 2018). Mesmo que as realidades das instituições apresentam lacunas, o trabalho coletivo, com envolvimento de equipes multidisciplinares, tendo participação, sobremaneira, de bibliotecários, de fato, é uma estratégia em andamento (ALMEIDA, 2019; SILVA, 2019).

Infere-se que os pesquisadores são comprometidos com a produção de novos conhecimentos, procuram facilitar o armazenamento de dados e arquivos por meio da aquisição de equipamentos de informática e arcam com os custos dos serviços em nuvem. Contudo, necessitam de aportes de outras áreas do conhecimento para que haja manutenção, curadoria e preservação dos dados científicos de maneira adequada. Além disso, carecem de orientações legais como o uso de uma política de preservação, na expectativa de que tal documento possa ser construído colaborativamente.

Mesmo que haja adesão dos pesquisadores em prol do compartilhamento e do reuso dos dados científicos, eles não disponibilizam os conjuntos de dados resultantes das coletas, pois se sentem inseguros quanto aos créditos da pesquisa, o que seria sanado quando da atribuição dos identificadores persistentes e das licenças de uso. Além disso, grande parte dos pesquisadores ainda registra os processos das pesquisas em cadernos de anotações e protocolos em meio físico. Embora essa seja uma prática comum em determinadas áreas, isso dificulta o armazenamento, o acesso e a localização da informação.

Esses resultados dialogam com as evidências do estudo de Dias, Anjos e Araújo (2019) que discorrem sobre a concordância dos pesquisadores em disponibilizar os dados de

suas pesquisas, mas que isso ainda não tem aplicação prática, sobretudo pela falta de políticas ou planos de preservação que orientem os pesquisadores. Outra constatação já proclamada na literatura e confirmada neste estudo diz respeito ao armazenamento restrito dos dados, ou seja, os resultados das pesquisas são disseminados, mas os dados que as geraram permanecem obscuros, localizados nos dispositivos pessoais dos pesquisadores (SALES; SAYÃO, 2018).

Referente ao uso do PGD, os pesquisadores analisados neste estudo mostraram-se favoráveis à adoção dos instrumentos que controlem as práticas de pesquisa, sejam por meio dos planos, das políticas ou de quaisquer materiais instrutivos e normativos. Mediante o uso desses instrumentos de controle, a gestão e a preservação dos objetos digitais, como evidenciado por Souza e Aganette (2020), serão um esforço que garantirá o armazenamento, por conseguinte, o acesso à produção intelectual das instituições de ensino e de pesquisa, de modo imediato, seguro e em longo prazo.

No que concerne ao diagnóstico das ações institucionais, a partir das análises aos documentos, registra-se o Quadro 27. O diagnóstico dos documentos é sistematizado com base nos seguintes itens: conjunto de dados, curadoria, governança digital, políticas e preservação digital.

QUADRO 27 - Diagnóstico da governança digital de dados científicos na UFMG

Diagnóstico de documentos institucionais	
Temática	Situação (abril 2022)
Diagnóstico	
Conjunto de Dados	Publicação, em 2021, do Plano de Dados Abertos (PDA), documento orientador para as ações de implementação e promoção da publicação de conjuntos de dados pela universidade. Disponibilização dos conjuntos de dados de interesse público produzidos pela UFMG e que possam ser disseminados como dados abertos em portal institucional, no ano de 2021.
Curadoria	Abertura de dados como sendo uma iniciativa que requer uma contínua e permanente curadoria sobre a publicação dos dados, encarregada de manter as ações de abertura, catalogação e publicação dos dados da UFMG.
Governança Digital	Institucionalização, em 2018, do Comitê de Governança Digital (CGD), responsável pela elaboração do Plano Diretor de Tecnologia da Informação (PDTI).
Políticas	Proposta de reestruturação, entre os anos de 2018 a 2023, da Política de Periódicos Científicos da UFMG para garantir, entre outras práticas, a preservação digital. Valorização, no ano de 2019, da política de gestão documental da UFMG para a preservação dos documentos, política essa indispensável à administração institucional, à memória da UFMG e à garantia dos direitos individuais e coletivos.
Preservação Digital	Ciclo de Palestras da UFMG, no ano de 2017, tendo como tema "O Sistema Eletrônico de Informações (SEI) e a Preservação Digital: o impacto desse sistema eletrônico na administração pública – mudanças e desafios".

Fonte: Dados da pesquisa (2022).

No tocante às políticas para viabilização da preservação digital e da gestão de dados científicos, o diagnóstico mostra que as ações institucionais ainda são incipientes. No âmbito da UFMG, é fundamental implementar uma política de preservação digital e fornecer documentos norteadores para a efetiva gestão de dados oriundos das pesquisas, assim como promover recursos para disponibilizar e salvaguardar esses conjuntos de dados científicos, de modo a contribuir para o desenvolvimento da ciência brasileira.

4.6 Diretrizes

Nesta subseção, são indicados os princípios e as diretrizes que foram colecionados ao longo da pesquisa. Eles poderão servir de insumos ou manifestarem-se como mecanismos para apoiar a UFMG na implementação de sua política de preservação digital.

4.6.1 Recomendações de diretrizes e princípios para preservar dados

Para apoiar as implementações de políticas de preservação digital e atender ao terceiro objetivo desta pesquisa “recomendar diretrizes de preservação digital de dados científicos”, foram listados os princípios e as diretrizes para a preservação digital e para a gestão de dados científicos.

A problemática da preservação deve fazer parte das preocupações das instituições. O plano ou o programa voltado à preservação precisam conter as normas do que fazer para garantir o acesso contínuo do objeto preservado, levando em conta as características físicas, o conteúdo do objeto como também questões de licenças de autor e de usos (AMORIM, 2008). Assim como Amorim (2008), Márdero Arellano (2008) concorda com as peculiaridades a serem inseridas nos documentos norteadores que indicarão as práticas da preservação, sendo esse documento elaborado de modo colaborativo e compartilhado entre os membros da instituição.

Desse modo, as entidades internacionais e nacionais apresentam diversas *guidelines* (diretrizes) no formato de guias ou cartilhas e contendo os princípios que norteiam as discussões. Tais documentos também apresentam soluções para subsidiar as atividades de curadoria, de preservação digital e de gestão. A temática vem sendo discutida amplamente, devido à importância da preservação digital, em face do crescente volume de dados produzidos e em meio à necessidade de adequação das entidades para preservar dados em longo prazo e retornar à sociedade os resultados dos investimentos em pesquisas. Assim, justifica-se a importância e a necessidade de institucionalizar a política de preservação, um documento formal que defina as práticas que viabilizarão o fluxo dos dados científicos (SAYÃO; SALES, 2016).

Posto isso, o Quadro 28 apresenta um conjunto de diretrizes e princípios que conduzirão as práticas da preservação digital e da gestão dos dados científicos, conforme os apontamentos da literatura.

QUADRO 28 - Recomendações e diretrizes para a política de preservação dos dados científicos

Preservação Digital		
Entidade/Órgão	Ação/Recomendação	Contribuições
- IBICT - Política de preservação digital do IBICT (2019); - CARINIANA – Política da Rede de Serviços de Preservação Digital (2014); - FIOCRUZ – Programa de Preservação Digital de Acervos (2020); - RNP.	- Adotar o modelo <i>Open Archival Information System</i> (OAIS), traduzido como Sistema Aberto de Arquivamento de Informação; - Considerar a norma internacional ISO 14721:2003.	- Padroniza as atividades de preservação digital; - Favorece o compartilhamento; - Mitiga o problema da obsolescência da informação digital; - Reduz os custos da operação.
- IBICT - Política de preservação digital do IBICT (2019).	- Estabelecer um sistema de preservação segundo o modelo de referência OAIS, traduzido como Iniciativa de Arquivos Abertos.	- Aplica-se a qualquer tipo de arquivo digital e compatível com o OAI.
- OPEN ARCHIVES.	- Adotar padrões de interoperabilidade que visam a facilitar a disseminação eficiente de conteúdo; - Adotar sistemas baseados no protocolo de coleta de metadados do OAI.	- Promove a interoperabilidade de conteúdo da web, garantindo o amplo acesso aos recursos digitais.
- FIOCRUZ – Programa de Preservação Digital de Acervos (2020).	- Adotar um Plano de Preservação Digital (PPD).	- Prevê que os acervos cumprem alguns requisitos relacionados ao armazenamento, à segurança da informação e à gestão de objetos digitais. Oferece a estrutura para se criar um PPD.
- Arquivo Nacional – Recomendações para a elaboração da política de preservação digital (2019).	- Implementar a política de preservação digital.	- Oferece roteiro para a elaboração/desenvolvimento de uma política de preservação de documentos arquivísticos digitais.
- IBICT/Rede Brasileira de Serviços de Preservação Digital (CARINIANA); - NDSA; - DCMI; - PREMIS; - METS.	- Adotar padrões de metadados adequados aos objetos digitais (administrativos, técnicos, descritivos, de preservação e estruturais), como: Dublin Core, EAD (descoberta e recuperação de recursos), Modelo PREMIS e METS (preservação em longo prazo).	- Oferece dados associados a um sistema de informação ou a um objeto de informação – localização, criação e preservação do objeto digital, podendo incluir informações de conteúdo e propriedade intelectual.
- Digital Preservation Coalition (DPC) (2015).	- Adotar diretrizes para o desenvolvimento da política de preservação digital.	- Aponta os procedimentos de elaboração da política de preservação digital.
- NDSA.	- Considerar a matriz dos níveis de preservação digital.	- Possibilita a avaliação do nível de preservação digital atingido pelas instituições.
Gestão de dados		
Entidade/Órgão	Ação/Recomendação	Contribuições

- IBICT.	- Adotar um sistema de identificadores persistentes (sistema PID) como DOIs, <i>Handles</i> e <i>ORCIDs</i> .	- Facilita a reprodutibilidade das pesquisas.
IBICT - Cartilha de Atribuição de identificadores digitais para publicações científicas: DOI para o SEER/OJS (2013).	- Adotar o sistema <i>Digital Object Identifier</i> (DOI), traduzido como Identificador Digital de um Objeto; - Valorizar a norma internacional ISO 26324.	- Oferece infraestrutura para ligar os usuários aos conteúdos dispostos pelos editores, gerenciando a comunicação entre eles.
Corporation for National Research Initiatives	- Adotar o <i>Handle System</i> , traduzido como Sistema <i>Handle</i> .	- Assinala, armazena, administra e resolve identificadores ou nomes persistentes de objetos digitais e facilita a localização de objetos em ambiente de rede, favorecendo a interoperabilidade.
- ORCID.	- Adotar o <i>Open Researcher and Contributor ID</i> (ORCID iD), traduzido como Identificador Único e Persistente.	- Conecta pesquisa e pesquisadores, viabilizando as conexões transparentes e confiáveis entre pesquisadores, suas contribuições e suas filiações, fornecendo um identificador único e persistente para os indivíduos envolvidos com a pesquisa.
- FAPESP; - FIOCRUZ; - UNESCO (2021).	- Adotar o PPG.	- Identifica quais tipos de dados serão gerados; - Indica como os dados serão preservados e disponibilizados; - Atenta à observância das questões éticas, legais, de confidencialidade, entre outras.
- FAPESP; - DCC.	- Recomendar o uso das ferramentas DMPTool< https://dmptool.org/ > e DMponline< dmponline.dcc.ac.uk >; < dmp.consorciomadrono.es/ >	- Oferece modelos e auxilia o pesquisador a gerar um PGD com perguntas em questionário on-line e diversos modelos que atendem aos requisitos institucionais e dos financiadores.
- InterPARES (2016); - GO FAIR BRASIL (2021); - UNESCO (2021).	- Adotar o conjunto de princípios FAIR.	- Aplicados à gestão de dados, melhoram a qualidade dos dados, tornando-os encontráveis, acessíveis, interoperáveis e reusáveis.
- OPENAIRE (2020) diretrizes para repositórios de dados.	- Utilizar o protocolo OAI-PMHv 2.0 para a coleta dos metadados dos conjuntos de dados; - Esquematizar os metadados <i>DataCite</i> .	- Coleta de conjuntos de dados; - Citação dos dados por meio do uso de identificadores persistentes e DOI.

<ul style="list-style-type: none"> - OGP/IBICT/CNPq; - RNP/RDP – Brasil. 	<ul style="list-style-type: none"> - Aderir à infraestrutura federada de repositório de dados científicos (piloto). 	<ul style="list-style-type: none"> - Melhora a governança dos conjuntos de dados da pesquisa brasileira por meio da atribuição do <i>Digital Object Identifier</i> (DOI). - Redução de custos.
<ul style="list-style-type: none"> - RNP/RDP – Brasil. 	<ul style="list-style-type: none"> - Adotar licenças de uso: licenças <i>Creative Commons</i>. 	<ul style="list-style-type: none"> - Assegura os direitos do autor e dos direitos conexos; - Observa as atribuições (licenças apropriadas).
<ul style="list-style-type: none"> - EPFL Library. 	<ul style="list-style-type: none"> - Adotar formatos de arquivos (<i>Open Formats</i>), traduzindo como formatos abertos, adequados à preservação. 	<ul style="list-style-type: none"> - Possibilita a codificação da informação para acesso ao conteúdo.
<ul style="list-style-type: none"> - eRSA - European Regional Science Association. 	<ul style="list-style-type: none"> - Adotar modelo de <i>checklist</i> para verificação do gerenciamento de dados. 	<ul style="list-style-type: none"> - Oferece 13 atividades-chave e requisitos para uma boa gestão de dados científicos.

