



UNIVERSIDADE FEDERAL DA PARAÍBA - UFPB
CENTRO DE EDUCAÇÃO / CENTRO DE CIÊNCIAS SOCIAIS APLICADAS
MESTRADO PROFISSIONAL EM GESTÃO NAS ORGANIZAÇÕES APRENDENTES

MARÍA PAULA FERNÁNDEZ JIMÉNEZ

**PRÁTICAS DE GESTÃO, COMPARTILHAMENTO E REÚSO DE DADOS DE
PESQUISA ENTRE PESQUISADORES COLOMBIANOS NA ÁREA DE ECONOMIA
E NEGÓCIOS**

JOÃO PESSOA

2020

MARIA PAULA FERNÁNDEZ JIMENÉZ

**PRÁTICAS DE GESTÃO, COMPARTILHAMENTO E REÚSO DE DADOS DE
PESQUISA ENTRE PESQUISADORES COLOMBIANOS NA ÁREA DE ECONOMIA
E NEGÓCIOS**

Dissertação apresentada ao Mestrado Profissional em
Gestão nas Organizações Aprendentes, da
Universidade Federal da Paraíba, como requisito para
obtenção do título de Mestre em Gestão nas
Organizações Aprendentes.

Linha de pesquisa: Gestão de projetos educativos e
tecnologias emergentes.

Orientador: Prof. Dr. Guilherme Ataíde Dias

JOÃO PESSOA

2020

**Catálogo na publicação Seção de
Catálogo e Classificação**

J61p Jiménez, María Paula Fernández.

Práticas de gestão, compartilhamento e reúso de dados de pesquisa entre pesquisadores colombianos na área de economia e negócios / María Paula Fernández Jiménez. - João Pessoa, 2020.

221 f.

Orientação: Guilherme Ataíde Dias.

Dissertação (Mestrado) - UFPB/Educação.

1. Gestão de dados de pesquisa. 2. Compartilhamento de dados de pesquisa. 3. Reúso de dados de pesquisa. I. Dias, Guilherme Ataíde. II. Título.

UFPB/BC

MARÍA PAULA FERNÁNDEZ JIMÉNEZ

**PRÁTICAS DE GESTÃO, COMPARTILHAMENTO E REÚSO DE DADOS DE
PESQUISA ENTRE PESQUISADORES COLOMBIANOS NA ÁREA DE ECONOMIA
E NEGÓCIOS**

Dissertação apresentada ao Mestrado Profissional em
Gestão nas Organizações Aprendentes da
Universidade Federal da Paraíba, como requisito para
obtenção do título de Mestre em Gestão nas
Organizações Aprendentes.

Data da defesa: 24/07/2020

BANCA EXAMINADORA



Prof. Dr. Guilherme Ataíde Dias
Orientador- MPGOA/UFPB



Profa. Dra. Marynice de Medeiros Matos Autran
Membro Externo - PPGCI/UFPB

Profa. Dra. Emília Maria Da Trindade Prestes
Membro Interno - MPGOA/UFPB

Prof. Dr. Marckson Roberto Ferreira De Sousa
Membro Interno Suplente – MPGOA/UFPB

Profa. Dra. Eliane Bezerra Paiva
Membro Externo Suplente – PPGCI/UFPB

RESUMO

No contexto da ciência aberta, a gestão, compartilhamento e reúso de dados de pesquisa tornou-se uma prática tão presente nos processos de pesquisa que atualmente é vista como elemento inerente ao ofício do pesquisador. Para a realidade da Colômbia trate-se de um tema relativamente novo na prática científica nacional. A ausência de trabalhos acadêmicos que revelem o estado atual do contexto científico colombiano perante esse tema, permite que se questione: Como os pesquisadores, pertencentes a grupos de pesquisa classificados pela Colciencias na área de economia e negócios, praticam a gestão, compartilhamento e reúso de dados de pesquisa? Com o objetivo de desvendar tal fenômeno, a pesquisa adotou uma abordagem quantitativa, de natureza descritiva, sob o método de pesquisa *Survey*. A análise dos dados foi realizada através de estatística descritiva e foi interpretada sob o arcabouço teórico estabelecido. Concluiu-se que os pesquisadores não aplicam sequencialmente o ciclo de vida dos dados, utilizando apenas parte de suas fases, bem como parte das atividades que compõem as fases implementadas. Nesse aspecto, evidencia-se que as práticas de gestão de dados que os pesquisadores realizam com mais frequência encontram-se implicitamente dentro do processo tradicional de pesquisa. Já as práticas exclusivas da curadoria, como documentação de um plano de gestão de dados, criação de metadados e preservação, tendem a ser realizadas ocasionalmente pelos pesquisadores. O grau de ocorrência com que a gestão de dados é implementada desvendou que as fases consideradas fundamentais para a geração de dados de pesquisa de qualidade, em termos de curadoria, preservação e acesso aberto, não são constantes no cotidiano investigativo dos professores. Por outro lado, a prática de compartilhamento de dados abrange inúmeros métodos que variam de pesquisador para pesquisador. O uso de dados próprios gerados em pesquisas anteriores e o reúso de dados de pesquisa sob autoria de terceiros não é uma prática comum entre os pesquisadores. Nessa direção, concebe-se que o acesso aberto a dados é desenvolvido de maneira informal e limitado ao contexto e à rede profissional do pesquisador. Finalmente, a presente pesquisa é uma contribuição para a elaboração futura de estratégias e ações que permitam maior produtividade investigativa e boa gestão dos recursos científicos no contexto estudado.

Palavras-chave: Dados de pesquisa. Gestão de dados de pesquisa. Compartilhamento de dados de pesquisa. Reúso de dados de pesquisa.

RESUMEN

En el contexto de la ciencia abierta, la gestión, el intercambio y el reúso de datos científicos se han convertido en una práctica tan presente en los procesos de investigación, que actualmente se considera un elemento inherente al oficio del investigador. Para la realidad de Colombia, este es un tema relativamente nuevo en la práctica científica nacional. La ausencia de trabajos académicos que revelen el estado actual del contexto científico colombiano en relación con este tema, generó el siguiente cuestionamiento: ¿cómo los investigadores, adjuntos a los grupos de investigación clasificados por Colciencias, en el área de economía y negocios, practican la gestión, intercambio y reúso de datos de investigación? Para desvendar este fenómeno, la investigación adoptó un enfoque cuantitativo, de naturaleza descriptiva, bajo el método de investigación *survey*. El análisis de datos se realizó mediante estadística descriptiva y se interpretó bajo el marco teórico establecido. Se concluyó que los investigadores no aplican secuencialmente el ciclo de vida de los datos, usando apenas parte de sus fases, al igual que parte de las actividades que conforman las fases implementadas. En ese aspecto, se evidenció que las prácticas de gestión de datos que con mayor ocurrencia realizan los investigadores están implícitamente dentro del proceso tradicional de investigación. Ya las prácticas exclusivas de la curaduría como la documentación de un plan de gestión de datos, creación de metadatos, y preservación tienden a ser realizadas por los investigadores ocasionalmente. El grado de ocurrencia con que se implementa la gestión de datos descifró que las fases consideradas fundamentales para generar datos de investigación de calidad, en términos de curaduría, preservación y acceso abierto, no son constantes en la cotidianidad investigativa de los profesores. Por otra parte, la práctica de intercambiar datos abarca inúmeras formas que varían de investigador a investigador. El uso de datos propios generados en anteriores investigaciones y el reúso de datos de investigación bajo autoría de terceros, no es una práctica común entre los investigadores. En esa dirección se concibe que el acceso abierto a datos se desarrolla de modo informal y limitado al contexto y red profesional del investigador. Finalmente, la presente investigación es una contribución para futuramente trazar estrategias y acciones que permitan una mayor productividad investigativa y una buena gestión de los recursos científicos en el contexto estudiado.

Palabras claves: Datos de investigación. Gestión de datos de investigación. Intercambio de datos de investigación. Reúso de datos de investigación.

LISTA DE FIGURAS

Figura 1 – Estrutura do trabalho.....	28
Figura 2 - Modelo de Ciclo de Vida do <i>DataONE</i>	52
Figura 3 – Modelo de ciclo de vida – <i>Digital Curation Centre</i> (DCC)	56
Figura 4 – Modelo de Ciclo de Vida do UKOLN – <i>Research360 Project</i>	60
Figura 5 – Modelo de Ciclo de Vida - Sant’Ana.....	62
Figura 6 - Significância bilateral do perfil do pesquisador.....	92

LISTA DE QUADRO

Quadro 1- Escolas de pensamento – <i>Open Science</i>	31
Quadro 2- Princípios da OECD para acesso a dados de pesquisa.....	35
Quadro 3 – O valor do compartilhamento de dados de pesquisa.....	46
Quadro 4 – Critérios de inclusão.....	67
Quadro 5 – População de estudo.....	68
Quadro 6 – Especialistas.....	71
Quadro 7 – Matriz de validação de conteúdo.....	73
Quadro 8 – W de Kendall: concordância entre expertos.....	73
Quadro 9 – Variável perfil do pesquisador.....	76
Quadro 10 – Variável dados de pesquisa.....	76
Quadro 11 – Variável gestão de dados de pesquisa.....	76
Quadro 12 – Variável compartilhamento de dados de pesquisa.....	76
Quadro 13 – Variável reúso de dados de pesquisa.....	77
Quadro 14 – Operacionalização perfil do pesquisador.....	77
Quadro 15 – Operacionalização dados de pesquisa.....	77
Quadro 16- Operacionalização gestão de dados de pesquisa.....	78
Quadro 17- Operacionalização compartilhamento de dados de pesquisa.....	79
Quadro 18- Operacionalização de reúso de dados de pesquisa.....	79
Quadro 19 - Escala Likert.....	81
Quadro 20 - Categorização dos resultados.....	82

Quadro 21 - Interpretação dos resultados.....	82
Quadro 22 – Conceitos estatísticos.....	82
Quadro 23 - Técnicas estatísticas da pesquisa.....	82
Quadro 24 . Categoria Colciencias.....	87
Quadro 25 - Linhas de pesquisa.....	89
Quadro 26 – Metadados.....	114
Quadro 27 - Critérios de qualidade dos dados.....	120

LISTA DE TABELAS

Tabela 1. Tipo de instituição.....	84
Tabla 2. Nivel de formação acadêmica.....	85
Tabla 3. Tempo de atuação como professor.....	85
Tabla 4. Tipo de instituição – nível de formação acadêmica.....	86
Tabela 5. Tipo de instituição - tempo como professor.....	86
Tabela 6. Categoria Colciencias.....	88
Tabela 7. Tempo de atuação como pesquisador.....	89
Tabela 8. Tipo de instituição - Categoria Colciencias.....	90
Tabela 9. Tipo de instituição – Tempo como pesquisador.....	91
Tabela 10. Pvalue teste Qui-quadrado de Pearson.....	91
Tabela 11. Perfil dos pesquisadores de acordo com suas características.....	93
Tabela 12. Natureza dos dados.....	94
Tabela 13. Tipos de dados.....	95
Tabela 14. Formatos de dados.....	95
Tabela 15. Tempo de armazenamento dos dados.....	96
Tabela 16. Tipo de armazenamento interno.....	97
Tabela 17. Tipo de armazenamento externo.....	98
Tabela 18. Tipo de armazenamento em Repositórios institucionais.....	99
Tabela 19. Tipo de armazenamento em Repositórios externos.....	100
Tabela 20. Pvalue teste Qui-quadrado de Pearson – Características dos dados.....	101
Tabela 21. Perfil dos pesquisadores segundo as particularidades dos dados.....	102
Tabela 22. Familiaridade gestão de dados.....	104

Tabela 23. Caracterização dimensão planejamento de gestão de dados.....	109
Tabela 24. Dimensão planejamento de gestão de dados – Pesquisadores de universidade particular.....	111
Tabela 25. Dimensão planejamento de gestão de dados – Pesquisadores de universidade pública.....	112
Tabela 26. Caracterização dimensão metadados.....	116
Tabela 27. Dimensão metadados – Pesquisadores de universidade particular.....	118
Tabela 28. Dimensão metadados – Pesquisadores de universidade pública.....	119
Tabela 29. Caracterização dimensão garantia e controle de qualidade dos dados.....	123
Tabela 30. Dimensão garantia e controle de qualidade dos dados – Pesquisadores de universidade particular.....	125
Tabela 31- Dimensão garantia e controle de qualidade dos dados – Pesquisadores de universidade pública.....	126
Tabela 32. Caracterização dimensão preservação de dados.....	132
Tabela 33. Dimensão preservação – Pesquisadores de universidade particular.....	136
Tabela 34. Dimensão preservação – Pesquisadores de universidade pública.....	137
Tabela 35. Quantidade de dados disponíveis.....	139
Tabela 36. Tipos de dados disponíveis.....	141
Tabela 37. Acesso a dados de pesquisa.....	142
Tabela 38. Canais de acesso a dados.....	144
Tabela 39. Caracterização disponibilidade de dados de pesquisa.....	144
Tabela 40. Qualidade dos dados de pesquisa compartilhados.....	147
Tabela 41. Qualidade dos dados de pesquisa compartilhados – Universidade particular.....	149
Tabela 42. Qualidade dos dados de pesquisa compartilhados– Universidade pública.....	149
Tabela 43. Tipo de licenças.....	151
Tabela 44. Caracterização tipos de licença.....	154
Tabela 45. Caracterização condições de uso dos dados disponibilizados.....	154

Tabela 46. Condições de uso dos dados disponibilizados – Universidade particular.....	156
Tabela 47. Condições de uso dos dados disponibilizados – Universidade pública.....	157
Tabela 48. Caracterização dimensão uso e reúso de dados próprios ou de terceiros.....	160
Tabela 49. Dimensão reúso de dados de pesquisa próprios e de terceiros – Pesquisadores de universidade particular.....	162
Tabela 50. Dimensão reúso de dados de pesquisa próprios e de terceiros – Pesquisadores de universidade público.....	162
Tabela 51. Atribuição de metadados a dados usados.....	163
Tabela 52. Dimensão qualidade dos dados de pesquisa– Pesquisadores de universidade particular.....	165
Tabela 53. Dimensão qualidade dos dados de pesquisa– Pesquisadores de universidade público.....	165
Tabela 54. Meios de acesso a dados de terceiros.....	166

LISTA DE SIGLAS

COLCIENCIAS	<i>Departamento Administrativo de Ciencia, Tecnología e Innovación</i>
DataONE	<i>Data Observation Network for Earth</i>
DCC	<i>Digital Curation Centre</i>
I2D	<i>Infraestrutura Institucional de Dados e Informação</i>
NSB	<i>US National Science Board</i>
OCyT	<i>Observatorio Colombiano de Ciencia y Tecnología</i>
OECD	<i>Organisation for Economic Co-operation and Development</i>
UKOLN	<i>UK Office for Library and Information Networking</i>
FAPESP	<i>Fundação de Amparo à Pesquisa do Estado de São Paulo</i>

SUMÁRIO

1	INTRODUÇÃO	17
1.1	PROBLEMA DE PESQUISA.....	22
1.2	OBJETIVOS	26
1.2.1	Objetivo geral.....	26
1.2.2	Objetivos Específicos.....	26
1.3	RELEVÂNCIA E JUSTIFICATIVA DO ESTUDO	26
1.4	ESTRUTURA DO TRABALHO.....	28
2	<i>E-SCIENCE</i> E CIÊNCIA ABERTA.....	30
2.1	CIÊNCIA ABERTA.....	30
2.2	E-SCIENCE.	36
3	DADOS DE PESQUISA	38
4	GESTÃO DE DADOS DE PESQUISA.....	44
5	MODELOS DE CICLO DE VIDA DOS DADOS.....	50
5.1.1	Data Life Cycle Model – DataONE	51
5.1.2	Curation Lifecycle Model – Digital Curation Centre (DCC).....	55
5.1.3	Research Lifecycle model (Research360 Project) – UKOLN.....	58
5.1.4	Ciclo de vida dos dados (CVD-CI) - Sant’ Ana.....	62
6	METODOLOGIA	65
6.1	TIPO DE PESQUISA.....	65
6.2	DELINEAMENTO DA PESQUISA.....	66
6.2.1	Contextualização	67
6.2.2	Características da população	67
6.2.3	Amostra da pesquisa.....	68
6.2.4	Elaboração do questionário	69
6.2.5	Validade e objetividade do questionário	71

6.2.6	Operacionalização das variáveis	75
6.2.7	Procedimento para a coleta de dados	79
6.2.8	Técnica de análise dos dados.....	80
7	APRESENTAÇÃO ANÁLISE E INTERPRETAÇÃO DOS DADOS	84
7.1	PERFIL DO PESQUISADOR	84
7.1.1	Dimensão trajetória profissional	84
7.1.2	Dimensão trajetória investigativa do pesquisador	86
7.2	CARATERIZAÇÃO DOS DADOS DE PESQUISA	93
7.2.1	Dimensão particularidades dos dados	94
7.3	PRÁTICAS DE GESTÃO DE DADOS DE PESQUISA	103
7.3.1	Dimensão conhecimento sobre gestão de dados	104
7.3.2	Dimensão planejamento de gestão de dados	105
7.3.3	Dimensão Metadados	113
7.3.4	Dimensão garantia e controle de qualidade dos dados	120
7.3.5	Dimensão preservação de dados.....	127
7.4	PRÁTICAS DE COMPARTILHAMENTO DE DADOS DE PESQUISA	138
7.4.1	Dimensão disponibilidade de dados de pesquisa.....	138
7.4.2	Dimensão qualidade dos dados de pesquisa compartilhados	145
7.4.3	Dimensão condições de compartilhamento e reúso dos dados de pesquisa	150
7.5	PRÁTICAS DE USO E REÚSO DE DADOS DE PESQUISA	158
7.5.1	Dimensão uso e reúso de dados de pesquisa	158
7.5.2	Dimensão qualidade dos dados de pesquisa.....	163
7.5.3	Dimensão meios de acesso a conjuntos de dados de terceiros	166
8	CONSIDERAÇÕES FINAIS	168
	REFERÊNCIAS	175
	APÊNDICES.....	183
	APÊNDICE A – QUESTIONARIO	184
	APÊNDICE B – MATRIZ DE CONSISTÊNCIA METODOLÓGICA.....	194

APÊNDICE C – FORMATO MATRIZ DE OPERACIONALIZAÇÃO DE VARIÁVEIS	196
APÊNDICE D- CARTA DE CONVITE A EXPERTOS	197
APÊNDICE E- FORMATO DE CERTIFICADO DE VALIDADE DE CONTEÚDO	200
ANEXOS	202
ANEXO A- AVALIAÇÃO DO INSTRUMENTO POR EXPERTOS.....	203
ANEXO B- TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO	222

1 INTRODUÇÃO

O uso de novos canais de comunicação massiva, como a internet, e a incorporação da tecnologia aos processos científicos, dentre outros, são fatores que tem levado a uma reinvenção da ciência. Dentre as características deste novo momento pelo qual a ciência passa, destaca-se a abertura dos processos investigativos pela comunidade acadêmica, como um passo fundamental à universalização, avanço e democratização da ciência.

Com os avanços tecnológicos em convergência com o desenvolvimento científico, o processo de investigação científica tem sido transformado, tendo em vista que nas últimas décadas, pela complexidade dos fenômenos, houve a necessidade de que eles fossem estudados a partir do uso intensivo de infraestrutura tecnológica, capazes de produzir enormes quantidades de dados. Assim, praticamente todas as fases do processo investigativo são permeadas pela utilização de ferramentas tecnológicas, desde a captura ou criação de grandes quantidades de dados, passando pelo seu processamento e análise, e finalizando com seu armazenamento em grande escala (GRAY, 2007).

Além destas características, a tecnologia tem permitido a integração da comunidade acadêmica, independente da sua localização geográfica, privilegiando e incentivando a ciência colaborativa. Este novo patamar de fazer ciência baseado no uso de infraestrutura informática para manipulação e processamento de grandes volumes de dados, em um ambiente cooperativo, é conhecido como *e-Science*. Os autores do artigo “*Recent Research Advances in e-Science*”, Yang, Wang, e Von Laszewki (2009), mencionam que o termo *e-Science* foi cunhado por John Taylor, diretor-geral do Escritório de Ciência e Tecnologia do Reino Unido em 1999. Taylor (1999, apud YANG, WANG, E VON LASZEWSKI, 2009, p. 353, tradução nossa) define *e-Science* da seguinte maneira: “a *e-Science* trata da colaboração global em áreas-chave da ciência, e a infraestrutura de computação da próxima geração permitirá isso”¹.

Em síntese, a *e-Science* pode ser concebida como “as tecnologias de informação em rede que apoiam atividades de pesquisa científica, como colaboração, compartilhamento de dados e disseminação de descobertas”² (RIBES; LEE, 2010, p. 231, tradução nossa), levando ao entendimento de que, além de gerenciar grandes quantidades de dados, de promover a ciência

¹ Texto original: “*e-Science is about global collaboration in key areas of science, and the next generation computing infrastructure will enable it*”.

² Texto original: Cyberinfrastructure (CI), eScience and eInfrastructure are the current terms of art for the networked information technologies supporting scientific research activities such as collaboration, data sharing and dissemination of findings.

colaborativa e de apresentar novos métodos de fazer ciência, privilegia a visão da ciência aberta, isto é, a abertura de todo o processo científico, significando o acesso livre aos recursos produzidos, especialmente aos dados de pesquisa, vistos como um recurso científico que se for reaproveitado pode gerar mais valor e maior eficiência.

Segundo o *National Research Council et al.* (1997, p. 10, tradução nossa), destaca-se que “o valor dos dados reside no seu uso”. A partir desta perspectiva, os dados se tornam relevantes na medida em que seu acesso se abre à comunidade acadêmica, possibilitando sua reutilização, sendo esta prática uma forma de fazer uso eficiente dos recursos científicos, especialmente no que se refere ao tempo e investimento econômico que requer uma pesquisa.

Os conjuntos de dados são indispensáveis para sustentar e validar as descobertas, sendo fontes primárias que possuem um tempo de vida maior que os projetos que os criaram (SAYÃO; SALES, 2015). Como consequência disso, caso sejam disponibilizados abertamente, estes dados podem ser reaproveitados por pesquisadores para: estudar novos fenômenos; promover diversidade de perspectiva na análise dos dados de modo que se construa novo conhecimento; explorar assuntos relacionados ao tópico de pesquisa dos dados que não foram abordados; evitar duplicidade de pesquisas, ou verificar os resultados obtidos mediante a replicação do experimento (UHLIR; SCHORÖEDER, 2007) entre outras inúmeras possibilidades que são fundamentais para o progresso científico.

O acesso aberto a dados já não é mais novidade, hoje, para a comunidade científica, porém, embora a adoção gradual de práticas abertas venha gerando resultados positivos, o tema ainda é envolto em muitas dúvidas e inseguranças quanto aos efeitos e consequências desta transformação pela qual a ciência está passando. Assim, ainda que exista certo consenso quanto aos benefícios que o compartilhamento de dados pode trazer, a ausência de segurança quanto aos detalhes mais práticos acaba conduzindo a um cenário no qual muitos pesquisadores querem poder receber os dados, mas poucos estão dispostos a compartilhar (RAMÍREZ; SAMOILOVICH, 2018).

É assim que se observa a rotina de pesquisadores na América Latina, por exemplo, na qual ao finalizar seus processos de investigação, apresentam à sociedade seus resultados através de artigos publicados em revistas científicas, que majoritariamente são de acesso aberto³.

³ A abertura de resultados de pesquisa mediante o acesso gratuito a artigos de revistas científicas é uma prática comum em América Latina. A plataforma e-revistas (Plataforma Open Access de Revistas Científicas Electrónicas españolas e latino-americanas) conta com 341 revistas de acesso aberto de âmbito ibero-americano. REDALYC (Rede de Revistas Científicas da América Latina e do Caribe, Espanha e Portugal) conta com 550 revistas de acesso aberto. Do mesmo modo entre as principais iniciativas de acesso aberto na América Latina, está a biblioteca eletrônica SciELO a qual atualmente possui 373 revistas. Por fim, LA Referencia é uma rede latino-americana de repositórios de acesso aberto a qual busca compartilhar e visibilizar a produção científica da América Latina.

Diferente desta prática, que já está naturalizada no meio acadêmico e que é percebida como um dever, o compartilhamento dos dados provenientes do processo científico ainda não foi incorporado à rotina dos pesquisadores, de modo que terminam, na maioria dos casos, sendo armazenados, sem a devida gestão, na base de dados pessoal do pesquisador, e consequentemente, sendo relegados à obsolescência tecnológica, impedindo seu reuso em investigações posteriores (SAYÃO; SALES, 2014).

Há, portanto, diversas preocupações e motivos que geram dúvidas nos pesquisadores e motivam esse comportamento de não disponibilizar nem compartilhar os dados provenientes de suas pesquisas, entre eles, o desconhecimento sobre como tratar e organizar os dados, ausência de estímulos para adotar esta prática, dúvidas sobre os direitos autorais e licenças, falta de reconhecimento suficiente ao pesquisador, mal uso de dados por parte dos usuários, falta de recursos para o tratamento de dados (BORGMAN, 2012 ; DIGITAL SCIENCE *et al.*, 2018), além da falta de um marco normativo que esclareça o processo, e possíveis conflitos de interesses entre os financiadores da pesquisa e o pesquisador. Contudo, cada vez mais o contexto científico atual está trabalhando em prol do compartilhamento e reuso de dados de pesquisa de modo que se torne uma prática natural e inerente ao ofício do pesquisador.

Na América Latina e no Caribe, é possível identificar atualmente um leque de iniciativas destinadas a gerar um cenário propício para o compartilhamento de dados de pesquisa. No que se refere às ações empreendidas pelos governos em relação aos sistemas de comunicação científica, como repositórios ou bases de dados, a Argentina mediante a Lei nº 26.899 de Dezembro de 2013, institui o *Sistema Nacional de Repositorios Digitales*⁴ formado por repositórios digitais institucionais de acesso aberto de organizações públicas e privadas que recebem financiamento de fundos públicos para o desenvolvimento de pesquisas. Do mesmo modo o Peru com o *Repositorio Nacional Digital de Ciencia, Tecnología e Innovación*⁵, “ALICIA”, estabelecido pela Lei nº 30.035/2013, contempla esta infraestrutura como um mecanismo de preservar e permitir o acesso aberto a produtos de investigação como dados de pesquisas, assim como também o México e seu Repositório Nacional⁶ e a iniciativa do estado de São Paulo, Brasil, com o lançamento da Rede de Repositórios de Dados Científicos do Estado de São Paulo, envolvendo universidades localizadas nessa região e a Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP. A FAPESP além de incentivar a construção do repositório criou outras ações relacionadas aos dados de pesquisa, entre elas, o

⁴ <http://repositoriosdigitales.mincyt.gob.ar/vufind/>

⁵ <https://portal.concytec.gob.pe/index.php/informacion-cti/alicia>

⁶ <https://www.repositorionacionalcti.mx>

compartilhamento de boas práticas em pesquisa, política de acesso aberto às publicações e dados resultantes de seu apoio, exigência de planos de gestão de dados para projetos de pesquisa financiados com recursos públicos procedentes da fundação, e políticas FAPESP para propriedade intelectual (FAPESP, online).

No mesmo sentido, a Colômbia registra iniciativas de entidades que têm sido reconhecidas nacionalmente por promoverem projetos de ciência aberta. Por exemplo, o *Instituto de Investigación de Recursos Biológicos Alexander von Humboldt*, pertencente ao *Sistema de Información sobre Biodiversidad de Colombia – SIB*, se destaca como líder na prática de acesso aberto a dados científicos com o projeto “*Infraestructura Institucional de Datos e Información*⁷ - I2D”. A I2D se apresenta como um repositório que visa a preservar os dados científicos provenientes das pesquisas desenvolvidas pelo instituto e seus parceiros com o fim de serem consultados e usados na construção de novos conhecimentos. Do mesmo modo, destaca-se o Jardim Botânico de Bogotá “*José Celestino Mutis*”, organização pública, com o repositório de dados: *Herbário JBB*⁸.

Estes tipos de canais de comunicação científica são constantemente alimentados por grupos de pesquisa adjuntos a universidades públicas e particulares, organizações privadas como ONGs, e entidades públicas que promovem a ciência aberta. Estes repositórios públicos contêm diversos produtos, de todas as áreas do conhecimento, derivados de processos de investigação, tais como artigos, dissertações, teses, projetos de pesquisa, e dados, os quais são administrados dentro de marcos jurídicos e políticas de compartilhamento e uso.

Nesse sentido, por trás dos esforços das diversas organizações para a construção de centros de dados de pesquisa, existem práticas de manutenção e preservação de dados, conhecidas como curadoria digital, as quais garantem o compartilhamento, acesso e potencial reúso desses dados de maneira sustentável.

A curadoria de dados envolve um conjunto de práticas de gestão de dados ao longo do seu ciclo de vida, a qual busca outorgar confiabilidade, autenticidade, e interpretação a este recurso, com o propósito de serem futuramente reusados em novos contextos científicos. Abbott ratifica a relevância da gestão de dados assegurando que a curadoria facilita “o acesso persistente a dados digitais confiáveis, melhora a qualidade dos dados em si e seu contexto de pesquisa, melhora a velocidade e a variedade de acesso, oportunidade de compartilhamento e

⁷ <http://datos.humboldt.org.co>

⁸ <http://coleccion.jbb.gov.co/herbario/site/index>

de análise de dados, entre outros benefícios de pesquisa”⁹ (ABBOTT, 2008, *online*, tradução nossa).

A curadoria, portanto, está ligada ao processo de gestão de dados, procurando que, mediante a constante preservação deste recurso, os dados sejam protegidos de ameaças como a perda e a obsolescência e se outorgue a contextualização para serem compreendidos no futuro, entre outros aspectos, que possibilitam e viabilizam seu acesso contínuo e reúso (ABBOTT, 2008, *online*), de modo que se maximize o potencial dos dados e consequentemente se dê um reaproveitamento dos recursos científicos, aspecto que ganha bastante relevância no ambiente atual.

Nessa perspectiva, a gestão de dados de pesquisa, torna-se a pedra angular no contexto de *e-Science* e ciência aberta, especialmente porque este conjunto de práticas “têm um papel crucial como facilitador nos processos de compartilhamento dos dados, na garantia da sustentabilidade e acessibilidade dos dados” (SAYÃO; SALES, 2015, p.5). Dentro deste contexto, os pesquisadores, que representam um dos principais atores no âmbito científico, responsáveis por gerar grandes quantidades de dados através da pesquisa, possuem o atual desafio de cada vez mais assimilar a cultura da gestão de dados e promover sua incorporação na rotina dos grupos de pesquisa sob sua tutela.

Por conseguinte, considerando a importância da gestão, compartilhamento e reúso de dados de pesquisa no contexto da *e-Science* e ciência aberta, esta pesquisa busca revelar as práticas de gestão, compartilhamento e reúso de dados de pesquisa¹⁰, realizadas pelos pesquisadores pertencentes aos grupos de pesquisa classificados por Colciencias, na área de conhecimento economia e negócios, localizados na cidade de Bucaramanga e sua região metropolitana.

⁹ Texto original: Digital curation facilitates: persistent access to reliable digital data, improved quality of the data itself and its research context, improved speed and range of access, data sharing and analysis opportunities, and other research benefits.

¹⁰ O Departamento Administrativo de Ciencia, Tecnología e Innovación –Colciencias-, agrupa os grupos de pesquisa por área de conhecimento, segundo a divisão da organização de cooperação e desenvolvimento econômico – OECD. A área a estudar pertence à grande área de conhecimento: ciências sociais, onde se encontra a área de conhecimento: economia e negócios, composta por carreiras relacionadas a economia, econometria, negócios, administração e afins

1.1 PROBLEMA DE PESQUISA

O gerenciamento, compartilhamento e reúso de dados de pesquisa tem se tornado uma prática tão presente nos processos de investigação em países desenvolvidos, que atualmente é visto como elemento inerente ao ofício do pesquisador. Os países associados à União Europeia, os Estados Unidos¹¹, dentre outros – que se destacam pelo seu contínuo progresso científico e tecnológico – são considerados hoje em dia referências em nível mundial em iniciativas que possibilitam o uso e reúso de dados de pesquisa (DIAS *et al*, 2019). Este panorama internacional tem sido possibilitado pelo interesse destes governos em construir políticas, estratégias e diretrizes que permitam que os dados se tornem acessíveis.

Neste contexto de acesso aberto a dados, embora para a Colômbia trate-se de tema relativamente novo na prática científica nacional, o país não está alheio às mudanças que a ciência está vivenciando. O Estado colombiano reconhece a importância de construir um ecossistema sólido que viabilize o acesso aberto a dados para seu potencial reúso (OBSERVATORIO COLOMBIANOS DE CIENCIA Y TECNOLOGIA, 2017), de modo que possa se inserir na tendência mundial da e-Science e ciência aberta.

Ante este panorama, observa-se na Colômbia a partir de 2017 empreendimentos que promovem e difundem a prática da ciência aberta primordialmente dentro do sistema de ensino superior e nas diversas organizações de pesquisa. No mesmo ano, a Colômbia por meio do *Observatorio Colombiano de Ciencia y Tecnología, OCyT*, procurou diagnosticar as dinâmicas e tendências dos pesquisadores em relação à ciência aberta¹² através de estudo intitulado “*estudio para identificar conocimientos, capacidades, percepciones y experiencias de los investigadores del país frente a la ciencia abierta*”, identificando que os pesquisadores apontam a importância de uma política pública que inclua o tema *open data*; fomenta projetos e ferramentas de ciência aberta; e esteja acompanhada por estratégias que incentivem a participação da comunidade acadêmica.

Como conclusão, o estudo da *OCyT* ressaltou que para a Colômbia gerar um ambiente robusto de ciência aberta, o país deve efetuar a construção de políticas estatais integradas às políticas institucionais de organizações pesquisadoras; bem como investimento em

¹¹ Como exemplo a esta afirmação, a Comissão Europeia desde o 2006 tem desenvolvido políticas e estratégias relacionadas ao acesso aberto. A partir destas, surgiram diversos empreendimentos que criaram um ecossistema de ciência aberta; entre elas se destaca o repositório de acesso aberto para a pesquisa em Europa, *OpenAIREplus*, o projeto *FOSTER+* o qual divulga as políticas europeias de ciência aberta, e desenvolve cursos de formação neste tema, entre outras iniciativas.