Fonte: Dados da pesquisa (2022).

5 CONSIDERAÇÕES FINAIS

Este último capítulo da dissertação é destinado às conclusões da pesquisa. Portanto, apresentam-se os resultados, as recomendações de boas práticas, as intercorrências no decorrer do estudo, assim como a perspectiva de estudos futuros.

A UFMG vem se consolidando como uma das melhores universidades federais do Brasil e da América Latina. Acredita-se que esse alcance seja devido ao esforço institucional, sobretudo da atuação dos pesquisadores. Essa afirmação pode ser assegurada mediante os esforços identificados neste estudo, em prol da prática da pesquisa, junto aos PPG, cujos pesquisadores estão abertos a novas práticas que fortaleçam o modo de se produzir conhecimento científico.

No âmbito das universidades públicas brasileiras, o tripé formado por ensino-pesquisa-extensão é indissociável e obrigatório, em virtude da importância de cada um deles para a sociedade. Nesse contexto, o pilar da pesquisa é a dimensão onde os conjuntos de dados científicos são gerados, possibilitando a produção de novos conhecimentos e de inovações. Assim, infere-se que gerir e preservar os dados científicos constitui um processo primordial para o avanço da ciência.

Com esta pesquisa, foi possível diagnosticar as práticas dos pesquisadores atuantes nos PPG, como também as ações institucionais no que concerne à curadoria, à preservação e à gestão dos dados científicos. Pelo diagnóstico realizado, o estudo revelou que esses pesquisadores lidam com dados heterogêneos e singulares, de tipologias e de suportes variados e, em muitos casos, salvos em formatos não adequados à preservação. Além disso, os metadados não são apropriados aos domínios disciplinares, cujos conjuntos de dados estão hospedados em nuvem, ou armazenados em mídias frágeis, sem monitoramento, sem migração e sem atualização tecnológica.

Inferiu-se que os pesquisadores estão comprometidos com a produção de novos conhecimentos, procuram facilitar o armazenamento dos dados e dos arquivos por meio da aquisição de equipamentos de informática para os laboratórios onde lideram as pesquisas e ainda arcam com os custos dos serviços em nuvem (terceirizados). Contudo, necessitam de aportes de outras áreas do conhecimento para que haja manutenção, curadoria e preservação dos dados científicos de maneira adequada. Além disso, carecem de orientações legais como o uso de uma política de preservação e acreditam que esse instrumento possa ser construído colaborativamente.

Mesmo sendo favoráveis ao compartilhamento e ao reuso dos dados científicos, não é comum os pesquisadores disponibilizarem os conjuntos de dados resultantes das coletas, por sentirem inseguros quanto aos créditos da pesquisa, o que seria sanado quando

da atribuição de identificadores persistentes (ORCID¹² para o autor e DOI¹³ para um documento digital), e da atribuição das licenças de uso. Além disso, grande parte dos pesquisadores ainda registra os processos das pesquisas em cadernos de anotações e em protocolos no formato físico. Mesmo sendo essa uma prática bastante comum em determinadas áreas, isso dificulta o armazenamento, o acesso e a localização da informação.

Referente ao uso do Plano de Gestão dos Dados, os sujeitos desta pesquisa mostraram-se favoráveis à adoção, bem como ao uso dos repositórios de dados científicos. Os resultados indicaram, conforme percepção dos pesquisadores, que a gestão e a preservação digital deveriam ser institucionalizadas para salvaguardar os dados científicos, tendo em vista garantir o acesso imediato, em médio e em longo prazo.

No tocante às políticas para viabilizar a preservação digital e a gestão de dados científicos, o diagnóstico mostrou que as ações institucionais ainda são incipientes, sendo fundamental implementar uma política de preservação digital, fornecendo documentos norteadores para a efetiva gestão dos dados oriundos das pesquisas. Ademais, é fundamental promover recursos que disponibilizem e salvaguardem esses conjuntos de dados científicos, contribuindo, assim, para o desenvolvimento da ciência brasileira.

Para a melhoria desse cenário, a universidade pode oferecer e divulgar orientações aos pesquisadores quanto à gestão dos dados (instrumentos, guias e planos). Também pode disponibilizar os serviços de gestão dos dados, com equipes multidisciplinares, compostas por pessoal especializado, como bibliotecários, estatísticos, tecnólogos, dentre outros. Além das equipes, a oferta adequada de recursos tecnológicos é condição *sine qua non* para a montagem de uma infraestrutura de Tecnologia da Informação, capaz de acompanhar e oferecer o devido suporte aos pesquisadores, durante todas as fases da pesquisa.

Salienta-se que, a princípio, as melhorias podem ser gradativas, a partir dos recursos que a universidade detém. No decorrer dos tempos, a oferta dos recursos vai se ampliando, em especial com a participação em consórcios e em convênios oferecidos pelo IBICT e por outras instituições. Acredita-se que essa ampliação da oferta de recursos se concretize, tendo em vista as projeções de alguns projetos públicos, sobretudo no âmbito da Parceria Governo Aberto (OGP), que menciona nove marcos e seus respectivos compromissos, assim como as estratégias para serem cumpridos de 2011 a 2023, em prol da Ciência Aberta.

¹² Em português, Identificador Único e Persistente.

¹³ Em português, Identificador de Objeto Digital.

Nesse sentido, muitas ações vêm evidenciando a preocupação governamental com os percursos para se chegar à Ciência Aberta. Dentre elas, destacam-se como principais e de influência direta na gestão dos dados de pesquisa: a implantação de uma rede interinstitucional, a realização de diagnósticos, a definição de diretrizes e princípios, a sensibilização e a participação cidadã, a articulação junto às agências de fomento e aos editores científicos, bem como a implantação de infraestrutura federada com a proposição de interoperabilidade para o repositório de dados de pesquisa do IBICT.

Diante do exposto, sugere-se observância às diretrizes e às recomendações nacionais e internacionais, sejam elas propostas ou seguidas pelas agências de fomento à ciência brasileira. A título de exemplo, as instituições podem manter ações articuladas junto ao IBICT, pois seus gestores e tecnologistas sinalizam quanto às tendências e auxiliam na tomada de decisões quanto aos projetos implementados e gerenciados nas instituições. Além disso, o Instituto une esforços no sentido de apoiar as demandas vinculadas à área da informação no contexto da ciência, da tecnologia e da inovação, antecedendo os instrumentos mandatórios provenientes das instâncias superiores como a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e o CNPq.

Assim, recomenda-se a adesão ao Consórcio Nacional para a Ciência Aberta, iniciativa do CNPq junto ao IBICT, para melhorar a governança dos conjuntos de dados da pesquisa brasileira por meio da atribuição do DOI. Esse registro permite a identificação, a localização e a recuperação dos objetos digitais de modo otimizado. Outras parcerias junto ao IBICT dizem respeito à utilização do repositório de dados da pesquisa brasileira, o Aleia, e à gestão dos registros dos pesquisadores, por meio do recurso ORCID, que possibilita estabelecer as afiliações institucionais.

Em síntese, espera-se que as análises e os resultados explanados neste estudo possam suscitar a reflexão no que tange à gestão e à preservação digital dos dados científicos, contribuindo para a melhoria das práticas dos pesquisadores atuantes nos PPG. Que os resultados práticos do estudo fortaleçam a governança dos dados resultantes das pesquisas desenvolvidas no interior dos laboratórios acadêmicos e auxiliem os gestores da UFMG na formulação e na implementação de planos, de políticas e de repositórios voltados à preservação de objetos digitais. O desejo é que esses ambientes sejam desenvolvidos, independentemente de sua tipologia documental, na concepção de que assegurem a produção intelectual da instituição, mantendo viva a história e a memória do fazer científico por ela mantido e, portanto, garantindo o acesso continuado, o uso e o reuso em longo prazo.

Dentre as dificuldades encontradas no desenvolvimento deste estudo, cita-se a consulta ao grupo de gestores da UFMG, identificados como grupo FPI. Esse grupo hesitou em responder às perguntas referentes à governança digital na UFMG. Considerando a

ausência de respostas, foi necessário recorrer à pesquisa documental, por meio da investigação aos documentos disponíveis nos sites da universidade e dos demais órgãos e setores ligados à Administração Central da UFMG.

Desse modo, a falta de organização da informação e de padronização nas nomenclaturas dos documentos consultados dificultou o levantamento dos documentos oficiais a serem analisados. Assim, foi obtida uma parte da pesquisa que seria coletada no questionário e direcionada ao grupo FPI. Portanto, fica o anseio de acessar o referido grupo em outros estudos mais profundos, sobretudo considerando o início de novas gestões, nos anos seguintes.

Acredita-se que este estudo poderá servir para inspirar outras investigações, cujo aprofundamento se dê quanto ao uso dos planos de gestão de dados e das atividades de curadoria e de preservação. Tais atividades precisam ser mais estudadas como elementos que permeiam todo o ciclo da investigação científica, visto que o uso dos planos de gestão de dados envolve as exigências por parte dos financiadores da pesquisa. Logo, as atividades de curadoria, de preservação e de gestão dos dados científicos necessitam ser incorporadas às práticas dos pesquisadores e serem orientadas pela universidade, antecipando-se à obrigatoriedade de organizar e de disponibilizar os conjuntos de dados em repositório específico.

A pesquisa mostrou que existe uma prática de preservação de dados de pesquisa realizada de modo individual pelos pesquisadores, o que demanda esforços por parte deles e nem sempre seguindo as melhores práticas recomendadas pela comunidade internacional. Existe, portanto, uma demanda por políticas, metodologias, ferramentas e serviços institucionais para essa finalidade que vai ao encontro das práticas dos bibliotecários realizadas junto ao Sistema de Bibliotecas da UFMG. À medida que sejam oferecidas políticas, metodologias, ferramentas e serviços institucionais, uma demanda latente e potencial por esses serviços poderá ser atendida, contribuindo para a qualidade da pesquisa na instituição. Se isso não é percebido pelos dirigentes da instituição, cabe aos setores envolvidos com a Biblioteconomia, como a Escola de Ciência da Informação e os seus cursos de Graduação e de Pós-Graduação identificarem a importância da atuação do bibliotecário, assumindo o compromisso de suprir essa demanda.

Há anseios, ainda, pesquisar sobre a percepção dos gestores acerca da importância das políticas institucionais para a preservação digital dos dados científicos. Essa abordagem tem relevância porque estimula e facilita a produção de novos conhecimentos, por meio do reuso dos dados, e amplia a visibilidade da ciência para a sociedade.

REFERÊNCIAS

- ALMEIDA, Fernanda Gomes. **Suporte à gestão de dados de pesquisa**: uma ampliação dos serviços oferecidos pelas bibliotecas. 2019. 250f. Tese (Doutorado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Belo Horizonte, 2019. Disponível em: <http://hdl.handle.net/1843/31596>. Acesso em: 30 nov. 2021.
- AMORIM, Cláudia. **Modelos de preservação para revistas eletrônicas científicas**. [S. l.], julho de 2008. Disponível em: <https://camorim.wordpress.com/2008/07/14/modelos-de-preservacao-para-revistas-electronicas-cientificas/>. Acesso em: 20 jan. 2021.
- ARQUIVO NACIONAL. **Recomendações para elaboração de política de preservação digital**. Rio de Janeiro: Ministério da Justiça e Segurança Pública, 2019. Disponível em: https://www.gov.br/arquivonacional/pt-br/servicos/gestao-de-documentos/orientacao-tecnica-1/recomendacoes-tecnicas-1/politica_presercacao_digital.pdf. Acesso em: 02 jul. 2022.
- BARDIN, Lauence. **Análise de conteúdo**. São Paulo: Edições 70, 2016.
- BARROZO, Viviane Lílian dos Santos; ALMEIDA, Carlos Henrique Marcondes de. Curadoria e preservação digital: proposta de diretrizes para a gestão de dados científicos. **Múltiplos Olhares em Ciência da Informação**, Belo Horizonte, v. 12, n. esp., p. 1-8, dez. 2021. Disponível em: <https://periodicos.ufmg.br/index.php/moci/article/view/37188/28960>. Acesso em: 07 jul. 2022.
- BOERES, Sonia Araújo de Assis. **Política de preservação da informação digital em bibliotecas universitárias brasileiras**. 2004. 180f. Dissertação (Mestrado em Ciência da Informação e Documentação) - Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2004. Disponível em: http://repositorio.unb.br/bitstream/10482/1693/1/Dissertacao_Sonia.pdf. Acesso em: 03 dez. 2020.
- BRANCO, Sérgio; BRITTO, Walter. **O que é Creative Commons? Novos modelos de direito autoral em um mundo mais criativo**. Rio de Janeiro: Editora FGV, 2013. 176 p. (Coleção FGV de bolso. Direito & Sociedade)
- BRASIL, **Constituição Federal de 1988**. Brasília, 05 de outubro de 1988. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 07 jul. 2022.
- BRASIL, **Lei dos Direitos Autorais de 1998**. Brasília, 19 de fevereiro de 1998. Disponível em: <http://www2.senado.leg.br/bdsf/handle/id/243240>. Acesso em: 07 jul. 2022.
- CAREGNATO, Sônia Elisa; ROCHA, Rafael Port da; GABRIEL JUNIOR, Rene Faustino. Reuso de dados: princípios FAIR e o ecossistema de pesquisa. In: SALES, Luana Farias; VEIGA, Viviane dos Santos; HENNING, Patrícia; SAYÃO, Luís Fernando (org.). **Princípios FAIR aplicados à gestão de dados de pesquisa**. Rio de Janeiro: IBICT, 2021. p. 187-200. Disponível em: https://ridi.ibict.br/bitstream/123456789/1182/2/IBICT_Principios%20FAIR%20aplicados%20a%20gest%c3%a3o%20de%20dados%20de%20pesquisa_2021.pdf. Acesso em: 02 jul. 2022.
- COMMUNITY OF SCHOLARS, LIBRARIANS, ARCHIVISTS, PUBLISHERS AND RESEARCH FUNDERS (FORCE11). **The FAIR Data Principles**. [S. l.], julho de 2021. Disponível em: <https://force11-org.translate.goog/info/the-fair-data-principles/? x tr sl=en& x tr tl=pt& x tr hl=pt-BR& x tr pto=sc>. Acesso em: 02 jul. 2022.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Carta para a preservação do patrimônio arquivístico digital**. Rio de Janeiro: [S. n.], 2005. Disponível em: https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/conarq_carta_preservacao_patrimonio_arquivistico_digital.pdf. Acesso em: 01 jul. 2022.