¹² O estudo define que a ciência aberta esta composta por diversos elementos que são fundamentais para sua consolidação. Entre os elementos ou pilares que menciona o estudo se encontra: o acesso aberto, dados abertos, investigação aberta, ferramentas tecnológicas abertas, políticas de ciência aberta, e avaliação aberta.

infraestrutura, financiamento, capacitação, avaliação dos fatores que implica está prática, e promoção de uma cultura em prol das novas formas de fazer ciência (*OBSERVATORIO COLOMBIANO DE CIENCIA Y TECNOLOGÍA*, 2017).

A partir das descobertas deste estudo, o *Departamento Administrativo de Ciencia, Tecnología e Innovación* – Colciencias – entidade do governo colombiano responsável pela construção de políticas públicas nesta temática, debateu sobre a necessidade de uma Política Nacional de Ciência Aberta na Colômbia, gerando como resultado a publicação em 2018 do documento “*Lineamientos para una política de ciencia abierta en Colombia*”. O documento propõe diretrizes para orientar a construção de uma política de ciência aberta no país, no qual se afirma que em “relação ao ético e à propriedade intelectual, a divulgação aberta de dados pode requerer ajustes na normatividade existente¹³” (COLCIENCIAS, 2018, p.5, tradução nossa).

Contudo, o governo não tem concretizado uma política nacional integral de ciência aberta, a qual substitua as políticas tradicionais que privilegiam processos científicos fechados. Por trás do desenvolvimento de uma política de ciência aberta, deve existir uma normatividade que sustente tais práticas. Neste aspecto, a Colômbia ainda não tem materializado um marco jurídico robusto que exija dos beneficiários de recursos públicos a disponibilização aberta dos dados procedentes de seus projetos de investigação, de modo que ofereça a segurança jurídica necessária para que os pesquisadores criem o hábito de compartilhamento e uso.

Atualmente os termos de propriedade intelectual entre a entidade pública financiadora, Colciencias, e a entidade executora da pesquisa são acordados a cada caso de acordo com as peculiaridades da pesquisa, portanto, segundo o artigo nº 169 do Plano de Desenvolvimento Nacional 2018-2022, as condições serão estabelecidas através de contrato, no qual o Estado como titular original dos direitos de propriedade intelectual poderá ceder ditos direitos ao executor do projeto de pesquisa (COLOMBIA, 2019). Esta imprevisibilidade da titularidade dos direitos de propriedade intelectual revela uma barreira às práticas de ciência aberta, na medida em que não determinam uma abertura compulsória, e concedem ampla margem de discricionariedade para que se estipule casuisticamente as regras do financiamento, sendo um potencial gerador de conflitos de interesses. Assim, é possível que o Estado e a organização executora na negociação dos termos contratuais estabeleçam restrições de divulgação e compartilhamento dos resultados e dados da pesquisa, afetando a possibilidade de sua reutilização em outras pesquisas públicas ou em propósitos de interesse público e social.

¹³ Texto original: En lo ético y relativo a la propiedad intelectual, la publicación abierta de datos puede requerir ajustes a regulaciones existentes

Países com maior trajetória em ambientes de ciência aberta, como os associados à União Europeia, favorecem o acesso aberto à produção científica através de um marco jurídico¹⁴ que exige seu compartilhamento por parte dos projetos de investigação financiados parcial ou totalmente por recursos públicos, independente do regime de propriedade intelectual que seja aplicado no caso, de modo que o ecossistema científico europeu não seja afetado nas suas práticas cotidianas de disponibilizar e reaproveitar recursos científicos.

Por outra parte, o governo tem realizado esforços, através do “*Premio Nacional de Ciencia Abierta*” outorgado por Colciencias, para divulgar as entidades que tenham incorporado por iniciativa própria o compartilhamento de dados aos seus processos de pesquisa a fim de torná-las referências de boas práticas. Apesar disso, são limitadas as entidades privadas, especialmente organizações de ensino superior, que divulgam a adoção por sua parte de práticas de acesso aberto a dados de pesquisa.

Dentre os poucos casos de sucesso divulgados, está a *Universidad de los Andes*, com o repositório CEDE¹⁵, onde se disponibilizam dados coletados pelo *Centro de Estudios de Desarrollo Económico*, pertencente à Faculdade de Economia da universidade, e outras entidades, tendo como propósito o acesso gratuito e reuso parcial ou total dos dados. Nesse sentido, desconhece-se como em nível nacional as demais universidades têm se desenvolvido perante o contexto científico de acesso aberto a dados, particularmente ante procedimentos como a gestão, compartilhamento e reuso de dados de pesquisa. Igualmente, questiona-se o quanto os pesquisadores têm incorporado tais práticas a sua rotina de pesquisa.

Embora o estudo “conhecimentos, capacidades, percepções e experiências dos pesquisadores referente ao tema da ciência aberta”¹⁶ tenha revelado a dinâmica da ciência aberta (incluindo o *Open Data*) na Colômbia, esta pesquisa adotou uma abordagem generalista¹⁷, sem se aprofundar em aspectos que envolvem a gestão, compartilhamento e reuso de dados de pesquisa. Este estudo, portanto, é considerado pioneiro, e fez surgir os primeiros questionamentos acerca de como os pesquisadores colombianos praticam aspectos de acesso aberto de dados de pesquisa.

¹⁴ Entre as iniciativas mais reconhecidas em relação à construção de um marco jurídico que incentive a ciência aberta se destaca o empreendimento da comissão europeia chamado de *Pilot on Open Research Data in Horizon 2020*, o qual tem com o propósito melhorar **e maximizar o acesso e a reutilização dos dados de investigação gerados pelos projetos do Horizonte 2020**

¹⁵ <https://datoscede.uniandes.edu.co/es/>

¹⁶ <http://repositorio.colciencias.gov.co/bitstream/handle/11146/21720/INFORME%20FINAL.pdf?sequence=1&isAllowed=y>

¹⁷ O foco da pesquisa foi analisar, especificamente, os componentes que integram a ciência aberta.

É assim que de maneira geral, na Colômbia se desconhece o modo como os pesquisadores lidam cotidianamente com as questões de gestão, compartilhamento e reúso de dados de pesquisa. Apesar disso, este desconhecimento não quer dizer que pesquisadores por iniciativa própria não estejam efetuando tais práticas. O estudo da OCyT, ainda que aborde de maneira superficial o *Open Data*, revela que os pesquisadores, têm noções sobre abertura de dados de científicos, consideram relevante uma cultura de *Open Data*, e, além disso verifica-se o uso e compartilhamento de dados pelos pesquisadores (*OBSERVATORIO COLOMBIANO DE CIENCIA Y TECNOLOGÍA*,2017).

Portanto, a ausência de trabalhos científicos ou de estatísticas que permitam desvendar as práticas dos pesquisadores colombianos em relação à gestão, compartilhamento e reúso de dados, é o que torna ainda mais relevante esta pesquisa. Sendo assim, questiona-se:

Como os pesquisadores, pertencentes a grupos de pesquisa classificados pela Colciencias na área de economia e negócios, praticam a gestão, compartilhamento e reúso de dados de pesquisa?

1.2 OBJETIVOS

Delimitada a problematização da pesquisa, estabelecem-se os seguintes objetivos para a investigação:

1.2.1 Objetivo geral

Revelar as práticas relacionadas à gestão, compartilhamento e reúso de dados de pesquisa, de cientistas vinculados a grupos de pesquisa classificados pela Colciencias na área de economia e negócios.

1.2.2 Objetivos Específicos

- a) Caracterizar o perfil dos pesquisadores da área de economia e negócios, localizados na cidade de Bucaramanga e sua região metropolitana;
- b) Descrever as particularidades dos dados produzidos pelos pesquisadores da área de economia e negócios, localizados na cidade de Bucaramanga e sua região metropolitana
- c) Caracterizar as práticas associadas à gestão, compartilhamento e reúso de dados de pesquisa por pesquisadores da área de economia e negócios, localizados na cidade de Bucaramanga e sua região metropolitana.

1.3 RELEVÂNCIA E JUSTIFICATIVA DO ESTUDO

A importância da gestão de dados para seu compartilhamento, acesso e potencial reúso em futuras pesquisas, não é uma questão nova para instituições e pesquisadores com uma trajetória mais longa. Esta conjuntura demonstra que há um consenso por parte das comunidades científicas mais experientes de que todo o conjunto de práticas que viabilizam o acesso a dados, contribui essencialmente ao exercício e progresso da ciência.

Universidades como a *University of Bath*, *University of Edinburgh* e *University of Glasgow* formaram em 2004 através de consórcio um centro de gestão de dados – o *Digital Curation Centre*¹⁸ – cuja missão inicialmente era solucionar desafios de preservação e curadoria digital e que, após se desenvolver através de algumas fases, hoje se consolidou como um centro especializado com atuação ampla em todos os aspectos que envolvem a gestão de dados. Particularmente destaca-se seu propósito de disponibilizar à comunidade acadêmica do Reino

¹⁸ Website do DCC: <http://www.dcc.ac.uk/resources/curation-lifecycle-model>

Unido, e de outros países, ferramentas e recursos necessários para adotar boas práticas de curadoria de dados de pesquisa, especialmente através dos seus programas de treinamento a produtores, usuários e curadores, visando atingir como resultado último uma mudança cultural mais alinhada com a ideia de acesso e disponibilização de dados (*DIGITAL CURATION CENTRE, online*).

Tendo em vista as iniciativas adotadas por parte de universidades internacionais de prestígio, como a mencionada anteriormente, estas impulsionam a construção e consolidação de sistemas *Open Data* em comunidades científicas que ainda estão incorporando tais práticas. Um ecossistema de acesso aberto a dados surge quando há um ambiente propício para tal, o que pressupõe aspectos de ordem jurídica e infraestrutura tecnológica, dentre outros. Já quanto aos pesquisadores, estes promovem a formação de um ecossistema quando possuem conhecimentos e capacidades técnicas que lhes permitem incorporar nos seus hábitos investigativos procedimentos fundamentais como a gestão de dados de pesquisa, de maneira que se garanta o eficiente compartilhamento e potencial reuso dos dados.

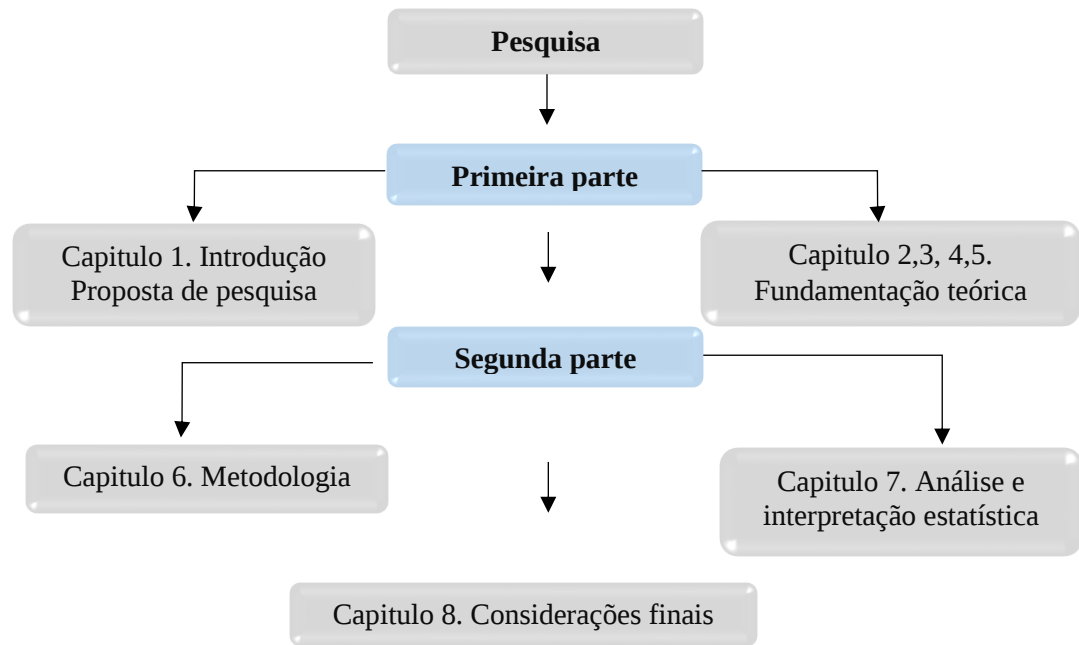
No contexto colombiano, observa-se que na comunidade científica nacional se desconhece se os pesquisadores nos seus processos de pesquisa possuem práticas relacionadas com a gestão, compartilhamento e reuso de dados de pesquisa. A importância de analisar estas práticas¹⁹, radica em que é indispensável identificar se os pesquisadores possuem experiência neste tema, de maneira que se esclareça a situação de acesso aberto a dados de pesquisa na Colômbia. Sobre este aspecto, mostra-se relevante explicitar as dinâmicas dos pesquisadores, tendo em vista que pela ausência de conhecimento sobre elas, e ao não expor como atualmente os pesquisadores se relacionam com seus dados, há uma repercussão negativa na construção de um entorno de acesso aberto a dados de pesquisa.

Neste sentido, identificar tais práticas, mostra-se como altamente pertinente, visto que a partir dos resultados poder-se-ia esclarecer as possíveis próximas fases que garantam a edificação de um ecossistema de acesso aberto a dados.

1.4 ESTRUTURA DO TRABALHO

Esta pesquisa é dividida em duas partes conforme expõe a figura 1. A primeira parte consiste na proposta de pesquisa e no levantamento teórico. A segunda parte apresenta os aspectos metodológicos, a análise e interpretação estatística, e as considerações finais.

Figura 1. Estrutura do trabalho



O primeiro capítulo apresenta o problema de pesquisa, a justificativa, e o objetivo geral e específicos. Este primeiro capítulo permite desvendar a situação objeto de estudo; o contexto em que se desenvolve, os elementos que a rodeiam, e a necessidade de resolver a situação problema por se tratar de um tema de relevância social.

O segundo capítulo apresenta o referencial teórico relacionado ao contexto atual da *e-Science* e Ciência aberta, dados de pesquisa, gestão de dados e seus modelos de ciclo de vida. O levantamento teórico, constrói-se a partir da literatura registrada em livros, dissertações, teses, artigos e *sites* especializados em curadoria e gestão de dados, entre outros. O arcabouço teórico se apresenta como necessário ao ser o referente para a construção do instrumento de coleta de dados e a lente para a análise e interpretação dos resultados.

O terceiro capítulo expõe a metodologia da pesquisa, isto é, o tipo e delineamento da pesquisa; contextualização; características da população a ser estudada; tamanho da amostra; elaboração, validade e fiabilidade do questionário; procedimento para a coleta de dados; procedimento ético, e técnica estatística de análise de dados.

O quarto capítulo expõe a análise e interpretação estatística dos resultados. Apresenta o perfil dos pesquisadores participantes, as particularidades dos dados produzidos pelos pesquisadores, e a descrição das dinâmicas de gestão, compartilhamento e reuso dos dados. Do mesmo modo, comprova-se a hipótese formulada.

O quinto capítulo apresenta as considerações finais referentes às discussões expostas, a contribuição da pesquisa na comunidade estudada, as limitações da pesquisa e as possíveis futuras linhas de estudo

2 E-SCIENCE E CIÊNCIA ABERTA

A maneira como se faz ciência ou como as pesquisas científicas são levadas a cabo esteve em constante evolução ao longo dos séculos, enfrentando processos revolucionários de mudanças de paradigma. Ultrapassados os paradigmas da ciência empírica, teórica e computacional, atualmente a ciência vive seu quarto paradigma, conhecido pelo termo “*e-Science*” e caracterizado pelo uso intensivo de dados em larga escala e de infraestruturas tecnológicas altamente modernas (GRAY, 2007).

Já o movimento pela ciência aberta (*Open Science*), por sua vez, numa perspectiva mais abrangente, busca garantir maior equilíbrio no acesso à ciência (sejam seus produtos, recursos, processos ou ferramentas) pela sociedade, tanto por profissionais quanto por amadores. O acesso livre a publicações, dados, ferramentas, software e hardware permite que os sujeitos colaborem e contribuam no processo científico, constituindo um ecossistema colaborativo necessário ao progresso científico (OLIVEIRA, 2016).

O quarto paradigma do desenvolvimento científico, ou *e-Science*, e o movimento pela ciência aberta, portanto, passaram a se retroalimentar e influenciar reciprocamente, posto que a *e-Science* requer livre acesso e ecossistemas colaborativos para seu desenvolvimento, ao mesmo passo que a ciência aberta requer técnicas e infraestruturas tecnológicas avançadas para tornar possível a manipulação aberta, transparente e acessível de processos, ferramentas e recursos científicos.

2.1 CIÊNCIA ABERTA

O termo “ciência aberta” é objeto de constantes debates no meio científico, de modo que a delimitação do alcance de seu conteúdo não é tarefa simples. No artigo “*Open Science: One Term, Five Schools of Thought*”, os autores Fecher e Friesike (2013) propõem abordar o conceito de ciência aberta além do comumente associado a discussões de produção e disseminação do conhecimento científico. Sob uma revisão literária especializada, Fecher e Friesike (2013) estudam a concepção de ciência aberta a partir de cinco ramos ou escolas de pensamento com diferentes perspectivas:

A escola da infraestrutura, que se preocupa com a arquitetura tecnológica; a escola pública, que se preocupa com o acesso da produção de conhecimento; a escola da medição, preocupada com meios alternativos de medição de impacto; a escola democrática, preocupada

com o acesso ao conhecimento; e a escola pragmática, preocupada com a realização de pesquisas colaborativas.

Quadro 1- Escolas de pensamento – *Open Science*

Escolas de pensamento	Ideias
Escola da Infraestrutura	Esta escola entende que pesquisas eficientes dependem das ferramentas e aplicativos disponíveis, e, portanto, busca criar plataformas, ferramentas e serviços para cientistas que sejam abertamente acessíveis.
Escola Pública	Esta escola entende que a ciência precisa ser acessível ao público, e, portanto, busca promover esse acesso aos cidadãos através de ferramentas e plataformas acessíveis de comunicação.
Escola da Medição	Esta escola entende que as contribuições científicas feitas nos dias de hoje precisam ter seu impacto medido através de mecanismos alternativos, e por isso busca desenvolver um sistema métrico.
Escola Democrática	Esta escola entende que o acesso ao conhecimento é distribuído entre os integrantes da sociedade de maneira desigual, e por isso busca torná-lo livremente acessível para todos.
Escola Pragmática	Esta escola entende que a produção do conhecimento poderia ser mais eficiente se vários cientistas trabalhassem juntos, em cooperação, e, portanto, propõe a abertura do processo de criação de conhecimento.

Fonte: Fecher & Friesike, (2013)

Nessa direção, a ciência aberta se desenvolve a partir da ideia central da Escola pragmática, na medida em que põe no centro das discussões a questão de abrir e disponibilizar o processo de produção do conhecimento científico, ou seja, o processo de pesquisa.

As escolas de pensamento mencionadas por Fecher e Friesike convergem na ideia de que a ciência necessariamente deve ser acessível e distribuída de maneira igualitária, a toda a sociedade. Para que a ideia da ciência aberta se desenvolva, as escolas de pensamento fazem ênfase em que a abertura deve estar sustentada pela infraestrutura tecnológica, de modo que se garanta a abertura do processo de criação de conhecimento.

Munafó *et al.* (2017, p.5) definem ciência aberta como “o processo de tornar o conteúdo e o processo de produção de evidências e reivindicações transparentes e acessíveis a outras pessoas”²⁰. Abrir o processo de pesquisa implica reivindicar os pilares da ciência de ser transparente, reproduzível, verificável e crível. Munafó *et al.* (2017) expressa a necessidade de levar em consideração, portanto, que os cientistas trabalham para acumular conhecimento como um bem comum, e a credibilidade de suas descobertas científicas está enraizada nas evidências que as sustentam, o que inclui a metodologia aplicada, os dados adquiridos, o processo de

²⁰ Texto original: Open science refers to the process of making the content and process of producing evidence and claims transparent and accessible to others

implementação da metodologia, a análise de dados e a interpretação dos resultados. Desse modo, as descobertas se tornam críveis na medida em que são revisadas, criticadas, ampliadas e tem suas evidências reproduzidas pela comunidade. Contudo, sem transparência, a credibilidade das descobertas ficará restrita à confiança ou autoridade do cientista (MUNAFÓ *et al*, 2017). Neste cenário, para que uma descoberta científica adquira credibilidade é necessário que o processo se desenvolva de maneira transparente, o que implica que o processo se de abertamente para garantir sua verificação e reprodução.

O movimento por uma ciência aberta, portanto, visa tornar o processo de investigação transparente e acessível com o propósito de que, mediante o constante questionamento das suas descobertas a ciência se torne mais robusta, confiável e crível.

Para Robert Merton (1973), a ciência aberta esta sustentada nos princípios: a) o Comunalismo o princípio que define o conhecimento científico como um bem público; b) o Universalismo a ideia de que qualquer cientista pode contribuir para o avanço científico e que sua contribuição será valorizada de maneira universal e impessoal; c) o Desinteresse referindo-se a busca pelo progresso científico como algo de interesse coletivo e não particular; e d) o Ceticismo como a necessidade de uma análise cuidadosa dos resultados científicos pela comunidade científica para que possam ser considerados válidos (apud CURTY, 2015).

Adicionalmente, são diversas as vantagens científicas e socioeconômicas resultantes de uma maior abertura do processo de pesquisa, especialmente no acesso aberto a dados de pesquisa. Uhler e Schröder (2007) expressam que ao residir o valor dos dados no seu reuso, a lógica de um regime de pesquisas financiadas com recursos públicos que permite acesso aberto e compartilhamento de dados de pesquisa apresenta muitas vantagens em comparação ao regime fechado, que inviabiliza a reutilização desses dados. Dentre tais vantagens, é possível mencionar:

a) reforço à investigação científica aberta; b) incentivo à diversidade de análises e opiniões; c) promoção de novas pesquisas e novos tipos de pesquisa; d) permite a aplicação online de ferramentas automatizadas de descoberta de conhecimento; e) permite a verificação de resultados anteriores; f) possibilita o teste de hipóteses e métodos de análise novos ou alternativos; g) apoio a estudos sobre métodos e medição de coleta de dados; h) facilidade na formação de novos pesquisadores; i) permite a exploração de tópicos não previstos pelos pesquisadores iniciais; j) permite a criação de novos conjuntos de dados, informações e conhecimento quando dados de várias fontes são combinados; k) ajuda na transferência de informações factuais e promoção da capacitação nos países em desenvolvimento; l) promoção de pesquisas interdisciplinares, intersetoriais, interinstitucionais e internacionais, e ajuda na maximização do potencial de pesquisa de novas tecnologias e redes digitais, proporcionando

assim maiores retornos do investimento público em pesquisa²¹. (UHLIR; SCHRÖDER, 2007, p. 203, tradução nossa).

Cumprе ressaltar que podem haver razões importantes e legítimas para que eventualmente dados de pesquisa produzidos com recursos públicos não se tornem abertamente acessíveis, sendo necessário mantê-los sob restrição, ainda que temporariamente e de maneira justificada. A discussão de tais exceções, embora seja complexa, é de grande relevância no desenvolvimento de regimes de acesso.

Embora a origem das ideias que envolvem o termo ciência aberta remontem a séculos passados, formas concretas de materializá-las apenas vieram a se apresentar nas últimas décadas. De acordo com Willinsky (2005) e Peters & Roberts (2012), conforme citado por Curty (2015, p.2), chegar a essa conclusão é necessário compreender que o termo “ciência aberta” possui um alcance relativamente amplo, genérico, que representa um modelo científico formado por três elementos interdependentes, que seriam: o *Open Access*, *Open Software* e *Open Data*.

Dentre as principais iniciativas dos últimos anos que tem materializado práticas de ciência aberta, encontra-se o movimento *Open Access* com a declaração da *Budapest Open Access Initiative* (BOAI), uma de suas ações mais emblemáticas. A declaração da *Budapest Open Access Initiative* (BOAI), define o conceito *Open Access* nos seguintes termos:

Por “acesso aberto” [à literatura científica com revisão por pares], queremos dizer a sua disponibilidade livre na Internet, permitindo a qualquer utilizador ler, fazer download, copiar, distribuir, imprimir, pesquisar ou referenciar o texto integral desses artigos, recolhe-los para indexação, introduzi-los como dados em software, ou usá-los para outro qualquer fim legal, sem barreiras financeiras, legais ou técnicas que não sejam inseparáveis do próprio acesso à Internet. As únicas restrições de reprodução ou distribuição, e o único papel para o copyright neste domínio, deveria ser dar aos autores controlo sobre a integridade do seu trabalho e direito de ser devidamente reconhecido e citado (BOAI, *online*).

²¹ Texto original: reinforces open scientific inquiry, encourages diversity of analysis and opinion, promotes new research and new types of research, enables the application of automated knowledge discovery tools online, allows the verification of previous results, makes possible the testing of new or alternative hypotheses and methods of analysis, supports studies on data collection methods and measurement, facilitates the education of new researchers, enables the exploration of topics not envisioned by the initial investigators, permits the creation of new data sets, information, and knowledge when data from multiple sources are combined, helps transfer factual information to and promote capacity building in developing countries, promotes interdisciplinary, inter-sectoral, interinstitutional, and international research, and generally helps to maximise the research potential of new digital technologies and networks, thereby providing greater returns from the public investment in research.

Para alcançar seu propósito, identificado como *Open Access*, o BOAI desenvolveu duas estratégias, chamadas de via “verde” e via “dourada”, relacionadas respectivamente com obter acesso aberto à literatura científica mediante repositórios e revistas.

A estratégia via verde, buscava outorgar ferramentas e assistência necessária para que os pesquisadores depositassem seus artigos em repositórios eletrônicos abertos em conformidade com padrões previamente estabelecidos. Já a via dourada, procurava promover uma nova geração de revistas comprometidas com o acesso aberto, de maneira que os usuários tivessem acesso à literatura compartilhada (BOAI, *online*). Estas estratégias incentivadas pela declaração de Budapeste, desencadearam a abertura dos resultados de pesquisa, levando a que a ciência retomasse seu propósito de se tornar universal e democrática.

Embora o movimento *Open Access* tenha seus pilares direcionados à forma como os pesquisadores divulgariam os resultados de suas produções. Outro elemento, contudo, surgiu para tornar esta discussão ainda mais complexa, que é a necessidade de que fossem disponibilizados não apenas os resultados, mas também os dados coletados e utilizados durante o processo investigativo, aspecto este que recebe o nome de movimento *Open Data*.

Cada vez mais os dados de pesquisa passam a ter um papel significativo dentro do cenário científico mundial. O acesso a um artigo e aos dados a ele relacionados torna possível a replicação dessas pesquisas e o desenvolvimento de novos estudos a partir do reúso dos dados de pesquisa. Esse processo de valorização do acesso aberto aos dados de pesquisa surge no longo caminho já pavimentado pelo acesso aberto aos periódicos científicos e neste cenário algumas questões merecem ser discutidas: todos os dados devem ser compartilhados? E os dados sensíveis, que geram patentes? Quem é o responsável pelo depósito dos dados a serem compartilhados e como fazê-lo? (CAVALCANTI; SALES, 2017, p.77)

Segundo Fienberg, Martin & Straf (1985), conforme citado por Curty (2015), o propósito perseguido pelo *Open Data* revela-se fundamental na agenda da ciência aberta, na medida em que promove maior disponibilidade e acesso de dados de pesquisa, potencializando os investimentos, maximizando o valor agregado a estes dados, ultrapassando as limitações de escopo ou de tempo dos pesquisadores originários, e ampliando a transparência dada ao processo de produção de conhecimento.

É sob essa perspectiva que os dados adquirem relevância no contexto da ciência aberta. A disponibilização aberta, gratuita, acessível e transparente destes dados promove um aumento na eficiência das produções científicas, posto que permite a sua reutilização para fins diversos daqueles para os quais foram coletados originariamente, algo que beneficia o reaproveitamento dos recursos públicos, o progresso científico e a sociedade de maneira geral.

Considerando a importância do progresso científico para os desafios globais, a *Organization for Economic Cooperation and Development* - OECD - (2007) desenvolveu treze princípios a partir de diretrizes e padrões internacionais para facilitar o acesso aberto a dados científicos procedentes de pesquisas financiadas por fundos públicos. As recomendações sobre acesso a dados de pesquisa dadas pela OECD são as seguintes:

Quadro 2- Princípios da OECD para acesso a dados de pesquisa

Princípios diretrizes para acesso a dados de pesquisa	
Abertura	Acesso aberto e gratuito em igualdade de condições para a comunidade científica internacional, preferencialmente através da internet.
Flexibilidade	Disponível para atender as mudanças rápidas e imprevisíveis que ocorrem nas tecnologias da informação.
Transparência	Disponibilizar informações sobre os dados de pesquisa, as organizações que os produziram, documentação sobre esses dados e especificação das condições associadas ao seu uso.
Conformidade Legal	Acordos de acesso aberto a dados deverão respeitar a legislação e os direitos das partes envolvidas, incluindo situações nas quais o acesso e uso de certos dados serão limitados, por exemplo, quando envolvam temas de segurança nacional, privacidade e confidencialidade, segredos de mercado e direitos de propriedade intelectual, proteção de espécies raras, ameaçadas ou em perigo, ou assuntos que configuram segredo de justiça.
Proteção da Propriedade Intelectual	Acordos de acesso aberto a dados deverão considerar a aplicação dos direitos de autor e a legislação referente a propriedade intelectual.
Responsabilidade Formal	Acordos de acesso aberto a dados deverão promover práticas institucionais explícitas e formais acerca da responsabilidade das diversas partes envolvidas em atividades relacionadas com os dados.
Profissionalismo	Acordos institucionais para o gerenciamento dos dados de pesquisa deverão estar baseados em padrões e valores profissionais refletidos nos códigos de conduta das comunidades científicas envolvidas.
Interoperabilidade	Permitir e promover o acesso e uso internacional e interdisciplinar dos dados de investigação.
Qualidade	Gerentes de dados e as organizações coletoras de dados de pesquisa devem ter especial atenção no atendimento às regras de qualidade mínima exigidas pelos padrões internacionais.
Segurança	Dedicar atenção especial ao uso de técnicas e instrumentos de segurança e integridade dos dados de pesquisa.
Eficiência	Melhorar a eficiência geral das pesquisas financiadas com recursos públicos para evitar o retrabalho caro e desnecessário dos esforços relacionados à coleta de dados.
Governança	O desempenho dos acordos de acesso aberto a dados deverá estar sob avaliação periódica por representantes das partes envolvidas.
Sustentabilidade	Permanentemente se deverá buscar a sustentabilidade do acesso aberto aos dados de pesquisa financiados com recursos públicos, assumindo a responsabilidade de administrar os meios necessários para garantir este acesso a longo prazo.

Fonte: OECD (2007, p. 15-22)

É assim que, cada um dos atores envolvidos no compartilhamento de dados científicos financiados por fundos públicos, mediante suas responsabilidades, garantem a construção de

um ecossistema de acesso aberto a dados. Embora nos países economicamente desenvolvidos tais práticas sejam mais comuns em certas áreas, muitos institutos e pesquisadores ainda enfrentam muita insegurança neste tema, principalmente nos países menos desenvolvidos, onde contam também com uma infraestrutura técnica e humana menos capacitada para garantir o acesso aberto a dados científicos (UHLIR; SCHRÖDER, 2007).

2.2 E-SCIENCE.

Inicialmente, a ciência abordava os fenômenos a partir do empírico, do experimental, o que posteriormente veio a ser chamado de primeiro paradigma da ciência. Com o avanço dos estudos, os fenômenos passaram a ser analisados mediante modelos teóricos, que constituíram o segundo paradigma. Já nas últimas décadas, a ciência tem se baseado majoritariamente no ramo computacional para o estudo de fenômenos complexos mediante simulações tecnológicas, o que foi concebido como o terceiro paradigma da ciência. Através desta técnica científica de simulação mediada pela computação, somado aos experimentos, tem-se produzido grandes volumes de dados, fato esse que provocou o surgimento do atual quarto paradigma, entendido como uma ciência baseada no uso intensivo de dados, e conhecida como *e-Science* (GREY, 2007; BELL; HEY; SZALAY, 2009).

O termo *e-Science* foi cunhado inicialmente por John Taylor, diretor geral do Conselho de Pesquisa do *Office of Science and Technology* (OST) do Reino Unido em 1999. Taylor designou o termo *e-Science* para descrever a ciência desenvolvida por meio de colaborações em escala global em áreas-chave da ciência, viabilizadas por infraestrutura e recursos computacionais (YANG, WANG, VON LASZEWSKI, 2009).

Por sua vez, Bell (2009), apoiado na concepção de Gray (2007) sobre o conceito de *e-Science*, ciência colaborativa, em rede, orientada por dados, expressa que o quarto paradigma da ciência consiste nas técnicas e tecnologias que possibilitam a pesquisa intensiva com dados em grande escala. Hey & Hey (2006, p. 2, tradução nossa), ratificam este conceito ao defini-lo como uma infraestrutura computacional que tem a “capacidade de acessar, mover, manipular e minerar dados”²², sendo essa estrutura tecnológica necessária para a ciência colaborativa em rede. Por último, Ribes e Lee (2010, p.232, tradução nossa), mencionam três características da *e-Science*: a) “colaboração comunitária e interdisciplinar; b) coleta, representação e análise de dados orientada computacionalmente; e c) integração de ponta a ponta”²³.