CRESWELL, John. **Projeto de pesquisa: métodos qualitativo, quantitativo e misto**. 3. ed. Porto Alegre: Artmed, 2010.

CURTY, Renata. Abordagens de reuso e a questão da reusabilidade dos dados científicos. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, p. 177-193, nov. 2019. Disponível em: <https://revista.ibict.br/liinc/article/view/4777/4315>. Acesso em: 02 jul. 2022.

DIAS, Guilherme Ataíde; ANJOS, Renata Lemos dos; ARAÚJO, Débora Gomes de. A gestão dos dados de pesquisa no âmbito da comunidade dos pesquisadores vinculados aos Programas de Pós-Graduação brasileiros na área da Ciência da Informação: desvendando as práticas e percepções associadas ao uso e reuso de dados. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, p. 5-31, jul./dez. 2019. Disponível em: <https://revista.ibict.br/liinc/article/view/4683/4327>. Acesso em: 01 jul. 2022.

DIGITAL CURATION CENTER (DCC). **A graphical, high-level overview of the stages required for successful management, curation and preservation of data**. Reino Unido, julho de 2022. Disponível em: <https://www.dcc.ac.uk/guidance/curation-lifecycle-model>. Acesso em: 05 jul. 2022.

DIGITAL CURATION CENTER (DCC). **Disciplinary Metadata**. Reino Unido, janeiro de 2020. Disponível em: <http://www.dcc.ac.uk/resources/metadata-standards>. Acesso em: 20 jul. 2020.

DIGITAL PRESERVATION COALITION (DPC). **What is digital preservation?** Reino Unido, dezembro de 2021. Disponível em: https://www.dpconline-org.translate.google.com/digipres/what-is-digipres?x_tr_sl=en&x_tr_tl=pt&x_tr_hl=pt-BR&x_tr_pto=sc. Acesso em: 01 jul. 2022.

DUDZIAK, Elisabeth. **Dados de Pesquisa agora devem ser armazenados e citados**. Sistema de Bibliotecas da USP. São Paulo, julho de 2016. Disponível em: <https://www.aquia.usp.br/noticias/dados-materiais-metodos-revistas-exigem-dados-pesquisa-estejam-disponiveis/#:~:text=A%20decis%C3%A3o%20baseou%2Dse%20em,no%20conhecimento%20cient%C3%ADfico%20e%20tecnol%C3%B3gico>. Acesso em: 02 jul. 2022.

E-ARQ Brasil. **Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos**. Câmara Técnica de Documentos Eletrônicos. Rio de Janeiro: Arquivo Nacional, 2020. Disponível em: https://www.gov.br/conarq/pt-br/assuntos/noticias/conarq-abre-consulta-publica-visualizando-a-atualizacao-do-e-arq-brasil/EARQ_v2_2020_final.pdf. Acesso em: 12 maio 2021.

FERREIRA, Miguel. **Introdução à preservação digital: conceitos, estratégias e actuais consensos**. Guimarães: Escola de Engenharia da Universidade do Minho, 2006. Disponível em: <http://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>. Acesso em: 26 jan. 2021.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO (FAPESP). **Fapesp começa a exigir Plano de Gestão de Dados**. São Paulo, novembro de 2017. Disponível

em: https://www.acessoaberto.usp.br/fapesp-comeca-a-exigir-plano-de-gestao-de-dados/?doing_wp_cron=1656728574.2968099117279052734375. Acesso em: 01 jul. 2022.

FUNDAÇÃO OSWALDO CRUZ (FIOCRUZ). **Dados Abertos**. Série 3, Curso 2. Rio de Janeiro, 2021. Disponível em: <https://mooc.campusvirtual.fiocruz.br/rea/ciencia-aberta/serie3/curso2/aula2.html>. Acesso em: 02 jul. 2022.

GLOBAL OPEN FAIR INITIATIVE BRASIL. **Documentos**. Rio de Janeiro, novembro de 2021. Disponível em: <https://www.go-fair-brasil.org/documentos>. Acesso em: 20 nov. 2021.

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. 2019. 244f. Tese (Doutorado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2019. Disponível em: <http://hdl.handle.net/1843/31349>. Acesso em: 30 nov. 2021.

HIGGINS, Sarah. Digital curation: the emergence of a new discipline. **International Journal of Digital Curation**, [S. l.], v. 6, n. 2, p. 78-88, jul./dez. 2011. Disponível em: <http://www.ijdc.net/article/view/184>. Acesso em: 02 jul. 2022.

HIGGINS, Sarah. The DCC Curation Lifecycle Model. **International Journal of Digital Curation**, [S. l.], v. 3, n. 1, p. 134-140, jan./jun. 2008. Disponível em: <http://www.ijdc.net/article/view/69/48>. Acesso em: 05 jul. 2022.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). **Dados Ibict**. Brasília, dezembro de 2021. Disponível em: <https://dados.ibict.br/>. Acesso em: 02 jul. 2022.

INTERNATIONAL RESEARCH ON PERMANENT AUTHENTIC RECORDS IN ELECTRONIC SYSTEMS (InterPARES). **Project Overview**. Vancouver, 2021. Disponível em: <http://www.interpares.org/welcome.cfm>. Acesso em: 01 jul. 2022.

ISOTANI, Seiji; BITTENCOURT, Ibert. **Dados abertos conectados**. São Paulo: Novatec Editora, 2015.

KIM, Jeonghyun. Growth and Trends in Digital Curation Research: the Case of the International Journal of Digital Curation. **ASIST**, Seattle, v. 51, n. 1, p. 1-4, jan./jun. 2015. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/meet.2014.14505101074>. Acesso em: 10 fev. 2022.

LATTESDATA. **Sobre o repositório**. Brasília, dezembro de 2021. Disponível em: <https://lattesdata.cnpq.br/>. Acesso em: 02 jul. 2022.

MARCONI, Marília Andrade; LAKATOS, Eva Maria. **Fundamentos de metodologia científica**. 8. ed. São Paulo: Atlas, 2017.

MARCONI, Marília Andrade; LAKATOS, Eva Maria. **Técnicas de pesquisa**. 7. ed. São Paulo: Atlas, 2016.

MÁRDERO ARELLANO, Miguel Ángel. **Critérios para a preservação digital da informação científica**. Tese (Doutorado em Ciência da Informação) – Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2008. Disponível em: <https://repositorio.unb.br/handle/10482/1518>. Acesso em: 05 nov. 2020.

MATRIZ DE PRESERVAÇÃO DIGITAL (LOP). **Matriz dos Níveis de Preservação Digital**. National Digital Stewardship Alliance, [S. l.], dezembro de 2019. Disponível em: <https://osf.io/3amnu/>. Acesso em: 03 dez. 2020.

MICHEL, Maria Helena. **Metodologia e pesquisa científica em ciências sociais**. 3. ed. São Paulo: Atlas, 2015.

MOORE, Reagan. Towards a Theory of Digital Preservation. **The International Journal of Digital Curation**, [S. l.], v. 3, n. 1, p. 63-75, jan./jun. 2008. Disponível em: <http://www.ijdc.net/article/view/63/42>. Acesso em: 02 jul. 2022.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA (UNESCO). **Open Science**. [S. l.], 2021. Disponível em: <https://en.unesco.org/science-sustainable-future/open-science>. Acesso em: 30 jan. 2022.

PAGANINE, Lucas Nóbrega *et al.* Desenvolvimento e aplicação de normas para interoperabilidade de repositórios de dados científicos: repositórios do IBICT e do CNPq. In: SALES, Luana Farias; VEIGA, Viviane dos Santos; HENNING, Patrícia; SAYÃO, Luís Fernando (org.). **Princípios FAIR aplicados à gestão de dados de pesquisa**. Rio de Janeiro: IBICT, 2021. p. 157-170.

PAGANINE, Lucas Nóbrega *et al.* **Padrões de interoperabilidade para repositórios de dados de pesquisa**. Brasília, DF: IBICT, 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, jan./jun. 2020. Disponível em: <https://periodicos.furg.br/biblos/article/view/11132>. Acesso em: 02 jul. 2022.

PEREIRA, Durval Vieira. **A Semiótica como teoria de base para a Ciência da Informação**. Projeto de Qualificação-Doutorado em Ciência da Informação. Niterói: PPGCI/UFF, 2019.

Phillips, M., Bailey, J., Goethals, A., & Owens, T. (2013). **Os Níveis NDSA de Preservação Digital: Uma Explicação e Usos**. In Proceedings of the 2013 IS&T Archiving Conference .

REDE BRASILEIRA DE SERVIÇOS DE PRESERVAÇÃO DIGITAL (REDE CARINIANA). **Lista de formatos preserváveis**. Brasília, junho de 2017. Disponível em: <https://cariniana.ibict.br/index.php/listas/lista-de-formatos-preservaveis>. Acesso em: 07 jul. 2022.

REDE BRASILEIRA DE SERVIÇOS DE PRESERVAÇÃO DIGITAL (REDE CARINIANA). **Preservação Digital**. Brasília, maio de 2014. Disponível em: <https://cariniana.ibict.br/index.php/institucional/pre-dig>. Acesso em: 01 jul. 2022.

REDE NACIONAL DE ENSINO E PESQUISA (RNP). **Dados de Pesquisa Abertos**. [S. l.], julho de 2019. Disponível em: <https://dadosdepesquisa.rnp.br/>. Acesso em: 02 jul. 2022.

RESENDE, Liliane Chaves de. **A curadoria de dados científicos na Ciência da Informação**: levantamento do cenário nacional. 2020. 134f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2019. Disponível em: <http://hdl.handle.net/1843/32413>. Acesso em: 30 nov. 2021.

REZENDE, Laura Vilela. **Níveis de preservação digital**. NDSA Levels 2019 (Tradução para Português). [S. l.], outubro de 2019. Disponível em: <https://osf.io/2mkwx/>. Acesso em: 03 dez. 2020.

RODRIGUES, Marcello Mundim. **Repositórios de dados científicos na América do Sul: uma análise da conformidade com os Princípios FAIR**. 2020. 111f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2020. Disponível em: <http://hdl.handle.net/1843/34774>. Acesso em: 30 nov. 2021.

SALES, Luana Farias; SAYÃO, Luís Fernando. A ciência invisível: revelando os dados da cauda longa da pesquisa. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19., 2018, Londrina. **Anais [...]**. Londrina: ANCIB, 2018. p. 4180-4199. Disponível em: <https://brapci.inf.br/index.php/res/download/124831>. Acesso em: 01 jul. 2022.

SANTOS, Thayse Natália Cantanhe de. Curadoria digital e preservação digital: cruzamentos conceituais. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 14, n. 3, p. 450-464, set./dez. 2016. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8646336/pdf>. Acesso em: 20 maio 2021.

SAYÃO, Luís Fernando; SALES, Luana Farias. Curadoria digital e dados de pesquisa. **Atoz**: novas práticas em informação e conhecimento, Curitiba, v. 5, n. 2, p. 67-71, jul./dez. 2016. Disponível em: <https://revistas.ufpr.br/atoz/article/view/49708/30179>. Acesso em: 19 jun. 2022.

SAYÃO, Luis Fernando; SALES, Luana Farias. **Guia de Gestão de Dados de Pesquisa para Bibliotecários e Pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. Disponível em: <https://aben.com.br/guia-de-gestao-de-dados-de-pesquisa-para-bibliotecarios-e-pesquisadores-2015g/>. Acesso em: 20 jun. 2021.

SAYÃO, Luís Fernando; SALES, Luana Farias. Um modelo de implementação para a internet de dados & serviços FAIR. *In*: SALES, Luana Farias; VEIGA, Viviane dos Santos; HENNING, Patrícia; SAYÃO, Luís Fernando (org.). **Princípios FAIR aplicados à gestão de dados de pesquisa**. Rio de Janeiro: IBICT, 2021. p. 215-242.

SIEBRA, Sandra de Albuquerque. Curadoria Digital: uma área em expansão. **Archeion Online**, João Pessoa, v. 6, n. 2, p. 1-6, jul./dez. 2019. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/archeion/article/view/47089>. Acesso em: 01 jul. 2022.

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

SOCIEDADE ESPANHOLA DE DOCUMENTAÇÃO E INFORMAÇÃO CIENTÍFICA (SEDIC). **Redesenhar o que somos para renovar o que fazemos**. Madri, julho de 2022. Disponível em: <https://www.sedic.es/>. Acesso em: 25 maio 2022.

SOUZA, Luciana Gonçalves Silva. **Preservação digital em Instituições de Ensino Superior**: instrumentos para a oficialização e operacionalização da preservação de documentos arquivísticos digitais. 2020. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2020. Disponível em: <http://hdl.handle.net/1843/34550>. Acesso em: 30 nov. 2021.

SOUZA, Luciana Gonçalves Silva; AGANETTE, Elisângela Cristina. A preservação digital em longo prazo amparada por planos de ações: uma revisão sistemática de literatura.

Revista Digital de Biblioteconomia e Ciência da Informação, Campinas, v. 18, n. 1, p. 1-25, jan./jun. 2020. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8661185>. Acesso em: 6 maio 2021.

STATE OF THE CLOUD. **A computação em nuvem**. [S. l.], janeiro de 2021. Disponível em: <https://materiais.matrix.com.br/lp-ebook-flexera-2021-state-of-the-cloud>. Acesso em: 05 jul. 2022.

TIMES HIGHER EDUCATION. **Latin American University Ranking 2022**: results announced. [S. l.], 2022. Disponível em: <https://www.timeshighereducation.com/world-university-rankings/latin-america-university-rankings-2022-results-announced>. Acesso em: 20 jul. 2022.

UNIVERSIDADE FEDERAL DE MINAS GERAIS (UFMG). **Resolução n. 12**, de 10 de outubro de 2019. Institui o Comitê Gestor do Repositório Institucional (COGERI) da Universidade Federal de Minas Gerais (UFMG) e estabelece a sua estrutura. Belo Horizonte, outubro de 2019.

VERGARA, Sylvia Constant. **Projetos e relatórios de pesquisa em administração**. 15. ed. São Paulo: Atlas, 2014.

VIDOTTI, Silvana Aparecida Borsetti Gregorio; TORINO, Emanuelle; CONEGLIAN, Caio Saraiva. #SejaJUSTOeCUIDADOSO: princípios FAIR e CARE na gestão de dados de pesquisa. *In*: SALES, Luana Farias; VEIGA, Viviane dos Santos; HENNING, Patrícia; SAYÃO, Luís Fernando (org.). **Princípios FAIR aplicados à gestão de dados de pesquisa**. Rio de Janeiro: IBICT, 2021. p. 201-214.

VIEIRA, Sônia. **Como elaborar questionários**. São Paulo: Atlas, 2009.

WILKINSON, Mark *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, London, v. 3, n. 1, p. 1-9, mar. 2016. Disponível em: <https://www.nature.com/articles/sdata201618#citeas>. Acesso em: 02 jul. 2022.

APÊNDICE A – ROTEIRO DO SURVEY - PCDC

QUESTIONÁRIO PARA PRODUTORES/CONSUMIDORES DE DADOS CIENTÍFICOS

Investigação destinada, precipuamente, aos professores pesquisadores atuantes na Pós-Graduação – níveis Mestrado e Doutorado da UFMG.

I. APRESENTAÇÃO

Prezado(a) Professor(a) pesquisador(a),

Esta pesquisa com professores atuantes nos cursos de Pós-Graduação da Universidade Federal de Minas Gerais (UFMG) tem por objetivo identificar as práticas de curadoria e preservação digital dos dados científicos adotadas pelos pesquisadores da UFMG. Por isso, gostaríamos de contar com sua participação, de modo voluntário, respondendo ao roteiro *survey* que contém 25 questões.

As questões foram elaboradas a partir da metodologia "Matriz dos Níveis de Preservação Digital", cujo objetivo é orientar a preservação de documentos digitais, propor o desenvolvimento de boas práticas e a formação profissional especializada. A metodologia também considera a reusabilidade que é determinada por características técnicas de conformidade, preconizadas pelo FAIR (do acrônimo *Findability, Accessibility, Interoperability e Reusability*).

Esclarecemos que os dados obtidos neste estudo serão utilizados como subsídio para o desenvolvimento da dissertação intitulada "*Curadoria e Preservação Digital: proposta de diretrizes para a gestão de dados de pesquisa na UFMG*" do Programa de Pós-Graduação em Gestão & Organização do Conhecimento (PPGGOC) da Escola de Ciência da Informação (ECI) da UFMG.

Os participantes que concordarem com os termos dispostos no Termo de Consentimento Livre e Esclarecido (TCLE) disponível no *link* <<https://drive.google.com/drive/>> aceitam participar do estudo mencionado e autorizam a realização e a divulgação dos dados obtidos neste estudo.

Contamos com sua colaboração e, desde já, agradecemos a sua disponibilidade. Em caso de dúvidas, entre em contato conosco pelos e-mails <vivianebarrozo@ufmg.br> e/ou <ch_marcondes@id.uff.br>.

Cordialmente,

Viviane Lílian dos Santos Barrozo (mestranda)
Carlos Henrique Marcondes (professor orientador)

II. CURADORIA E PRESERVAÇÃO DIGITAL DE DADOS CIENTÍFICOS GRUPO PRODUTORES/CONSUMIDORES DE DADOS CIENTÍFICOS (PCDC)

CONSENTIMENTO

1. Termo de Consentimento Livre e Esclarecido (TCLE).

- a. Sim, eu concordo em participar do estudo acima mencionado. b. Não.

ATUAÇÃO NA PÓS-GRADUAÇÃO

2. Em qual área do conhecimento você atua?

- a. Ciências Exatas e da Terra. e. Ciências Agrárias.
b. Ciências Biológicas. f. Ciências Sociais Aplicadas.
c. Engenharias. g. Ciências Humanas.
d. Ciências da Saúde. h. Lingüística, Letras e Artes.

3. Há quanto tempo atua na Pós-Graduação?

- a. Entre 01 (um) e 05 (cinco) anos. c. Entre 11 (onze) e 15 (quinze) anos.
b. Entre 06 (seis) e 10 (dez) anos. d. Mais de 15 (quinze) anos.

PRODUÇÃO/CONSUMO DE DADOS

4. Você gera dados como resultados de suas pesquisas?

- a. Sim. b. Não.

5. Você usa dados de terceiros em suas pesquisas?

- a. Sim. b. Não.

6. Quais instituições fornecem esses dados? [Informe os nomes das instituições e se pertencem à esfera pública ou privada].

ARMAZENAMENTO/CÓPIAS

7. Se você gera dados, mantém mais de uma cópia de cada um de seus arquivos de dados?

- a. Sim. b. Não.

8. Se sim, essas cópias estão em lugares distintos?

- a. Sim. b. Não.
opções. Em caso de marcar a opção "Outro", indique-a].

9. Em quais dispositivos encontram-se essas cópias? [Marque uma ou mais

- a. *Pendrive* pessoal.
- b. *Pendrive* institucional.
- c. HD externo pessoal.
- d. HD externo institucional.
- e. Notebook pessoal.
- f. Notebook institucional.
- g. Computador do Laboratório de Pesquisa da instituição.
- h. Computador pessoal.
- i. Servidor de arquivos da instituição.
- j. Outro. Indique:

10. Em quais formatos/extensões estão cada uma das cópias dos seus arquivos de dados? [Indique-os].

INTEGRIDADE

11. As cópias dos seus arquivos de dados foram verificadas contra vírus?

- a. Sim.
- b. Não.

METADADOS

12. Você mantém uma descrição dos campos de cada um dos seus arquivos de dados?

- a. Sim.
- b. Não.

13. Esta descrição dos campos está em qual idioma? [Em caso de marcar a opção "Outro", indique-a].

- a. Alemão.
- b. Espanhol.
- c. Francês.
- d. Inglês.
- e. Português.
- f. Outro. Indique:

14. Você mantém descrições nos campos em mais de um idioma?

- a. Sim.
- b. Não.

15. Você cria metadados para seus arquivos? [Relativo aos arquivos de dados como um todo: identificação do projeto de pesquisa, descrição dos dados, fonte de financiamento, data ou período de coleta, ferramenta ou instrumento de coleta].

- a. Sim.
- b. Não.

CONTROLE/REUSO

16. Você cede seus arquivos de dados para que terceiros possam reutilizá-los?

- a. Sim.
- b. Não.

17. Você mantém uma cópia on-line dos seus arquivos de dados para que terceiros possam utilizá-los?

a. () Sim.

b. () Não.

18. Se sim, sob que tipo de licença você disponibiliza seus arquivos de dados? [Em caso de marcar a opção "Outra", indique-a].

a. () CC0 1.0 Universal - Dedicção ao Domínio Público - sem direitos autorais.

b. () *Open Data Commons* - ODC-BY 1.0 - A Licença de Atribuição *Open Data Commons* é um contrato de licença destinado a permitir que os usuários compartilhem, modifiquem e usem livremente esse banco de dados.

c. () Outra. Indique:

19. Se sim, em qual repositório de dados ou site você disponibiliza esses arquivos de dados? [Informe o nome do repositório ou site].

20. Em qual formato você disponibiliza os arquivos de dados? [Em caso de marcar a opção "Outro", indique o formato].

a. () PDF.

d. () RDF.

b. () XLS.

e. () LOD.

c. () CSV.

f. () Outro. Indique:

21. Você mantém as versões e um controle de versionamento desses arquivos de dados? [Em caso de marcar a opção "Outros", indique a versão ou o controle de versionamento].

a. () Sim.

b. () Não.

c. () Outros. Indique:

22. Se a UFMG disponibilizasse um repositório de dados que oferecesse serviços de armazenamento, cópia de segurança e acesso a esses dados, facilitaria o compartilhamento e o reuso dos dados de pesquisas? [Avalie, atribuindo notas de 0 a 10, sendo 0 o valor mínimo (nada provável) e 10 o máximo (extremamente provável)].

0 1 2 3 4 5 6 7 8 9 10

ENTREVISTA

23. Você poderia conceder uma breve entrevista remota para enriquecer os detalhes de suas respostas?

a. () Sim.

b. () Não.

24. Se sim, informe o seu nome e o meio para contato (*WhatsApp* (DDD) XXXXX - XXXX ou e-mail institucional para utilização da Conferenciaweb - RNP ou *Microsoft Teams*).

25. Qual é a melhor data para o agendamento da entrevista? [Indique o período entre XX/XX/2022 e XX/XX/2022].

APÊNDICE B – ROTEIRO DO SURVEY - FPI

QUESTIONÁRIO PARA FORMULADORES DE POLÍTICAS INSTITUCIONAIS

Investigação destinada, precipuamente, aos assessores, diretores, pró-reitores e tecnologistas, atuantes nos órgãos e setores ligados à Administração Central da UFMG.

I. APRESENTAÇÃO

Prezado(a) Pró-reitor(a), Diretor(a) e Membro(a) do COGERI,

No intuito de consultar os Dirigentes e os Membros do Comitê Gestor do Repositório Institucional da Universidade Federal de Minas Gerais (UFMG), gostaríamos de contar com sua participação, de modo voluntário, respondendo ao questionário que contém 14 (quatorze) questões sobre as Políticas e Planos de Curadoria e de Preservação digital.

Considerando que a "preservação digital refere-se à série de atividades de gestão necessária para garantir o acesso contínuo aos materiais digitais pelo tempo que for necessário" (DPC, 2021, p. 1) e que os "dados científicos de pesquisa são informações registradas ou produzidas através de qualquer forma ou meio durante o decurso de uma pesquisa" (SILVA, 2019, p. 21), apresentamos esta pesquisa. Pretendemos com este inquérito identificar as ações institucionais voltadas à Curadoria Digital, à Preservação Digital e à Gestão de Dados Científicos no âmbito da UFMG, pois os conjuntos de dados de pesquisa necessitam de tratamento especializado para que possam ser acessados em longo prazo. Isso é necessário, uma vez que os "dados resultantes de financiamento público devem ser gerenciados e compartilhados de forma a garantir o maior benefício possível para o avanço científico e tecnológico" (FAPESP, 2017, não paginado).

Esclarecemos que esta pesquisa fará parte de uma dissertação do Programa de Programa de Pós-Graduação em Gestão & Organização do Conhecimento (PPGGOC) da Escola de Ciência da Informação (ECI) da UFMG. Os dados recolhidos serão utilizados, apenas, para fins científicos e os resultados serão disponibilizados para essa instituição. Será assegurada a confidencialidade das respostas.

Contamos com sua colaboração e, desde já, agradecemos a disponibilidade.

Em caso de dúvidas, entre em contato com a pesquisadora Viviane Barrozo pelo e-mail <vivianebarrozo@ufmg.br> ou com o orientador professor doutor Marcondes pelo e-mail <ch_marcondes@id.uff.br>.

O TCLE está disponível em: <https://drive.google.com/file/d/15p-xWHBQ66-Hifyl0u9u4Bmwd-5u07zN/view?usp=sharing>

Cordialmente,

Viviane Lílian dos Santos Barrozo (mestranda)
Carlos Henrique Marcondes (professor orientador)

ASPECTOS INSTITUCIONAIS DA GESTÃO E DA PRESERVAÇÃO DIGITAL DE DADOS

CONSENTIMENTO

1. Termo de Consentimento Livre e Esclarecido (TCLE).
 - a. Sim, eu concordo em participar do estudo acima mencionado.
 - b. Não.

2. Há algum Plano de Gestão de Dados (PGD) em desenvolvimento ou implementado na UFMG? [Em caso de marcar a opção "Outra", indique-a].
 - a. Sim.
 - b. Não.
 - c. Outra. Indique:

3. Há algum Plano de Preservação Digital (PPD) em desenvolvimento ou implementado na UFMG? [Em caso de marcar a opção "Outra", indique-a].
 - a. Sim.
 - b. Não.
 - c. Outra.