²² Texto original: The ability to access, move, manipulate and mine data

²³ Texto original : Three features characterize cyberinfrastructure’s promised transformations in the sciences: community wide and cross-disciplinary collaboration, computationally driven collection, representation and analysis of data, and end-to-end integration

As expectativas relacionadas aos novos métodos de produzir ciência, sustentada no uso e reúso intensivo de dados, na cooperação, na incorporação de ciberinfraestrutura para o suporte dos processos científicos, e principalmente na abordagem multidisciplinar, são percebidas como altamente positivas para o progresso científico, isto devido a que, este entorno possibilita a cada vez mais a integração da ciência, levando ao surgimento de inovações e descobertas nas diversas áreas do conhecimento (OLIVEIRA, 2016).

Neste sentido, a ciência intensiva de dados ou *e-Science* se constitui como uma nova forma de fazer ciência, em que os problemas, cada vez mais complexos, são abordados mediante novas práticas e procedimentos científicos caracterizados por estarem sustentados por infraestrutura tecnológica, além de serem abordados multidisciplinarmente por equipes de pesquisadores que atuam de forma colaborativa.

Albagli, Appel e Maciel (2013) mencionam que apesar do contexto científico ser um entorno aberto e de colaboração, contudo, existe um movimento constante no sentido da privatização de informações e conhecimentos que passam pela tentativa de restringir o acesso ou controlar espaços de interação digital, como é o caso da internet.

Nesse sentido, é importante compreender que embora existam diversas ferramentas da *e-Science* que estimulam e apoiam projetos de acesso aberto, não é possível concluir automaticamente que toda pesquisa dentro do contexto de *e-Science* pode ser qualificada como ciência aberta (ALBAGLI; APPEL; MACIEL, 2013), isto é, como um processo de investigação aberto e acessível para a comunidade acadêmica. Portanto, um contexto de *e-Science* sob a ideia de ciência aberta, envolve o acesso aberto a dados, infraestrutura, software, processos, metodologias, etc., levando a que se de um intercâmbio de recursos científicos.

Desde a perspectiva tecnológica no contexto *e-Science*, Oliveira (2016) ressalta que é a ciberinfraestrutura o fator que suporta a abertura da pesquisa. Para atender a necessidade de lidar com enormes quantidades de dados, faz-se uso de alta tecnologia computacional, tendo como finalidade precisamente fornecer aos pesquisadores o apoio necessário para processar seus dados no processo de pesquisa, bem como para lidar com sua preservação. Assim, entende-se que a transição gradual das práticas científicas para este quarto paradigma passam por três elementos chave:

- a) a incorporação de novas práticas a partir da adoção de um modelo de ciclo de vida de dados que proporcione seu gerenciamento; b) o estabelecimento de uma ciberinfraestrutura sustentável que permita o uso, reúso e compartilhamento dos dados de pesquisa; e c) a consolidação de um ecossistema colaborativo, que fomentem a comunicação e a interação

entre diferentes pesquisadores, profissionais e instituições, tanto nacionais quanto estrangeiras. (OLIVEIRA, 2016, p. 42,43)

Nesse contexto, Bell (2009, p. xiii, tradução nossa) indica três ações ou atividades que são elementares para a *e-Science*, que são a “captura, curadoria e análise dos dados²⁴”. Dentre as práticas atuais incorporadas pela *e-Science* ao processo de pesquisa, encontra-se a curadoria dos dados. A curadoria se refere a um conjunto de práticas de gestão e preservação dos dados durante seu ciclo de vida, tendo como finalidade a manutenção da integridade dos dados no tempo, a fim de poder colocá-los à disposição das gerações futuras (SAYÃO; SALES, 2012). O cenário da *e-Science* possibilita o tratamento pelos quais os dados devem ser submetidos para conservar seu valor e ser potencialmente usados no futuro. A plataforma tecnológica de *e-Science*, os novos procedimentos metodológicos, e os valores promovidos de colaboração e integração da ciência, convergem com o objetivo de concretizar e materializar as ideias promovidas pela ciência aberta.

O desenvolvimento da *e-Science*, portanto, insere-se num cenário de grandes desafios que perpassam questões de diversas ordens, como os aspectos técnicos, tecnológicos, legais, regulatórios, de propriedade intelectual, governança, cultural, de transparência, dentre outros. (ALBAGLI; APPEL; MACIEL, 2013).

A convergência e o contínuo diálogo entre *e-Science* e o movimento de ciência aberta constituem mecanismos que possibilitam a concretização de um ecossistema rico em dados de pesquisa, e que permitem o compartilhamento e reúso deste recurso em diversos contextos acadêmicos, desta forma garantindo o progresso científico.

3 DADOS DE PESQUISA

O atual paradigma da ciência, centrado no uso intensivo de dados, outorgou valor a este recurso científico ao reconhecer seu potencial informacional. Os dados, vistos como o componente central da ciência moderna ou *e-Science*, começaram a ser concebidos por parte da comunidade acadêmica como uma fonte inestimável de recursos informacionais, que devidamente preservados, tratados e gerenciados, poderiam ser reaproveitados em futuras pesquisas científicas (SAYÃO; SALES, 2013, 2015).

A preservação dos dados no atual entorno científico é indispensável visto que estes apresentam grandes benefícios para a ciência se forem reutilizados (*NATIONAL RESEARCH COUNCIL et al.*, 1997; UHLIR; SCHRÖDER, 2007). Dentre os benefícios que os dados

²⁴ Texto original: Data-intensive science consists of three basic activities: capture, curation, and analysis.

devidamente curados, preservados e acessíveis oferecem, identificam-se: a construção, a partir de um conjunto de dados já capturados, de novas perguntas e formas de pesquisa; a minimização de fraude e de má conduta ao ser requerida a citação por parte dos usuários para o reuso dos dados; a aceleração do ritmo da pesquisa por não precisar executar a fase de coleta; e o aproveitamento dos investimentos públicos em pesquisa e educação (DAVID; DEN BESTEN; SCHROEDER, 2010; ESANU;UHLIR, 2004; NIELSEN, 2011; BOULTON *et al.*, 2012; UHLIR; SCHRÖDER, 2007 *apud* BORGMAN, 2015).

No contexto científico, em que os dados são considerados valiosos, aprofundar no estudo do seu significado e características, como sua proveniência, natureza, o propósito da sua geração, dentre outros, concede um panorama que revela a heterogeneidade destes mesmos e que, portanto, a partir das suas propriedades, possibilita desvendar a forma em que devem ser tratados para garantir seu potencial.

Nessa direção, definir o termo de dados de pesquisa se torna um desafio para os acadêmicos, já que o seu conceito vai depender da forma como os dados se produzem, se usam, do seu propósito e da área de conhecimento, e do pesquisador que os criou. Assim, o termo “dados de pesquisa” terá um significado amplo e diversos que se construirá a partir dos próprios processos de investigação científica de cada comunidade acadêmica.

A *Organization for Economic Cooperation and Development* (OECD), no relatório *Principles and Guidelines for Access to Research Data from Public Funding*, conceitua dados de pesquisa como “registros factuais usados como fontes primárias de pesquisa científica e que são comumente aceitos na comunidade científica como necessários para validar resultados da pesquisa”²⁵ (OECD, 2007, p.13, tradução nossa).

Por sua vez, Borgman (2015) confirma que pela complexidade de definir o termo, o contexto científico ainda não chega a um consenso do seu significado. Para Borgman (2015, p. 42, tradução nossa) a definição de dados mais abrangente “é dizer que os dados são representações de observações, objetos ou outras entidades usadas como evidência de fenômenos para fins de pesquisa”²⁶.

Portanto, os dados são concebidos no contexto da pesquisa como aqueles recursos científicos que são gerados ou coletados no processo de investigação e que são necessários para

²⁵ Texto original: “research data” are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings.

²⁶ Texto original: The most inclusive summary is to say that data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research

servirem como base da evidência sobre a qual os pesquisadores vão construir o novo conhecimento.

A Universidade de Melbourne na sua política institucional de gestão de dados e registro de pesquisa para fins de garantir sua preservação no futuro, define os dados de pesquisa como representações de fatos, observações e experiências que sustentaram a teoria, o experimento ou argumento construído. Indica a inerência da heterogeneidade dos dados pela sua diversa natureza numérica, textual, visual ou auditiva; pelo seu estado no processo de investigação, podendo-se encontrar em bruto ou analisados; e assina as diversas formas que os dados podem assumir – documentos, cadernos de laboratórios ou campo, questionários, fotografias, filmes, algoritmos, modelos, equações, entre outros –. Como conclusão, a universidade confirma que os dados em si possuem características variadas e peculiares, e, portanto, devem ser considerados como um conjunto mutável e em constante revalidação de possibilidades emergentes, materialidades e ideias que encontram personificação na arte final (*The University of Melbourne*, 2013, *online*).

Nesse contexto de diversidade e particularidade dos dados de pesquisa produzidos e da atual importância que este recurso científico tem na ciência pela sua capacidade de adquirir valor ao longo do tempo, manter, preservar e adicionar valor aos dados mediante seu tratamento e gestão, ele tem se apresentado no mundo acadêmico como um processo de alta relevância.

Entretanto, para que os dados possam ser tratados de forma eficiente, é necessário esclarecer as peculiaridades que os envolvem e procurar categorizá-los, para decidir o melhor modo de geri-los. Os tipos de dados de pesquisa podem ser categorizados a partir da sua natureza, origem, estado na fase de pesquisa, dentre outras características.

A *US National Science Board* (NSB) categoriza os dados de pesquisa a partir da sua origem e propõe três classes: observacional, computacional e experimental. A NSB (2005 *apud* BORGMAN, 2015, p.38, tradução nossa) classificou os dados científicos a partir do contexto de pesquisa em que foram gerados ou coletados, já que sua proveniência influencia nas “decisões operativas sobre que dados são dignos de preservação e por quanto tempos devem ser curados”²⁷.

Os dados observacionais procedem de anotações, registros de fatos ou de ocorrências de fenômenos. A coleta dos dados observacionais abarca instrumentos diretos ou indiretos e que podem capturar o recurso de maneira manual ou mecanicamente, como por exemplo através

²⁷ Texto original: The origins of data may influence operational decisions on what data are worthy of preservation and for how long they should be curated.

de satélites ou rede de sensores na área das ciências exatas, ou mediante anotações no caderno de campo, entrevistas etnográficas ou de auto relato de empresas na área das ciências sociais e humanas. O estudo dos fenômenos mediante observação está relacionado a espaços e tempos específicos, sendo os dados pouco replicáveis e, portanto, tornando a preservação destes de grande relevância (BORGMAN, 2015).

Os dados computacionais advêm da execução de simulações ou de modelos computacionais. Este tipo de dados comumente se gera nas áreas das ciências exatas, porém também podem ser produzidos nas áreas das ciências sociais e humanas. A preservação deste tipo de dados ao longo prazo não implica um grande esforço devido a que sua natureza permite que sejam reproduzidos ao longo do tempo. No entanto, a replicação da simulação ou de modelo como um todo, no futuro demanda informações sobre os elementos envolvidos no procedimento, dentre deles, o hardware e software utilizados, dados de entrada e de saída, e outras características técnicas (SALES; SAYÃO, 2015; BORGMAN, 2015). Neste sentido, determinar quais dos elementos vão ser preservados estará sujeito ao modelo ou simulação que será reproduzida, bem como a conveniência do pesquisador.

Os dados experimentais procedem de situações em condições controladas que possibilitam testar, estabelecer hipóteses e descobrir novas leis. Este tipo de dado pode ser gerado ou coletado tanto nas áreas das ciências exatas, biológicas, quanto das ciências sociais, humanas, dentre outras. Os experimentos em certos casos podem ser facilmente reproduzidos no futuro, no entanto, como na prática dificilmente as condições experimentais são as mesmas para serem reproduzidas, a preservação destes dados se torna adequada (UK DATA ARCHIVE, 2015 *apud* SAYÃO; SALES, 2015; BORGMAN, 2015).

As diferenças entre as três classes de origem dos dados de pesquisa se fazem evidentes quando se reconhece que suas particularidades demandam distintas estratégias de preservação. Borgman (2015) aponta que existem outros tipos de dados que não se encaixam nas três categorias apresentadas pela NBS – observacional, experimental e computacional –. Segundo ele,

Os registros de qualquer fenômeno estudado podem ser considerados como dados de pesquisa. Esses registros podem incluir documentação de atividades governamentais, comerciais, públicas e privadas, livros e outros textos, matérias de arquivo, documentação na forma de gravações de áudio e vídeo, placas de vidro, papiros, cuneiformes, bambu e assim por diante (BORGMAN, 2015, p.39, tradução nossa)

Sayão e Sales (2015) também mencionam estes tipos de registros e os classificam segundo sua natureza: numérica, imagens, vídeos ou áudio, software, algoritmo, animações,

dentre outros. Neste sentido, identificar a natureza dos dados ganha relevância para conceber a que tipo de tratamentos devem ser submetidos, bem como seu adequado arquivamento para sua preservação ao longo do tempo.

Por último, Green *et al.* (2005 *apud* SAYÃO; SALES, 2015), apresenta a classificação dos dados a partir do seu estado no processo de investigação. O relatório *Data information Specialists Committee- UK* indica que os dados científicos no processo de investigação percorrem diversas fases. Inicialmente são dados brutos, crus ou preliminares oriundos do instrumento científico que os coletou ou gerou; logo, os dados passam pela fase de validação e pode haver uma seleção ou filtração destes; e, por fim, os dados são submetidos a certos procedimentos padrões que os traduzem num esquema mais geral, de maneira que são consolidados e se encontram no seu estado final para serem analisados. Nesta direção, é indispensável identificar o estado dos dados procedentes da investigação, já que o seu arquivamento em centros de dados, repositórios etc., está sujeito ao estado em que estes se encontram, podendo ser dados não processados ou preliminares, dados prontos para serem utilizados por usuários selecionados, dados finais completos, resumo de dados, dados tabulares, ou dados derivados.

Por outra parte, Sayão e Sales (2015) mencionam os tipos de dados que têm sido produzidos por entidades governamentais sem ser coletados para fins de pesquisa, contudo possíveis de se tornarem dados de pesquisa, uma vez que sejam processados para tais fins. Estes dados produzidos sem fins investigativos são do mesmo modo considerados valiosos para o contexto acadêmico, pelo potencial de reutilização para a abertura de novos processos investigativos em que se questionam novas perguntas e surgem novos conhecimentos.

Caracterizar os dados de pesquisa é, portanto, umas das atividades iniciais que devem ser desenvolvidas pelo pesquisador quando se trata da preservação destes. No contexto da ciência em que o centro são os dados e se enfatiza a abertura deste recurso para seu compartilhamento e reutilização, estratégias de preservação eficientes são indispensáveis para contrariar sua fragilidade e perda no longo prazo.

Apesar das instituições, financiadoras de pesquisa e a comunidade acadêmica em geral reconhecerem o potencial informacional dos dados de pesquisa, por serem fontes inesgotáveis, há uma certa dificuldade para determinar o quão valioso os conjuntos de dados podem ser, bem como as maneiras pelas quais eles podem se tornar valiosos (BORGMAN, 2015).

Somado a esta dificuldade, os pesquisadores encaram outros desafios que surgem ao tratar os dados. Essas tensões em relação a eles advêm das inquietudes sobre temas de propriedade intelectual, políticas, tecnologias, práticas de tratamento e questões de acesso,

especialmente porque ao ser compartilhados, os dados devem conservar o contexto em que foram criados.

Abrir os dados de pesquisa não é uma questão simples pela sua própria natureza. O *Open Knowledge Foundation (online)* define o conceito de *Open Data* ou dados abertos, descrevendo aberto como aquilo que “qualquer pessoa pode acessar, usar, modificar e compartilhar livremente para qualquer finalidade”²⁸, na medida em que se preserve a proveniência. Dados abertos, assim, são “dados ou conteúdos abertos que alguém livremente pode usar, modificar, redistribuir para diversos fins”²⁹ (*OPEN KNOWLEDGE FOUNDATION*, online).

O propósito do *Open Data* é em si o compartilhamento e reuso dos dados abertos e, para que isso aconteça, deve-se garantir que os dados sejam de qualidade, encontráveis, acessíveis, interoperáveis, reusáveis (FAIR). Além dos aspectos técnicos relacionados ao acesso dos dados de pesquisa, o ecossistema de acesso aberto a dados está suportado por aspectos como infraestrutura tecnológica, normatividade, política, ética e direitos autorais.

Dentre os aspectos controversos que geram questionamentos aos pesquisadores no panorama do acesso aberto a dados, encontra-se a propriedade intelectual, já que o reuso deste recurso científico está condicionado à normatividade que o rege. Determinar quem possui a autoria dos dados, que tipo de dados serão compartilhados, as condições de acesso destes, dentre outras questões legais e éticas, é considerado fundamental para nortear as práticas científicas abertas (BORGMAN, 2015; OLIVEIRA, 2016).

Nessa direção, Oliveira (2016, p.53) expressa que o acesso aos dados científicos³⁰ fundamentalmente deve estar “conduzida por princípios políticos e éticos, regidos por instrumentos regulatórios, normativos e códigos de conduta que assegurem a fidedignidade, sigilosidade, veracidade e anonimização”. Isto se apresenta porque a abertura de dados implica que este recurso científico seja reusado livre e gratuitamente, sob a condição de outorgar ao autor o devido reconhecimento e de reaproveitar os dados de acordo com os preceitos legais.

A *e-Science* é considerada como a infraestrutura tecnológica que possibilita que os pesquisadores possam acessar dados de pesquisa. A ciberinfraestrutura a serviço da ciência, criou ferramentas capazes de possibilitar não só a produção, análise e armazenamento de dados de pesquisa, mas também permitiu o compartilhamento e reutilização deste recurso. A *e-Science*, pela sua natureza, reivindica a produção científica a partir da cooperação, colaboração

²⁸ Texto original: “Open means anyone can freely access, use, modify, and share for any purpose.”

²⁹ Texto original: “Open data and content can be freely used, modified, and shared by anyone for any purpose”

³⁰ Sinônimo de dados de pesquisa

e a interdisciplinaridade, abordando problemas e fenômenos altamente complexos, e, portanto, contribuindo para o avanço da ciência (MEDEIROS; CAREGNATO, 2012).

O valor dos dados no contexto acadêmico se dá porque estes “possuem um tempo de vida maior que os projetos de pesquisa que os criaram” (SAYÃO; SALES, 2015). Esta característica dos dados permite que sejam reaproveitados em um ambiente científico aberto que promove o compartilhamento e reúso deste recurso, de forma que as comunidades acadêmicas se comprometam a tratar seus dados de pesquisa para torná-los acessíveis e potencializar seu valor. Além dos aspectos tecnológicos, políticos, normativos e éticos que viabilizam e influenciam no acesso aos dados de pesquisa, há outros aspectos técnicos que devem ser abordados para garantir sua abertura.

Os dados de pesquisa são a base do conhecimento e para serem reutilizados precisam ser estruturados, organizados, descritos e disponibilizados de forma que sejam compreendidos no tempo atual e futuramente. Para assegurar o acesso, compartilhamento e reúso de dados, estes devem ser submetidos a certas atividades conhecidas como gestão do ciclo de vida – planejamento, coleta, criação de metadados, preservação em repositórios, integração com outros dados e análise – de modo que se outorgue valor e qualidade ao dado científico. Garantir o acesso integral aos dados de pesquisa implica tratá-los, curá-los, e preservá-los mediante seu gerenciamento (DUDZIAK, *online*). Assim sendo, o tópico seguinte passará a detalhar o conceito e as práticas de gestão de dados.

4 GESTÃO DE DADOS DE PESQUISA

A sociedade contemporânea tem incorporado como meio de apoio em todas as suas esferas o uso constante de tecnologia, fazendo com que, no exercício do seu uso, produzam-se e consumam grandes quantidades de dados com potencial informacional. No contexto científica, suas transformações têm sido mais incisivas devido à incorporação de infraestruturas tecnológicas que abarcam todo o processo de investigação, como a criação de instrumentos tecnológicos automatizados para a coleta de dados, a construção da realidade por meio de simulações para estudar fenômenos complexos, a criação e captura de grandes quantidades de dados pelas condições em que estes são produzidos, o compartilhamento destes dados para seu reaproveitamento e a necessidade de um ambiente cooperativo (SAYÃO; SALES, 2016).

Segundo Gray (2007) a atual ciência, mais conhecida como o quarto paradigma da ciência ou *e-Science*, caracteriza-se em si pela produção e uso intensivo de dados. A ciência começa a considerar que este recurso possui um potencial informacional que pode ser reaproveitado para diversos fins, passando assim a ser percebido como um produto valioso

procedente do processo de pesquisa (SAYÃO; SALES, 2015). Somado a isso, os pressupostos da ciência aberta que consideram a ciência como um bem universal que deve estar acessível a toda a sociedade, promove e exige a abertura de todo o processo de investigação – método, instrumentos, dados, softwares etc.– de forma a 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).

Nesse contexto, o movimento de acesso aberto a dados de pesquisa ganha força na medida que a comunidade acadêmica vai reconhecendo o valor intrínseco que este recurso possui. Segundo Spinak (2015), conforme citado por Cavalcanti e Sales (2017) o movimento de dados abertos recomenda a construção de diretrizes direcionadas ao tratamento e preservação dos dados procedentes de pesquisa; procura boas práticas de preservação e gestão de dados para expandi-las nas comunidades científicas; busca formas eficientes para visibilizar a produção científica; e incentiva o compartilhamento e reúso dos dados para reaproveitar o investimento financeiro da pesquisa.

O acesso aos dados de pesquisa vem-se consolidando graças aos movimentos sociais, às plataformas de comunicação e informação, e às transformações no fazer científico, dentre outros que se focaram em fazer possível a disponibilização dos dados para reaproveitar seu valor. Neste sentido, “um dos maiores desafios que a pesquisa científica apresenta atualmente para criar conhecimento não é mais a ausência de dados e informação, é ao contrário, o excesso deles” (SAYÃO; SALES, 2013, p. 24).

No processo de pesquisa, a produção ou coleta de dados de pesquisa vai depender dos diversos propósitos para sua criação, da natureza da pesquisa, da área do conhecimento em que é produzido, da perspectiva em que as comunidades acadêmicas percebem este recurso, entre outros aspectos, que concedem aos dados essas características particulares, fazendo que estes sejam concebidos como diversos e heterogêneos no entorno científico. Diante do excesso de dados heterogêneos e o valor que este recurso oferece, na medida em que são compartilhados e reusados, a formulação de estratégias para submeter os dados a processos de preservação e curadoria se torna uma questão central no âmago da ciência.

A preservação dos dados de pesquisa está enraizada na ciência contemporânea ao reconhecer que estes podem ser compartilhados e reusados. No contexto da *e-Science* e ciência aberta, um dos seus propósitos no referente aos dados, é garantir seu compartilhamento e reúso. Para isso, é necessário mediante o tratamento de dados, adicionar certas descrições aos dados sobre seu contexto, tempo e espaço em que foram criados e produzidos, além de outras informações técnicas indispensáveis.

Este objetivo da atual ciência de disponibilizar os dados de pesquisa abertamente se explica pelo fato de que, ao reutilizá-los, se está dinamizando novas pesquisas ao eliminar a fase que já foi realizada por outros pesquisadores. Do mesmo modo, o “impacto também ocorre na economia de recursos com a diminuição de pesquisas duplicadas e com o direcionamento destes recursos para novas pesquisas” (CAVALCANTI; SALES, 2017).

Benefícios pelo compartilhamento de dados de pesquisa devem ser estendidos à sociedade. Dentro do contexto científico e social, o compartilhamento de dados de pesquisa para propósito de reúso, atinge e concede benefícios tanto para agências de financiamento de pesquisa, comunidades científicas, pesquisadores, quanto ao público em geral.

Quadro 3 – O valor do compartilhamento de dados de pesquisa

Instituições	Benefícios
Agência financiadora de pesquisa	Valor aprimorado dos investimentos, maximizando o retorno dos investimentos em pesquisa.
Comunidade científica	Melhor capacidade de construir uma nova pesquisa sobre outra já avançada ao invés de duplicar a pesquisa. Capacidade de realizar meta-análises sobre o conjunto de dados. Maior transparência, reprodutibilidade e comparabilidade dos resultados. Capacidade de expandir a avaliação da metodologia, recomendações e melhorias. Melhor educação para novos pesquisadores sobre descobertas mais atuais e significativas.
Cientista	Os benefícios que os pesquisadores recebem por compartilhar seus dados de pesquisa: Reconhecimento da agência financiadora da pesquisa como uma fonte autorizada e verificável de conhecimento e bom investimento. Dados de qualidade ao serem gerenciados antes do seu compartilhamento concedendo o valor do reúso em outros contextos. Maior oportunidade para troca de dados. Estabelecimento de conexões de rede científica e possíveis colaborações.
Público	Um melhor acesso à informação leva a uma melhor compreensão e contribuições para a tomada de decisões públicas e pessoais.

Fonte: DataONE Education Module, 2016.

Comumente, ao finalizar os processos de investigação, os pesquisadores armazenam os dados em computadores ou mídias pessoais, podendo levar à perda dos dados devido à obsolescência tecnológica das mídias em que são registrados, somado à falta de intenção de tratá-los para preservá-los adequadamente (SAYÃO; SALES, 2012). Para que esse cenário não aconteça, os pesquisadores necessariamente devem incorporar aos seus processos científicos atividades que possibilitam salvar os dados, bem como potencializar os benefícios que intrinsecamente eles possuem.

É nesse entorno que surge a concepção da curadoria de dados, cujo principal desafio recai na necessidade de se preservar o conjunto de dados e a capacidade que ele possui de

transmitir conhecimento para uso futuro das comunidades interessadas (SAYÃO; SALES, 2012).

Atualmente a ciência baseada no uso intensivo de dados, centra-se em três atividades essenciais: captura, curadoria e análise (GRAY, 2007). Tendo em vista que a ciência moderna concede aos dados científicos uma relevância maior no transcorrer do processo de pesquisa, o desafio que se apresenta para a manipulação dos dados é procurar mecanismos e técnicas de gestão que possibilitem a preservação de forma adequada. Este tipo de mecanismo para Gray (2007) é chamado de curadoria digital e abrange uma série de atividades, como por exemplo, a contextualização (metadados) do conjunto de dados, e designação de uma localização única, entre outros elementos.

Sayão e Sales (2015, p.80) definem a curadoria de dados como “ações voltadas para o gerenciamento de dados de pesquisa durante seu ciclo de vida, envolvendo manter, preservar e adicionar valor aos dados”. Neste sentido, o conjunto de práticas direcionadas para o tratamento dos dados de pesquisa é conhecido por gestão de dados e se refere a um processo que implicitamente se encontra dentro do conceito de curadoria digital. A gestão de dados é, portanto, o processo operacional que envolve um conjunto de ações gerenciais e técnicas que procuram tratar os dados para manter essa capacidade, ao longo do tempo, de transmitir informação e conhecimento.

A gestão de dados está constituída por diversas etapas que formam um ciclo complexo, o qual inicia com o planejamento de como os dados serão gerenciados, e encerra com a disponibilização dos dados já tratados e preparados para serem reusados. Neste contexto, a gestão de dados procura fazer com que este recurso esteja organizado, estruturado, documentado, preservado, acessível e validado. Uma vez que a gestão de dados se desenvolve de maneira eficiente, os resultados obtidos deste processo são: a) dados de alta qualidade fáceis de compartilhar e reusar; b) retorno de investimento c) credibilidade e reconhecimento por meio de citações, ao pesquisador que compartilhou seus dados para serem usados (DATAONE, 2016).

Na perspectiva do pesquisador, a importância de gerenciar os dados radica em que este processo permite que seus dados se mantenham organizados e adequadamente arquivados; possibilita ter melhor controle das versões do seu conjunto de dados e da sua qualidade; realiza backups para evitar sua perda; documenta seus dados para serem compreendidos em outros contextos e por outros pesquisadores; e os submete a certos procedimentos para poderlos compartilhar e obter credibilidade e reconhecimentos por seus descobrimentos científicos (DATAONE, 2016).

Já desde uma perspectiva institucional, a gestão de dados é cada vez mais demandada pelas agências, instituições públicas que estimulam e financiam pesquisas (SAYÃO; SALES, 2013). Atualmente há certas exigências, por parte de entidades públicas, de submeter aos dados a atividades de preservação para fins de arquivá-los em repositórios abertos, especialmente quando procedem de pesquisas financiadas com fundos públicos (SAYÃO; SALES, 2012, 2015; CAVALCANTI; SALES, 2017). Esta demanda das instituições, especialmente públicas, procura “assegurar o compartilhamento e reúso dos dados; aumentar a visibilidade da pesquisa; promover um ambiente de colaboração; maximizar a transparência do processo de pesquisa; reduzir os custos e evitar a duplicação na coleta de dados” (OLIVEIRA, 2016, p.64).

As instituições financiadoras atualmente estão delineando políticas, diretrizes e estratégias que pautam as práticas de gerenciamento de dados de forma que os pesquisadores as incorporem nos seus processos de investigação (SAYÃO; SALES, 2013). Nessa direção, um aspecto determinante para consolidar a cultura de compartilhamento de dados nas comunidades acadêmicas é a adoção das práticas de gestão de dados nas pesquisas, de modo que se torne natural.

As universidades também têm tido um papel indispensável no que se refere a salvaguardar os dados de pesquisa e a disseminar boas práticas de gestão de dados, conseguindo que outras comunidades acadêmicas consigam se desenvolver nesta questão. Universidades prestigiosas como *University of Edinburgh*, *University of Bath*, e *University of Glasgow* criaram centros que oferecem modelos de ciclo de vida de dados, os quais “propiciam uma visão gráfica dos estágios exigidos para que os dados sejam preservados e cuidados adequadamente” (SEGUNDO; ARAÚJO, p.135). Dentre os centros de curadoria digital criados pelas universidades, encontra-se o *Digital Curation Centre* (DCC). Este centro propõe um modelo de ciclo de vida de dados que atenda às necessidades que as universidades identificam.

O conjunto de práticas propostas pelo DCC (*online*), inicia com o planejamento da coleta dos dados e suas possíveis opções de armazenamento; a captura dos dados e atribuição de metadados de tipo administrativo, descritivo, estrutural e técnico; a avaliação e seleção dos dados que vão ser curados e preservados; o armazenamento em repositórios digitais confiáveis sob orientações políticas e de requisitos legais; verificação do acesso dos dados para seu reúso por parte dos usuários; e avaliação constante do estado do formato técnico em que foi arquivado o conjunto de dados.

O processo de gestão de dados abrange não só as práticas técnicas de preservação mencionadas anteriormente, mas também questões de infraestrutura tecnológica, investimentos econômicos, políticas, questões éticas e capacitações de profissionais no tema referente

(SAYÃO; SALES, 2014). Como guarda-chuva do conjunto de práticas de gestão de dados, e pelas quais devem estar necessariamente ancoradas, apresentam-se as políticas e estratégias que direcionam todo o procedimento de gestão deste recurso.

Essas políticas são projetadas com o propósito de orientar o fluxo de trabalho de gestão, desde o planejamento com a constituição de um “Plano de Gestão de Dados de Pesquisa” onde se definem os formatos de arquivamento, preservação, segurança, licenças, direito de propriedade intelectual, metadados, até questões éticas como o compartilhamento de dados sensíveis ou de interesse comercial, de forma que se assegure e garanta a essência da gestão de dados, o compartilhamento e reúso deste recurso (SAYÃO; SALES, 2016).

Os aspectos éticos, de licenças e de direito de propriedade intelectual, tornam-se questões prioritárias já que, ao estabelecer as pautas sobre estes temas, define-se o que é indicado disponibilizar para toda a comunidade acadêmica e o que não é recomendável ou necessário disponibilizar, bem como quem possuirá a autoria dos dados. Eticamente, determinados dados de pesquisa devem contar com restrições para serem compartilhados devido à sua característica sensível. Um dos exemplos de dados especialmente sensíveis pertencem às áreas que tratam com o ser humano, como é o caso da área de saúde que estudam pacientes em condições de vulnerabilidade, ou áreas como a biologia que envolvem estudos com espécimes ou habitats que enfrentam elevado risco de extinção (SAYÃO; SALES, 2015, 2016).

Do mesmo modo, o direito de propriedade intelectual define quem vai obter os direitos autorais dos dados de pesquisa, para manipulá-los e abri-los à sociedade. Este instrumento de autoria não se apresenta mais de forma tradicional. O atual contexto de ciência aberta e *e-Science* alterou os tradicionais mecanismos autorais e os reconfigurou para adaptar-se às novas condições científicas de colaboração e compartilhamento sem deixar de garantir o direito de autoria aos pesquisadores (OLIVEIRA, 2016). Nesta direção, as novas propostas de licenciamentos autorais, como a licença *creative commons*, apresentam-se como meios para reafirmar o propósito da abertura de dados de pesquisa permitindo a distribuição e reúso gratuito e garantindo o devido reconhecimento autoral.

Já desde uma perspectiva operacional, as práticas de gestão de dados estão suportadas por infraestruturas tecnológicas que são construídas para fins de gestão, desde o início com a coleta de dados, sua manipulação, avaliação, seleção até os formatos de arquivamento e repositórios ou centros adequados para sua preservação.

Os repositórios digitais no contexto da gestão de dados se apresentam como ferramentas indispensáveis, especialmente em dois momentos. Um deles se dá no sentido de que estes

centros digitais possibilitam o acesso aos dados de pesquisa em escala global a partir da integração e interoperabilidade entre repositórios. O segundo momento se dá com a padronização das ações de preservação e curadoria para possibilitar que os dados sejam compreendidos, interpretados, e reusados em qualquer contexto científico (OLIVEIRA, 2016; SAYÃO; SALES, 2014).

A interoperabilidade técnica, semântica, política e legal entre os sistemas de troca de dados ou entre repositórios digitais se tornam essenciais para a construção de ambientes globais, cooperativos e abertos. Isto significa que normas, padrões, protocolos, técnicas e diretrizes devem, necessariamente, ser homogêneas para sustentar o acesso e reuso dos dados (OLIVEIRA, 2016)

Os repositórios e centros digitais de dados, além de oferecer um serviço de armazenamento e preservação de dados, também possuem um propósito educativo. Neste sentido, além de apresentar na comunidade acadêmica mecanismos eficientes para a gestão de dados, estes incentivam uma formação profissional nesta questão.