4. A UFMG participa de algum consórcio para a Gestão de Dados Científicos? Qual?

5. Há algum Repositório de Dados Científicos na UFMG? [Em caso de marcar a opção "Outra", indique-a].
 - a. Sim.
 - b. Não.
 - c. Outra. Indique:

6. Se sim, em qual Unidade Acadêmica o Repositório se encontra?

7. Há alguma iniciativa para institucionalização de um Repositório de Dados de Pesquisa para a UFMG? [Em caso de marcar a opção "Outra", indique-a].
 - a. Sim.
 - b. Não.
 - c. Outra. Indique:

8. Há alguma iniciativa para implementação de uma Política de Preservação Digital na UFMG? [Em caso de marcar a opção "Outra", indique-a].
 - a. Sim.
 - b. Não.
 - c. Outra.

9. A UFMG adota alguma(s) estratégia(s) de Preservação Digital? [Indique-a(s)].
 - a. Sim. Indique:
 - b. Não.
 - c. Outra. Indique:

10. Utilize este espaço para compartilhar alguma documentação sobre as estratégias de Preservação Digital na UFMG.

11. Utilize este espaço para relatar a experiência da UFMG referente à Curadoria e à Preservação de Dados Científicos.

ENTREVISTA

12. Você poderia conceder uma breve entrevista em formato remoto para enriquecer os detalhes de suas respostas?
 - a. Sim.
 - b. Não.

13. Se sim, informe o seu nome e o meio para contato (WhatsApp ou e-mail).

14. Qual é a melhor data para o agendamento da entrevista?

APÊNDICE C – TERMO DE CONSENTIMENTO DA PESQUISA - PCDC

Prezado(a) Professor(a) Pesquisador(a),

Eu sou Viviane Lílian dos Santos Barrozo, mestranda na Escola de Ciência da Informação (ECI) da Universidade Federal de Minas Gerais (UFMG), atuando na linha de pesquisa “Arquitetura & Organização do Conhecimento (AOC)”, e venho convidá-lo(a) a participar, de modo voluntário, da pesquisa que pretendo desenvolver, cuja finalidade é diagnosticar o estágio em que a UFMG encontra-se quanto à gestão de dados científicos e sua conformidade às atividades e aos processos de curadoria e de preservação digital consagrados, internacionalmente. A pesquisa também pretende apresentar diretrizes para subsidiar a constituição de uma Política de Preservação de Dados Científicos para a UFMG.

A pesquisa intitulada “Curadoria e preservação digital: proposta de diretrizes para a gestão de dados científicos” contará com um grupo de interesse denominado Produtores/Consumidores de Dados Científicos (PCDC) para obtenção dos dados. O PCDC será composto por pesquisadores vinculados às vinte e uma (21) unidades acadêmicas e atuantes nos noventa e dois (92) cursos de Pós-Graduação nos níveis de Mestrado e Doutorado da UFMG.

A presente pesquisa possui caráter teórico-metodológico efetivado pelo referencial teórico e pelo levantamento de dados por meio de pesquisa de campo. Para tanto, usará uma abordagem indutiva, generalizando suas conclusões a partir dos dados levantados.

De natureza aplicada e de abordagem qualitativa, tentará compreender as práticas dos pesquisadores da referida instituição quanto à curadoria e à preservação dos dados científicos. Quanto aos objetivos, será exploratória e explicativa. Por fim, quanto aos procedimentos técnicos, será bibliográfica, documental e pesquisa de campo, sendo que essa usará a técnica de coleta de dados por meio de um *survey*, com questionários semiestruturados e aplicará uma entrevista semiestruturada.

O questionário será respondido por meio de um formulário eletrônico disponibilizado no *Forms Microsoft Office 365* (disponível para membros da comunidade MinhaUFMG) e enviado para o e-mail institucional dos participantes selecionados. Quanto à entrevista, ocorrerá em dia e horário agendado com o participante (de acordo com sua disponibilidade), terá duração prevista de até 30 minutos e será realizada de modo não presencial, por meio da(s) plataforma(s) Conferênciaweb da Rede Nacional de Ensino e Pesquisa (RNP), *Microsoft Teams* ou aplicativo multiplataformas *WhatsApp* de acordo com a preferência de cada participante. A entrevista será registrada em áudio e/ou vídeo, mediante autorização dos participantes e, posteriormente, transcrita.

Todo o material (questionários e entrevistas) obtido na coleta de dados será utilizado apenas para fins de pesquisa e será mantido em arquivo digital, sob guarda e responsabilidade da pesquisadora, por um período de cinco anos após o término da pesquisa.

A identidade dos participantes será mantida em sigilo, de modo a garantir o anonimato dos mesmos e para atender à Lei Geral de Proteção de Dados Pessoais (LGDP), Lei n. 13.709, de 2018.

Os participantes não terão nenhum tipo de despesa com esta pesquisa e não haverá remuneração pela participação no estudo. Além disso, será garantida a liberdade dos participantes de se retirarem da pesquisa quando bem lhe aprouver.

Durante a pesquisa, é possível surgir dúvidas acerca da curadoria de dados científicos que serão dirimidas pela pesquisadora. Além disso, poderão ocorrer incômodos aos participantes devido à interferência em sua rotina ou cansaço no uso dos ambientes virtuais durante a abordagem. Esses desconfortos serão mitigados ao assegurar o cumprimento da agenda escolhida pelo participante, dando-lhe opção em não responder alguma pergunta do questionário, durante sua entrevista, ou mesmo de interromper a abordagem.

Caso sejam identificados e comprovados danos provenientes desta pesquisa, o(a) participante terá assegurado o direito à indenização.

Os procedimentos adotados nesta pesquisa obedecem aos Critérios da Ética em Pesquisa com Seres Humanos, conforme preconiza a Resolução n. 466, de 2012, do Conselho Nacional de Saúde. Em respeito ao que determina o item IV dessa Resolução, que trata de pesquisas envolvendo seres humanos, apresento este Termo de Consentimento Livre e Esclarecido para que, caso as condições acima lhes convenham, deem o seu “aceite” no questionário on-line para anuência das condições constantes deste documento bem como sua participação no estudo mencionado.

Recomenda-se, fortemente, que os participantes guardem em seus arquivos uma cópia deste documento eletrônico que será disponibilizado para download por meio do link <<https://forms.office.com/>>.

Esperamos que este estudo gere contribuições importantes para a gestão dos dados científicos, incentive a adoção de boas práticas de curadoria e possa subsidiar as políticas de preservação digital da UFMG.

A pesquisadora se compromete a divulgar os resultados da pesquisa. Sempre que os participantes quiserem pedir mais informações sobre a pesquisa, ou seja, que haja interesse do participante quanto à retirada do consentimento de utilização dos dados da pesquisa, basta contatar pelo e-mail da pesquisadora.

Tendo em vista as informações apresentadas, de forma livre e esclarecida, os participantes que manifestarem seu consentimento em participar da pesquisa autorizam a realização e a divulgação dos dados obtidos neste estudo.

Na expectativa de contar com sua participação, agradeço antecipadamente.

Belo Horizonte, _____ de _____ de 2022.

Assinatura do professor orientador

Assinatura da pesquisadora

ENDEREÇO DA PESQUISADORA PARA CONTATOS:

Pesquisadora principal: Viviane Lílian dos Santos Barrozo
E-mail: vivianebarrozo@ufmg.br – Celular (31) 9 8802-6846 WhatsApp (31) 99357-7734

Orientador: Prof. Dr. Carlos Henrique Marcondes de Almeida
E-mail: ch_marcondes@id.uff.br

Gabinete: DMTE 1545 – Fone: 31 3409-6203

Departamento Organização e Tratamento da Informação – DOTI

Escola de Ciência da Informação – sala 3004

Universidade Federal de Minas Gerais (UFMG)

Av. Antônio Carlos, 6.627 – Campus Pampulha - CEP 31.270-901

Escola de Ciência da Informação – sala 3004

Belo Horizonte – Minas Gerais – Brasil – Homepage: <https://ppggoc.eci.ufmg.br/>

Dúvidas de natureza ética sobre a pesquisa poderão ser esclarecidas junto ao Comitê de Ética em Pesquisa – UFMG

Av. Antônio Carlos, 6.627 – Unidade Administrativa II – 2º andar

Campus Pampulha – Belo Horizonte- MG – Brasil – CEP: 31270 901

Contato: coep@prpq.ufmg.br – Telefone: 3409-4592

APÊNDICE D – TERMO DE CONSENTIMENTO DA PESQUISA - FPI

Prezado(a) Dirigente e Membro(a) do COGERI,

Eu sou Viviane Lílian dos Santos Barrozo, mestranda na Escola de Ciência da Informação (ECI) da Universidade Federal de Minas Gerais (UFMG), linha de pesquisa “Arquitetura & Organização do Conhecimento (AOC)”, e venho convidá-lo(a) a participar, de modo voluntário, na pesquisa que pretendo desenvolver, cuja finalidade é diagnosticar o estágio em que a UFMG encontra-se quanto à gestão de dados científicos e sua conformidade às atividades e aos processos de curadoria e de preservação digital consagrados, internacionalmente. A pesquisa também pretende apresentar diretrizes para subsidiar a constituição de uma Política de Preservação de Dados Científicos para a UFMG.

A pesquisa intitulada “Curadoria e preservação digital: proposta de diretrizes para a gestão de dados científicos” contará com um grupo de interesse denominado “Formuladores de Políticas Institucionais (FPI)” para obtenção dos dados. O FPI será composto por pró-reitores(as), diretores(as) administrativos(as), assessores(as), bem como os bibliotecários(as) e tecnologistas membros(as) do Comitê Gestor do Repositório Institucional (COGERI) da UFMG.

A presente pesquisa possui caráter teórico-metodológico efetivado pelo referencial teórico e pelo levantamento de dados por meio de pesquisa de campo. Para tanto, usará uma abordagem indutiva, generalizando suas conclusões a partir dos dados levantados.

De natureza aplicada e de abordagem qualitativa, tentará compreender os aspectos institucionais da gestão e da preservação digital de dados científicos e as políticas e estratégias de preservação adotadas por essa instituição. Quanto aos objetivos, será exploratória e explicativa e, quanto aos procedimentos técnicos, será bibliográfica, documental e pesquisa de campo, sendo que essa usará a técnica de coleta de dados por meio de um *survey*, com questionários semiestruturados e aplicará uma entrevista semiestruturada.

O questionário será respondido por meio de um formulário eletrônico disponibilizado *no Forms Microsoft Office 365* (disponível para membros da comunidade MinhaUFMG) e enviado para o e-mail institucional dos participantes selecionados. Quanto à entrevista, ocorrerá em dia e em horário agendado com o participante (de acordo com sua disponibilidade), terá duração prevista de até 30 minutos e será realizada de modo não presencial, por meio da(s) plataforma(s) Conferênciaweb da Rede Nacional de Ensino e Pesquisa (RNP), *Microsoft Teams* ou aplicativo multiplataformas *WhatsApp* de acordo com a preferência de cada participante. A entrevista será registrada em áudio e/ou vídeo, mediante autorização dos participantes e, posteriormente, transcrita.

Todo o material (questionários e entrevistas) obtido na coleta de dados será utilizado apenas para fins de pesquisa e será mantido em arquivo digital, sob guarda e responsabilidade da pesquisadora, por um período de cinco anos após o término da pesquisa.

A identidade dos participantes será mantida em sigilo, de modo a garantir o anonimato dos mesmos e para atender à Lei Geral de Proteção de Dados Pessoais (LGPD), Lei n. 13.709, de 2018.

Os participantes não terão nenhum tipo de despesa com esta pesquisa e não haverá remuneração pela participação no estudo. Além disso, será garantida a liberdade dos participantes de se retirarem da pesquisa quando bem lhe aprouver.

Durante a pesquisa, é possível surgir dúvidas acerca da curadoria de dados científicos que serão dirimidas pela pesquisadora. Além disso, poderão ocorrer incômodos aos participantes devido à interferência em sua rotina ou cansaço no uso dos ambientes virtuais durante a abordagem. Esses desconfortos serão mitigados ao assegurar o cumprimento da agenda escolhida pelo participante, dando-lhe a opção em não responder alguma pergunta do questionário, durante sua entrevista, ou mesmo de interromper a abordagem.

Caso sejam identificados e comprovados danos provenientes desta pesquisa, o(a) participante terá assegurado o direito à indenização.

Os procedimentos adotados nesta pesquisa obedecem aos Critérios da Ética em Pesquisa com Seres Humanos, conforme preconiza a Resolução n. 466, de 2012, do Conselho Nacional de Saúde. Em respeito ao que determina o item IV dessa Resolução, que trata de pesquisas envolvendo seres humanos, apresento este Termo de Consentimento Livre e Esclarecido para que, caso as condições acima lhes convenham, deem o seu “aceite” no questionário on-line para anuência das condições constantes deste documento, bem como sua participação no estudo mencionado.

Recomenda-se, fortemente, que os participantes guardem em seus arquivos uma cópia deste documento eletrônico que será disponibilizado para download por meio do link <<https://forms.office.com/>>.

Esperamos que este estudo gere contribuições importantes para a gestão dos dados científicos, incentive a adoção de boas práticas de curadoria e possa subsidiar as políticas de preservação digital da UFMG.