Esta finalidade educacional, como a percebida nos centros digitais *DataONE* (*Data Observation Network for Earth*) e *DCC* (*Digital Curation Centre*), é de grande relevância na medida em que há uma necessidade por parte dos pesquisadores de executar as fases de gestão de dados durante o processo de pesquisa. Como mencionado anteriormente, vários repositórios têm proposto modelos de ciclo de vida de dados que padronizam e contemplam as diversas fases que envolvem a gestão de dados. A seguinte seção se aprofundará nesses modelos.

5 MODELOS DE CICLO DE VIDA DOS DADOS

A multiplicação na produção de dados resultante da fase intensivamente tecnológica vivenciada pela ciência despertou a necessidade de que fosse aplicado a estes objetos digitais o conceito de curadoria, historicamente utilizado em museus e coleções de arte. O significado do termo curadoria se associa às funções de cuidar, manter e preservar, o qual aplicado ao contexto dos dados de pesquisa, assume um papel fundamental ao permitir o reuso deste recurso (ALBUQUERQUE, 2018).

A possibilidade de reutilização futura dos dados representa a principal maneira de dar-lhes valor (*NATIONAL RESEARCH COUNCIL et al.*, 1997; UHLIR; SCHRÖDER, 2007). A curadoria digital de dados, sendo um conjunto de práticas de gestão e preservação, assegura a integridade deste recurso ao longo do tempo e garante seu potencial reuso em diversos espaços e tempos. O ciclo de vida de um dado no processo de investigação, coincide, portanto, com as fases de curadoria que lhe será aplicado, sendo de fundamental importância identificar cada

uma dessas etapas e as tarefas de gestão correspondentes, responsáveis por assegurar a sustentabilidade deste recurso.

Neste sentido, foram desenvolvidos diversos modelos de ciclo de vida de dados, cada um pensado para atender aspectos peculiares de determinada área do conhecimento, ou a depender das necessidades das instituições. Para ilustrar o formato destes modelos, ressaltando a importância da identificação das atividades de gestão e dos sujeitos responsáveis por cada uma delas, apresentaremos a seguir, de maneira objetiva, as principais características de quatro modelos que são frequentemente mencionados pela literatura.

5.1.1 Data Life Cycle Model – DataONE

O *Data Observation Network for Earth (DataONE)* é uma entidade norte-americana mantida pela U.S. National Science Foundation (NSF), responsável por uma ciberinfraestrutura sustentável capaz de atender às necessidades da ciência e da sociedade em relação ao acesso aberto, constante, robusto e seguro, a dados provenientes de observação do planeta terra que estejam bem descritos e facilmente descobertos (DATAONE, *online*).

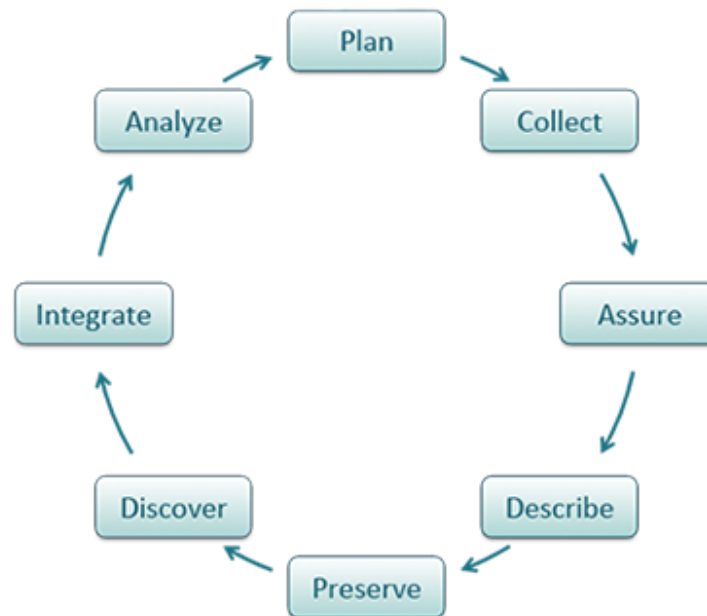
Nesse sentido, *DataONE* se apresenta como uma proposta de fazer ciência de maneira inovadora, a qual garantirá a preservação, acesso, uso e reúso de dados de pesquisa a nível multiescala e multidisciplinar, mediante aspectos oferecidos pelo observatório que devem suportar e estar presentes no processo de tratamento dos dados: ciberinfraestrutura e programa de capacitação e divulgação.

O observatório *DataONE*³¹ propõe um modelo de ciclo de vida de dados que fornece uma visão geral das etapas envolvidas no processo de gestão e preservação de dados. O propósito do modelo *DataONE* é viabilizar o acesso, uso e reúso de dados de pesquisa, além de ter uma finalidade educativa por ser uma estrutura de referência para o desenvolvimento de ferramentas, serviços e materiais educacionais que orientarão as comunidades acadêmicas a incorporar tais práticas nos seus processos de pesquisa.

Algumas atividades de pesquisa podem usar apenas parte do ciclo de vida indicado no modelo (Figura 2), outras podem não seguir o caminho linear apresentado, outras ainda, necessitam repetir o ciclo diversas vezes. Além disso, existem casos em que pesquisadores criam novos dados no processo de descoberta, integração, análise e síntese de dados existentes. O modelo, dessa forma, adota um formato circular com oito componentes, como se apresenta a seguir:

³¹ <https://www.dataone.org/data-life-cycle>

Figura 2 – Modelo de Ciclo de Vida do *DataONE*



Fonte: *The DataONE data life cycle (ONLINE)*

A primeira etapa, planejar (*plan*), consiste na criação do Plano de Gerenciamento de Dados (DMP), o qual é um documento formal que procura garantir que os dados procedentes de pesquisas estejam seguros para serem usados no futuro. *DataONE* (2016) propõe cinco componentes que são imprescindíveis para estarem no plano de gestão. O primeiro componente está relacionado com informações sobre os dados, como eles serão produzidos, coletados, processados, em que formato serão armazenados e qual procedimento será aplicado para garantir, através do controle, a qualidade dos dados.

O segundo componente trata da documentação sobre os dados, chamados de metadados, que consiste em contextualizar técnica e semanticamente os dados de forma que estes sejam compreensíveis no futuro, independente do espaço e tempo. Já o terceiro componente se refere às políticas para acesso, compartilhamento e reutilização de dados, o que implica obrigação ética, questões de privacidade, licenciamento e direitos autorais.

O quarto componente a ser definido é o armazenamento e gestão ao longo prazo, onde o pesquisador deve se perguntar quais dados serão preservados, onde serão preservados, quais formatos serão utilizados, e quem será o responsável de preservá-los no decorrer do tempo. Finalmente, o último componente se concentra da disponibilização de recursos econômicos para executar as atividades relacionadas ao gerenciamento de dados (*DATAONE*,2016).

Além da descrição do processo de gestão de dados, também se define os atores responsáveis de cada atividade que faz parte da gestão, desde quem coletará, analisará e manipulará os dados, até questões mais focadas na curadoria deles. Nessa direção, o plano de gestão de dados se apresenta como o documento que norteará o gerenciamento e compartilhamento dos dados de pesquisa, assegurando a eficiência na execução e evitando a perda de recursos e o retrabalho.

A segunda etapa, coleta (*collect*), refere-se às práticas padrões que *DATAONE* recomenda para capturar, manipular e arquivar os dados. Padronizar a manipulação dos dados, especialmente na coleta e arquivamento destes, é uma questão relevante a ser estabelecida, visto que as ferramentas usadas na entrada e validação de dados devem ser consistentes para evitarem erros no processo que afetem a leitura ou perda dos dados.

As recomendações que a *DATAONE* (2016) realiza se concentram na manipulação eficiente do arquivo que contém os dados. Por exemplo, no caso da utilização de planilhas do Excel como ferramenta de entrada de dados, estas devem estar adequadamente formatadas, as colunas nomeadas corretamente, evitando misturar diferentes tipos de dados em uma única coluna; os dados ainda devem ser mantidos brutos e salvaguardados por *backups*, dentre outros aspectos.

A terceira etapa, assegurar (*assure*), trata de garantir a qualidade dos dados através de verificação, inspeção e controle. Nessas atividades, segurança e controle de qualidade, busca-se “impedir que erros entrem ou permaneçam em um conjunto de dados”³². O controle de qualidade se apresenta tanto antes dos dados serem coletados, manipulados e analisados, quanto no decorrer do seu ciclo de vida.

O Observatório *DataONE* também recomenda evitar certos erros na manipulação dos dados, tais como imprecisão dos dados inseridos, dados incorretos, instrumentos de coleta com defeitos, dados sem metadados, dados contaminados, dentre outros. As práticas sugeridas para o controle de qualidade na coleta de dados, consistem em evitar a duplicação de dados na sua entrada, em projetar bem o armazenamento dos dados, utilizar terminologia padrão e consistente para uma adequada leitura, e documentar alterações realizadas nos dados para evitar a dupla verificação (*DATAONE*, 2016).

A etapa descrever (*describe*) é uma das atividades mais relevantes do ciclo de vida de dados porque possibilita que o conjunto de dados compartilhados seja confiável, contribui ao

³² Texto original: Activities that prevent errors from entering or remaining in a data set.

seu fácil acesso e à sua adequada avaliação e controle, e fornece as diretrizes sobre como adquirir e usar os dados.

Essa fase centra-se na descrição precisa e completa dos dados, a partir de padrões de metadados já estabelecidos. Esses padrões de metadados se referem ao uso de termos, definições e estruturas comuns que possibilitam que os dados sejam facilmente interpretáveis no futuro. O padrão de metadados a ser aplicado vai depender das características e peculiaridades do conjunto de dados a descrever, bem como dos aspectos organizativos, técnicos, normativos e de políticas associadas ao modelo utilizado.

DataONE (2016) aconselha aos pesquisadores descrever como os dados foram criados ou coletados, explicar o seu significado, como foram processados e manipulados, que tipo de ferramentas foram utilizadas no seu procedimento, quais lacunas e limitações eles possuem, dentre outros aspectos técnicos e semânticos que permitem a compreensão do conjunto de dados.

Na etapa preservar (*preserve*), executa-se duas ações principais, *backups* e arquivamento, e outras ações necessárias para preservar os dados, como sua reformatação e conversão que evitam, sobretudo, a obsolescência digital. Os *backups*, conforme Araújo e Segundo (2019), referem-se a cópias de segurança periódicas indispensáveis para proteger a integridade dos dados. Esse procedimento tem como propósito evitar o desperdício de tempo e dinheiro; reduzir ou impedir a perda de dados; resguardar-se em caso de acidentes inesperados; reduzir riscos, entre outros (DATAONE, 2016).

A próxima etapa do ciclo é a descoberta (*discover*). Ela envolve tanto a localização de outros dados que podem agregar valor ao se associarem aos dados atuais (WIGGINS, 2013), como a disponibilização dos mesmos. Para isso, Wiggins *et al.* (2013, p. 3, tradução nossa) cita algumas estratégias possíveis de serem utilizadas, como por exemplo, "registrar o projeto em um site de diretório do projeto, depositar dados em um repositório aberto e adicionar descrições de dados às câmaras de compensação de metadados"³³. Allard (2012) aponta ainda para a importância dos profissionais de biblioteconomia neste estágio, ao sugerir que estes podem auxiliar no acesso e identificação dos dados úteis ao pesquisador.

Já a etapa de integração (*integrate*), consiste em agrupar em um mesmo conjunto, dados coletados em fontes diversas, que podem servir tanto de sustentação para a análise, como de complemento para a interpretação e verificação dos dados disponíveis, além de ser uma etapa

³³ Texto original: Strategies to help ensure the data have maximum impact include registering the project on a project directory site, depositing data in an open repository, and adding data descriptions to metadata clearinghouses

fundamental nos casos em que se investigam problemas complexos (WIGGINS, *et al.*, 2013). De acordo os autores, para que esta combinação seja bem-sucedida, é importante que os dados tenham sido documentados e descritos adequadamente, desse modo, facilitando o seu acesso e reutilização por outros usuários (WIGGINS, *et al.*, 2013).

A análise (*analyze*) dos dados é a última fase do ciclo e é essencial que esteja alinhada com os objetivos da pesquisa ou projeto e com a necessidade de dados para atingir os mesmos. Além disso, ela também deve corresponder às expectativas do público de interesse (WIGGINS, *et al.*, 2013). Segundo Wiggins *et al.* (2013, p. 13), nessa etapa da análise, busca-se "identificar padrões, testar hipóteses e ilustrar descobertas"³⁴.

Salienta-se que a colaboração de especialistas multidisciplinares, com seus conhecimentos em áreas específicas, pode enriquecer a análise dos dados. Para isso, atualmente existem ferramentas eficazes, gratuitas ou de baixo custo, que podem facilitar a comunicação entre os participantes e profissionais envolvidos na pesquisa (WIGGINS, *et al.*, 2013, tradução nossa).

5.1.2 Curation Lifecycle Model – Digital Curation Centre (DCC)

O *Digital Curation Centre* (DCC) é uma instituição britânica que foi criada pela *University of Edinburgh*, pelo departamento *UK Office for Library and Information Networking* (UKOLN) pertencente à *University of Bath*, e pelo *Humanities Advanced Technology and Information Institute* (HATII), instituto adjunto à *University of Glasgow*. Este centro de curadoria digital é reconhecido internacionalmente por haver desenvolvido um modelo de ciclo de vida de dados que descreve detalhadamente as fases pelas quais os dados devem ser submetidos de forma a garantir sua preservação, compartilhamento e reuso (SEGUNDO, ARAÚJO, 2019).

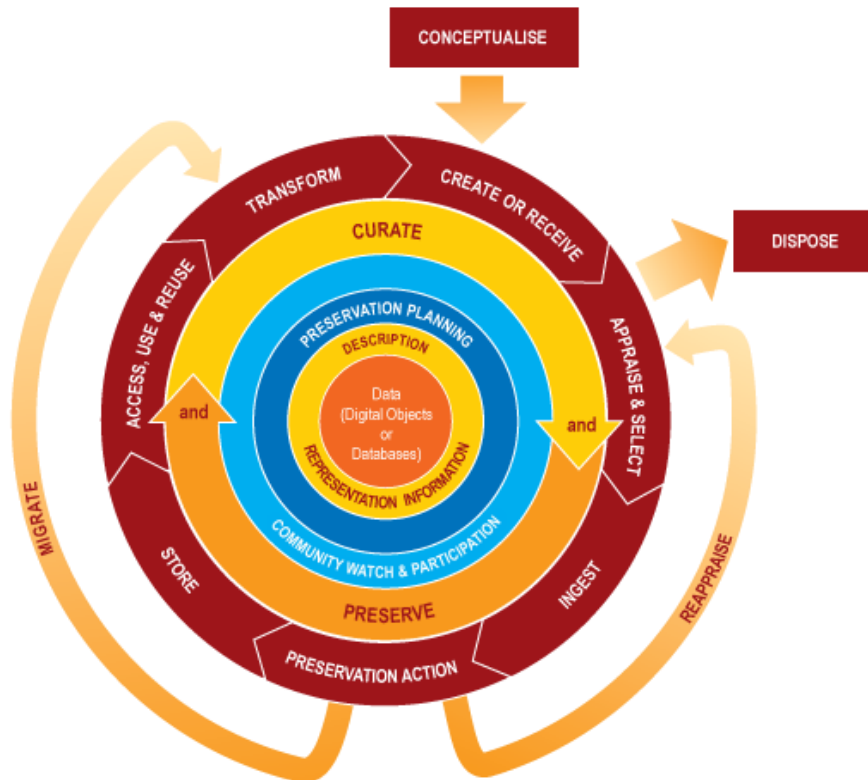
Do mesmo modo, seu reconhecimento se deve ao caráter educativo concentrado em capacitar e desenvolver entre os pesquisadores habilidades necessárias para a gestão e compartilhamento de dados. Além de capacitação, o DCC oferece ainda consultoria a organizações que desejam gerenciar, preservar e compartilhar dados de pesquisa (*DIGITAL CURATION CENTRE*, online).

O modelo de ciclo de vida de dados proposto pelo DCC oferece uma visão geral gráfica de alta qualidade sobre todas as etapas necessárias de curadoria para a preservação exitosa dos dados. O modelo da DCC é disponibilizado ao público em geral e pode ser utilizado tanto por

³⁴ Texto original: Identify patterns, test hypotheses, and illustrate findings.

centros de pesquisa, instituições acadêmicas, quanto por organizações privadas ou públicas que procurem garantir que todos os passos do ciclo de vida sejam dados de acordo com a sequência correta (*DIGITAL CURATION CENTRE, online*). As fases que envolvem o modelo *DCC Curation Lifecycle* se encontram na figura 3.

Figura 3 – Modelo de ciclo de vida – *Digital Curation Centre (DCC)*



Fonte: *Curation Lifecycle Model (ONLINE)*

Neste modelo que adota formato espiral, os dados, localizados no centro, abrangem tanto objetos digitais simples e complexos – texto, imagem, som, ou combinações de diversos objetos digitais, como bases de dados estruturadas.

As ações representadas no modelo podem ser divididas entre **a) completas**, ilustradas pelos círculos concêntricos ao redor dos dados no centro do modelo; **b) sequenciais**, representadas pelo eixo externo de cor vinho; e **c) ocasionais**, correspondentes às flechas exteriores laranjas.

As ações completas ocorrem a qualquer momento durante o ciclo de vida e são relevantes para várias ações sequenciais. As ações sequenciais, por sua vez, são etapas realizadas repetidas vezes para garantir que os dados sejam gerenciados conforme as melhores

práticas. O ciclo de ações sequenciais não é realizado uma única vez, formando, em realidade, a base da cadeia de gestão que deverá continuar durante todo o processo de curadoria. Já as ações ocasionais são aquelas que interrompem ou alteram a ordem das ações sequenciais como resultado de uma tomada de decisão. Acontecem a cada certo tempo para manter a preservação dos dados ao longo prazo.

O modelo inicia com a camada de ações completas e o desenvolvimento dos estágios que a conformam. Integram as ações permanentes: a) *Description and Representation Information* (descrição e representação da informação): é a primeira fase pela qual são submetidos os dados. Neste estágio são atribuídos aos dados de pesquisa seus metadados descritivos, técnicos, estruturais e de preservação, a partir de padrões apropriados, para garantir uma contextualização e controle adequado no longo prazo; b) *Preservation Planning* (Planejamento da Preservação): refere-se ao planejamento da gestão e preservação durante todo o ciclo de vida dos dados; c) *Community Watch & Participation* (vigilância e participação da comunidade): uma comunidade designada, mantém sob observação, alterações sobre os dados, além de participar no desenvolvimento de padrões comuns e de outras ferramentas compartilhadas; d) *Curate & Preserve* (curadoria e preservação): neste estágio realizam-se as ações de gestão, planejadas para a promoção da curadoria e da preservação durante o ciclo de vida (DIGITAL CURATION CENTRE, *online*, BALL, 2012).

A segunda camada são as ações sequenciais, constituídas por oito estágios que não são consideradas exclusivamente curadoria, mas representam o ciclo de vida dos dados no processo de investigação, os quais devem incorporar certas características de curadoria (BALL, 2012). As ações sequenciais são as seguintes:

a) *Conceptualise* (conceituação): concebe o planejamento de como os dados serão criados, sob que metodologias serão coletados e que opções de armazenamento são adequadas dada a natureza dos dados, além de abordar aspectos financeiros direcionados para garantir a curadoria; b) *Create or Receive* (criação ou recebimento): cria dados e metadados administrativos, descritivos, estruturais, técnicos e de preservação, ou recebe dados, conforme política de coleta, de outros criadores de dados, arquivos, repositórios ou centros e aloca metadados, caso necessário. Esta fase procura garantir que os dados estejam acompanhados de metadados independente do procedimento, política ou padrão pelo qual foram descritos; c) *Appraise & Select* (avaliação e seleção): avalia e seleciona os dados que serão curados e preservados a longo prazo. d) *Ingest* (ingestão): nesta fase os dados serão transferidos a arquivos, repositórios, centros de dados ou outros curadores.

Após avaliar, descartar, transferir e selecionar os dados que serão preservados se inicia a e) *Preservation action* (ação de preservação): realiza-se ações para assegurar a preservação de longo prazo dos dados, mantendo-os autênticos, confiáveis e usáveis sem comprometer sua integridade. Inclui limpeza e validação de dados, controle de qualidade, classificação e alocação de metadados de preservação. Caso no processo de verificação e controle de qualidade dos dados este procedimento identifique uma falha, os dados devem ser retornados ao remetente para uma reavaliação, de forma que se realize uma nova seleção e descarte e, portanto, uma melhoria na qualidade dos dados.

Posteriormente à preservação se seguem o f) *Store* (armazenamento): armazena os dados de maneira segura conforme padrões relevantes. Para garantir a segurança dos dados há várias atividades associadas, como atualizar a mídia, backups, etc; g) *Access, User & Reuse* (acesso, uso e reúso): esta fase garante que os dados sejam acessíveis a usuários e “reusuários” diariamente, fazendo uso de controles de acesso robusto e processos de autenticação; h) *Transform* (transformação): sendo esta a última fase, procura a criação de novos dados a partir dos originais, através de migração ou da criação de novos subconjuntos, por exemplo. O produto final é um novo conjunto de dados que iniciará um novo ciclo de vida. (DIGITAL CURATION CENTRE, *online*; BALL, 2012)

Finalmente, nas flechas se apresentam as ações ocasionais, as quais se desenvolvem esporadicamente e em certos estágios da camada de ações sequenciais, e procura manter e preservar ao longo do tempo os dados. As ações ocasionais são: a) *Dispose* (descarte): descarte de dados que não tenham sido selecionados para curadoria e preservação de longo prazo. Via de regra os dados são transferidos para outros arquivos, repositórios, centros de dados ou curadores, mas eventualmente podem ser destruídos por razões legais. Esta ação se encontra envolvida no estágio de avaliação e seleção; b) *Reappraise* (reavaliação): retorna dados que tenham falhado em procedimentos de validação para posterior reavaliação. Esta ação acontece nos estágios de seleção e avaliação, ingestão e ação de preservação; c) *Migrate* (migração): migram dados para um formato diferente, o que pode ocorrer em razão do ambiente de armazenamento ou para proteger os dados das obsolescências de hardwares ou softwares. A migração de dados se encontra nos estágios de ação de preservação, armazenamento, acesso, uso e reúso, e transformação (DIGITAL CURATION CENTRE, *online*)

5.1.3 Research Lifecycle model (Research360 Project) – UKOLN

O “UKOLN” é um centro especializado de consultoria em infraestrutura digital, política de informação e gestão de dados, que oferece recursos e serviços para diferentes setores

educacionais de nível superior. É composto por dois grupos, o *Informatics Research Group* (IRG), encarregado de entregar contribuições ao *Digital Curation Centre* (DCC), e o *Innovation Support Centre* (ISC), encarregado de prestar serviços de consultoria (UKOLN, *online*).

O UKOLN pertence à *University of Bath* e ao *Digital Curation Centre*. Seu modelo de ciclo de vida de dados é fruto de um projeto chamado “*The Research360 Project*”³⁵, que teve como escopo “oferecer suporte de alta qualidade para pesquisas da *University of Bath* e seus parceiros, de modo a extrair a maior quantidade de valor possível dos dados gerados a partir de suas pesquisas”³⁶. (UKOLN, *online*)

Este projeto se concentrou, em especial, nos dados gerados a partir de pesquisas colaborativas entre a *University of Bath* e a indústria, aplicando o modelo de ciclo de vida apenas às faculdades de ciências e engenharia inicialmente, mas logo esta aplicação se estendeu por toda a instituição. (UKOLN, *online*)

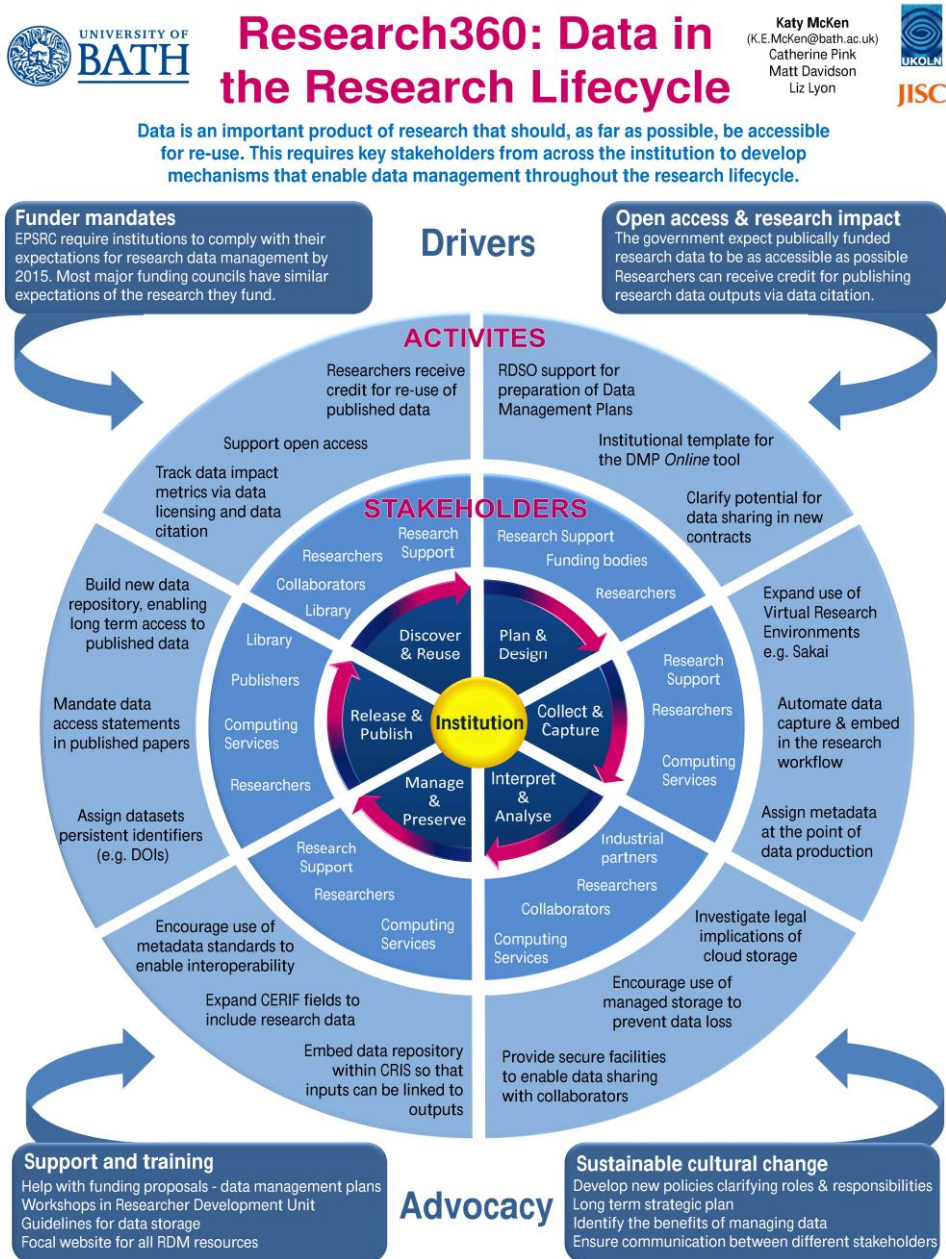
O modelo de ciclo de vida de dados do Research360 possui um foco mais direcionado a estabelecer quais partes interessadas (*Stakeholders*) estão envolvidas e são responsáveis por cada uma das atividades necessárias para a preservação dos dados durante as etapas do modelo. Adotando uma forma geométrica circular, o modelo coloca a instituição responsável pela pesquisa no centro, enquanto o primeiro eixo circular ao seu redor, em azul escuro, representa as ações elementares na gestão dos dados.

O segundo eixo, em um tom médio de azul, representa as partes interessadas em cada ação; e o último eixo externo, em azul claro, representa grupos de três atividades para cada conjunto de partes interessadas em uma ação específica. Adicionalmente, o modelo faz referência à condutores (*drivers*) mencionando as expectativas de financiadores e do governo, e à defesa (*advocacy*), mencionando elementos de suporte, capacitação e de uma mudança cultural sustentável.

Figura 4 – Modelo de Ciclo de Vida do UKOLN – *Research360 Project*

³⁵ <http://www.ukoln.ac.uk/projects/research360/>

³⁶ Texto original: provide high-quality support for research, to ensure that the University of Bath and its partners get the best possible value from data generated by research.



Fonte: Mcken et al. (2012)

No primeiro círculo, o mais interno, estão previstas as ações elementares ou fases da gestão de dados: a) planejamento e design; b) coleta e captura; c) interpretação e análise; d) gestão e preservação; e) liberar e publicar; e f) descoberta e reutilização. Para cada uma dessas ações, conforme passaremos a tratar, são indicadas as partes interessadas e as atividades correspondentes.

Na fase inicial, planejamento e design (*Plan & Design*) em que se constitui o plano de gestão de dados, as partes interessadas são a) os órgãos financiadores, sendo especialmente as

entidades públicas que demandam esta documentação e registros; b) os pesquisadores, ao considerar o planejamento um mecanismo eficiente para executar otimamente as práticas de gestão de dados; e por último c) o pessoal de apoio. Nesta fase, a atividade principal desenvolvida é a preparação do plano de gestão de dados que orienta a preservação dos dados, bem como estabelece as diretrizes e a definição de como se compartilhará os dados já tratados.

A seguinte fase, é a coleta e captura (*collect & capture*) dos dados de pesquisa, em que os envolvidos nesta atividade são os próprios pesquisadores, seu pessoal de apoio, e o serviço de informática. As atividades que compõem esta etapa são: expansão do uso de ambientes virtuais de pesquisa, a automatização da captura de dados, incorporação de dados no fluxo de trabalho de pesquisa, e atribuição de metadados aos dados durante sua geração (MCKEN *et al.*, 2012).

A terceira fase, interpretação e análise (*Interpret & Analyse*) tem como grupo de interesse os parceiros industriais, colaboradores, serviço de informática e pesquisadores. As atividades de gestão de dados desenvolvidas são: a investigação de implicações legais do armazenamento em nuvens; a promoção do uso de armazenamento adequado para evitar a perda de dados; e o fornecimento de instalações seguras para possibilitar o compartilhamento de dados de forma aberta (MCKEN *et al.*, 2012).

Na quarta fase, gestão e preservação (*Manage & Preserve*), as partes interessadas são o serviço de informática, pesquisadores e seu pessoal de apoio. Esses atores cumprem o rol nesta etapa de “incentivar o uso de padrões de metadados garantindo a interoperabilidade, incluir os dados nos formatos padrão estabelecidos e incorporar os repositórios de dados a um sistema unificado” (SEGUNDO; ARAÚJO, 2019; MCKEN *et al.*, 2012).

A quinta fase, liberar e publicar (*Release & Publish*), envolve bibliotecas, editores, serviço de informática e pesquisadores, atores que possibilitam a publicação dos dados. O papel principal dos atores nesta etapa é criar novos repositórios de dados que permitam o acesso a longo prazo aos dados disponibilizados, possibilitar o acesso a dados publicados mediante a declaração do autor para abri-los e reusá-los, e atribuir identificadores únicos ao conjunto de dados.

A última fase, descoberta e reúso (*Discover & Reuse*), trata do acesso aberto e da reutilização dos dados de pesquisa. Dentre os atores responsáveis por garantir a disponibilização, acesso e potencial reúso dos dados de pesquisa, encontram-se as bibliotecas, colaboradores, pesquisadores e pessoal de apoio. O papel do pessoal de apoio se concentra em possibilitar o acesso aberto aos dados de pesquisa. Por sua vez, as bibliotecas e pesquisadores devem medir o impacto que os dados estão causando mediante a quantidade de vezes que estão

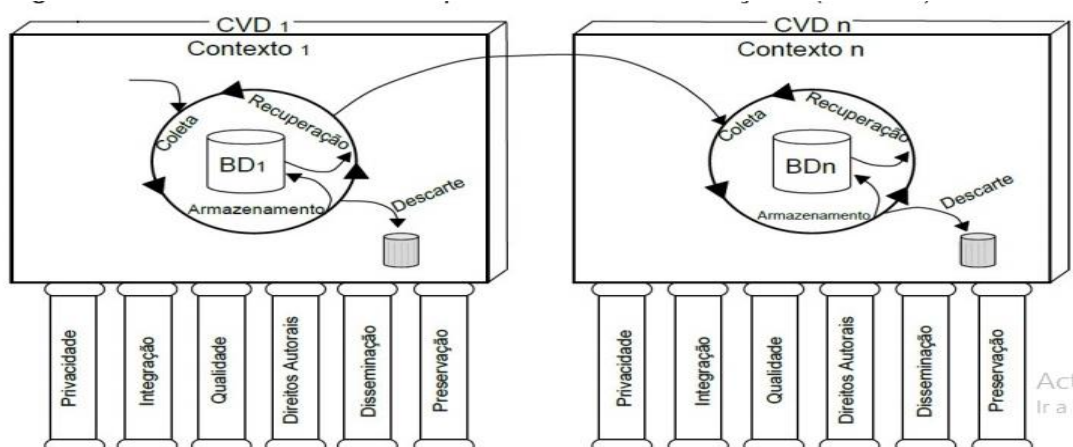
sendo citados. Neste sentido, as questões de licenciamento e citação de dados são indispensáveis para outorgar aos pesquisadores o devido reconhecimento.

5.1.4 Ciclo de vida dos dados (CVD-CI) - Sant'Ana

No contexto de acesso e uso de dados de pesquisa, Sant'Ana (2016) a partir da revisão bibliográfica relacionada a modelos de ciclo de vida de dados, propõe uma nova estrutura do processo de gestão de dados, formada por fases em que os dados são submetidos e fatores que necessariamente devem estar envolvidos para que a gestão, acesso e uso do recurso seja garantido.

O modelo de Sant'Ana (2016) apresenta as fases de coleta, armazenamento, recuperação e descarte, e fatores transversais presentes nas quatro fases: privacidade, integração, qualidade, direito autoral, disseminação e preservação.