A pesquisadora se compromete a divulgar os resultados da pesquisa. Sempre que os participantes quiserem pedir mais informações sobre a pesquisa, ou seja, se houver interesse do participante quanto à retirada do consentimento de utilização dos dados da pesquisa, basta contatar pelo e-mail da pesquisadora.

Tendo em vista as informações apresentadas, de forma livre e esclarecida, os participantes que manifestarem seu consentimento em participar da pesquisa autorizam a realização e a divulgação dos dados obtidos neste estudo.

Na expectativa de contar com sua participação, agradeço antecipadamente.

Belo Horizonte, _____ de _____ de 2022.

Assinatura do professor orientador

Assinatura da pesquisadora

ENDEREÇOS PARA CONTATOS:

Pesquisadora principal: Viviane Lílian dos Santos Barrozo

E-mail: vivianebarrozo@ufmg.br – Celular (31) 9 8802-6846 WhatsApp (31) 9 9357-7734

Orientador: Prof. Dr. Carlos Henrique Marcondes de Almeida

E-mail: ch_marcondes@id.uff.br

Gabinete: DMTE 1545 – Fone: 31 3409-6203

Departamento Organização e Tratamento da Informação – DOTI

Escola de Ciência da Informação – sala 3004

Universidade Federal de Minas Gerais (UFMG)

Av. Antônio Carlos, 6.627 – Campus Pampulha - CEP 31.270-901

Escola de Ciência da Informação – sala 3004

Belo Horizonte – Minas Gerais – Brasil – Homepage: <https://ppggoc.eci.ufmg.br/>

Dúvidas de natureza ética sobre a pesquisa poderão ser esclarecidas junto ao Comitê de Ética em Pesquisa – UFMG

Av. Antônio Carlos, 6.627 – Unidade Administrativa II – 2º andar

Campus Pampulha – Belo Horizonte- MG – Brasil – CEP: 31270 901

Contato: coep@prpq.ufmg.br – Telefone: 3409-4592

APÊNDICE E – ROTEIRO COMPLETO DA ENTREVISTA - PCDC

ENTREVISTA COM PRODUTORES/CONSUMIDORES DE DADOS CIENTÍFICOS

Investigação destinada, precipuamente, aos professores atuantes na Pós-Graduação – níveis Mestrado e Doutorado da UFMG.

I. IDENTIFICAÇÃO

GRUPO PRODUTORES/CONSUMIDORES DE DADOS CIENTÍFICOS (PCDC) PROFESSOR(A) PESQUISADOR(A)

Dados pessoais

Entrevistado(a): _____

E-mail: _____

Área do conhecimento: _____

Nível: _____

Telefone: _____ Data: _____

II. CURADORIA E PRESERVAÇÃO DIGITAL DE DADOS CIENTÍFICOS

SEÇÃO: PRODUÇÃO/CONSUMO DE DADOS

- 1) Qual(ais) tipo(s) de dados científicos são produzidos em suas pesquisas?
- 2) Quais tecnologias são utilizadas?

SEÇÃO: ARMAZENAMENTO/CÓPIAS

- 3) Você já perdeu arquivos de dados importantes para suas pesquisas? Se sim, comente o fato.

SEÇÃO: INTEGRIDADE

- 4) Algum arquivo se corrompeu ou foi recuperado parcialmente? De qual tipo ou formato?
- 5) Foi necessário coletar os dados, novamente? Quais as dificuldades encontradas durante a coleta?

SEÇÃO: METADADOS

- 6) Você mantém algum tipo de organização para encontrar seus dados, com facilidade, como ordenar uma lista ou agrupá-los em pastas por tipo de dados?

- 7) Há algum responsável pelas coleções de dados de pesquisa? Que cargo ele ocupa?

SEÇÃO: CONTROLE/REUSO

- 8) Você compartilha dados coletados em suas pesquisas com terceiros?
Se sim ->
- 9) Com quem? Seus colegas de laboratório, do Programa de Pós-Graduação ou com seus orientandos ou pesquisadores externos à UFMG?
- 10) Neste caso, existe algum termo de cessão ou obrigação de citar você como o produtor dos dados?
Se não ->
- 11) Você cederia seus dados se fosse solicitado? Por que não cederia?

SEÇÃO: PLANO DE GESTÃO DE DADOS

- 12) Você utiliza algum template fornecido pela UFMG, agências de fomento ou ferramenta on-line para registrar as etapas (ciclo de vida) de suas pesquisas?
Se sim ->
- 13) Qual template ou ferramenta on-line? De qual instituição?
- 14) Se não utiliza templates, como registra o processo de pesquisa?

SEÇÃO: BOAS PRÁTICAS

- 15) Os dados produzidos durante suas pesquisas são publicados ou ficam restritos aos dispositivos de armazenamento? Em qual plataforma? Essa plataforma é nacional ou internacional?
- 16) Relate sua experiência na UFMG referente à curadoria e à preservação de dados científicos e indique:
=> Dificuldades:
=> Boas práticas:
=> Sugestões:
- 17) Relate a experiência do seu campo de atuação ou área do conhecimento referente à curadoria e à preservação de dados científicos, seja no âmbito nacional ou internacional.

APÊNDICE F – ROTEIRO SIMPLIFICADO DA ENTREVISTA - PCDC

Identificação (ID)	Questões da entrevista semiestruturada
Produção/Consumo de dados (PC)	
1PC	- Qual(ais) tipo(s) de dados científicos são produzidos em suas pesquisas?
2PC	- Quais tecnologias são utilizadas?
Armazenamento/cópias (AC)	
1AC	- Você já perdeu arquivos de dados importantes para suas pesquisas? Se sim, comente o fato.
Integridade (IN)	
1IN	- Algum arquivo se corrompeu ou foi recuperado parcialmente? De qual tipo ou formato?
2IN	- Foi necessário coletar os dados do arquivo, novamente? Quais as dificuldades encontradas durante a coleta?
Metadados (ME)	
1ME	- Você mantém algum tipo de organização para encontrar seus dados, com facilidade, como uma lista ou a ordenação em pastas por tipo de dados?
2ME	- Há algum responsável pelas coleções de dados de pesquisa? Que cargo ele ocupa?
Controle/Reuso (CR)	
1CR	- Você compartilha dados coletados em suas pesquisas com terceiros?
2CR	- Se sim, com quem? Com seus colegas de laboratório, do Programa de Pós-Graduação ou com seus orientandos ou pesquisadores externos à UFMG?
3CR	- Neste caso, existe algum termo de cessão ou obrigação de citar você como o produtor dos dados?
4CR	- Se não, você cederia seus dados se fosse solicitado? Por que não cederia?
Plano de Gestão de Dados	
1PGD	- Você utiliza algum <i>template</i> fornecido pela UFMG, agências de fomento ou ferramenta on-line para registrar as etapas (ciclo de vida) de suas pesquisas?
2PGD	- Se sim, qual <i>template</i> ou ferramenta on-line? De qual instituição?
3PGD	- Se não utiliza <i>templates</i> , como registra o processo de pesquisa?
Boas práticas (BP)	
1BP	- Os dados produzidos durante suas pesquisas são publicados ou ficam restritos aos dispositivos de armazenamento? Em qual plataforma? Essa plataforma é nacional ou internacional?
2BP	- Relate sua experiência na UFMG referente à curadoria e à preservação de dados científicos e indique: dificuldades, boas práticas e/ou sugestões.
3BP	- Relate a experiência de seu campo de atuação ou área de conhecimento referente à curadoria e à preservação de dados científicos, seja no âmbito nacional ou internacional.

APÊNDICE G - ENDEREÇOS VIRTUAIS DOS DOCUMENTOS INSTITUCIONAIS DA UFMG

DOCUMENTOS INSTITUCIONAIS DA UFMG	
Documento	Endereço eletrônico
Plano de Desenvolvimento Institucional (PDI) - (2018-2023) Administração Central	URL https://ufmg.br/storage/c/e/c/9/cec964e64ae9ba1b073e4c169c50165f_15525812888029_1368629454.pdf
Plano de Dados Abertos (PDA) - (2021-2022) Administração Central	URL https://ufmg.br/storage/b/c/1/e/bc1e2dbd77397b83b20b4d79326925bf_16147770205789_1577796177.pdf
Relatório Gerencial - (2017) Biblioteca Universitária (BU)	URL https://www.bu.ufmg.br/bu_atual/wp-content/uploads/2019/01/017-Relatorio-anual-2017-vfinal.pdf
Política de Desenvolvimento do Acervo - (2015) Biblioteca Universitária (BU)	URL https://www.bu.ufmg.br/bu_atual/wp-content/uploads/2018/02/Pol%C3%ADtica-de-Desenvolvimento-do-Acervo_2015.pdf
Relatório Anual de Gestão - (2017) Diretoria de Arquivos Institucionais (DIARQ)	URL https://www.ufmg.br/diarq/index.php/relatorios-anuais/
Manual de Gestão de Documentos Arquivísticos da UFMG - (2019)	URL https://www.ufmg.br/diarq/index.php/download/manual-de-gestao-de-documentos-arquivisticos-da-universidade-federal-de-minas-gerais/
Plano Estratégico Institucional (PROPLAN) - (semestral)	URL https://www.ufmg.br/proplan/wp-content/uploads/Plano-Estrat%C3%A9gico-Institucional-UFMG.pdf
Plano Diretor de Tecnologia da Informação e Comunicação da UFMG (PDTIC) - (2021-2024)	URL https://www.ufmg.br/dti/wp-content/uploads/2021/07/PDTIC-UFMG-2021-2024-1.pdf
Projeto de Desenvolvimento Institucional da DTI – (2020)	URL https://www.ufmg.br/dti/wp-content/uploads/2021/02/peti-v7-sintetico-19-20.pdf
Relatório de Gestão da DGI – (2018-2021)	URL https://www.ufmg.br/dgi/wp-content/uploads/2021/09/Relatorio-Gestao-2018-2021.docx
Plano de Gestão da PROEX - (2018-2022)	URL https://www.ufmg.br/proex/wp-content/uploads/2022/01/PROEX-2018-2022_-alterado.pdf
Resolução n. 01, de 28 de janeiro de 2021 - Infraestruturas de Apoio Institucional à Pesquisa (AIPq) Pró-Reitoria de Pesquisa (PRPq)	URL https://www.ufmg.br/prpq/wp-content/uploads/2021/02/01resrpq2021FINAL.pdf
Resolução n. 1, de 06 de fevereiro de 2018 - Infraestruturas de pesquisa da Universidade Federal de Minas Gerais Pró-Reitoria de Pesquisa (PRPq)	URL https://www.ufmg.br/prpq/i2pq/wp-content/uploads/2018/02/Resolucao01-2018.pdf

APÊNDICE H – ENDEREÇOS ELETRÔNICOS DOS CURSOS DE PÓS-GRADUAÇÃO DA UFMG

FACE-CEPEAD/ADM/FACE/UFMG@UFMG, ICA-Secretaria de Pos-graduacao/ADM/ICA/UFMG@UFMG, ARQ-Programa de Pós-Graduação em Ambiente Construído e Patrimônio Sustentável/ADM/ARQ/UFMG@UFMG, IGC-Mestrado em Análise e Modelagem de Sistemas Ambientais/ADM/IGC/UFMG@UFMG, FAR-Programa de Pos Graduacao em Análises Clínicas e Toxicológicas/ADM/FAR/UFMG@UFMG, antro_pos@fafich.ufmg.br, ARQ-Colegiado de Pos-Graduacao/ADM/ARQ/UFMG@UFMG, EBA-Programa de Pos-Graduacao em Artes/ADM/EBA/UFMG@UFMG, EBA-Mestrado Profissional em Artes/ADM/EBA/UFMG@UFMG, ICB-Colegiado de Pos-Graduacao em Bioinformatica/ADM/ICB/UFMG@UFMG

ICB-Colegiado de Pos-Graduacao em Biologia Celular/ADM/ICB/UFMG@UFMG, ICB-Colegiado de Pos-Graduacao em Biologia Vegetal/ADM/ICB/UFMG@UFMG, ICB-Colegiado de Pos-Graduacao em Bioquimica/ADM/ICB/UFMG@UFMG, VET-Colegiado de Pos-Graduacao em Veterinaria/ADM/VET/UFMG@UFMG, ppgcc@dcc.ufmg.br, ppgcc.dcc@ufmg.br, ECI-Colegiado de Pos-Graduacao/ADM/ECI/UFMG@UFMG, FAR-Colegiado de Pos-Graduacao em Ciencia dos Alimentos/ADM/FAR/UFMG@UFMG, FAFICH-Pos Graduacao em Ciencia Politica/ADM/FAFICH/UFMG@UFMG, cpg@medicina.ufmg.br, MED-Programa de Pos-Graduacao em Ciencias Aplicadas a Saude do Adulto/ADM/MED/UFMG@UFMG

ICB-Colegiado de Pos-Graduacao em Fisiologia e Farmacologia/ADM/ICB/UFMG@UFMG, EEEFFTO-PROGRAMA DE POS GRADUACAO EM CIÊNCIAS DA REAB.MESTRADO E DOUTORADO/ADM/DFIT/EEFFTO/UFMG@UFMG, cpg.ppg.sca@medicina.ufmg.br, cpg@medicina.ufmg.br, EEEFFTO-Programa de Pós-Graduação em Ciências do Esporte/ADM/EEFFTO/UFMG@UFMG, poscctn@nuclear.ufmg.br, FAR-Colegiado de Pos-Graduacao em Ciencia Farmaceutica/ADM/FAR/UFMG@UFMG, ppgfono@medicina.ufmg.br, ICA-Secretaria de Pos-graduacao/ADM/ICA/UFMG@UFMG, ppgcom@fafich.ufmg.br

mestrado@demc.ufmg.br, FACE-CEPCON-Centro de Pós-Graduação e Pesquisas em Contabilidade e Controladoria/ADM/CIC/FACE/UFMG@UFMG, pgdemografiaeconomia@cedepiar.ufmg.br, DIR-Pos Graduacao da Escola de Direito/ADM/DIR/UFMG@UFMG, DIR-Informacoes da Pos Graduacao da Escola de Direito/ADM/DIR/UFMG@UFMG, ICB-Colegiado de Pos-Graduacao em ECMVS/ADM/ICB/UFMG@UFMG, pgdemografiaeconomia@cedepiar.ufmg.br, colpgsec@fae.ufmg.br, colpgsec@fae.ufmg.br, EEEFFTO-Secretaria de Pos-Graduacao-PROEF/ADM/EEFFTO/UFMG@UFMG, ENF-Colegiado de Pos-Graduacao/ADM/ENF/UFMG@UFMG, secpromestre@gmail.com

posees@dees.ufmg.br, ppgep@dep.ufmg.br, coordenacao-ppgep@dep.ufmg.br, ppgee@cpdee.ufmg.br, ENG-Colegiado de Pos-Graduacao em Engenharia Mecanica/ADM/ENG/UFMG@UFMG, ppgem@demet.ufmg.br, mestradoprofissional@demet.ufmg.br, ENG-Colegiado de Pos-Graduacao em Engenharia Quimica/ADM/ENG/UFMG@UFMG, ICB-Programa de Mestrado Profissional em Ensino de Biologia BH/ADM/ICB/UFMG@UFMG, profbionacional@ufmg.br, ICEX-DEST-Colegiado de Pós-Graduação/ADM/EST/ICEX/UFMG@UFMG

EEFFTO-DTO-Mestrado em Estudos da ocupacao/ADM/DTO/EEFFTO/UFMG@UFMG, EEEFFTO-Programa Pós Graduação em Estudos do Lazer/ADM/EEFFTO/UFMG@UFMG, LETRAS-Programa de Pos-Graduacao em Estudos Linguisticos/ADM/LETRAS/UFMG@UFMG,

FAFICH-Programa de Pos-graduacao em Filosofia/ADM/FIL/FAFICH/UFMG@UFMG,
pgfisica@fisica.ufmg.br, ICB-Colegiado de Pos-Graduacao em Genetica/ADM/ICB/UFMG@UFMG,
IGC-Colegiado Pos-Graduacao Geografia/ADM/IGC/UFMG@UFMG, IGC-Colegiado Pos-Graduacao
Geologia/ADM/IGC/UFMG@UFMG, posgrad@etg.ufmg.br, ENF-Colegiado de Pós Graduação em
Gestão de Serviços de Saúde/ADM/ENF/UFMG@UFMG

ECI-PPGGOC/ADM/ECI/UFMG@UFMG, FAFICH-Programa de Pos-graduacao em
Historia/ADM/HIS/FAFICH/UFMG@UFMG, ICB-Mestrado- Inovacao
Biofarmaceutica/ADM/ICB/UFMG@UFMG, dout-inovacao@qui.ufmg.br, LETRAS-Mestrado
Profissional/ADM/LETRAS/UFMG@UFMG, LETRAS-Colegiado de Pós-Graduação em Estudos
Literários/ADM/LETRAS/UFMG@UFMG, ICEx-DMAT-Colegiado de
Pos-Graduacao/ADM/MAT/ICEX/UFMG@UFMG, FAR-Programa de Pos Graduacao em
Medicamentos e Assistencia Farmaceutica-PPGMAF/ADM/FAR/UFMG@UFMG,
ppgmm@medicina.ufmg.br, ICB-Colegiado de Pos-Graduacao em
Microbiologia/ADM/ICB/UFMG@UFMG

MUSICA-Colegiado de Pos Graduacao/ADM/MUSICA/UFMG@UFMG, ICB-Colegiado de
Pos-Graduacao em Neurociencias/ADM/ICB/UFMG@UFMG, ENF-Colegiado de Pos Graduacao em
Nutricao e Saude/ADM/ENF/UFMG@UFMG, ODONTO-Colegiado de
Pos-Graduacao/ADM/ODONTO/UFMG@UFMG, ICB-Colegiado de Pos-Graduacao em
Parasitologia/ADM/ICB/UFMG@UFMG, ICB-Programa de Pos-Graduacao em
Patologia/ADM/ICB/UFMG@UFMG, ICA-Secretaria de Pos-graduacao/ADM/ICA/UFMG@UFMG,
ppgpspv@medicina.ufmg.br

FAFICH-Colegiado de Pos-Graduacao em Psicologia/ADM/FAFICH/UFMG@UFMG,
ppg.psi.cogcom@gmail.com, posgradpsi@gmail.com, pgquimic@qui.ufmg.br,
posgrad@desa.ufmg.br, ppgsm@medicina.ufmg.br, ppgsp@medicina.ufmg.br, ICA-Secretaria de
Pos-graduacao/ADM/ICA/UFMG@UFMG, FAFICH-Programa de Pos Graduacao em
Sociologia/ADM/FAFICH/UFMG@UFMG, ICB-Colegiado de Pos Graduacao em
Zoologia/ADM/ZOO/ICB/UFMG@UFMG, pgzootecnia@gmail.com

ANEXO A – ARTICULAÇÃO METODOLÓGICA

ARTICULAÇÃO METODOLÓGICA PROPOSTA POR DURVAL VIEIRA PEREIRA			
TÍTULO:			
PROBLEMA:			
OBJETO:			
QUESTÕES:			
•Q1-			
•Q2-			
•Q3-			
OBJETIVOS:			
Objetivo geral:			
Objetivos específicos:	Procedimentos metodológicos:	Autores e temáticas:	Órgãos regulamentadores:
1)			
2)			
3)			

Fonte: Adaptado de Pereira (2019).

ANEXO B – MATRIZ DOS NÍVEIS DE PRESERVAÇÃO DIGITAL (LOP)

Área funcional	Nível 1 (conheça o seu conteúdo)	Nível 2 (proteja o seu conteúdo)	Nível 3 (supervisione o seu conteúdo)	Nível 4 (sustente o seu conteúdo)
Armazenamento	Ter duas cópias completas em localização distinta.	Ter três cópias completas com no mínimo uma cópia em uma localização geográfica distinta.	Ter no mínimo uma cópia em uma localização geográfica com distinta ameaça de desastres em relação ao local das outras cópias.	Ter no mínimo três cópias em localizações geográficas com diferentes ameaças de desastres.
	Documentar todos os suportes de armazenamento onde o conteúdo está armazenado.	Documentar o(s) seu(s) sistema(s) e os suportes de armazenamento indicando os recursos e as dependências requeridas para o seu funcionamento.	Ter no mínimo uma cópia em um tipo de suporte de armazenamento distinto.	Maximizar a diversidade de armazenamento visando a evitar pontos concretos de falha.
	Inserir o conteúdo em um suporte de armazenamento estável.		Monitora a obsolescência de armazenamento e de suportes.	Ter um plano e executar ações para tratar da obsolescência de hardware, de software e de suportes de armazenamento.
Integridade	Verificar a informação de integridade caso tenha sido disponibilizada juntamente com o conteúdo.	Verificar a informação de integridade ao mover ou copiar o conteúdo.	Verificar a informação de integridade de conteúdo em intervalos fixos.	Verificar a informação de integridade em resposta a situações ou a atividades específicas.
	Gerar a informação integral caso não tenha sido disponibilizada junto com o conteúdo.	Usar dispositivos com escrita bloqueada quando se utilizam suportes originais.	Documentar os processos de verificação da informação de integridade e dos seus resultados.	Substituir ou corrigir conteúdos corrompidos, quando necessário.
	Escanear vírus em todo o conteúdo; Isolar o conteúdo em quarentena, caso seja necessário.	Fazer um backup da informação de integridade e armazenar a cópia em uma localização distinta do conteúdo.	Realizar as auditorias de informação de integridade sob demanda.	
Controle	Determinar os agentes humanos e de software que devem ter autorização para ler, gravar, mover e eliminar conteúdo.	Documentar e pôr em prática quais os agentes humanos e de software estão autorizados a ler, gravar, mover e eliminar conteúdo.	Manter os registros de ações (logs) e identificar quais os agentes humanos e software realizaram ações junto ao conteúdo.	Realizar as revisões periódicas de ações/ registros de acesso.
Metadados	Criar inventário de conteúdo e documentar sua localização atual e o armazenamento.	Armazenar metadados suficientes visando a identificar o conteúdo (poderiam estar incluídas combinações de metadados administrativos, técnicos, descritivos, de preservação e estruturais).	Determinar quais os padrões de metadados que serão aplicados.	Registrar as ações de preservação associadas ao conteúdo, incluindo quando essas ações ocorrerem.
	Fazer um backup do inventário e armazenar no mínimo uma cópia separada do conteúdo.		Encontrar e corrigir os problemas nos metadados para cumprir com os padrões por eles especificados.	Implementar padrões de metadados escolhidos.
Conteúdo	Documentar os formatos de arquivos e outras características essenciais do conteúdo, incluindo quando e como foram identificados.	Verificar formatos de arquivos e outras características essenciais de conteúdo.	Monitora a obsolescência e as mudanças nas tecnologias das quais o conteúdo seja dependente.	Realizar migrações, normalizações, emulação e atividades similares que garantam o acesso ao conteúdo.
		Estabelecer as relações/vínculos com os criadores de conteúdo visando a incentivar uma escolha sustentável de arquivos (formatos abertos).		

Fonte: Rezende (2019).

ANEXO C – PARECER CONSUBSTANCIADO DO COMITÊ DE ÉTICA EM PESQUISA DA UFMG

UNIVERSIDADE FEDERAL DE
MINAS GERAIS



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: CURADORIA E PRESERVAÇÃO DIGITAL: PROPOSTA DE DIRETRIZES PARA A GESTÃO DE DADOS DE PESQUISA NA UFMG

Pesquisador: CARLOS HENRIQUE MARCONDES DE ALMEIDA

Área Temática:

Versão: 2

CAAE: 53235021.5.0000.5149

Instituição Proponente: Escola de Ciências da Informação

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 5.272.984

Apresentação do Projeto:

Trata-se da 2a. versão do projeto de pesquisa que responde diligências do parecer de número: 5.180.416.

A Ciência Aberta (Open Science) visa tornar a pesquisa e os dados científicos acessíveis a todos (UNESCO, online, 2021). No contexto da Ciência Aberta, além da abertura dos periódicos, o fazer científico implica na abertura da pesquisa científica de forma ampla, incluindo, assim, a gestão de dados de investigação ou curadoria de dados, as anotações e a revisão por pares, possibilitando o reuso de dados e a reprodutibilidade da pesquisa. Considerando a crescente relevância da gestão e preservação de dados digitais, o alinhamento entre processos e técnicas de curadoria e preservação de dados de investigação, bem como a participação institucional, são fundamentais no tratamento e na disseminação do conhecimento científico. Assim, este estudo pretende elaborar um diagnóstico da gestão de dados científicos e elucidar as diretrizes para a constituição de uma política de preservação digital na UFMG.

A pesquisa adota abordagem qualitativa e de natureza aplicada, e em relação aos seus objetivos será exploratória e descritiva e, se valerá de um estudo de caso (UFMG). Para realizar a pesquisa fará análise bibliográfica e documental e utilizará como técnica de coleta de dados a aplicação de questionários e entrevistas semiestruturadas, ambos a serem realizados através de ambiente

Endereço: Av. Presidente Antonio Carlos, 6627 2º. Andar 2ª Sala 2005 2 Campus Pampulha
Bairro: Unidade Administrativa II **CEP:** 31.270-901
UF: MG **Município:** BELO HORIZONTE
Telefone: (31)3409-4592 **E-mail:** coep@prpq.ufmg.br

Continuação do Parecer: 5.272.984

virtual (online).

Para diagnosticar o estágio em que estágio a UFMG encontra-se quanto à gestão de dados científicos e sua conformidade às atividades e processos de curadoria e preservação digital, foi selecionado 01 (um) grupo de interesse para obtenção dos dados da pesquisa. Este grupo identificado com "Produtores/Consumidores de Dados Científicos" (PCDC), será composto por professores pesquisadores vinculados às vinte e uma (21) unidades acadêmicas e os coordenadores dos noventa (90) cursos de pós-graduação nos níveis de mestrado e doutorado da UFMG. Com este estudo, espera-se contribuir para reflexão e melhoria das práticas de curadoria de dados científicos e fornecer subsídios para a formulação de uma política de preservação de dados de pesquisa digitais para a UFMG.

Os participantes serão contatados por meio de mensagens de correio eletrônico em suas contas institucionais. Sendo que o contato de cada colegiado dos programas de pós-graduação foi coletado em março de 2021, disponibilizados pela Plataforma Integrada de Ouvidoria e Acesso à Informação "Fala.BR" e obtidos no site institucional da UFMG. Os aspectos relacionados à governança digital para preservação de dados científicos, serão obtidos por meio de pesquisa documental e análise do Plano de Desenvolvimento Institucional - PDI (2018-2023), o Plano Diretor de Tecnologia da Informação e Comunicação da UFMG (2021-2024) entre outros documentos disponíveis nas páginas institucionais dos órgãos vinculados à administração direta da UFMG, identificados como "Formuladores de Políticas Institucionais" (FPI).