Figura 5 – Modelo de Ciclo de Vida - Sant'Ana



Fonte: Sant'Ana, 2016.

Segundo Sant'Ana (2016) a primeira fase consiste na coleta de dados. Esta fase define quais dados serão coletados; sob que estratégias serão avaliados e localizados os dados; e sob que metodologia e instrumento os dados serão coletados, isto é, estabelece-se o plano de ação para a obtenção dos dados. A coleta de dados considera o grau de privacidade dos dados, principalmente quando estão envolvidos indivíduos e informação confidencial procedente de instituições. A integridade dos dados se refere à definição de requisitos sobre a base de dados e os mecanismos necessários para vincular a outras bases. A qualidade dos dados envolve aspectos como descrição de procedência, métodos de coleta, integridade e preservação física dos dados, além de outros elementos que garantem a qualidade. Desde o início da pesquisa, na

fase de coleta deve ser estabelecido a autoria ou responsável dos dados que se pretendem obter. A disseminação e preservação dos dados é contemplada desde o início de tal forma que se planeje a atribuição de informação ao conjunto de dados, garantindo a viabilidade de acesso futuro a este recurso.

Uma vez coletados os dados, inicia a fase de armazenamento. Nesta fase surgem questões a serem refletidas, dentre elas, quais dados serão disponibilizados e armazenados? Qual estrutura digital será utilizada para seu armazenamento? Como os dados serão acessados? Quem poderá acessar ao conjunto de dados? Qual padrão de formatação vai ser adotado? Onde estarão armazenados? Além de outras especificações técnicas e tecnológicas. Outros fatores a serem levados em conta são a privacidade dos indivíduos, quem poderá acessar, manipular e usar o conjunto de dados? Como o conjunto de dados será integrado a outros dados? Como garantir a qualidade física e lógica dos dados? Possuir o direito autoral para armazenar e compartilhar os dados; atribuir aos dados elementos que possibilitem sua disseminação e acesso futuro; e prever elementos que propiciem a preservação dos dados. A fase de armazenamento, portanto, foca-se em aspectos estruturais técnicos, tendo uma participação ativa de profissionais da área da ciência da informação (SANT'ANA, 2016).

A terceira fase, recuperação, segundo Sant'ana, foca-se em tornar os dados acessíveis e com potencial de uso. Nesta fase, o grupo de apoio adquire um papel fundamental ao ter como propósito, mediante estratégias e ações, a manutenção dos dados coletados e armazenados, bem como procurar mecanismos para ampliar a disponibilização, acesso, e uso dos dados. A fase de recuperação leva em consideração a privacidade dos dados, especialmente quando são sensíveis e deve existir a anonimização; a integração entre diversos conjuntos de dados; os elementos que sustentam a qualidade dos dados que estão sendo disponibilizados; a definição dos direitos autorais, bem como a permissão para usar os dados; a definição de estratégias para garantir o acesso e disseminação; a disponibilização de como os dados devem ser usados; os aspectos semânticos envolvidos; e a garantia de interpretação ao longo do tempo (SANT'ANA, 2016).

A última fase, descarte, segundo Sant'ana (2016) consiste em selecionar e eliminar os conjuntos de dados que não são mais necessários ou eficientes e, portanto, devem ser excluídos dos repositórios. No que diz respeito à privacidade, o indivíduo tem o direito de solicitar o retiro dos seus dados quando considerar necessário. A integração dos dados na fase de descarte torna-se um assunto sensível devido ao fato de que excluir um conjunto de dados em determinada base pode impactar negativamente na degeneração das outras bases em que os dados se encontram ligados. A qualidade dos dados no processo de descarte considera que se deve manter detalhadamente registrada a eliminação destes. Questões de direito autoral indicam que

deve ser mantida a informação relacionada à autoria, mesmo se os dados não estiverem mais disponíveis, evitando desta maneira perder a origem dos dados. Por fim, a preservação deve estar presente, incluso nesta fase de descarte, devido a que podem surgir novas necessidades que venham a solicitar os dados eliminados.

6 METODOLOGIA

O capítulo apresenta o processo de pesquisa do ponto de vista metodológico e seus procedimentos operacionais, especificando suas etapas e as atividades adotadas em cada uma delas, desde a classificação no tipo de pesquisa, definição da população e seleção da amostra, assuntos éticos, elaboração e validação do instrumento, até as técnicas de coleta e análise de dados.

6.1 TIPO DE PESQUISA

A pesquisa científica é definida por Andrade (2001) como um conjunto de procedimentos sistemáticos, baseados no raciocínio lógico, que tem como fim procurar soluções aos problemas estudados, mediante a aplicação de métodos científicos. No contexto científico, é necessário a definição metodológica por parte do pesquisador para esclarecer os mecanismos pelos quais o processo de investigação vai ser desenvolvido, as técnicas de coleta e os critérios de análise e interpretação de dados a serem aplicadas, de forma que os resultados procedentes da pesquisa sejam confiáveis e verazes. A pesquisa adquire um caráter científico quando seu desenvolvimento se dá mediante o “emprego de procedimentos intelectuais e técnicos” (GIL, 2008, p.26) que possibilita ao pesquisador descobrir repostas para o fenômeno estudado e atingir um novo conhecimento.

A presente pesquisa está caracterizada por ser, em relação aos seus objetivos, de tipo exploratória e descritiva. Uma vez que o objetivo geral é revelar as práticas de gestão, compartilhamento e reúso dos dados de pesquisa, dos pesquisadores, justifica-se o enquadramento desta, como pesquisa de tipo exploratória devido a uma ausência de conhecimento sobre a área estudada no contexto científico colombiano. Gil (2007) ratifica que as pesquisas de caráter exploratório têm como propósito proporcionar uma visão geral acerca do fenômeno quando a problemática a ser estudada tem sido pouco explorada.

Do mesmo modo, encaixa-se no tipo descritiva ao ter como propósito descrever as características de determinado fenômeno (GIL, 2002). A pesquisa procura identificar e caracterizar como os pesquisadores cotidianamente praticam a gestão, compartilhamento e reúso de dados de pesquisa no contexto do acesso aberto a dados.

Em relação à abordagem, a pesquisa é concebida de natureza quantitativa visto que se procura a mensuração do fenômeno, mediante a coleta de dados numéricos e sua análise estatística. (COLLIS; HUSSEY, 2005). A abordagem quantitativa adequa-se à natureza do

fenômeno a ser estudado, ao objetivo de pesquisa a ser atingido, e ao número de informantes que se quer abordar.

O procedimento ou técnica escolhida para realizar a investigação é do tipo levantamento ou pesquisa tipo *survey*. Essa técnica é adequada para estudos exploratórios e descritivos (GIL, 2002; BABBIE, 2003). Segundo Babbie (2003), a pesquisa tipo *survey*, possibilita o desenvolvimento rigoroso, mediante um pensamento lógico, do procedimento científico que buscar abordar diversas variáveis em entornos complexo. Este tipo de técnica tem como propósito generalizar uma descoberta a partir dos resultados obtidos em uma amostra. Portanto, para atingir esse caráter científico genérico na pesquisa *survey*, deve-se selecionar e delimitar a população que se pretende estudar, para, em seguida, definir uma amostra representativa da população, realizar um levantamento de dados desta amostra, e por último, mediante sua interpretação realizar conclusões.

6.2 DELINEAMENTO DA PESQUISA

O procedimento técnico, pesquisa *survey*, caracteriza-se, segundo Gil por:

(...)pela interrogação direta das pessoas cujo comportamento se deseja conhecer. Basicamente, procede-se à solicitação de informações a um grupo significativo de pessoas acerca do problema estudado para em seguida, mediante análise quantitativa, obter as conclusões correspondentes dos dados coletados (...). (2002.p 50):

O procedimento de uma pesquisa de tipo *survey* está formado pelas seguintes fases: **a)** estudo do estado atual do fenômeno a ser abordado; **b)** construção do referente teórico a fim de determinar as variáveis a serem estudadas; **c)** definição das técnicas de coleta de dados (instrumentos); **d)** determinação da população e amostra; **e)** técnicas de registro desses dados e, **f)** tipo de análises que será aplicado. (MARCONI; LAKATOS, 2010).

Do mesmo modo, Gil propõe um delineamento similar ao estabelecido por Marconi e Lakatos para executar o tipo de pesquisa *Survey*. Gil (2002) determina as fases na seguinte sequência: **a)** definição dos objetivos a fim de determinar o que se pretende atingir para solucionar ou dar resposta ao fenômeno estudado; **b)** determinação e operacionalização dos conceitos e variáveis, construídos a partir do referencial teórico levantado; **c)** elaboração do instrumento de coleta de dados de acordo com os objetivos estabelecidos ; **d)** pré-teste do instrumento para validar sua pertinência e confiabilidade; **e)** seleção rigorosa da amostra, garantindo uma alta representação do universo estudado; **f)** coleta e verificação dos dados, avaliando sua clareza, coerência e precisão; **g)** análise e interpretação dos dados, envolvendo

codificação, tabulação dos dados, e cálculos estatísticos; e por fim, **h)** a apresentação dos resultados.

6.2.1 Contextualização

Seguindo o procedimento de pesquisa proposto por Gil (2002), Marconi e Lakatos (2010), inicialmente a pesquisa contextualizou o estado em que se encontra a Colômbia em relação aos tópicos gestão, compartilhamento, reúso de dados de pesquisa, isto revelando a problemática a ser abordada, e como consequência, definindo os objetivos a serem atingidos. Posteriormente, construiu-se o referencial teórico e, a partir de seu levantamento, foram estabelecidas as variáveis estudadas e sua matriz de operacionalização (apêndice C).

6.2.2 Características da população

A população a ser observada está constituída pelos professores pesquisadores vinculados aos grupos de pesquisa classificados por Colciencias, e pertencentes a Instituições de Ensino Superior do tipo universidade. Colciencias, como agência financiadora colombiana, categoriza os grupos de pesquisa a partir das áreas e subáreas de conhecimento estabelecidas pela OECD. A área de conhecimento a ser abordada é a área de economia e negócios, formada pelas subáreas de negócios, administração, economia, finanças, contabilidade, e outros afins. Esta área é considerada uma fonte geradora de grande quantidade de dados devido à natureza dos fenômenos que pesquisam. Geograficamente o estudo se concentra na cidade de Bucaramanga e sua região metropolitana. A cidade concentra um significativo número de universidades de prestígio em nível nacional, além de ter um grupo de pesquisadores com um nível alto de formação e experiência em investigação³⁷. A população, conforme apresentado no quadro 4:

Quadro 4 – Critérios de inclusão

Características dos pesquisadores participantes	
Colciencias:	Classificados pela agência financiadora
Grande área	Ciências Sociais
Área	Economia e Negócios (Ciências econômicas, administrativas, contábeis e afins)
Departamento	Santander
Cidade	Bucaramanga e região metropolitana
Universidade	Particular e Pública

Fonte: elaboração própria, 2019

³⁷ O estudo “*estudio para identificar conocimientos, capacidades, percepciones y experiencias de los investigadores del país frente a la ciencia abierta*” identificou que o nível de formação, a experiência em investigação, e os estados do país com melhor capacidade investigativa, repercutem no grau de conhecimento que o pesquisador tem sobre ciência aberta

A unidade de observação se refere ao mundo a ser estudado, podendo ser pessoas, famílias, cidades, estados, nações, indústrias, etc. (BABBE, 2003). Neste sentido, para esta pesquisa a unidade apropriada de observação é o indivíduo (pesquisador), em conformidade com o objetivo que se pretende atingir.

De acordo com as informações da população de estudo registrada por Colciencias nas plataformas *cvLAC* e *GrupLAC*³⁸, e verificadas nos sites dos grupos de pesquisa e das universidades nas quais os pesquisadores se vinculam, identificam-se 138 professores pesquisadores que cumprem com os critérios de inclusão: a) Professor pesquisador vinculado a um grupo de pesquisa, da área de economia e negócios, classificado por Colciencias; b) Grupo de pesquisa vinculado a uma universidade particular ou pública, localizado na cidade de Bucaramanga e região metropolitana

Quadro 5 – População de estudo

Universidade	Grupos de pesquisa de economia, negócios e afins	Número de pesquisadores
Pública	3	38
Particular	3	19
Particular	3	37
Particular	3	24
Particular	1	8
Particular	1	12
Número de pesquisadores		138

Fonte: Elaboração própria, 2019.

6.2.3 Amostra da pesquisa

A amostragem realizada foi do tipo intencional, definida como a que ocorre quando os elementos ou indivíduos selecionados para a amostra são escolhidos de acordo com os critérios do pesquisador, pois atendem a condições de interesse do estudo (ARIAS, 2006).

Atualmente registram-se 138 professores- pesquisadores que cumprem com os critérios de inclusão. O método de amostragem a ser aplicada é amostragem aleatória simples. A continuação se calcula a amostra a partir da formula proposta por Miquel (1997):

$$n = \frac{z^2 \times p \times q \times N}{e^2 (N - 1) + z^2 \times q \times p}$$

38

https://sba.minciencias.gov.co/tomcat/Buscador_Grupos/busqueda?q=SANTANDER&pagenum=1&start=0&type=load&inmeta=AREA_CONOCIMIENTO_s!CIENCIAS%20SOCIALES&inmeta=AREA_s!ECONOMIA%20Y%20NEGOCIOS&inmeta=ubi_departamento_s!SANTANDER

N: tamanho da população

z: Valor crítico que corresponde ao grau de confiança desejado

p: Proporção populacional de indivíduos que pertence a categoria que estamos interessados em estudar.

q: Proporção populacional de indivíduos que NÃO pertence à categoria que estamos interessados em estudar ($q = 1 - p$).

e: Margem de erro ou erro máximo de estimativa.

Em que:

N: 138

Grau de confiança: 95%

Valor crítico z: 1.96

Prevalência positiva p: 0.5

Prevalência negativa p: 0.5

Erro amostral: 5% ou 10%

Considerando um erro de 5%, o tamanho da amostra é de 102 casos. Se um erro de 10% for considerado, o tamanho da amostra corresponderá a 57 casos. Levando em conta que o número de casos coletados foi de 88, o erro será estimado usando o seguinte cálculo (MIQUEL, 1997):

$$e = \sqrt{\frac{z^2 pqN - z^2 pqn}{n(N-1)}}$$

De acordo com a formula mencionada, o erro para os casos coletados, 88, é de 5,74%. O referido erro de acordo com a literatura deve estar entre 5% e 10% para garantir fiabilidade das descobertas.

6.2.4 Elaboração do questionário

O instrumento de coleta de dados selecionado, tendo em conta a natureza quantitativa da pesquisa, foi o questionário. Esse instrumento se apresenta como adequado devido a uma série de vantagens que oferece esta técnica. Dentre as vantagens, menciona-se “economia de tempo, atingir um maior número de pessoas simultaneamente, menos risco de distorção, pela não influência do pesquisado, mais uniformidade na avaliação obter respostas rápidas e precisas” (MARCONI; LAKATOS, 2002). O questionário é considerado pertinente por duas

principais razões: a) atinge a um número maior de participantes e b) possibilita a medição de diversas variáveis de maneira rápida, segura e confiável.

O instrumento é construído a partir do referencial teórico proposto nesta pesquisa, e de questionários de pesquisas internacionais, validados, referentes ao compartilhamento, e reuso de dados de pesquisa. Antes da construção do questionário, foi necessário formular a matriz de operacionalização das variáveis (Apêndice C). O objetivo desta matriz é tornar as variáveis abstratas em variáveis concretas, de modo que, as variáveis a serem observadas sejam medidas quantitativamente. Uma vez determinado o as variáveis, dimensões e indicadores a observar e medir, o questionário foi formulado.

O instrumento “ Questionário: práticas de gestão, compartilhamento e reuso de dados de pesquisa” (apêndice A) está baseado nos questionários intitulados “Acesso Aberto a Dados de Pesquisa (AADP) ” desenvolvido pela Rede de Dados de Pesquisa Brasileira (RDP Brasil) (VANZ *et al.*, 2018); “A gestão de dados científicos: Desvendando as práticas dos pesquisadores brasileiros na área da Ciência da Informação” desenvolvido pelos pesquisadores do Grupo de Pesquisa Web Representação do Conhecimento e Ontologias – WRCO da Universidade Federal da Paraíba (DIAS;ANJOS; ARAÚJO,2017) ; “*Scientists and research data: Continuing to build an understanding of your data needs*” desenvolvido por Data Observation Network for Earth (CAROL *et al.*, 2015); “*Data sharing in environmental sciences: A survey of CNR researchers*” desenvolvido pelos pesquisadores Luzi *et al* (2012); “*Gestión de datos de Investigación en universidades en base al ciclo de vida de los datos. Caso de estudio: Área de Ciencias de la Salud*” desenvolvido por Briceño (2017); módulos de educação: “*Data sharing*”, “*Data Management Planning*”, “*Data Entry and Manipulation*”, “*Data Quality Control and Assurance*”, “*Protecting your Data*”, “*Metadata*”, “*Legal and Policy Issues*”, desenvolvido por Data Observation Network for Earth (2016).

O questionário (apêndice A) é composto por 61 perguntas do tipo fechadas de múltipla escolha, dicotômicas, resposta única e matriz de resposta única, e uma pergunta de tipo aberta, agrupadas em 5 blocos, com os seguintes conteúdo específicos:

Bloco I. Perfil do pesquisador: refere-se à trajetória do pesquisador como nível de formação, tempo de atuação como professor universitário, área de conhecimento investigada, tempo dedicado à pesquisa.

Bloco II. Características dos dados de pesquisa: refere-se ao caráter dos dados da pesquisa, sua natureza, formato em que são arquivados, tipo de armazenamento e seu tempo de conservação.

Bloco III. Práticas de gestão de dados de pesquisa: refere-se ao conhecimento sobre o conceito de gestão de dados, e às práticas de planejamento, criação de metadados, garantia da qualidade e preservação dos dados.

Bloco IV. Compartilhamento de dados de pesquisa: refere-se à disponibilidade e abertura dos dados, a qualidade dos dados compartilhados, e as políticas, condições e licenças usados que possibilitam seu compartilhamento.

Bloco V. Reúso de dados de pesquisa: refere-se ao acesso e reúso de dados próprios e de outros pesquisadores, qualidade dos dados reusados e os meios de comunicação científica pelos quais foram acessados.

6.2.5 Validade e objetividade do questionário

Entende-se por validade como o “grau em que um instrumento realmente mede a variável a ser medida”³⁹ (HERNÁNDEZ; FERNÁNDEZ; BAPTISTA, 2010, p. 197, traduções nossa). A medição da validade do instrumento se realiza mediante a validação de conteúdo por expertos. A validação de conteúdo por expertos é uma prática que requer de rigorosidade metodológica e estatística para evitar erros de suficiência, clareza, coerência, relevância e pertinência nas perguntas que constituem o questionário. Este método é definido como a apreciação de pessoas expertas sobre o assunto a ser avaliado e que, portanto, tem as condições para fornecer julgamentos, informações ou evidências (PEREZ; MARTINEZ, 2008).

Objetivou-se validar o conteúdo do questionário construído para atingir os propósitos da pesquisa. Os critérios estabelecidos para a seleção de especialistas são: a. experiência relacionada à gestão, compartilhamento e reúso de dados de pesquisa, b. formação acadêmica, c. trajetória profissional, d. trajetória em investigação, e. disponibilidade para participar. O número de expertos que participaram da avaliação foram 3. A continuação se apresentam suas informações:

Quadro 6- Especialistas

Especialistas	Formação acadêmica	Trajetória profissional	Trajetória em investigação	Experiência no assunto
1	Pós-doutorado Universitat De Valencia Universidade Federal de Sergipe	16 anos como professor titular	16 anos como professor – pesquisador	Arquiteturas de Integração Web/Bases de Dados Tecnologias da informação e comunicação
2	Mestrado em Ética e Filosofia política.	22 anos de experiência como editora de revistas, coordenadora de gestão em	13 anos como professora-pesquisadora.	Biblioteconomia. Documentação.

³⁹ La validez en términos generales se refiere al grado en que un instrumento realmente mide la variable que se busca medir.

Especialistas	Formação acadêmica	Trajetória profissional	Trajetória em investigação	Experiência no assunto
	Graduação Biblioteconomia Universidad de Antioquia	bibliotecas, professora-pesquisadora, gestora acadêmica, consultora		Gestão compartilhamento e reuso de dados científicos.
3	Mestrado em Administração de Empresas Universidad Del Mar Colegio mayor nuestra señora del rosario Especialista en finanzas privadas Universidad Santo Tomas Economista	20 anos Diretor do Centro de Formação em Tecnologias e administração pública.	30 anos como professor	Planejamento e gestão económica, administração pública, engenharia de produção, tecnologia.

Fonte: Elaboração própria, 2020.

O processo iniciou com a elaboração dos documentos a enviar para a avaliação do questionário. Os documentos enviados são: carta de convite para participar na validação do instrumento (apêndice D), resumo da pesquisa, matriz de consistência metodológica (apêndice B), matriz de operacionalização das variáveis (apêndice C), e certificado de validade de conteúdo do questionário (apêndice E). Estas informações são compartilhadas com os expertos para fornecer um panorama completo da pesquisa, e assim avaliar eficientemente o instrumento.

Compartilhar a matriz de consistência metodológica e a matriz de operacionalização das variáveis com os expertos torna-se indispensável já que permite identificar claramente a coerência e suficiência entre as dimensões, os indicadores e os itens do questionário. Proporcionar o resumo da pesquisa possibilita contextualizar os expertos sobre o objetivo da pesquisa, e verificar como os itens que compõem o questionário respondem a estes, desta maneira garantindo a pertinência.

O certificado de validade de conteúdo do questionário contém as categorias a serem avaliadas para cada item do questionário. As categorias se referem ao que será avaliado para cada item do questionário. Tais categorias são (PEREZ; MARTINEZ, 2008):

Suficiência: itens que pertencem à mesma dimensão são suficientes para obter sua medida.

Clareza: o item é facilmente entendido, isto é, suas sintáticas e semânticas são adequadas.

Coerência: o item está logicamente relacionado à dimensão ou indicador que está medindo.

Relevância: o item é essencial ou importante, isto é, deve ser incluído.

De acordo com os indicadores por categoria, classifica-se cada um dos itens conforme o experto perceba apropriado:

Quadro 7- Matriz de validação de conteúdo

Categoria	Calificação	Indicador
Suficiencia Los ítems que pertenecen a una misma dimensión bastan para obtener la medición de ésta	1. No cumple con el criterio 2. Bajo nivel 3. Moderado nivel 4. Alto nivel	1. Los ítems no son suficientes para medir la dimensión 2. Los ítems miden algún aspecto de la dimensión pero no corresponden con la dimensión total 3. Se deben incrementar algunos ítems para poder evaluar la dimensión complementemente. 4. Los ítems son suficientes
Claridad El ítem se comprende fácilmente, es decir, su sintáctica y semántica son adecuadas	1. No cumple con el criterio 2. Bajo nivel 3. Moderado nivel 4. Alto nivel	1. El ítem no es claro 2. El ítem requiere bastantes modificaciones o una modificación muy grande en el uso de las palabras de acuerdo con su significado o por la ordenación de las mismas 3. Se requiere una modificación muy específica de algunos de los términos de ítem 4. El ítem es claro, tiene semántica y sintaxis adecuada.
Coherencia El ítem tiene relación lógica con la dimensión o indicador que está midiendo.	1. No cumple con el criterio 2. Bajo nivel 3. Moderado nivel 4. Alto nivel	1. El ítem no tiene relación lógica con la dimensión 2. El ítem tiene una relación tangencial con la dimensión 3. El ítem tiene una relación moderada con la dimensión que está midiendo. 4. El ítem se encuentra completamente relacionado con la dimensión que está midiendo.
Relevancia El ítem es esencial o importante, es decir debe ser incluido	1. No cumple con el criterio 2. Bajo nivel 3. Moderado nivel 4. Alto nivel	1. El ítem puede ser eliminado sin que se va afectada la medición de la dimensión. 2. El ítem tiene alguna relevancia, pero otro ítem puede estar incluyendo lo que mide éste. 3. El ítem es relativamente importante. 4. El ítem es muy relevante y debe ser incluido.

Fonte: PEREZ; MARTINEZ, 2008.

A avaliação do instrumento (anexo A) foi realizada em 3 semanas. A partir dos seus resultados foi estimado o grau de concordância mediante o coeficiente de Kendall.

O coeficiente de concordância W de Kendall expressa o grau de associação ou nível de concordância entre as avaliações dos expertos sobre um determinado item. O coeficiente de concordância de Kendall pode variar de 0 a 1. Quanto maior o valor de Kendall, mais forte será a concordância (PEREZ; MARTINEZ, 2008). A continuação se expressam os resultados obtidos mediante o software SPSS:

Quadro 8- W de Kendall: concordância entre expertos

Dimensões	Itens	Total	Suficiência	Clareza	Coerência	Relevância
Trajetória profissional	1-3	0,857	1,000	1,000	1,000	0,905
Trajetória de pesquisa	3-8	0,565	1,000	0,815	0,815	0,708
Particularidades dos dados	9-12	0,853	1,000	0,905	1,000	1,000
Pesquisadores envolvidos na coleta de dados	13	0,667	1,000	0,667	0,667	0,905
Conhecimento sobre gestão de dados	14-15	0,783	1,000	0,933	0,933	1,000
Assistência para a gestão de dados	16-17	1,000	1,000	1,000	1,000	1,000
Gestão de dados de pesquisa	18	1,000	1,000	1,000	1,000	1,000
Planejamento da gestão de dados	19-28	0,564	1,000	0,594	1,000	1,000

Dimensões	Itens	Total	Suficiência	Clareza	Coerência	Relevância
Coleta, processamento, armazenamento de dados	29-39	0,841	1,000	0,622	0,841	1,000
Metadados	40-43	0,647	1,000	0,699	0,810	1,000
Garantia da qualidade dos dados	44-51	0,836	1,000	0,842	1,000	1,000
Preservação dos dados	52-61	0,739	1,000	1,000	1,000	1,000
Disponibilidade dos dados de pesquisa	62-64	0,857	1,000	1,000	1,000	1,000
Qualidade dos dados de pesquisa compartilhados	65-67	0,482	1,000	0,681	0,800	0,681
Políticas de acesso e intercâmbio de dados de pesquisa	68-78	0,449	1,000	0,472	1,000	1,000
Reúso de dados próprios e de outros pesquisadores	79-80	0,867	1,000	1,000	1,000	1,000
Conjunto de dados formados a partir dos próprios e de outros pesquisadores	81	0,667	0,667	0,667	0,667	0,667
Meios de acesso a dados de outros pesquisadores	82	1,000	1,000	1,000	1,000	1,000

Fonte: Elaboração própria, com base em SPSS 22®, 2020.

Como resultado do nível de concordância entre os especialistas, o instrumento obteve em média Coeficiente de Kendall w de 0,759, o qual deve ser maior que 0,5 para todo o teste. O valor de Kendall de w obtido nos permite esclarecer que há concordância entre os juízos feitos pelos expertos.

Observa-se 3 dimensões com concordância absoluta, e 4 com baixo nível de concordância, indicando que há divergências na sua avaliação. As dimensões e categorias (relevância, clareza, suficiência, coerência) abaixo de 0,79, foram intervindas.

O instrumento avaliado estava estruturado com 18 dimensões, as quais foram reduzidas a 14 por recomendação dos expertos. Inicialmente o questionário estava composto por 82 perguntas, sendo reduzidas a 61 perguntas. A eliminação de perguntas é causada por falta de relação entre o item e a dimensão, e falta de relevância para o que se pretende observar. Além disso, foi necessário reajustar semanticamente determinadas perguntas indicadas pelos expertos, pois apresentavam problemas de clareza. Seguindo as observações propostas pelos especialistas, decidiu-se criar novas dimensões, indicadores e itens (questões) de acordo com os objetivos e hipótese estabelecida. A avaliação de expertos teve como resultado ajustes na matriz de consistência metodológica, matriz de operacionalização das variáveis, e questionário. Emitidas as conclusões da validação, obteve-se o questionário final (apêndice A).

Cabe esclarecer que os questionários bases, utilizados para a criação do questionário aplicado, são instrumentos válidos e fiáveis, aplicados em diversas investigações realizadas por instituições com grande experiência em gestão, compartilhamento e reúso de dados de pesquisa.

Nesse contexto, em que os instrumentos de base para a criação do questionário aplicado nesta pesquisa estão em várias línguas, foi necessário fazer uma adaptação cultural e de linguagem para evitar interpretações errôneas. Nessa direção, era essencial realizar a validade do conteúdo por meio de expertos medindo a suficiência, clareza, coerência e relevância das questões, e verificar o nível de concordância entre eles por W de Kendall para fazer os respectivos ajustes.

A objetividade é entendida como o “grau em que o instrumento é permeável à influência dos vieses e tendências dos pesquisadores que o administram, qualificam e interpretam”⁴⁰ (MORALES *et al*, 2018, p. 4). Como algumas respostas são carregadas de juízos de valor, devido à multiplicidade de opções que o indivíduo deve escolher, determinou-se que o questionário apresentava objetividade menos que perfeita, considerada objetividade média.

6.2.6 Operacionalização das variáveis

A operacionalização de variáveis é um processo metodológico que consiste em decompor as variáveis que constituem o problema de pesquisa, começando do mais geral ao mais específico (CARRASCO, 2009). O propósito de operacionalizar as variáveis, é transformar um conceito abstrato em concreto, capaz de ser medido mediante um instrumento.

A matriz é essencial para a medição quantitativa das variáveis, pois orientará a execução de todo o processo de pesquisa, especialmente no processamento dos dados.

O processo de operacionalização das variáveis indicado Villavicencio-Caparó *et al* (2019) foi aplicado para a construção da matriz de operacionalização das variáveis da presente pesquisa (apêndice C). O processo inicia determinando, a partir do referencial teórico, o problema de pesquisa, os objetivos, as hipóteses, e por último, as variáveis a serem estudadas. Tais variáveis ao serem abstratas, são definidas conceitual e operacionalmente de forma a comunicar com precisão o que se pretende medir. Segundo Carrasco (2009), a definição conceitual se refere à descrição da variável, enquanto que, a operacional é a decomposição da descrição conceitual em referentes empíricos.

A partir dos resultados obtidos no teste de validade e confiabilidade do instrumento, fizeram-se as respectivas modificações na matriz de operacionalização das variáveis. Tais modificações foram feitas nas dimensões, indicadores e perguntas. A continuação, observa-se a matriz final:

⁴⁰ It is the degree to which this is or is not permeable to the influence of the biases and tendencies of the researchers who administer, qualify and interpret it.

Quadro 9 – Variável perfil do pesquisador

Variável	Definição	
	Conceitual	Operacional
Perfil do pesquisador	Expressa o conjunto de características e traços que uma pessoa tem, de acordo com a atividade em que está envolvida (Castaño, Calderón, & Posada, 2014, p. 47-48, tradução nossa).	Trajetória profissional e como pesquisador: formação acadêmica, tempo como docente universitário, categoria como pesquisador por Colciencias, área de conhecimento investigada, tempo de atuação como pesquisador.

Fonte: Elaboração própria, 2020.

Quadro 10 – Variável dados de pesquisa

Variável	Definição	
	Conceitual	Operacional
Dados de pesquisa	Unidades de informação criadas ou coletadas no curso da pesquisa científica, e que são frequentemente formatadas de maneira a torná-las adequadas à comunicação, interpretação e processamento por computador. (SAYÃO, SALES, 2015, p.80)	Natureza e carácter dos dados. Formatos e armazenamento dos dados.

Fonte: Elaboração própria, 2020.

Quadro 11 – Variável gestão de dados de pesquisa

Variável	Definição	
	Conceitual	Operacional
Gestão de dados de pesquisa	Conjunto de práticas de gestão voltadas para o tratamento de dados de pesquisa durante o seu ciclo de vida; inclui todos os aspectos de manutenção, compartilhamento, segurança e preservação (SAYÃO, SALES, 2015, p.80-81)	Ciclo de vida dos dados de pesquisa DataOne (2016): Planear Coletar Assegurar Descrever Preservar

Fonte: Elaboração própria, 2020.

Quadro 12 – Variável compartilhamento de dados de pesquisa

Variável	Definição	
	Conceitual	Operacional
Compartilhamento de dados de pesquisa	Compartilhamento de dados é a liberação de dados de pesquisa para uso de outras pessoas. A liberação pode assumir muitas formas, desde troca privada sob pedido até depósito em uma coleção pública de dados. A publicação dos conjuntos de dados em um site público ou seu fornecimento a uma revista como material suplementar também é considerada uma troca. Contudo, o grau de utilidade, confiabilidade e valor dos dados compartilhados varia muito. Algumas podem ser ricamente estruturadas e curadas. Outros podem ser arquivos em bruto com o mínimo de documentação. Da mesma forma, os usuários previstos podem variar entre investigadores de uma determinada área e o público em geral. (BORGMAN, 2012, p. 3, tradução nossa)	Intercambio de dados de pesquisa através de canais de comunicação científicos ou de forma privada: Disponibilidade de dados. Qualidade dos dados compartilhados. Canais de comunicação científica. Condições de intercâmbio de dados.

Fonte: Elaboração própria, 2020.

Quadro 13 – Variável reúso de dados de pesquisa

Variável	Definição	
	Conceitual	Operacional
Reúso de dados de pesquisa	Uso dos dados para propósitos diferentes do qual eles foram coletados, geralmente por outros pesquisadores que não os autores dos dados (SAYÃO, SALES, 2015, p.82)	Reúso de dados de pesquisa em novas pesquisas. Qualidade dos dados reusados. Canais de comunicação científica para aceder aos dados de pesquisa

Fonte: Elaboração própria, 2020.

Estabelecido a definição conceitual e operativas das variáveis, determina-se sua operacionalização mediante a descrição das dimensões, indicadores, perguntas, tipo estatístico, e escala de medição.