Como resultado, apresentar diretrizes para subsidiar a constituição de uma Política de Preservação de Dados Científicos para a UFMG.

Objetivo da Pesquisa:

Objetivo Primário:

Em função das questões levantadas, temos como objetivo geral, por meio de diretrizes, subsidiar a constituição de uma política de preservação de dados científicos na UFMG.

Objetivo Secundário:

1) diagnosticar o estágio em que a UFMG encontra-se quanto à gestão de dados científicos e sua conformidade às atividades e processos de curadoria e preservação digital consagrados

Endereço: Av. Presidente Antonio Carlos, 6627 2º. Andar Sala 2005 Campus Pampulha
Bairro: Unidade Administrativa II **CEP:** 31.270-901
UF: MG **Município:** BELO HORIZONTE
Telefone: (31)3409-4592 **E-mail:** coep@prpq.ufmg.br

UNIVERSIDADE FEDERAL DE
MINAS GERAIS



Continuação do Parecer: 5.272.984

internacionalmente;

2) elucidar as diretrizes para subsidiar uma política de preservação de dados científicos da/na UFMG.

Avaliação dos Riscos e Benefícios:

Riscos:

Durante a pesquisa é possível surgir dúvidas acerca da curadoria de dados científicos que serão dirimidas pela pesquisadora. Além disso, poderá ocorrer incômodos aos participantes devido à interferência em sua rotina ou cansaço no uso dos ambientes virtuais durante a abordagem. Esses desconfortos serão mitigados ao assegurar o cumprimento da agenda escolhida pelo participante, dando-lhe opção em não responder alguma pergunta do questionário, durante sua entrevista ou mesmo de interromper a abordagem.

Benefícios:

Com este estudo, espera-se contribuir com a gestão de dados científicos, suscitar a reflexão dos pesquisadores para a melhoria das práticas de curadoria de dados científicos, além de fornecer subsídios para a formulação de uma política de preservação de dados de pesquisa digitais para a UFMG.

Comentários e Considerações sobre a Pesquisa:

Trata-se de uma pesquisa de mestrado sendo realizada no Programa de Pós-Graduação de Gestão & Organização do Conhecimento da Escola de Ciência da Informação. A pesquisa será realizada com recursos próprios.

Considerações sobre os Termos de apresentação obrigatória:

Consideração sobre os termos:

No parecer constubstanciado de número 5.180.416 foi indicado que os seguintes termos já tinham sido entregues e considerados adequados:

1) Folha de rosto preenchida e assinada: OK

2) Aprovação ad referendum da Câmara Departamento de Organização e Tratamento da Informação do parecer consubstanciado que atesta o mérito e relevância científica da proposta, assim como a adequação da metodologia adotada.

Endereço: Av. Presidente Antonio Carlos, 6627 2º. Andar Sala 2005 Campus Pampulha
Bairro: Unidade Administrativa II **CEP:** 31.270-901
UF: MG **Município:** BELO HORIZONTE
Telefone: (31)3409-4592 **E-mail:** coep@prpq.ufmg.br

UNIVERSIDADE FEDERAL DE
MINAS GERAIS



Continuação do Parecer: 5.272.984

Para atender as diligências deste mesmo parecer foram enviados nesta 2ª versão do projeto:

- 1) Carta de encaminhamento, indicando as adequações feitas na pesquisa, esclarecendo como será o processo de obtenção para a pesquisa virtual, assim como as alterações feitas nos termos apresentados e lista de novos documentos encaminhados.
- 2) Projeto completo: enviada nova versão, atualizado de acordo com solicitações feitas – esclarecendo as coletas de dados que serão realizadas e como;
- 3) Instrumentos de coleta de dados: foram enviados o questionário para produtores de dados científicos (revisto conforme solicitações) e roteiro para entrevista semiestruturada;
- 4) TCLE revisto de acordo com solicitações. A versão enviada do TCLE está na forma de carta convite, resguardando a confidencialidade dos dados, o anonimato, o direito à recusa, e desistir do projeto a qualquer momento sem qualquer prejuízo. Foi informado sobre a metodologia, o objetivo e o armazenamento de 05 anos dos dados, salvaguardando a sua consulta. Esclarece que não haverá qualquer forma de pagamento, mas garante o direito a solicitar indenização por danos decorrentes da pesquisa, nos termos da Lei. Informa sobre o registro de áudio/vídeo conforme autorização para fins de transcrição e utilizados apenas para fins de pesquisa. Informa sobre plataformas a serem usadas e duração prevista da coleta.
- 5) Cronograma revisto.
- 6) Formulário de informações básicas revisto em relação às adequações metodológicas, objetivos, riscos e benefícios.

Conclusões ou Pendências e Lista de Inadequações:

Tendo em vista que os pesquisadores fizeram as alterações necessárias e atenderam às diligências que haviam solicitadas, somos, S.M.J., favoráveis à aprovação do projeto.

Considerações Finais a critério do CEP:

Tendo em vista a legislação vigente (Resolução CNS 466/12), o CEP-UFMG recomenda aos

Endereço: Av. Presidente Antonio Carlos, 6627 2º Andar Sala 2005 Campus Pampulha
Bairro: Unidade Administrativa II **CEP:** 31.270-901
UF: MG **Município:** BELO HORIZONTE
Telefone: (31)3409-4592 **E-mail:** coep@prpq.ufmg.br

**UNIVERSIDADE FEDERAL DE
MINAS GERAIS**



Continuação do Parecer: 5.272.984

Pesquisadores: comunicar toda e qualquer alteração do projeto e do termo de consentimento via emenda na Plataforma Brasil, informar imediatamente qualquer evento adverso ocorrido durante o desenvolvimento da pesquisa (via documental encaminhada em papel), apresentar na forma de notificação relatórios parciais do andamento do mesmo a cada 06 (seis) meses e ao término da pesquisa encaminhar a este Comitê um sumário dos resultados do projeto (relatório final).

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1807714.pdf	15/01/2022 22:41:25		Aceito
Projeto Detalhado / Brochura Investigador	BROCHURA_PESQUISA_REVISADA_VIVIANEBARROZO_15012022.pdf	15/01/2022 22:40:40	Viviane Lílian dos Santos Barrozo	Aceito
Outros	CARTA_ENCAMINHAMENTO_ADEQUACOES_PARECER_5180416_VIVIANEBARROZO_15012022.pdf	15/01/2022 22:38:19	Viviane Lílian dos Santos Barrozo	Aceito
Outros	CRONOGRAMA_REVISADO_VIVIANEBARROZO_15012022.pdf	15/01/2022 00:18:29	Viviane Lílian dos Santos Barrozo	Aceito
Outros	ROTEIRO_ENTREVISTA_APENDICE_B_GPCDC_VIVIANEBARROZO_14012022.pdf	14/01/2022 23:45:55	Viviane Lílian dos Santos Barrozo	Aceito
Outros	QUESTIONARIO_APENDICE_A_GPCDC_VIVIANEBARROZO_14012022.pdf	14/01/2022 23:45:26	Viviane Lílian dos Santos Barrozo	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_REVISADO_VIVIANEBARROZO_GPCDC_14012022.pdf	14/01/2022 23:29:42	Viviane Lílian dos Santos Barrozo	Aceito
Parecer Anterior	Parecer_consultado_Mestranda_Viviane_Barroso_Relatora_Elisangela_Aganette.pdf	12/11/2021 14:18:47	Viviane Lílian dos Santos Barrozo	Aceito
Folha de Rosto	Folha_1012643_Folha_de_Rosto_Assinada.pdf	08/10/2021 15:10:15	CARLOS HENRIQUE MARCONDES DE ALMEIDA	Aceito
Brochura Pesquisa	Brochura_Pesquisa.docx	16/08/2021 14:59:37	CARLOS HENRIQUE MARCONDES DE ALMEIDA	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Endereço: Av. Presidente Antonio Carlos, 6627 2º. Andar Sala 2005 2 Campus Pampulha
Bairro: Unidade Administrativa II **CEP:** 31.270-901
UF: MG **Município:** BELO HORIZONTE
Telefone: (31)3409-4592 **E-mail:** coep@prpq.ufmg.br

UNIVERSIDADE FEDERAL DE
MINAS GERAIS



Continuação do Parecer: 5.272.984

Não

BELO HORIZONTE, 04 de Março de 2022

Assinado por:
Críssia Carem Paiva Fontainha
(Coordenador(a))

Endereço: Av. Presidente Antonio Carlos, 6627 2º. Andar Sala 2005 Campus Pampulha
Bairro: Unidade Administrativa II **CEP:** 31.270-901
UF: MG **Município:** BELO HORIZONTE
Telefone: (31)3409-4592 **E-mail:** coep@prpq.ufmg.br

ANEXO D – MANIFESTAÇÃO DE ACESSO À INFORMAÇÃO – FALA.BR

24/06/2022 23:07 Fala.BR - Plataforma Integrada de Ouvidoria e Acesso à Informação

CONTROLADORIA-GERAL DA UNIÃO

Fala.BR - Plataforma Integrada de Ouvidoria e Acesso à Informação
(../..//Principal.aspx)

☰ Viviane Lilian dos Santos Barrozo ⓘ (../..//Login/Logout.aspx)
Usuário

Consultar Manifestação

Respostas ▼

22/03/2021
08:45

Tipo
Resposta Conclusiva

Responsável
Pró-reitora Adjunta de Pós-graduação

Decisão
Acesso Concedido

Especificação da decisão
Resposta solicitada inserida no Fala.Br

Destinatário Recurso 1ª
Pró-reitor de Pós-graduação

Prazo para recorrer
01/04/2021

Anexos Mestrado e Doutorado -NUP.doc

<https://falabr.cgu.gov.br/publico/Manifestacao/DetalleManifestacao.aspx> 1/4

24/06/2022 23:07

Fala.BR - Plataforma Integrada de Ouvidoria e Acesso à Informação

Prezado (a) Sr. (a),

Em relação ao seu pedido de informação, segue em anexo resposta da Pró-reitoria de Pós-graduação da UFMG.

Nota: O SIC UFMG informa que nos termos do art. 15 da Lei 12.527/11, V.Sa. poderá interpor recurso contra a resposta recebida no prazo de 10 (dez) dias a contar da sua ciência. O recurso será dirigido à autoridade hierarquicamente superior a que exarou a resposta em questão, que deliberará no prazo de 05 (cinco) dias corridos.

Aproveitamos o ensejo para manifestar nossos votos de elevada estima e consideração,

Equipe SIC UFMG.



Teor



Resumo

Relação dos Programas de Pós-graduação constantes das Unidades Acadêmicas

Fale aqui

Prezados,
gentileza relacionar todos os Programas de Pós-graduação constantes das Unidades Acadêmicas e os e-mails institucionais dos colegiados dos programas de pós-graduação da UFMG. Caso não seja possível, indicar em qual plataforma poderei fazer a busca.

Estes dados são para fins de pesquisa acadêmica.

Atenciosamente,

Anexos Originais

Não foram encontrados registros.

Manifestação



Tipo de manifestação

Acesso à Informação

Número

23546.012771/2021-11

Esfera

24/06/2022 23:07

Fala.BR - Plataforma Integrada de Ouvidoria e Acesso à Informação

Federal

Órgão destinatário

UFMG – Universidade Federal de Minas Gerais

Serviço

-

Órgão de interesse

-

Assunto

Acesso à informação

Subassunto**Tag**

-

Data de cadastro

15/02/2021

Prazo de atendimento

22/03/2021

Situação

Concluída

Registrado por

032643

Modo de resposta

Pelo sistema (com avisos por email)

Canal de entrada

Internet

Anexos**Anexos da Manifestação** Origem**Nome****Extensão** Anexo Resposta

Mestrado e Doutorado -NUP.doc

doc

[Download](#)**Históricos de ações****Histórico de ações****Data/Hora****Ação****Responsável****Informações Adicionais**15/02/2021
17:37

Cadastro

032643

Registro dos dados da manifestação

10/03/2021
07:43

Prorrogação

Órgão

Resposta de manifestação prorrogada de 10/03/2021 para 22/03/2021

24/06/2022 23:07

Fala.BR - Plataforma Integrada de Ouvidoria e Acesso à Informação

Data/Hora	Ação	Responsável	Informações Adicionais
22/03/2021 08:45	Registro Resposta	Órgão	Resposta Conclusiva
19/11/2021 12:10	Resposta a Pesquisa	032643	Referente a resposta conclusiva de 22/03/2021 08:45:57

Encaminhamentos

Não foram encontrados registros.

Prorrogações

Data/Hora	Prazo Original	Novo Prazo	Responsável	Motivo	Justificativa
10/03/2021 07:43	10/03/2021 23:59	22/03/2021 23:59	Órgão	Outros motivos	<p>Prezado (a),</p> <p>O SIC UFMG aguarda resposta do setor competente. Por isso, prorroga prazo de resposta.</p> <p>Att.,</p> <p>SIC UFMG.</p>

Respostas as pesquisas de satisfação

Data/Hora	Referência	Respostas
19/11/2021 12:10	22/03/2021 08:45	<p>A resposta fornecida atendeu plenamente ao seu pedido? 5</p> <p>A resposta fornecida foi fácil de compreender? Muito fácil de compreender</p> <p>Você está satisfeito(a) com o atendimento prestado? 😊 Muito Satisfeito</p>

[🏠 Voltar à Página Inicial](#)
[🖨 Imprimir](#)
[Voltar ao Topo ^](#)