As dimensões referem-se às categorias que serão observadas e descritas. Cada categoria está relacionada a indicadores, os quais determinam especificamente a característica a ser observada e medida. O tipo estatístico refere-se à natureza quantitativa ou qualitativa da pergunta, orientando sob que técnica estatística é possível processar e analisar os dados. A escala de medição indica o nível em que é possível a medição matemática da variável.

A partir do tipo estatístico da pergunta e a escala de medição, determina-se a técnica estatística pela qual será observada, descrita e medida a variável, de tal modo que os objetivos sejam atingidos e a hipótese seja testada.

A finalidade da matriz de operacionalização das variáveis (apêndice C) é responder coerentemente à matriz metodológica da pesquisa (apêndice B); desenhar e elaborar o instrumento de coleta de dados (apêndice A); e determinar as técnicas estatísticas a serem aplicadas. A continuação, observa-se a operacionalização das variáveis a partir dos ajustes feitos no teste de validade.

Quadro 14 – Operacionalização perfil do pesquisador

Variável	Dimensões	Indicadores	Questionário Pergunta	Tipo estatístico	Escala de medição
Perfil do pesquisador	Trajetória profissional	Instituição em que labora	1	Qualitativa	Nominal
		Grau de formação acadêmica	2	Qualitativa	Ordinal
		Tempo de atuação como professor (a) universitário (a)	3	Quantitativa	Intervalo
	Trajetória de pesquisa	Categoria atual como pesquisador classificado por Colciencias	4	Qualitativa	Ordinal
		Tempo de atuação como pesquisador	5	Quantitativa	Intervalo
		Área de conhecimento investigada	6	Qualitativa	Nominal

Fonte: Elaboração própria, 2020.

Quadro 15 – Operacionalização dados de pesquisa

Variável	Dimensões	Indicadores	Questionário Pergunta	Tipo estatístico	Escala de medição
		Natureza dos dados	7	Qualitativa	Nominal

Variável	Dimensões	Indicadores	Questionário Pergunta	Tipo estatístico	Escala de medição
Dados de pesquisa	Particularidades dos dados	Caráter dos dados	8	Qualitativa	Nominal
		Tipo de formato que arquiva os dados	9	Qualitativa	Nominal
		Tempo que armazena os dados	10	Qualitativa	Nominal
		Tipo de armazenamento dos dados	11	Qualitativa	Nominal

Fonte: Elaboração própria, 2020.

Quadro 16 – Operacionalização gestão de dados de pesquisa

Variável	Dimensões	Indicadores	Questionário Pergunta	Tipo estatístico	Escala de medição
Gestão de dados de pesquisa	Conhecimento sobre gestão de dados	Conhecimento de práticas de gestão de dados de pesquisa	12	Qualitativa	Nominal
	Planejamento da gestão de dados	Planejamento da coleta, processamento, análise, descrição e preservação dos dados	13-15	Qualitativa	Ordinal
		Propriedade Intelectual	16	Qualitativa	Ordinal
		Documento plano de gestão de dados	17	Qualitativa	Ordinal
		Assistência para a criação de um plano de gestão de dados	18-19	Qualitativa	Ordinal
	Metadados	Descrição dos dados da pesquisa	20-23	Qualitativa	Ordinal
		Orienta a descrição de dados a partir de metadados padrão	24	Qualitativa	Ordinal
		Assistência na criação de metadados	25-26	Qualitativa	Ordinal
	Controle e garantia da qualidade dos dados	Inspeção e monitoramento da qualidade do conjunto de dados	27-30	Qualitativa	Ordinal
		Diretrizes que verifiquem, inspecionem e controlem a qualidade dos dados	31	Qualitativa	Ordinal
		Assistência para garantir a qualidade dos dados	32	Qualitativa	Ordinal
	Preservação de dados	Seleção de dados para preservar	33	Qualitativa	Ordinal
		Preservação de dados físicos em versão digital	34	Qualitativa	Ordinal
		Arquivamento em formatos legíveis a médio e longo prazo	35	Qualitativa	Ordinal
		Realiza migração de dados quando for necessário e verifica a obsolescência do formato periodicamente.	36-37	Qualitativa	Ordinal
		Backups de dados	38-39	Qualitativa	Ordinal
		Inspeção periódica de backups	40	Qualitativa	Ordinal
		Diretrizes padrão para preservar dados	41	Qualitativa	Ordinal
		Assistência para preservar dados	42	Qualitativa	Ordinal

Quadro 17 –Operacionalização compartilhamento de dados de pesquisa

Variável	Dimensões	Indicadores	Questionário Pergunta	Tipo de estatístico	Escala de medição
Compartilhamento de dados de pesquisa	Disponibilidade de dados de pesquisa	Quantidade de dados disponíveis	43	Qualitativa	Nominal
		Tipos de dados disponíveis	44	Qualitativa	Nominal
		Usuários que têm acesso aos dados de pesquisa do pesquisador	45	Qualitativa	Nominal
		Meios para acessar os dados de pesquisa	46	Qualitativa	Nominal
	Qualidade dos dados de pesquisa compartilhados	Dados de pesquisa estruturados e gerenciados sob diretrizes	47	Qualitativa	Ordinal
		Dados de pesquisa não estruturados ou documentados.	48	Qualitativa	Ordinal
	Condições de compartilhamento e reuso de dados de pesquisa	Condições de uso dos dados de pesquisa disponibilizados pelo pesquisador	49-56	Qualitativa	Ordinal
		Tipos de licenças	57	Qualitativa	Nominal

Fonte: Elaboração própria, 2020.

Quadro 18– Operacionalização de reuso de dados de pesquisa

Variável	Dimensões	Indicadores	Questionário Pergunta	Tipo de estatístico	Escala de medição
Reuso de dados de pesquisa	Reuso de dados próprios e de outros pesquisadores	Reuso, em uma nova pesquisa, de dados coletados em pesquisas anteriores	58	Qualitativa	Ordinal
		Reuso, em uma nova pesquisa, de dados de pesquisa sob autoria de terceiros	59	Qualitativa	Ordinal
	Qualidade dos dados	Dados de pesquisa reusados possuem metadados.	60	Qualitativa	Ordinal
	Meios de acesso a conjuntos de dados de outros pesquisadores	Canais de comunicação para acessar a dados de terceiros	61	Qualitativa	Nominal

Fonte: Elaboração própria, 2020.

6.2.7 Procedimento para a coleta de dados

A presente pesquisa envolve seres humanos, portanto, a questão ética ganha relevância. No Brasil, os protocolos de pesquisa que envolvem seres humanos necessariamente devem ser submetidos ao Comitê de Ética em Pesquisa (CEP). O principal propósito do CEP é assegurar que as pesquisas cumpram com as normas éticas estabelecidas e dentre suas responsabilidades estão a avaliação e acompanhamento dos fatores éticos a fim de salvaguardar a dignidade, os direitos, e o bem-estar dos participantes (CENTRO DE CIÊNCIAS DA SAÚDE, *online*).

O projeto foi submetido e aprovado na banca de qualificação do programa de pós-graduação em gestão nas organizações aprendentes, e no Comitê de Ética em Pesquisa do Centro de Ciências da Saúde da Universidade Federal da Paraíba CEP/CCS/UFPB. Prévio à aplicação do questionário, se explicou aos participantes, os objetivos da pesquisa, a justificativa, a importância de sua participação, os riscos e benefícios da pesquisa e, por último, declararam sua participação mediante o Termo de Consentimento Livre e Esclarecido (anexo B). Após da declaração, o questionário foi aplicado por meio da plataforma *google docs*⁴¹, sem nenhuma interferência que prejudicasse a coleta de dados.

A presente pesquisa oferece riscos considerados “mínimos” de caráter intelectual ou emocional como invasão a dados confidenciais, ou desconforto por tomar o tempo do sujeito para responder o questionário. A pesquisadora associada se compromete a prestar assistência. Para evitar tais riscos, assegura-se a confidencialidade e a privacidade das informações compartilhadas, e garante-se o sigilo e anonimato dos participantes.

6.2.8 Técnica de análise dos dados

Esta pesquisa busca caracterizar a amostra com base em cinco variáveis: 1) trajetória profissional do pesquisador 2) dados de pesquisa produzidos pelos pesquisadores 3) práticas de gestão de dados de pesquisa 4) práticas de compartilhamento de dados de pesquisa 5) práticas de uso e reúso de dados de pesquisa.

As variáveis descritas derivam das informações obtidas através do questionário aplicado, da pergunta 1 à pergunta 61, considerando que as variáveis extraídas são de dois tipos: escala nominal e escala Likert.

A caracterização das variáveis foi feita em função do tipo de instituição à qual os pesquisadores pertencem, particulares ou públicas. A caracterização é abordada mediante técnicas de estatística descritiva. A estatística descritiva se apresenta adequada para o que se pretende atingir devido a que visa descrever o fenômeno a partir do conjunto de observações registradas, tendo como propósito esclarecer a realidade.

Assim, no caso das variáveis da escala nominal, foram elaboradas tabelas de contingência em que os dados são apresentados em termos de frequência absoluta e relativa.

A distribuição de frequência trata-se das repetições agrupadas por valores da variável (MARCONI; LAKATOS, 2002). Este indicador estatístico permite compreender a oscilação

⁴¹ <https://forms.gle/rv1XXhRZnpQC6sbK7>

dos dados, isto é, quais dados são os que mais se apresentam e quais menos acontecem. A frequência se classifica em absoluta e relativa:

Frequência absoluta: número de vezes em que se dá a repetição de uma observação

Frequência relativa: representação do número de vezes em que se repete uma observação, em termos de percentagem.

É importante verificar nas variáveis nominais se existe uma relação de dependência ou associação entre o tipo de universidade e os indicadores de interesse. Para isso, é realizado um teste de qui-quadrado de Pearson. Como é conhecido, o teste de Pearson procura determinar a existência ou não de relação ou associação entre variáveis nominais. O nível de confiança utilizado para este teste é de 95%, ou seja, foi selecionado um nível de significância (α) de 5%. Lembrando que o sistema de hipóteses do teste Pearson é (LIND; MARCHAL; WATHEN, 2008):

H0: As variáveis não possuem relação ou associação estatisticamente significativa.

H1: As variáveis têm um relacionamento ou associação estatisticamente significativa.

A regra de decisão utilizada é:

Se $p\text{value} > \alpha$, a hipótese nula não é rejeitada, ao contrário se $p\text{value} < \alpha$, a hipótese nula é rejeitada, o que implica a existência de uma relação entre as variáveis.

No caso das variáveis da escala Likert, foi estabelecido um procedimento para sua interpretação. Nessa direção, optou-se por utilizar, para fins de interpretação, os indicadores estatísticos média, mediana, variância, e desvio padrão. Isso ocorre porque essas três medidas permitem apontar a atuação média dos indivíduos e a homogeneidade ou heterogeneidade presente entre eles. As outras medidas, como a moda, curtose, o coeficiente de assimetria, o máximo e o mínimo, são apresentadas com uma finalidade meramente informativa.

No quadro 19 expõe escala ordinal Likert⁴² usada para medir as respostas.

Quadro 19 - Escala Likert

Nunca	Raramente	Ocasionalmente	Frequentemente	Sempre
1	2	3	4	5

Fonte: Adaptado de Salkind, 1999.

Devido a que os resultados são analisados em termos de média, foram atribuídos os pontos indicados com a finalidade de realizar uma análise quantitativa dos resultados através uma escala de valoração e, com base nela, foram determinados os seguintes intervalos para a categorização dos resultados e sua interpretação.

⁴² A escala Likert utilizada foi adaptada conforme proposto por Salkind, 1999.

Quadro 20 - Categorização dos resultados

Categoria	Pontos
Extremamente forte	4,1 – 5
Forte	3,1 – 4,09
Neutro	2,1 – 3,09
Fraco	1,1 – 2,09
Extremamente fraco	0 – 1,09

Fonte: Adaptado de Verdugo, 2005.

As categorias são interpretadas conforme expõe o quadro 21.

Quadro 21 - Interpretação dos resultados.

Categoria	Interpretação
Extremamente forte	Em média, os pesquisadores sempre realizam a prática em questão.
Forte	Em média, os pesquisadores frequentemente realizam a prática em questão.
Neutro	Em média, os pesquisadores ocasionalmente realizam a prática em questão.
Fraco	Em média, os pesquisadores raramente realizam a prática em questão.
Extremamente fraco	Em média, os pesquisadores muito raramente realizam a prática em questão.

Fonte: Elaboração própria, 2020

Para a análise das dimensões, é importante fixar os conceitos estatísticos usados:

Quadro 22 – Conceitos estatísticos

Conceito estatístico	
Média	Medida afetada pelos valores extremos da distribuição. Representa o centro de gravidade da distribuição de frequências, considerado o valor característico em estudo que todos os elementos da amostra teriam se não houvesse diferença entre eles.
Mediana	Dados centrais que dividem a amostra em duas partes iguais. Identifica a tendência central da amostra sem ser afetada por valores extremos.
Moda	Valor da variável mais repetida em uma população ou amostra.
Variância	Soma dos quadrados dos desvios dos valores da variável em relação à sua média aritmética. É maior que ou igual a zero. É a medida da dispersão absoluta.
Desvio padrão	Raiz quadrada positiva da variância. Expressa-se na mesma unidade de medida da variável.
Coefficiente de curtose	Mede a concentração maior ou menor de dados em torno da média.
Coefficiente de assimetria	Mede o grau de distribuição dos dados em torno de sua média.
Mínimo	Valor mais baixo que a variável leva na distribuição.
Máximo	Maior valor que a variável leva na distribuição.

Fonte: Lind; Marchal, Wathen, (2008).

O processamento estatístico dos dados é realizado no software SPSS 22 ®. Convém mencionar que a interpretação estatística é analisada sob o arcabouço teórico estabelecido.

Quadro 23- Técnicas estatísticas da pesquisa

Analises	Técnica estatística
Caracterização do perfil do pesquisador	Estatística descritivo: tabela de contingência
Caracterização dos dados de pesquisa	Estatística descritivo: tabela de contingência

Analises	Técnica estatística
Caracterização das práticas de gestão, compartilhamento e reúso de dados de pesquisa	Estatística descritivo: tabela de contingência Medidas de tendência central Medidas de dispersão

Fonte: Elaboração própria, 2020.

Do ponto de vista estatístico, metodologicamente é importante destacar:

- ✓ A amostra é bastante homogênea, o que indica que dados muito diferentes não foram encontrados (valores extremos ou extremos).
- ✓ Os resultados são válidos para a população indicada, não para todos os professores de universidades públicas e particulares.
- ✓ Os dados são analisados em termos de média devido à pouca dispersão nos dados. Quando os dados são heterogêneos, o melhor indicador é a mediana, porém, neste estudo, sendo dados relativamente homogêneos, o melhor indicador é a média.

7 APRESENTAÇÃO ANÁLISE E INTERPRETAÇÃO DOS DADOS

O capítulo expõe os resultados obtidos no processamento estatístico dos dados, bem como sua interpretação através da lente teórica levantada.

7.1 PERFIL DO PESQUISADOR

Os critérios de inclusão descritos no capítulo metodológico estabelecem a participação de professores pesquisadores vinculados a instituições de ensino superior, especificamente universidades públicas e particulares. Esta delimitação deve-se às características jurídicas das universidades⁴³, nas quais têm como princípio a produção e transmissão do conhecimento através de processos de pesquisa científica de alto nível.

Conforme referido anteriormente, a amostra calculada é de 138 pesquisadores, dos quais 88 concordaram participar. Considerando que a população estudada é de 38 professores para a universidade pública, identifica-se que houve um 97% de participação. Para as universidades particulares, a população estudada é de 100 professores, obtendo uma aceitação de 51% em participação.

A fim de traçar o perfil do pesquisador, a operacionalização desta variável gerou duas dimensões, a dimensão trajetória profissional e trajetória investigativa do pesquisador. As questões que respondem às dimensões referidas são as questões 1-6 do questionário aplicado, tendo em vista que as questões extraídas são de dois tipos: escala nominal.

7.1.1 Dimensão trajetória profissional

A dimensão trajetória profissional do pesquisador é composta pelos indicadores: a) tipo de instituição a que está vinculado, b) grau de formação acadêmica, e c) tempo de atuação como professor (a) universitário (a).

Na Tabela 1, observa-se que a participação de pesquisadores de uma universidade pública é de 37 professores, abrangendo 42% do número total de participantes. Enquanto a participação de pesquisadores vinculados a universidades particulares é de 51 professores, cobrindo 58% do número total de participantes.

Tabela 1- Tipo de instituição

Itens	Frequência	Porcentagem	Porcentagem Válida	Porcentagem Acumulada
Particular	51	58,0%	58,0%	58,0%
Pública	37	42,0%	42,0%	100,0%

⁴³ A Lei 30 de 1992, que organiza o serviço público de Educação Superior, estabelece no artigo 19 que as universidades são as instituições que credenciam sua atuação nas seguintes atividades: Pesquisa científica ou tecnológica; formação acadêmica em profissões ou disciplinas, e produção, desenvolvimento e transmissão de conhecimentos e cultura universal e nacional

Itens	Frequência	Percentagem	Percentagem Válida	Percentagem Acumulada
Total	88	100,0%	100,0%	

Fonte: Dados de pesquisa, com base em SPSS 22®, 2020.

Em relação ao último grau de formação acadêmica dos participantes, a Tabela 2 registra que 3,4% têm apenas graduação, 47,7% têm mestrado completo, 1,1% estão fazendo mestrado, 12,5% estão fazendo doutorado, 26,1% têm doutorado e, finalmente, 9,1% têm pós-doutorado.

Tabela 2- Nível de formação acadêmica

Itens	Frequência	Percentagem	Percentagem Válida	Percentagem Acumulada
Doutorado completo	23	26,1%	26,1%	26,1%
Doutorado em curso	11	12,5%	12,5%	38,6%
Mestrado completa	42	47,7%	47,7%	86,4%
Mestrado em curso	1	1,1%	1,1%	87,5%
Pós-doutorado	8	9,1%	9,1%	96,6%
Graduação	3	3,4%	3,4%	100,0%
Total	88	100,0%	100,0%	

Fonte: Dados de pesquisa, com base em SPSS 22®, 2020.

O tempo de atuação como professor, conforme exposto na Tabela 3, indica que 31,8% dos participantes têm entre 5 e 10 anos atuando, 25% entre 11 e 15 anos, 18,5% têm mais de 20 anos, 12,5% têm entre 11 e 15 anos e os 12,5% restantes têm menos de 5 anos.

Tabela 3- Tempo de atuação como professor

Itens	Frequência	Percentagem	Percentagem Válida	Percentagem Acumulada
Entre 11 a 15 anos	22	25,0	25,0	25,0
Entre 16 a 20 anos	11	12,5	12,5	37,5
Entre 5 a 10 anos	28	31,8	31,8	69,3
Más de 20 anos	16	18,2	18,2	87,5
Menos de 5 anos	11	12,5	12,5	100,0
Total	88	100,0	100,0	

Fonte: Dados de pesquisa, com base em SPSS 22®, 2020.

É de interesse traçar o perfil dos pesquisadores de acordo com o tipo de instituição à qual está vinculado profissionalmente, por este motivo, foi tomada a decisão de relacionar o tipo de instituição com as variáveis de nível de formação acadêmica, o tempo de atuação como professor universitário, através do uso de técnicas de contingência.

A Tabela 4 mostra a relação entre o tipo de instituição e o nível de formação acadêmica do pesquisador. Nas universidades particulares, 13,6% possuem doutorado completo, 6,8% possuem doutorado em curso, 28,4% possuem mestrado completo, 1,1% cursando mestrado, 4,5% pós-doutorado e 3,4% tem graduação. Na universidade pública, 12,5% possuem doutorado completo, 5,7% possuem doutorado em curso, 19,3% mestrado completo, 1,1% estão

cursando mestrado, 4,5% possuem pós-doutorado. Conclui-se que na universidade particular e pública, a maioria dos professores têm mestrado completo, seguido de doutorado.

Tabela 4 - Tipo de instituição – Nível de formação acadêmica

Itens	Estatística	Particular	Público	Total
Pós-doutorado	Frequência	4	4	8
	% do total	4,5%	4,5%	9,1%
Doutorado completo	Frequência	12	11	23
	% do total	13,6%	12,5%	26,1%
Doutorado em curso	Frequência	6	5	11
	% do total	6,8%	5,7%	12,5%
Mestrado completo	Frequência	25	17	42
	% do total	28,4%	19,3%	47,7%
Mestrado em curso	Frequência	1	0	1
	% do total	1,1%	0,0%	1,1%
Graduação	Frequência	3	0	3
	% do total	3,4%	0,0%	3,4%
Total	Frequência	51	37	88
	% do total	58,0%	42,0%	100,0%

Fonte: Dados de pesquisa, SPSS 22®, 2020

A Tabela 5 reflete a relação entre o tipo de instituição e o tempo atuando como professor. Na universidade particular, 21,6% têm entre 5 e 10 anos, seguidos por 13,6% entre 11 e 15 anos, 9,1% têm mais de 20 anos, enquanto outros 9,1% menos de 5 anos, e 4,5% têm entre 16 e 20 anos. Na universidade pública, 10,2% tinham entre 5 e 10 anos, seguido por 11,4% entre 11 e 15 anos, 9,1% tinham mais de 20 anos, 8% tinham entre 16 e 20 anos e 3,4% tinham menos de 5 anos.

Tabela 5 - Tipo de instituição - Tempo como professor

Itens	Estatística	Particular	Público	Total
Más de 20 anos	Frequência	8	8	16
	% do total	9,1%	9,1%	18,2%
Entre 16 a 20 anos	Frequência	4	7	11
	% do total	4,5%	8,0%	12,5%
Entre 11 a 15 anos	Frequência	12	10	22
	% do total	13,6%	11,4%	25,0%
Entre 5 a 10 anos	Frequência	19	9	28
	% do total	21,6%	10,2%	31,8%
Menos de 5 anos	Frequência	8	3	11
	% do total	9,1%	3,4%	12,5%
Total	Frequência	51	37	88
	% do total	58,0%	42,0%	100,0%

Fonte: Dados de pesquisa, SPSS 22®, 2020

Concluiu-se que, segundo o tipo de instituição e do tempo como professor, a universidade particular e pública difere. Na universidade particular, a maioria dos professores têm entre 5 e 10 anos, enquanto na universidade pública, a maioria dos professores têm entre 11 e 15 anos.

7.1.2 Dimensão trajetória investigativa do pesquisador

A dimensão da trajetória de pesquisa do professor é determinada a partir de a) categoria atual como pesquisador classificado pela Colciencias, b) no tempo de atuação como

pesquisador e, finalmente, c) na área de conhecimento que o pesquisador explora em sua produção acadêmica.

Colciencias, como entidade pública administrativa de ciência, tecnologia e inovação, através de chamadas anuais, aplica o modelo de reconhecimento e medição de grupos de pesquisa e pesquisadores vinculados contratualmente a instituições de ensino superior e centros científicos. O modelo de medição procura tornar visíveis as informações atualizadas dos grupos de pesquisa referentes a perfis científicos dos integrantes do grupo e produção científica. (COLCIENCIAS, 2018).

Da mesma forma, na chamada para medição de grupos de pesquisa, os pesquisadores atualizam sua tipologia, ou seja, sua categoria de acordo com suas características profissionais e de pesquisa. Neste sentido, o quadro 24 indica as categorias concedidas aos pesquisadores vinculados aos grupos de pesquisa reconhecidos e classificados pela Colciencias com base em sua produção e trajetória acadêmico-científica:

Quadro 24. Categoria Colciencias

PESQUISADORES				
Características	Investigador Emérito	Investigador Sênior	Investigador Associado	Investigador Junior
Nível de formação	Doutorado concluído ou 15 produtos científicos de novo conhecimento ⁴⁴ ou atividades de desenvolvimento tecnológico, e inovação ⁴⁵ de alto nível Tipo A.	Doutorado concluído ou 15 novos produtos científicos ou de desenvolvimento tecnológico e inovação.	Doutorado ou mestrado concluído, ou 7 novos produtos de conhecimento ou de desenvolvimento tecnológico e inovação.	Doutorado, mestrado ou graduação, com 7 produtos de novo conhecimento e de desenvolvimento tecnológico e inovação.
Produção mínima	Produtos tipo A	10 produtos tipo A ⁴⁶ nos últimos 10 anos.	Mínimo 2 do tipo A ao longo da trajetória científica.	Mínimo 1 produto científico tipo A nos últimos cinco anos
Produtos de formação	Diretor de tese de doutorado ou mestrado.	Diretor de 4 projetos de mestrado ou 1 de doutorado concluídos nos últimos 10 anos.	Diretor de 1 tese de doutorado, ou 2 de mestrados, ou 8 projetos de graduação nos últimos cinco anos	
MEMBROS VINCULADOS				
Características	Membro com doutorado	Membro com mestrado	Membro vinculado com especialização	Sem categoria

⁴⁴ Artigos de Pesquisa (Q1, Q2, Q3, Q4). Notas Científicas Públicas em Periódicos. Livros de resultados de pesquisa. Capítulos em livros resultantes de pesquisas. Produtos resultantes de novas atividades de geração de conhecimento: produtos tecnológicos patenteados ou em processo de fabricação.

⁴⁵ Produtos tecnológicos certificados ou validados. Produtos de negócios. Regulamentos, normas ou legislação. Conceitos técnicos e relatórios técnicos. Registros de contratos de licença para exploração de obras protegidas por direitos autorais.

⁴⁶ Nomenclatura para se referir a produtos científicos de alto nível.

PESQUISADORES				
Características	Investigador Emérito	Investigador Sênior	Investigador Associado	Investigador Junior
Nível de formação	Doutorado concluído. Vinculado a um grupo de pesquisa.	Mestrado concluído. Vinculado a um grupo de pesquisa.	Especialização concluída. Vinculado a um grupo de pesquisa.	Não é reconhecido por Colciencias como pesquisador.

Fonte: Colciencias, 2018.

A categorização dos pesquisadores é realizada a partir das informações registradas em seus currículos digitais (CvLAC). Ao verificar se os requisitos atribuídos para cada categoria são atendidos, os pesquisadores são classificados como emérito, sênior, associado e júnior.

É importante mencionar que a categoria mais alta que pode ser obtida é emérito, seguida de sênior, associado e finalmente júnior, determinado pela qualidade de sua produção científica.

Por outro lado, se o perfil do pesquisador não atender aos requisitos das categorias mencionadas, verifica-se se ele aplica para os outros tipos de categoria: a) estudante de mestrado ou doutorado, b) jovem pesquisador ou c) membro vinculado.

A diferença entre pesquisadores e membros vinculados é sua produção científica. Os pesquisadores têm uma trajetória acadêmica que se reflete nos produtos mencionados em seu currículo. Já os membros vinculados não possuem a mesma quantidade e qualidade de produtos científicos sob sua autoria que um pesquisador tem.

No caso de pesquisadores (emerito, sênior, associados, júnior), como requisito para pertencer a essa categoria, eles devem possuir novos produtos de conhecimento ou produtos tecnológicos ou de inovação do tipo A, ou seja, avaliados e com critérios de alta qualidade.

Com relação a este contexto, conforme é possível observar na Tabela 6, 27,3% dos participantes são membros vinculados com mestrado, 20,5% são pesquisadores associados, 19,3% são pesquisadores Junior, 17% classificam como pesquisadores Sênior, 12,5% não estão categorizados, e 3,4% dos pesquisadores são reconhecidos como membro vinculado com doutorado.

Tabela 6- Categoria Colciencias

Itens	Frequência	Percentagem	Percentagem Válida	Percentagem Acumulada
Associado	18	20,5%	20,5%	20,5%
Membro vinculado com doutorado	3	3,4%	3,4%	23,9%
Membro vinculado com mestrado	24	27,3%	27,3%	51,2%
Júnior	17	19,3%	19,3%	70,5%
Sênior	15	17,0%	17,0%	87,5%
Sem categorização	11	12,5%	12,5%	100,0%
Total	88	100,0%	100,0%	

Fonte: Dados de pesquisa, SPSS 22 ®, 2020.

Quanto ao tempo atuando como pesquisador, 35,2% têm entre 5 e 10 anos, 30,7% têm menos de 5 anos, 17% têm um tempo como pesquisador entre 11 e 15 anos, 10,2% têm entre 16 e 20 anos, enquanto 6,8% têm mais de 20 anos, como expõe a Tabela 7.

Tabela 7- Tempo de atuação como pesquisador

Itens	Frequência	Percentagem	Percentagem Válida	Percentagem Acumulada
Entre 11 a 15 anos	15	17,0%	17,0%	17,0%
Entre 16 a 20 anos	9	10,2%	10,2%	27,3%
Entre 5 a 10 anos	31	35,2%	35,2%	62,5%
Más de 20 anos	6	6,8%	6,8%	69,3%
Menos de 5 anos	27	30,7%	30,7%	100,0%
Total	88	100,0%	100,0%	

Fonte: Dados de pesquisa, SPSS 22 ®, 2020.

Colciencias adere-se à classificação das principais áreas de conhecimento desenvolvidas pela OCDE para organizar os grupos de pesquisa e seus pesquisadores de acordo com o campo de ação que investigam. A área de economia e negócios, segundo a OCDE (2007), abrange as disciplinas de economia, econometria, relações industriais, finanças, contabilidade, administração, negócios e gestão. As linhas de pesquisa desenvolvidas pelos pesquisadores podem ser vistas no quadro 25.

Quadro 25- Linhas de pesquisa

Área	Linhas de pesquisas
Administração	Gestão do conhecimento Marketing Estratégia de negócios Talentos humanos Gestão organizacional Gestão tecnológica Empreendedorismo e sustentabilidade Inovação Desenvolvimento empresarial Eficiência e Produtividade Competitividade Responsabilidade social Tecnologias da informação e comunicação Segurança e saúde ocupacional Finanças Consumidor Indústria
Economia	Economia social Economia ambiental Economia agrícola Economia do trabalho Economia da educação Desigualdade e pobreza Desenvolvimento Econômico Estatísticas
Negócios	Operações, logística e transporte Estratégias e internacionalização Agronegócio Clima de negócios Sobrevivência nos negócios

Fonte: Dados de pesquisa, 2020

A continuação se traça o perfil dos pesquisadores de acordo com o tipo de instituição e a relação com as variáveis categoria como pesquisador classificado pela Colciencias, e o tempo de atuação como pesquisador, a partir de tabelas de contingência.

A relação entre o tipo de instituição e a categoria em Colciencias é exposto na Tabela 8. Na universidade particular, 19,3% dos professores estão classificados como membros com Mestrado, 12,5% são pesquisadores Júnior, 10,2% não são categorizados, 6,8% são pesquisadores Sênior, e outros 6,8% são pesquisadores Associados, por último, 2,3% são membros com Doutorado. Na universidade pública, 13,6% são pesquisadores Associados, 10,2% são pesquisadores Sênior, 8% são classificados membros vinculados com Mestrado, 6,8% são pesquisadores Júnior, 2,3% não são categorizados, e 1,1% são membros com Doutorado.

Tabela 8. Tipo de instituição - Categoría Colciencias

Itens	Estatística	Particular	Público	Total
Associado	Frequência	6	12	18
	% do total	6,8%	13,6%	20,5%
Membro vinculado com Doutorado	Frequência	2	1	3
	% do total	2,3%	1,1%	3,4%
Membro vinculado com Mestrado	Frequência	17	7	24
	% do total	19,3%	8%	27,3%
Júnior	Frequência	11	6	17
	% do total	12,5%	6,8%	19,3%
Sênior	Frequência	6	9	15
	% do total	6,8%	10,2%	17,0%
Sem categorização	Frequência	9	2	11
	% do total	10,2%	2,3%	12,5%
Total	Frequência	51	37	88
	% do total	58,0%	42,0%	100,0%

Fonte: Dados de pesquisa, SPSS 22 ®, 2020.

O perfil dos professores da universidade pública e particular em relação à categorização segundo Colciencias apresenta diferenças. Na universidade particular, os professores são principalmente membros com mestrado, seguidos por pesquisadores Júnior e Sênior. Na universidade pública, são principalmente associados, seguidos por pesquisadores Sênior e Júnior. Isso indica que, em termos de qualidade na produção acadêmica, os pesquisadores da universidade pública estão melhor posicionados no contexto científico.

A relação entre tipo de instituição e tempo como pesquisador de acordo com a Tabela 9 evidencia que nas universidades particulares 22,7% dos professores têm entre 5 e 10 anos, 19,3% têm menos de 5 anos, 8% têm entre 11 e 15 anos, 4,5% têm entre 16 e 20 anos, e apenas 3,4% têm mais de 20 anos. Na universidade pública, 12,5% têm entre 5 e 10 anos, 11,4% têm menos de 5 anos, 9,1% têm entre 11 e 15 anos, 5,7% têm entre 16 e 20 anos, e 3,4% são pesquisadores com mais de 20 anos na sua área.

Tabela 9. Tipo de instituição – Tempo como pesquisador

Itens	Estatística	Particular	Público	Total
Más de 20 anos	Frequência	3	3	6
	% do total	3,4%	3,4%	6,8%
Entre 16 a 20 anos	Frequência	4	5	9
	% do total	4,5%	5,7%	10,2%
Entre 11 a 15 anos	Frequência	7	8	15
	% do total	8,0%	9,1%	17,0%
Entre 5 a 10 anos	Frequência	20	11	31
	% do total	22,7%	12,5%	35,2%
Menos de 5 anos	Frequência	17	10	27
	% do total	19,3%	11,4%	30,7%
Total	Frequência	51	37	88
	% do total	58,0%	42,0%	100,0%

Fonte: Dados de pesquisa, SPSS 22 ®, 2020

De acordo com o tempo de pesquisador, os professores da universidade, pública e particular, têm um perfil semelhante. A maioria tem entre 5 e 10 anos, seguida por menos de 5 anos, e entre 11 e 15 anos.

1. Considerações das dimensões trajetória profissional e investigativa do pesquisador

É importante verificar se existe uma relação de dependência ou associação entre as variáveis de interesse. Para isso, é realizado um teste de qui-quadrado de Pearson. O nível de confiança utilizado para este teste é de 95%, ou seja, foi selecionado um nível de significância (α) de 5%. Se $P\text{value} > \alpha$, a hipótese nula não é rejeitada, ao contrário se $P\text{value} < \alpha$, a hipótese nula é rejeitada, o que implica a existência de uma relação entre as variáveis. A Tabela 10 expõe os Pvalues associados ao estatístico Qui-quadrado de Pearson da relação entre variáveis pares.

Tabela 10- Pvalue teste Qui-quadrado de Pearson

ITENS	Universidade vinculado	Formação acadêmica	Tempo como professor universitário	Categoria Colciencias	Tempo como investigador
Universidade vinculado	-	0,620	0,315	0,016	0,665
Formação acadêmica	0,620	-	0,053	0,000	0,000
Tempo como professor universitário	0,315	0,053	-	0,114	0,003
Categoria Colciencias	0,0016	0,000	0,114	-	0,019
Tempo como pesquisador	0,665	0,000	0,003	0,019	-

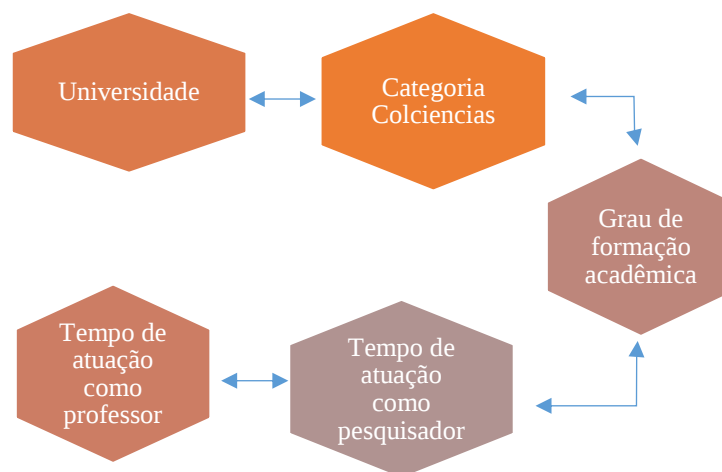
Fonte: Dados de pesquisa, SPSS 22 ®, 2020

- ✓ O tipo de instituição e as categorias de Colciencias rejeitam a hipótese nula, ou seja, essas variáveis têm uma relação de dependência ou associação.

- ✓ As variáveis nível de formação acadêmica e as categorias de Colciencias também rejeitam a hipótese nula, ou seja, que as variáveis em questão têm uma relação estatisticamente significativa.
- ✓ As variáveis nível de formação acadêmica e tempo atuando como pesquisador apresentam um valor menor que o nível de significância selecionado, o que implica em rejeitar a hipótese nula, resultando em associação estatisticamente significativa entre essas variáveis.
- ✓ Encontrou-se uma relação ou associação estatisticamente significativa entre o tempo como professor universitário e o tempo como pesquisador, caso em que a hipótese nula é rejeitada porque o Pvalue é inferior ao nível de significância de 5%.

A Figura 6 expõe as características do pesquisador que têm significância bilateral, ou seja, há uma relação nas duas direções. Essas associações fazem sentido, pois o fato de estarem registradas em uma das categorias de Colciencias implica que cumpre com os requisitos de estar vinculado contratualmente a uma instituição de ensino superior, e ter formação acadêmica superior. Da mesma forma, a relação entre o nível de graduação e pós-graduação e o tempo atuando como pesquisador é gerada pelo seu propósito de formar e qualificar em pesquisa. O tempo atuando como professor e o tempo como pesquisador estão associados ao fato de serem tarefas próximas e indissolúveis na academia.

Figura 6- Significância bilateral do perfil do pesquisador



Fonte: Elaboração própria, 2020.

Com base no exposto, os pesquisadores podem ser perfilados de acordo com o tipo de universidade a que pertencem e as características predominantes.

Tabela 11- Perfil dos pesquisadores de acordo com suas características

ITENS	Pesquisadores de Universidade Particular	Pesquisadores de Universidade Pública
Formação acadêmica	Mestrado completo	Mestrado completo
Tempo como professor universitário	Entre 5 e 10 anos	Entre 11 e 15 anos
Categoria Colciencias	Membro vinculado com mestrado	Associado
Tempo como pesquisador	Entre 5 e 10 anos	Entre 5 e 10 anos

Fonte: Elaboração própria, 2020.

A partir das características predominantes, assume-se que:

Os pesquisadores de universidade pública possuem mais experiência como docentes universitários, além de ter um maior reconhecimento investigativo por sua carreira acadêmica e seus produtos de alto impacto científico. Os professores de universidade particular possuem menos tempo como professores universitários. Sua categoria predominante, membro vinculado com mestrado, indica que os produtos científicos, avaliados pela Colciencias, têm menos impacto em comparação aos professores de universidade pública. Em termos de formação acadêmica e tempo atuando como pesquisador, as características são as mesmas para os pesquisadores de universidade privada e pública, no entanto apresenta-se, conforme mencionado, diferenças notáveis na quantidade e qualidade da produção científica.

A universidade pública e particular, dadas as particularidades de cada uma, estabelecem critérios de trajetória profissional e investigativa, que diferem e se manifestam no tempo de experiência como professor e na sua categoria de Colciencias.

7.2 CARATERIZAÇÃO DOS DADOS DE PESQUISA

Os dados de pesquisa das disciplinas de contabilidade, finanças, economia, indústria, administração, negócios e afins têm certas peculiaridades decorridas das particularidades dos seus métodos de pesquisa. Tais características dos dados determinam como esse recurso deve ser curado, protegido e preservado. Assim, o reconhecimento da proveniência, natureza e caráter desse recurso torna possível identificar os métodos tecnológicos e de gerenciamento apropriados pelos quais eles devem passar para manter intactas suas características.

Para conhecer as características dos dados produzidos pelos pesquisadores, são levadas em consideração as questões 7 a 11 do questionário aplicado. É de interesse traçar o perfil dos pesquisadores de acordo com o tipo de instituição a que pertencem e as particularidades dos dados que produzem e tratam. A decisão foi tomada para relacionar o tipo de instituição com os indicadores associados à natureza dos dados, tipos de dados, formatos, e tempo de armazenamento.

7.2.1 Dimensão particularidades dos dados

A diversidade dos dados de pesquisa, como consequência dos propósitos da investigação e as diferentes metodologias aplicadas para sua coleta e processamento, irá determinar os métodos pelos quais os dados serão gerenciados. Segundo Sayão e Sales (2015), reconhecer os vários tipos de dados torna-se crucial para o estabelecimento dos processos de preservação e curadoria dos dados. A dimensão particularidades dos dados tem como objetivo caracterizar as seguintes questões relacionadas aos dados:

- ✓ Natureza dos dados
- ✓ Caráter dos dados
- ✓ Tipo de formato que arquiva os dados
- ✓ Tempo que armazena os dados
- ✓ Tipo de armazenamento dos dados

A fim de caracterizar as particularidades dos dados são levadas em consideração as questões 7 a 11 do questionário aplicado, tendo em vista que as questões extraídas são de tipo escala nominal. O propósito da dimensão é explicitar a relação que o pesquisador tem com seus dados.

1. Natureza dos dados

A relação entre o tipo de instituição e a natureza dos dados utilizados pelos pesquisadores observa-se na Tabela 12. Na universidade particular, 8% dos pesquisadores usam dados qualitativos, 12% usam dados quantitativos e 80% usam dados mistos. Na universidade pública, 6% dos pesquisadores produzem dados qualitativos, 24% produzem dados quantitativos e 70% produzem dados mistos. Concluiu-se que os pesquisadores majoritariamente produzem dados mistos.

Tabela 12- Natureza dos dados

Itens	Estatística	Particular	Público	Total
Dados qualitativos	Frequência	4	2	6
	Porcentagem	8%	6%	6,8%
Dados quantitativos	Frequência	6	9	15
	Porcentagem	12%	24%	17%
Mistos	Frequência	41	26	67
	Porcentagem	80%	70%	76,1%
Total	Frequência	51	37	88
	% do total	100%	100%	100,0%

Fonte: Dados de pesquisa, SPSS 22 ®, 2020

2. Caráter dos dados

A relação entre o tipo de instituição e o tipo de dados produzidos pelo pesquisador, observa-se na Tabela 13. Como em algumas categorias apenas um ou dois pesquisadores

indicaram usá-las, a continuação mostra-se os principais resultados. Na universidade particular, 29% indicam que produzem dados numérico-textuais, 29% somente texto e 22% apenas dados numéricos. Na universidade pública, 22% indicam produzir dados numéricos, 27% dados textuais e outros 27% produzem dados numéricos-textuais. Conclui-se que comumente os pesquisadores produzem dados de tipo numérico e textual.

Tabela 13- Tipos de dados

Itens	Particular	Pública	Total
Auditivos-Software	2%		1,1%
Numéricos	22%	22%	21,6%
Numéricos-Auditivos	2%		1,1%
Numéricos-Software	6%	5%	5,7%
Numéricos-Textuais	29%	27%	28,4%
Numéricos-Textuais-Auditivos	4%		2,3%
Numéricos-Textuais-Software	2%	6%	3,4%
Numéricos-Textuais-Visuais-Software		5%	2,3%
Textuais	29%	27%	28,4%
Textuais-Auditivos	2%		1,1%
Textuais-Software		3%	1,1%
Textuais-Visuais-Auditivos-Software		5%	2,3%
Visuais-Software	2%		1,1%
Total	100%	100%	100,0%

Fonte: Dados de pesquisa, SPSS 22 ®, 2020

Como pode ser visto, os tipos de dados produzidos tendem a ser heterogêneos, o que implica na necessidade de estabelecer estratégias de gestão diversificadas para garantir seu adequado status. O esclarecimento da natureza e tipo de dados dita os mecanismos de organização e estruturação dos dados, o método de descrição dos dados, formatos e transformação dos arquivos, o tipo mais apropriado de armazenamento, técnicas de garantia, entre outros fatores que fazem parte do gerenciamento dos dados.

3. Tipo de formato que arquiva os dados

Em relação ao formato, a Tabela 14 mostra a relação entre o tipo de instituição e o tipo de formato em que os pesquisadores armazenam seus dados. Tanto em universidade particular quanto pública, os pesquisadores lidam com documentos de texto em arquivos Doc, Txt, pdf, planilhas em arquivos Excel, XML, CSV, XLS, e dados estatísticos em arquivos do tipo SPSS, SAS, SAV, SPV, sendo os formatos de maior uso.

A utilização desses tipos de formatos deve-se a que o formato é selecionado de acordo com o conteúdo dos dados, sendo assim que os pesquisadores produzem, em sua maioria, dados textuais e numéricos, como expõe a Tabela 13.

Tabela 14- Formatos de dados

Itens	Particular	Pública	Total
Documento de texto (Doc, Txt, pdf)	37%	34%	36%

Itens	Particular	Público	Total
Planilhas (Excel, XML, CSV, XLS)	20%	20%	20%
Dados estatísticos (SPSS, SAS, SAV, SPV)	18%	15%	17%
Imagens / gráficos (JPEG, GIF)	11%	12%	12%
Áudio (mp3, wav)	8%	3%	5%
Software	4%	9%	6%
Web (html)	2%	7%	4%
Total	100%	100%	100%

Fonte: Dados de pesquisa, SPSS 22 ®, 2020

A seleção de formatos de arquivo apropriados garantirá o acesso e a usabilidade dos dados a médio e longo prazo. A fragilidade dos dados não é apenas uma característica dos formatos físicos, ela está presente até mesmo na obsolescência dos formatos digitais ou nas diferentes ameaças informáticas.

De acordo com *The University of Edinburgh (ONLINE, 2020)*, o depósito de dados em formatos padrão certifica que eles podem ser abertos em diferentes sistemas operacionais e com diferentes programas ou aplicações, além de garantir que não haja perda de informações. Neste caso, pode-se observar que os formatos digitais utilizados são padrão e legíveis em diversos programas, o que legitima a máxima longevidade da informação arquivada. Entretanto, apesar dos formatos utilizados pelos pesquisadores serem adequados, as constantes mudanças tecnológicas a que os formatos estão sujeitos, tornam necessário que a migração de formatos seja um mecanismo indispensável na atividade de pesquisa, um tema que será mais explorado na variável práticas de gestão.

4. Tempo que armazena os dados

A Tabela 15 indica a relação entre o tipo de instituição e o tempo de armazenamento dos dados pelos pesquisadores. Os pesquisadores de universidade particular, 41% armazenam dados indefinidamente, 29% o fazem entre 1 e 5 anos, 24% mantêm seus dados por mais de cinco anos, 4% declararam não saber e 2% armazena essas informações por menos de 1 ano. Na universidade pública, 8% armazenam dados por menos de 1 ano, 8% não sabem, 30% armazenam dados por mais de 5 anos, 32%, entre 1 a 5 anos, e 22% as armazenam indefinidamente.

Tabela 15- Tempo de armazenamento dos dados

Itens	Estatística	Particular	Pública	Total
Mais de 5 anos	Frequência	12	11	23
	Porcentagem	24%	30%	26%
Entre 1 a 5 anos	Frequência	15	12	27
	Porcentagem	29%	32%	31%
Menos de 1 ano	Frequência	1	3	4
	Porcentagem	2%	8%	4%
Indefinidamente	Frequência	21	8	29

Itens	Estatística	Particular	Pública	Total
	Porcentagem	41%	22%	33%
Não sei	Frequência	2	3	5
	Porcentagem	4%	8%	6%
Total	Frequência	51	37	88
	% do total	58,0%	42,0%	100,0%

Fonte: Dados de pesquisa, SPSS 22 ®, 2020

Observa-se que os pesquisadores de universidade particular armazenam indefinidamente os dados, enquanto os pesquisadores de universidade pública o armazenamento de dados é entre 1 a 5 anos. A avaliação do tempo pertinente de retenção de dados é uma questão que pode depender das políticas institucionais, repositórios onde são depositados os dados, ou dos argumentos pessoais ou acadêmicos do pesquisador.

A estimativa do tempo de preservação se determina a partir de diversos critérios como, o grau de dificuldade do processo de pesquisa para obter os dados; os resultados da pesquisa e da necessidade de replicá-los no futuro; o alto valor informativo e científico que os dados possuem; ou as experiências subjetivas de cada pesquisador. Nesse aspecto, não há um adequado parâmetro que determine o tempo correto para reter os dados, porém os pesquisadores devem estabelecer, de acordo com suas necessidades e as exigências das instituições ou repositórios, o tempo em que serão armazenados, evitando que os dados sejam armazenados por tempo indefinido. Armazenar os dados sem um tempo definido causa desperdício de espaço, e aumento na possibilidade de obsolescência e abandono dos dados. Considera-se que, por precaução, os dados devem ser armazenados entre 1 a 5 anos, ou mais de 5 anos.

5. Tipo de armazenamento dos dados

Em relação ao tipo de armazenamento utilizado para arquivar os dados, esta questão teve várias opções possíveis que são apresentadas a continuação. O tipo de armazenamento selecionado determina a capacidade de acesso aos dados por outros pesquisadores, bem como indica a possível fragilidade de perda dos dados.

Observa-se na Tabela 16 que 47% dos pesquisadores não armazenam dados em forma física, 45% armazenam parte deles. 70% dos pesquisadores armazenam todos os seus dados no computador pessoal, 24% armazenam parte de seus dados no computador pessoal. Em referente ao armazenamento de dados no computador do trabalho, 51% armazenam parte dos dados dessa maneira, 32% armazenam parte dos dados no computador do trabalho, enquanto 17% não armazenam nenhum dos dados desta forma.

Tabela 16- Tipo de armazenamento interno

Armazenamento	Itens	Particular	Público	Total
Físico	Todos os meus dados	8%	8%	8%
	Parte dos meus dados	49%	41%	45%

Armazenamento	Itens	Particular	Público	Total
	Nenhum dos meus dados	43%	51%	47%
	Total	100%	100%	100%
Computador pessoal	Todos os meus dados	72%	68%	70%
	Parte dos meus dados	20%	30%	24%
	Nenhum dos meus dados	8%	2%	6%
	Total	100%	100%	100%
Computador de trabalho	Todos os meus dados	35%	27%	32%
	Parte dos meus dados	55%	46%	51%
	Nenhum dos meus dados	10%	27%	17%
	Total	100%	100%	100%

Fonte: Dados de pesquisa, SPSS 22®, 2020

Os resultados expressados na Tabela 16, indicam que é comum que os pesquisadores armazenem todos os seus dados no computador pessoal, bem como parte de seus dados no computador do trabalho. Em relação ao armazenamento físico, evidencia-se que metade dos pesquisadores não armazena nenhum dos seus dados por esse meio, e a outra metade armazena parte de seus dados. Esses três locais de armazenamento são considerados de alto risco de perda de dados, bem como meios de acesso restritos para uma possível abertura e compartilhamento de dados.

Conforme a Tabela 17, em relação ao armazenamento de dados em um disco rígido externo, 40% dos pesquisadores armazenam todos os dados, 34% armazenam parte deles e 26% não armazena nenhum dos seus dados nesse meio.

Em relação ao armazenamento de dados em dispositivos USB, 47% dos pesquisadores armazenam parte de seus dados dessa maneira, 37% não armazenam nenhum dado em USB e 16% armazenam todos os dados em USB.

Já o armazenamento de dados na nuvem (*Dropbox, Drive, etc.*), 50% dos pesquisadores armazenam parte dados dessa maneira, 39% armazenam todos os dados na nuvem, enquanto 11% dos pesquisadores não armazenam nenhum dos seus dados na nuvem.

Tanto os dispositivos USB quanto os discos rígidos externos são considerados locais de armazenamento fáceis e baratos; no entanto, do ponto de vista da preservação a longo prazo e do acesso aberto aos dados, não é considerado o meio mais apropriado já que pode implicar problemas de obsolescência, ameaças informáticas, e acesso restrito. Seu uso deve ser acompanhado por outro método de armazenamento.

Tabela 17- Tipo de armazenamento externo

Armazenamento	Itens	Particular	Público	Total
Disco rígido externo	Todos os meus dados	35%	46%	40%
	Parte dos meus dados	35%	32%	34%
	Nenhum dos meus dados	30%	22%	26%
	Total	100%	100%	100%

Armazenamento	Itens	Particular	Público	Total
USB/pendrive	Todos os meus dados	8%	27%	16%
	Parte dos meus dados	47%	46%	47%
	Nenhum dos meus dados	45%	27%	37%
	Total	100%	100%	100%
Nuvem (Dropbox, Google Drive)	Todos os meus dados	43%	32%	39%
	Parte dos meus dados	45%	57%	50%
	Nenhum dos meus dados	12%	11%	11%
	Total	51	37	88

Fonte: Dados de pesquisa com base em SPSS 22®, 2020

Em relação ao armazenamento em nuvem, ele é considerado um espaço relativamente seguro e adequado em comparação com um dispositivo físico, no entanto, dependerá das necessidades do pesquisador se os critérios de uso do servidor, suas políticas de segurança e a quantidade de espaço que oferta é apropriado ou não. Do ponto de vista do acesso aberto aos dados, esse meio só terá impacto nos usuários que o pesquisador permita acessar, tendo em vista que sua visibilidade é limitada.

Em relação ao armazenamento de dados em repositórios, observa-se na Tabela 18 que 48% dos pesquisadores não armazenam nenhum dado no repositório de seu grupo de pesquisa, 40% armazenam parte dos dados, enquanto 12% dos pesquisadores armazenam todos os dados. Também se observa que 50% dos pesquisadores não armazenam nenhum de seus dados no repositório de sua instituição, 40% armazenam parte de seus dados sob essa metodologia, e 10% armazenam todos os seus dados.

Tabela 18 -Tipo de armazenamento em Repositórios institucionais

Armazenamento	Itens	Particular	Público	Total
Repositório do meu grupo de pesquisa.	Todos os meus dados	6%	22%	12%
	Parte dos meus dados	49%	27%	40%
	Nenhum dos meus dados	45%	51%	48%
	Total	100%	100%	100%
Repositório da minha Instituição.	Todos os meus dados	6%	16%	10%
	Parte dos meus dados	43%	35%	40%
	Nenhum dos meus dados	51%	49%	50%
	Total	100%	100%	100%

Fonte: Dados de pesquisa, SPSS 22®, 2020

Os resultados gerais indicam que a metade dos pesquisadores depositam em parte ou todos seus dados no repositório do grupo de pesquisa e da instituição, enquanto que a outra metade não deposita nenhum dos seus dados. No entanto, observa-se que prevalece o uso do repositório do grupo de pesquisa nas universidades particulares. Na universidade pública predomina o uso do repositório da instituição.

O armazenamento de dados no repositório da área de conhecimento, conforme exposto na Tabela 19, 66% não armazenam nenhum dado nessa modalidade, 24% armazenam parte de seus dados, e 10% armazenam todos os dados. Por outro lado, em relação ao armazenamento

em repositórios de redes científicas, 67% não armazenam nenhum dado nessa modalidade, 27% dos pesquisadores armazenam parte de seus dados, enquanto 6% armazenam todos os dados.

Em relação aos repositórios nacionais e internacionais, observa-se que 67% dos pesquisadores não armazenam seus dados nesses meios de comunicação científica, 25% armazenam parte de seus dados, e 8% armazenam todos os dados.

Tabela 19- Tipo de armazenamento em Repositórios externos

Armazenamento	Itens	Particular	Público	Total
Repositórios Temáticos	Todos os meus dados	8%	14%	10%
	Parte dos meus dados	25%	22%	24%
	Nenhum dos meus dados	67%	65%	66%
	Total	100%	100%	100%
Repositórios de Redes científicas	Todos os meus dados	6%	5%	6%
	Parte dos meus dados	27%	27%	27%
	Nenhum dos meus dados	67%	68%	67%
	Total	100%	100%	100%
Repositórios Nacionais	Todos os meus dados	10%	5%	8%
	Parte dos meus dados	23%	27%	25%
	Nenhum dos meus dados	67%	68%	67%
	Total	100%	100%	100%
Repositórios Internacionais	Todos os meus dados	6%	11%	8%
	Parte dos meus dados	27%	22%	25%
	Nenhum dos meus dados	67%	68%	67%
	Total	100%	100%	100%

Fonte: Dados de pesquisa, SPSS 22®, 2020

Os repositórios são considerados como o local de armazenamento mais eficiente para garantir a preservação e o acesso aos dados a um maior público. Entre os vários benefícios de armazenar os dados em repositórios temáticos, redes científicas, nacionais e internacionais, são mencionados a adequada preservação a longo prazo, a maximização de sua visibilidade devido à interoperabilidade dos repositórios, e o fácil acesso aos dados mediante um motor de busca como o *google*.

Os resultados mostram que mais de 66% dos participantes não armazenam seus dados em repositórios temáticos, de redes científicas, nacionais e internacionais, implicando que os dados não têm alta visibilidade e preservação a longo prazo.

Conclui-se que os pesquisadores utilizam diversos dispositivos para armazenar suas cópias de segurança. No entanto, apesar da diversificação do armazenamento de dados, observa-se que os espaços comumente usados são computador pessoal, computador de trabalho, disco rígido externo, dispositivos USB, e nuvem, locais que pelas suas características não garantem a preservação a longo prazo e o acesso aberto aos dados. Além disso, esses tipos de armazenamento são considerados frágeis devido à sua rápida obsolescência, espaço limitado e riscos de segurança. Por esses motivos, considera-se que os pesquisadores usem, além desses locais, outros métodos para proteger os dados. No referente ao armazenamento por meios

físicos é considerado um dos meios mais vulneráveis e obsoletos, portanto, é essencial migrar esses dados para formatos e armazenamento digitais.

Os repositórios que apresentam uso ocasional por parte dos pesquisadores têm um limitado acesso, portanto sua consulta se restringe ao grupo de pesquisa e à comunidade acadêmica da universidade. Já os repositórios temáticos, redes científicas, nacionais, e internacionais são usados por aproximadamente 34% dos pesquisadores.

Esse tipo de meio de armazenamento é considerado o mais eficaz em termos de preservação e acesso a dados. Apesar do cenário de baixo uso de repositórios, o repositório institucional é utilizado por 50% dos pesquisadores, implicando que há uma rota já percorrida e que poderia no futuro levar a um aumento na abertura e visibilidade dos dados. Por outro lado, os repositórios temáticos e as redes científicas, embora pouco utilizados, representam grande potencial para a área de economia e negócios, uma vez que possibilita o acesso a dados de diversas disciplinas.

6. Considerações da dimensão particularidades dos dados

Conforme foi apresentado os resultados das questões relacionadas às particularidades dos dados produzidos pelos pesquisadores, a continuação se expõe sua caracterização. É importante para a descrição do perfil do pesquisador a partir da perspectiva das particularidades dos dados, verificar se existe associação entre as variáveis de interesse. Para isso, é realizado um teste Qui-quadrado de Pearson.

A Tabela 20 indica que não há relação de dependência entre as variáveis: tipos de instituição – natureza dos dados; tipo de universidade- tipos de formato; tipo de universidade- tipo de dados; tipo de universidade- tempo de armazenamento; natureza dos dados-tipos de formato; tipo de formato- tempo de armazenamento; tipos de dados-tempo de armazenamento, uma vez que em todos os casos a hipótese nula não é rejeitada.

As relações entre natureza dos dados- tipos de dados; natureza dos dados-tempo de armazenamento, tipos de formato-tipos de dados, mostram que existe uma relação ou associação entre estes pares de variáveis porque em cada caso o Pvalue é inferior ao nível de confiança de 5% selecionado, portanto a hipótese nula do teste Qui-quadrado de Pearson é rejeitada.

Tabela 20- Pvalue teste Qui-quadrado de Pearson – características dos dados

ITENS	Tipo Universidade	Natureza dos dados	Tipo de formato	Tipo de dados	Tempo de armazenamento
Tipo universidade	-	0,292	0,299	0,439	0,257
Natureza dos dados	0,292	-	0,439	0,006	0,020
Tipo de formato	0,299	0,439	-	0,000	0,240
Tipo de dados	0,439	0,006	0,000	-	0,938

Tempo de armazenamento	0,257	0,020	0,240	0,938	-
------------------------	-------	--------------	-------	-------	---

Fonte: dados de pesquisa, SPSS 22 ®, 2020

As associações encontradas fazem sentido porque a classificação dos dados pode ser dada com base em sua natureza e tipo de dados, variáveis que se encontram relacionadas, pois os dados quantitativos podem ser apresentados em diversas versões, como equações, algoritmos, estatísticas, software etc., enquanto os dados qualitativos podem ser textuais, visuais ou auditivos. Da mesma forma, o tipo de dados determinará os formatos apropriados nos quais eles serão arquivados, como por exemplo os dados textuais serão salvos em formatos do tipo Doc., Txt ou pdf. A associação entre a natureza dos dados e o tempo em que eles devem ser armazenados pode ocorrer porque, dependendo da complexidade do processo pelo qual foram coletados e do valor que o pesquisador atribui, será determinado o tempo de preservação.

Com base no exposto, os pesquisadores podem ser perfilados de acordo com o tipo de universidade à qual pertencem e as particularidades dos dados que manipulam, como mostra a Tabela 21.

Tabela 21- Perfil dos pesquisadores segundo as particularidades dos dados

ITENS	Professores de Universidade Particular	Professores de Universidade Pública
Natureza de dados	Dados de mistos	Dados mistos
Tipos de dados	Numéricos- Textuais	Numéricos- Textuais
Tipos de formato	Documentos texto: Doc, Txt, pdf Planilhas: Excel, XML, CSV, XLS Dados estatísticos: (SPSS, SAS, SAV, SPV)	Documentos texto: Doc, Txt, pdf Planilhas: Excel, XML, CSV, XLS Dados estatísticos: (SPSS, SAS, SAV, SPV)
Tempo de armazenamento	Indefinidamente	Entre 1 a 5 anos Mais de 5 anos
Tipo de armazenamento	Computador pessoal Computador de trabalho Disco rígido externo Nuvem	Computador pessoal Computador de trabalho Disco rígido externo USB/pendrive Nuvem

Fonte: Dados de pesquisa, SPSS 22 ®, 2020

A partir das características predominantes, assume-se que:

Os pesquisadores de universidade particular e pública, em relação aos dados produzidos, compartilham características semelhantes. As pesquisas produzem principalmente dados de natureza mista, de tipo numérico e textual. Os pesquisadores selecionam os formatos de arquivo acorde ao conteúdo a ser arquivado. Além disso, os pesquisadores usam formatos padrão e legíveis em vários sistemas operacionais. Em relação ao armazenamento de dados, predominam o uso de computador pessoal e de trabalho, disco rígido externo, dispositivos USB e nuvem, o que significa que, em termos de preservação e acesso aberto a longo prazo, os dados encontram-se em espaços frágeis, em potencial ameaça de perda e em acesso limitado ou restrito.

No caso do tempo de armazenamento, os pesquisadores de universidade pública e particular diferem. Nas universidades particulares, os pesquisadores armazenam seus dados indefinidamente, enquanto nas universidades públicas, predomina o armazenamento de dados entre 1 a 5 anos ou por mais de 5 anos. O estabelecimento do tempo varia de acordo com os critérios do local de armazenamento, as necessidades do pesquisador, e o potencial informativo com o qual o pesquisador perceba seus dados. Recomenda-se determinar desde o início o tempo em que serão preservados os dados.

7.3 PRÁTICAS DE GESTÃO DE DADOS DE PESQUISA

A gestão de dados é apresentada como um método eficiente para a organização, manutenção, garantia e preservação desse recurso no atual contexto científico em que seu valor informativo ao longo do tempo é reconhecido. O principal objetivo da gestão de dados é ditar estratégias de proteção contra os riscos de integridade, perda ou inacessibilidade, de acordo com as necessidades e a natureza dos dados a serem curados (SEGUNDO, ARAÚJO, 2019). Da mesma forma, o gerenciamento de dados garante práticas promovidas pela ciência aberta - intercâmbio e reutilização de dados de pesquisa – facilitando que os dados sejam acháveis, acessíveis, interoperáveis e reutilizáveis (princípios FAIR).

Diante do cenário atual da ciência aberta, a aplicação de boas práticas de gestão de dados implica o uso de protocolos padronizados no ciclo de gerenciamento, de modo que o processo seja executado de forma estruturada e produtiva, economizando tempo e evitando o esquecimento de atividades essenciais de organização, descrição, garantia e preservação.

Além disso, os protocolos ou diretrizes garantem o uso padronizado de linguagem e vocabulário claro, evitando uma interpretação parcial, errônea ou descontextualizada dos dados. Contudo, a gestão de dados, como parte inerente ao processo de pesquisa, pode ser desenvolvida sem a aplicação de protocolos padrão, com base nas experiências e necessidades de pesquisa do cientista. Cabe referir que, dependendo das necessidades do pesquisador ou dos requisitos dos repositórios, as fases de gestão de dados podem ser aplicadas na sua totalidade ou em parte.

Para revelar o conhecimento e as práticas de gestão de dados dos pesquisadores, são levadas em conta as questões 12 a 42 do questionário aplicado, considerando que as questões extraídas são de dois tipos: escala nominal e escala Likert.

7.3.1 Dimensão conhecimento sobre gestão de dados

Na ausência de informação relacionadas ao grau de presença ou ocorrência de gestão de dados de pesquisa no contexto colombiano, levanta-se a questão sobre o grau de familiaridade que os pesquisadores têm com o conceito de gestão de dados. Esta questão é de tipo nominal.

A Tabela 22 indica a familiaridade dos pesquisadores com o conceito de gestão de dados, revelando que os pesquisadores de universidade pública têm um conhecimento maior sobre gestão de dados do que pesquisadores de universidade particular. 53% dos pesquisadores de universidade particular declaram ter conhecimento parcial sobre o assunto. Enquanto 32,4% dos pesquisadores de universidade pública têm conhecimento parcial.

Aproximadamente 33,3% dos pesquisadores particulares declaram ter pouco ou nenhum conhecimento sobre gestão de dados, enquanto 24,3% dos pesquisadores de universidade pública declaram ter pouco ou nenhum conhecimento.

Tabela 22- Familiaridade gestão de dados

Itens	Estatística	Particular	Pública	Total
Desconheço	Frequência	5	3	8
	%	9,8%	8,1%	9,0%
Conhecimento completo	Frequência	7	16	23
	%	13,7%	43,3%	26,1%
Conhecimento parcial	Frequência	27	12	39
	%	53,0%	32,4%	44,3%
Pouco conhecimento	Frequência	12	6	18
	%	23,5%	16,2%	20,4%
Total	Frequência	51	37	88
	%	100,0%	100,0%	100,0%
Qui-Quadrado de Pearson	Valor		Significância Bilateral	
		df		
	9,812	3	0,020	

Fonte: dados de pesquisa, SPSS 22 ®, 2020

No final da tabela 22, há informações sobre o teste do qui-quadrado de Pearson, sendo o nível de significância de 5% (0,05) e o Pvalue foi de 0,020, evidencia-se uma associação estatisticamente significativa entre o tipo de universidade e a familiaridade com o conceito.

Conclui-se que aqueles que estão mais familiarizados com a gestão de dados são os professores de universidade pública. Nesse aspecto, independente do grau de familiaridade com gestão de dados, seus estágios estão implicitamente presentes, em diferentes graus, no processo de pesquisa.

7.3.2 Dimensão planejamento de gestão de dados

A dimensão planejar a gestão de dados, menciona que esta fase determina o que será feito com os dados antes de iniciar a investigação. O planejamento define os tipos de dados que serão produzidos, processados e analisados; determina como os dados serão documentados; métodos de preservação e manutenção; políticas de propriedade intelectual, entre outros aspectos.

De uma perspectiva mais formal, o plano de gestão de dados é um documento formal, desenvolvido antes, durante e após a investigação, o qual contém aspectos anteriormente referidos, desde atividades relacionadas à pesquisa até questões mais centradas na curadoria, com o propósito de assegurar os dados. Além disso, as ferramentas digitais ou suporte por parte de especialistas que auxiliem na criação de um documento são, por si só, um mecanismo que permite, de maneira fácil e rápida, estabelecer um processo formal de gestão de dados.

A dimensão, é, portanto, medida a partir de quatro indicadores:

- ✓ “Planejam a coleta, processamento, análise, descrição e preservação dos dados”
- ✓ “Estabelecem critérios de propriedade intelectual”
- ✓ “Documentam o plano de gestão de dados”
- ✓ “Recebem assistência de instituições especializadas em dados para a criação de um plano de gestão”

Em geral, a dimensão Planejamento da Gestão de Dados para universidade particular, revela que, em média 3,23, os pesquisadores frequentemente planejam como os dados serão gerenciados. Observa-se que a variância de 0,90 e desvio padrão de 0,95 indicam pouca dispersão dos dados em relação à média, de forma que valida a representatividade da média em relação ao comportamento do grupo.

A mediana reforça o que a média indica, e revela que 50% dos pesquisadores em relação ao planejamento, realizam essa prática com um grau de ocorrência igual ou superior a frequentemente, sempre, e, o outro 50%, o fazem igual ou inferior a frequentemente, sendo ocasionalmente, raramente ou nunca.

Nos pesquisadores de universidade pública, observa-se uma média de 3,07, indicando que ocasionalmente os pesquisadores planejam a gestão de dados. Em relação às medidas de variabilidade, observa-se uma variação de 1,10 e um desvio padrão de 1,05, indicando pouca dispersão dos dados em relação à média. Os valores mencionados confirmam a representatividade da média, e reforçam que os pesquisadores atuam de forma próxima a essa.

A mediana reforça o que a média indica e revela que 50% dos pesquisadores realizam essa prática com um grau de ocorrência igual ou superior a ocasionalmente, sendo frequente e sempre, e o outro 50%, o fazem igual ou inferior a ocasionalmente, sendo raramente ou nunca.

1. Planejamento da coleta, processamento, análise, descrição e preservação dos dados:

Este indicador mede o grau de ocorrência com o qual os pesquisadores planejam as seguintes questões. A primeira questão diz respeito a informações sobre os dados, como os dados serão coletados, processados e analisados. A segunda questão refere-se ao planejamento da construção de metadados, consistindo em como se descreverá em termos de conteúdo, técnico e administrativo os dados. A terceira questão trata da preservação de dados, ou seja, o estabelecimento de estratégias para proteger os dados, mantendo seu estado e evitando possíveis danos.

Em geral, a média 3,68 para pesquisadores de universidade particular, e 3,91 para pesquisadores de universidade pública, indicam que, em termos de categorização, o indicador tem forte presença, implicando que este tipo de planejamento é uma prática, em média, presente de forma frequente entre os pesquisadores. A mediana 3,67 para universidades particulares, e 4,0 para universidade pública, reforça o afirmado pelas médias, ao ser localizado no mesmo intervalo.

Já em relação às questões que compõem o indicador, observa-se que os pesquisadores frequentemente planejam como os dados serão coletados, processados, analisados, atividades inerentes à pesquisa, bem como planejam os métodos de preservação e metadados que serão utilizados, sendo atividades mais de curadoria.

No referente aos pesquisadores de universidade particulares, observa-se uma variância de 0,92 e um desvio padrão de 0,96, quanto que para os pesquisadores de universidade pública se registra uma variância de 0,55 e desvio padrão de 0,74, valores considerados pequenos, o que indica pouca dispersão dos dados em relação à média. Esses valores expressam que os pesquisadores de universidade particular e pública, atuam de forma semelhante à média, o que implica que a média é confiável em termos de que suas características são representativas.

2. Propriedade Intelectual:

Toda pesquisa é atravessada por conceitos de privacidade, confidencialidade e aspectos éticos; bem como por termos normativos relacionados à propriedade primária dos dados e ao direito de terceiros de usá-los sob certas condições. O estabelecimento da propriedade intelectual à qual os dados estarão sujeitos interfere na maneira como os dados serão gerenciados e sua possível abertura, daí a importância de seu planejamento. Nesse sentido, este

indicador mede se os pesquisadores estabelecem critérios de propriedade intelectual sobre os dados, desde o início de suas pesquisas.

Entre os pesquisadores de universidade particular, a média 3,33, indica que, em termos de categorização, esta prática é forte. A forte categorização significa que os pesquisadores frequentemente projetam os critérios de propriedade intelectual pelos quais os dados serão gerenciados.

A variância 1,51 e desvio padrão 1,23 para os pesquisadores de universidade particular, indica pouca dispersão dos dados em relação à média. A dispersão baixa expressa que os pesquisadores atuam de maneira semelhante à média, ou seja, frequentemente estabelecem critérios de propriedade intelectual em seu processo de pesquisa. Considera-se que devido ao comportamento homogêneo dos dados, a média representa a amostra estudada.

Em relação aos pesquisadores de universidade pública têm uma média de 3,03. Em termos de categorização, a prática de estabelecer critérios de propriedade intelectual tem presença neutra, ou seja, ocorre ocasionalmente nas pesquisas dos docentes. A mediana reforça o que a média indica.

Os pesquisadores de universidade pública obtiveram uma variância de 2,19 e um desvio padrão de 1,48, indicando uma maior dispersão dos dados em relação à média. Os resultados nas medidas de dispersão expressam que os pesquisadores atuam de forma relativamente heterogênea em relação ao grau de ocorrência em que planejam os critérios de propriedade intelectual.

3. Documento plano de gestão de dados:

De acordo com *Data Observación Network for Earth* (2016), o plano de gestão de dados é um documento formal que descreve o que o pesquisador fará com seus dados durante e após a conclusão da pesquisa. O objetivo deste documento formal é estabelecer as fases da gestão de dados para garantir que os dados estejam seguros a médio e longo prazo. A criação desse documento e seu conteúdo dependerá dos protocolos estabelecidos pelos diversos repositórios, instituições, aos quais o pesquisador pretende se aderir.

Pesquisadores de universidade particular e universidade pública, de acordo com seus respectivas médias, 3,08 e 2,76, em termos de categorização, indicam que a presença dessa prática é neutra. A categoria neutra significa que, em geral, a maioria de pesquisadores de universidade particular e pública, realizam ocasionalmente um plano de gestão de dados. A mediana 3 para os dois grupos de pesquisadores, reforça o que é indicado pela média.

Os pesquisadores de universidade particular registram uma variância de 1,95 e um desvio padrão de 1,4 indicando pouca dispersão dos dados em relação à média, ou seja, os pesquisadores atuam de maneira semelhante à média. A média é considerada representativa.

Enquanto que os pesquisadores de universidade pública registram uma variação de 2,36 e um desvio padrão de 1,53, indicando uma maior dispersão dos dados em relação à média. Os valores obtidos nas medidas de variabilidade expressam que embora a média seja representativa, os pesquisadores atuam de uma forma relativamente heterogênea em relação ao grau de ocorrência em que se realiza o documento.

4. Assistência para a criação de um plano de gestão de dados:

A assistência no planejamento da gestão de dados refere-se aos serviços que as instituições fornecem aos pesquisadores para realizar o documento em conformidade com os padrões estabelecidos. A assistência possibilita que a construção do documento seja realizada de forma sistematizada, portanto, fácil e rapidamente. A assistência de instituições especializadas em gerenciamento de dados ocorre de duas maneiras: a) apoio por parte de especialistas em curadoria e, b) ferramentas digitais que possibilitam que os pesquisadores construam os documentos de forma autônoma.

Pesquisadores de universidade particular e universidade pública, de acordo com seus respectivas médias, 2,83 e 2,49, em termos de categorização, indicam que a presença dessa prática é neutra. A categoria neutra significa que os pesquisadores ocasionalmente usam serviços de assistência ou ferramentas digitais oferecidas pelas instituições para criar seus planos de gestão de dados. As medianas para os dois grupos de pesquisadores reforçam o resultado da média, ao ser localizada no mesmo intervalo.

No referente aos pesquisadores de universidade particulares, observa-se uma variância de 1,79 e um desvio padrão de 1,34, enquanto que para os pesquisadores de universidade pública se registra uma variância de 1,47 e desvio padrão de 1,21, o que indica pouca dispersão dos dados em relação à média. Os valores obtidos indicam que os pesquisadores atuam, em relação à assistência, de forma homogênea e semelhante à média, assim, confirmando a representatividade das características enunciadas pela média no grupo de pesquisadores.

5. Considerações da dimensão

O resultado geral da dimensão, revela que o planejamento da gestão de dados entre pesquisadores de universidade particular e pública diferem. No entanto, observando o resultado de cada indicador, revela-se que os pesquisadores atuam de forma semelhante, e apenas divergem no planejamento da propriedade intelectual.

Tabela 23 - Caracterização dimensão planejamento de gestão de dados

ITENS	Pesquisadores de universidade particular	Pesquisadores de universidade pública
Dimensão		
Planejam a gestão de dados	Frequentemente	Ocasionalmente
Indicadores		
Planejam a coleta, processamento, análise, descrição e preservação dos dados	Frequentemente	Frequentemente
Estabelecem critérios de propriedade intelectual	Frequentemente	Ocasionalmente
Documentam o plano de gestão de dados	Ocasionalmente	Ocasionalmente
Recebem assistência de instituições especializadas em dados para a criação de um plano de gestão	Ocasionalmente	Ocasionalmente

Fonte: Elaboração própria, 2020.

No que diz respeito a determinar a propriedade intelectual dos dados, os pesquisadores de universidade particular realizam frequentemente referida prática na fase de planejamento da gestão de dados, enquanto que os pesquisadores de universidade pública o fazem ocasionalmente. Planejar desde o início de uma investigação os critérios de propriedade intelectual, isto é, autoria principal dos dados, e as limitações e restrições de uso por terceiros, reduz ou elimina a probabilidade de que se apresentem incertezas e ambiguidades sobre o que pode e o que não pode ser feito com eles. Essa imprevisibilidade na titularidade dos direitos de propriedade intelectual, contemplado desde o acesso aberto, é um impedimento para exercer, no decorrer ou no fim da pesquisa, práticas de compartilhamento e reúso de dados.

Cabe assinalar que os dados da pesquisa têm valor informativo e frequentemente comercial, assim como sua produção é comumente realizada em conjunto, o que leva a que seja indispensável, antes de iniciar uma investigação, determinar entre os envolvidos, pesquisadores e instituições, a propriedade intelectual, de modo que não se gere conflitos de interesse no futuro.

Já, em relação ao planejamento de como os dados serão coletados, processados, analisados, descritos e preservados, os dois grupos de pesquisadores indicam que realizam tais práticas frequentemente. Esse resultado se deve a que, como mencionado no referencial teórico, o ciclo de vida de um dado na pesquisa coincide com as fases de curadoria, o que significa que essa prática de pesquisa também é concebida como uma atividade pertencente à gestão de dados, pois garante sua qualidade.

No referente à construção do documento, plano de gestão de dados, o que será feito com os dados durante e após a conclusão da pesquisa, é uma prática feita ocasionalmente por pesquisadores de universidade particular e pública. Gerar um plano formal de gestão, desde o início do planejamento da pesquisa, é considerada uma fase fundamental, pois seu objetivo é que a curadoria e a preservação sejam desenvolvidas de forma eficiente, de modo a garantir que os dados não se percam com o tempo ou não sejam aproveitados para outros fins.

Pode-se considerar que documentar esporadicamente a gestão de dados se deve a que não é uma prática geralmente obrigatória no contexto científico da amostra estudada, e quando acontece se dá porque a instituição patrocinadora da pesquisa ou os repositórios que armazenarão os dados, o determinarem como necessário.

A não criação de um plano documentado significa que a gestão de dados continua sendo um processo não estruturado que é implicitamente ou inconscientemente realizado pelos pesquisadores em suas investigações. Em geral, considerar o planejamento através de um exercício formal, como a documentação, garantirá a execução eficiente das práticas de gerenciamento de dados.

Do ponto de vista da abertura dos dados, não fazer um plano formal pode causar a reutilização indevida dos dados, uma vez que não existe documentação que legitime e esclareça o contexto de pesquisa em que foram criados, as características e documentação dos dados, quem tem a titularidade dos direitos de propriedade intelectual e sob quais condições eles podem ser reusados, entre outros aspectos de relevância.

Por outro lado, somado à documentação do plano de gestão de dados, é medido a ocorrência com a qual os pesquisadores recebem assistência para sua criação, seja através de ferramentas digitais ou apoio especializado, observando que os pesquisadores ocasionalmente recebem esse tipo de suporte. O uso mínimo dessas ferramentas pode gerar que os documentos que são ocasionalmente construídos não sejam completos e padronizados, aumentando a probabilidade de que seu conteúdo seja ambíguo e não responda ao seu objetivo de garantir uma integral contextualização de aspectos que envolvem como os dados são curados e preservados, bem como sua autoria.

Em geral, os pesquisadores frequentemente planejam como os dados serão tratados, mas ocasionalmente documentam, através de um plano formal, essa prática. Em relação ao planejamento da propriedade intelectual dos dados, pesquisadores de universidades públicas o fazem ocasionalmente, gerando possíveis conflitos de interesse nos dados devido ao seu valor informacional e possivelmente comercial.

Tabela 24- Dimensão planejamento de gestão de dados – pesquisadores de universidade particular

ITENS	Média	Mediana	Moda	Desvio padrão	Variância	Curtose	Coefficiente de assimetria	Mínimo	Máximo
Dimensão planejamento de gestão de dados	3,23	3,29	5	0,95	0,9	-0,88	0,12	1,67	5
Planejamento da coleta, processamento, análise, armazenamento, garantia, controle, descrição e preservação dos dados	3,68	3,67	4,33	0,96	0,92	-0,22	-0,58	1	5
P. 13: planeja, antes de iniciar a pesquisa, como os dados serão coletados, processados e analisados.	3,96	4	5	1	1	0,24	-0,8	1	5
P.14: planeja a descrição dos dados –metadados-.	3,49	4	4	1,22	1,49	-0,88	-0,35	1	5
P.15: planeja a preservação dos dados.	3,59	4	4	1,15	1,33	-0,67	-0,51	1	5
Propriedade Intelectual	3,33	3	4	1,23	1,51	-0,9	-0,27	1	5
P.16: estabelece critérios de propriedade intelectual sobre os dados	3,33	3	4	1,23	1,51	-0,9	-0,27	1	5
Documento plano de gestão de dados	3,08	3	4	1,4	1,95	-1,26	-0,1	1	5
P.17: cria um documento formal - plano de gestão de dados -, a partir de diretrizes propostas pelas instituições especialistas.	3,08	3	4	1,4	1,95	-1,26	-0,1	1	5
Assistência para a criação de um plano de gestão de dados	2,83	3	1	1,34	1,79	-1,35	0	1	5
P.18: utiliza ferramentas digitais que assistem na criação do plano de gestão de dados	3,2	4	4	1,44	2,08	-1,3	-0,36	1	5
P.19: tem algum tipo de assistência, por parte de um grupo de especialistas em dados, para construir seu plano de gestão de dados	2,47	2	1	1,47	2,17	-1,31	0,41	1	5

Fonte: Dados de pesquisa, SPSS 22 ®, 2020

Tabela 25- Dimensão planejamento de gestão de dados – pesquisadores de universidade pública

ITENS	Média	Mediana	Moda	Desvio padrão	Variância	Curtose	Coefficiente de assimetria	Mínimo	Máximo
Dimensão planejamento de gestão de dados	3,05	3,00	5,00	1,03	1,06	-0,56	0,35	1,33	5,00
Planejamento da coleta, processamento, análise, armazenamento, garantia, controle, descrição e preservação dos dados	3,91	4,00	4,00	0,74	0,55	0,30	-0,51	2	5
P. 13: planeja, antes de iniciar a pesquisa, como os dados serão coletados, processados e analisados.	4,16	4,00	5,00	1,04	1,08	1,69	-1,43	1	5
P.14: planeja a descrição dos dados –metadados-.	3,78	4,00	4,00	1,08	1,17	0,67	-0,93	1	5
P.15: planeja a preservação dos dados.	3,78	4,00	4,00	0,92	0,84	3,06	-1,37	1	5
Propriedade Intelectual	3,03	3,00	5,00	1,48	2,19	-1,42	-0,05	1	5
P.16: estabelece critérios de propriedade intelectual sobre os dados	3,03	3,00	5,00	1,48	2,19	-1,42	-0,05	1	5
Documento plano de gestão de dados									
P.17: cria um documento formal - plano de gestão de dados -, a partir de diretrizes propostas pelas instituições especialistas.	2,76	3,00	1,00	1,53	2,36	-1,52	0,09	1	5
Assistência para a criação de um plano de gestão de dados	2,49	2,50	1,00	1,21	1,47	-0,50	0,51	1	5
P.18: utiliza ferramentas digitais que assistem na criação do plano de gestão de dados	2,89	3,00	1,00	1,49	2,21	-1,39	0,03	1	5
P.19: tem algum tipo de assistência, por parte de um grupo de especialistas em dados, para construir seu plano de gestão de dados	2,08	2,00	1,00	1,30	1,69	-0,10	0,97	1	5

Fonte: Dados de pesquisa, SPSS 22®, 2020

7.3.3 Dimensão Metadados

Os metadados se referem a uma descrição detalhada do contexto metodológico, técnico, e administrativo do conjunto de dados. O objetivo desta fase é a documentação de aspectos que interferem na compreensão presente ou futuro dos dados. Portanto, sua boa prática garantirá que os dados sejam interpretados corretamente, e possivelmente reusados, se o autor pretender colocá-los em acesso aberto.

Geralmente os metadados são construídos a partir de metadados padrão propostos pelos repositórios ou centros de dados. A padronização de metadados segundo *DataONE* (2016) refere-se a que tais descrições sejam realizadas a partir de termos, definições e estruturas comuns que proporcionem coerência na documentação dos dados e uma maior compreensão do seu conteúdo pela comunidade acadêmica. Neste contexto, tais organizações fornecem serviços de assistência especializada e ferramentas digitais para a construção de seus metadados de forma padronizada.

A dimensão, é, portanto, medida a partir de três indicadores:

- ✓ “Descrevem os dados de pesquisa –metadados-“.
- ✓ “Orientam a descrição de dados – metadados- a partir de metadados padrão”
- ✓ “Recebem assistência na criação de metadados”

Em termos gerais, para a dimensão metadados, tanto para pesquisadores de universidade particular, quanto pesquisadores de universidade pública, de acordo com suas respectivas médias, 2,98 e 3,07, em termos de categorização, indicam que a presença dessa prática é neutra.

A categoria neutra para essa dimensão indica que, em geral, os pesquisadores ocasionalmente constroem metadados. No entanto, se for explorado os indicadores que o compõem, observa-se que os dados são descritos e contextualizados frequentemente pelos pesquisadores, porém, ocasionalmente sua criação se faz sob metadados padrão ou mediante assistência de especialistas ou com suporte de ferramentas digitais.

Em relação às medidas de variabilidade da dimensão, variância e desvio padrão, tanto em pesquisadores de universidade particular quanto pública, revelam pouca dispersão dos dados em relação à média. Os valores obtidos indicam que, devido ao comportamento homogêneo dos dados, as médias representam as características dos grupos em relação à criação de metadados.

Por outro lado, a medianas da dimensão, tanto para pesquisadores de universidade particulares e pública, reforçam o que as médias indicam e revela que 50% dos pesquisadores realizam essa prática com um grau de ocorrência igual ou maior que ocasionalmente, sendo

frequente e sempre, e os outros 50%, fazendo o igual ou inferior que ocasionalmente, raramente ou nunca.

1. Descrição dos dados de pesquisa

Como já foi referido, o termo metadados refere-se à documentação dos dados, na qual são descritas as características consideradas pertinentes, tendo como finalidade que esse recurso seja compreendido contextualmente ao longo do tempo.

O objetivo do indicador é revelar se os pesquisadores constroem metadados e o que eles contêm. Para isso, questiona-se se os pesquisadores descrevem seus dados no nível de pesquisa, nível de base de dados, nível de variável e no nível administrativo.

Quadro 26 - Metadados

Metadados	
Nível de pesquisa	Descrição dos dados em relação ao contexto, metodologia, procedência dos dados, processamento, análise e garantia e controle da qualidade dos dados
Nível de base de dados	Descrição dos dados em relação às tabelas de base de dados e seu formato.
Nível de variáveis	Descrição dos dados em relação às variáveis, categorias, nomes, etiquetas.
Nível administrativo	Descrição dos dados em relação ao técnico, preservação, direitos de autor

Fonte: Adaptado de Gilliland - Swetland, P. 3-4, 1999

Em geral, a média 3,54 para pesquisadores de universidade particular, e 3,55 para pesquisadores de universidade pública, indicam que, a descrição dos dados frequentemente é realizada em suas pesquisas. Nos dois grupos de pesquisadores se registra uma variância e desvio padrão com pouca dispersão dos dados em relação à média. A pouca dispersão significa que a média é representativa, ao ser um grupo que se comporta homogeneamente.

Já em relação aos níveis que compõem o indicador de descrição, se revela que os pesquisadores de universidade particular e pública, efetivamente realizam com frequência a descrição dos dados nos níveis de pesquisa, base de dados e variáveis. No entanto a ocorrência difere no nível administrativo –técnica, preservação, direitos autorais-, em que os pesquisadores de universidade particular ocasionalmente documentam seus dados em tais termos, enquanto que o pesquisador de universidade pública o realiza frequentemente.

2. Orienta a descrição de dados a partir dos metadados padrão

As diversas instituições especializadas em gestão de dados propõem metadados padrão de acordo com as áreas de conhecimento para as quais o serviço é direcionado, no entanto algumas instituições estabelecem metadados padrão genéricos que aplicam a todos os dados.

Em relação a orientar a descrição de dados a partir de metadados padrão, para os pesquisadores de universidade particular a média é 2,88, e para pesquisadores de universidade pública a média é 3, indicando que, em termos de categorização, esta prática é neutra. As medianas para os dois grupos de pesquisadores reforçam o indicado pela média, ao ser localizada no mesmo intervalo. A categórica neutra significa que os pesquisadores ocasionalmente criam seus metadados a partir de metadados padrão.

No referente aos pesquisadores de universidade particular, observa-se uma variância de 1,47 e um desvio padrão de 1,21, quanto que para os pesquisadores de universidade pública se registra uma variância de 1,94 e desvio padrão de 1,39, o que indica pouca dispersão dos dados em relação à média. Os resultados das medidas de variabilidade indicam que a ocorrência com que realizam metadados padrão é homogêneo, próximo à característica traçada pela média. O desvio padrão para cada grupo indica que a média é representativa.

3. Assistência na criação de metadados

Além dos metadados padrão estabelecidos pelas instituições especializadas, estas organizações geralmente fornecem assistência para que o processo ocorra com eficiência. Este indicador revela se os pesquisadores recebem assistência de expertos ou usam ferramentas digitais para criar seus metadados.

Em relação a receber assistência ou usar ferramentas digitais para a descrição dos dados, os pesquisadores de universidades privadas, obtiveram uma média de 2,52, expressando que, em termos de categorização, a ocorrência desta prática é neutra. A mediana classifica a prática como fraca. O resultado revela que os pesquisadores ocasionalmente usam ou são assistidos para criar metadados. Somado a isto a mediana indica que 50% dos pesquisadores em um grau de ocorrência igual ou inferior raramente realizam essa prática, enquanto 50% recebem apoio nessa atividade igual ou superior raramente, sendo de uma forma mais presente.

Pesquisadores de universidade pública, a média é de 2,68, em termos de categoria, indicam que a ocorrência desta prática é neutra. A mediana categoriza a prática no mesmo intervalo. Tais medidas de tendência central indicam que ocasionalmente os pesquisadores usam esse tipo de serviços para criar seus metadados, em que a metade dos pesquisadores realizam tais práticas em um grau de ocorrência igual ou maior a ocasionalmente, e a outra metade o realiza igual ou inferiormente a ocasionalmente.

Por outro lado, com relação à variância e desvio padrão, os dois grupos de pesquisadores revelam pouca dispersão dos dados em relação à média. As medidas de variabilidade indicam que a ocorrência com que os pesquisadores são assistidos se dá de forma homogênea no grupo, próximo à média. Os desvios padrão para cada grupo indica que a média é representativa.

4. Considerações da dimensão

Em geral, não existem diferenças entre pesquisadores de universidade particular e pública em relação à categorização da dimensão e aos indicadores que a compõem.

Tabela 26- Caracterização dimensão metadados

ITENS	Pesquisadores de universidade particular	Pesquisadores de universidade pública
Dimensão		
Metadados	Ocasionalmente	Ocasionalmente
Indicadores		
Descreve os dados de pesquisa a nível de pesquisa, base de dados, variáveis e administrativo	Frequentemente	Frequentemente
Orienta a descrição de dados a partir dos metadados padrão propostos por instituições especialistas	Ocasionalmente	Ocasionalmente
Recebem assistência na criação dos metadados	Ocasionalmente	Ocasionalmente

Fonte: Elaboração própria, 2020.

Observa-se que a descrição dos dados nos níveis de pesquisa, banco de dados, variáveis e administrativo, nos dois grupos de pesquisa, é realizada com frequência. Esse resultado positivo permite que os dados da pesquisa sejam compreendidos e interpretados ao longo do tempo, seja pelos pesquisadores que os criaram ou por outros pesquisadores interessados em sua reutilização. Do ponto de vista do acesso aberto, quando o autor pretende compartilhar seus dados, essa atividade é concebida com maior relevância, pois fornece um contexto adequado, garantindo a reutilização correta dos dados.

No entanto, ao investigar os resultados pelos níveis em que os dados são descritos, observa-se que, embora os pesquisadores criam metadados frequentemente, ocorre uma exceção em pesquisadores de universidade pública, os quais afirmam que ocasionalmente criam metadados de caráter administrativo, isto é, técnico, preservação e direitos autorais.

Os metadados administrativos são indispensáveis porque contêm as informações necessárias para, em termos técnicos, decifrar a estruturação em que se encontra a base de dados; em preservação, necessário para a gestão e possível migração a longo prazo; e de propriedade intelectual para conhecer a autoria dos dados e seu grau de abertura.

Por outra parte, atualmente, instituições especializadas oferecem modelos e assistência aos pesquisadores para criar metadados padrão. Nesse sentido, pesquisadores de universidade particular e pública ocasionalmente criam metadados a partir de metadados padrão, assim como ocasionalmente recebem algum tipo de assistência. Os pesquisadores geralmente aplicam metadados padrão para atender aos requisitos fornecidos pelos repositórios ou data centers em onde compartilharão ou armazenarão dados.

Não realizar a descrição dos dados a partir de metadados padrão indica que os dados podem estar desestruturados, com informações inconsistentes ou incompletas, ou com linguagem não

padronizada que dificulte a interpretação. Embora os pesquisadores frequentemente criem metadados, a qualidade deles pode variar de pesquisador para pesquisador, pois não é criada a partir de metadados padrão, com assistência especializada ou apoiada por ferramentas digitais que garantem que os dados sejam descritos, contextualizados corretamente e com informações completas.

Em geral a dimensão metadados expressa que frequentemente os pesquisadores documentam seus dados, porém, de forma ocasional o fazem a partir de metadados padrão, assistidos por especialistas ou usando ferramentas digitais, o que implica que pode haver metadados incompletos que prejudiquem a utilidade dos dados no futuro.

Tabela 27- Dimensão metadados – pesquisadores de universidade particular

ITENS	Média	Mediana	Moda	Desvio padrão	Variância	Curtose	Coefficiente de assimetria	Mínimo	Máximo
Dimensão Metadados	2,98	3	2,25	0,92	0,85	-0,42	-0,01	1	5
Descrição de dados	3,54	3,5	3,75	0,9	0,81	0,81	-0,77	1	5
P.20: descreve os dados a nível de pesquisa	3,59	4	3	1,06	1,13	-0,25	-0,34	1	5
P.21: descreve os dados a nível de base de dados	3,75	4	4	1,04	1,07	-0,41	-0,47	1	5
P.22: descreve os dados a nível de variável	3,78	4	4	1,08	1,17	0,06	-0,73	1	5
P.23: descreve os dados a nível administrativo	3,06	3	3	1,22	1,5	-0,73	-0,12	1	5
Orienta a descrição de dados a partir dos metadados padrão	2,88	3	4	1,21	1,47	-1,04	-0,19	1	5
P.24: orienta a descrição de seus dados com base em metadados padrões determinados por instituições especialistas	2,88	3	4	1,21	1,47	-1,04	-0,19	1	5
Assistência na criação de metadados	2,52	2	1	1,26	1,59	-1,1	0,39	1	5
P.25: usa ferramentas digitais para criar metadados	2,92	3	3	1,37	1,87	-1,14	0	1	5
P.26: tem algum tipo de assistência de um grupo de especialistas em dados para criar metadados	2,12	1	1	1,34	1,79	-0,78	0,77	1	5

Fonte: Dados de pesquisa, SPSS 22 ®, 2020

Tabela 28- Dimensão metadados – pesquisadores de universidade pública

ITENS	Média	Mediana	Moda	Desvio padrão	Variância	Curtose	Coefficiente de assimetria	Mínimo	Máximo
Dimensão Metadados	3,07	3,00	5,00	1,05	1,10	-0,37	0,20	1	5
Descrição de dados	3,55	3,50	3,00	0,88	0,77	0,80	-0,21	1	5
P.20: descreve os dados a nível de pesquisa	3,41	4,00	4,00	1,30	1,69	-0,82	-0,42	1	5
P.21: descreve os dados a nível de base de dados	3,84	4,00	4,00	1,07	1,14	0,92	-0,96	1	5
P.22: descreve os dados a nível de variável	3,78	4,00	4,00	1,08	1,17	-0,23	-0,65	1	5
P.23: descreve os dados a nível administrativo	3,16	3,00	3,00	1,21	1,47	-0,68	-0,13	1	5
Orienta a descrição de dados a partir dos metadados padrão	3,00	3,00	3,00	1,39	1,94	-1,20	-0,13	1	5
P.24: orienta a descrição de seus dados com base em metadados padrões determinados por instituições especialistas	3,00	3,00	3,00	1,39	1,94	-1,20	-0,13	1	5
Assistência na criação de metadados	2,68	2,50	1,00	1,33	1,77	-1,07	0,19	1	5
P.25: usa ferramentas digitais para criar metadados	3,03	3,00	1,00	1,55	2,42	-1,49	-0,09	1	5
P.26: tem algum tipo de assistência de um grupo de especialistas em dados para criar metadados	2,32	2,00	1,00	1,47	2,17	-1,14	0,61	1	5

Fonte: Dados de pesquisa, SPSS 22 ®, 2020

7.3.4 Dimensão garantia e controle de qualidade dos dados

A fase de garantia da qualidade abrange atividades relacionadas à inspeção, verificação e controle no conjunto de dados de pesquisa. O propósito desta fase é melhorar a qualidade dos dados, identificando possíveis inconsistências neles durante seu ciclo de vida (DATAONE, 2016). A dimensão é composta por três indicadores:

- ✓ “Inspeccionam, verificam, e monitoram a qualidade dos dados”
- ✓ “Aplicam diretrizes, propostas por instituições especializadas, para orientar a verificação, inspeção e controle no conjunto de dados”.
- ✓ “Recebem assistência por parte de especialistas para garantir a qualidade dos dados”.

Em termos gerais, a Dimensão de Controle e Garantia de Qualidade dos Dados, para pesquisadores de universidade particular mostra uma média de 3,2, enquanto em pesquisadores de uma universidade pública a média é 3,40, indicando uma presença forte dessa prática. Esse resultado significa que as práticas que compõem a dimensão são, em média, realizadas com frequência pelos pesquisadores.

Por outra parte, a mediana respectiva de cada grupo reforça o que é indicado pela média ao ser localizado no mesmo intervalo. Somado a isso revela que 50% dos pesquisadores realizam essa prática com um grau de ocorrência igual ou superior a frequentemente, sendo sempre, e o outro 50%, o fazem igual ou inferior a frequente, sendo ocasionalmente, raramente ou nunca.

Em relação à variância e desvio padrão dos dois grupos, eles podem ser considerados relativamente pequenos, indicando pouca dispersão dos dados em relação à média. As medidas de variabilidade revelam que a atuação dos pesquisadores é homogênea, próximo à média, e, portanto, a característica traçada pela média é considerada representativa.

1. Inspeção e monitoramento da qualidade do conjunto de dados

Atividades relacionadas à inspeção e verificação dos dados no seu ciclo de vida são práticas implícitas no processo de pesquisa e que ao mesmo tempo fazem parte das fases de gestão de dados. A realização de tais atividades torna os dados seguros e confiáveis. O objetivo do indicador é revelar se os pesquisadores inspecionam e monitoram a qualidade dos dados mediante a verificação da confiabilidade, consistência e legibilidade dos dados. Os critérios de qualidade medidos são os seguintes:

Quadro 27- Critérios de qualidade dos dados

Qualidade dos dados		Questões
Confiabilidade	Os dados fornecidos são precisos. A representação dos dados (ou valor) reflete bem o estado real das informações de origem.	Verificar a fiabilidade dos dados

Qualidade dos dados		Questões
Consistência	Após o processamento dos dados, seus conceitos, domínios e formatos de valor ainda coincidem. Todos os dados são consistentes ou verificáveis.	Verificar a consistência dos dados
Legibilidade	O conteúdo dos dados é claro e compreensível.	O conteúdo dos dados é claro e compreensível
	Os metadados são fáceis de entender.	Realiza metadados e verifica sua fácil compreensão

Fonte: Adaptado de PowerData,2014.

Em relação à inspeção, verificação e monitoramento da confiabilidade, consistência e legibilidade dos dados, pesquisadores de universidade particular e universidade pública, de acordo com seus respectivas médias, 4,1 e 4,4, em termos de categorização, indica uma presença extremamente forte dessa prática entre os pesquisadores. A mediana para os dois grupos reforça o indicado pela média ao ser localizado no mesmo intervalo.

A categoriza extremamente forte, revela que, em média, os pesquisadores sempre inspecionam e controlam que os dados produzidos nas suas pesquisas sejam confiáveis, consistentes e compreensíveis, bem como verificam que os metadados criados sejam fáceis de entender.

Em relação à variância e desvio padrão dos dois grupos, eles revelam muito pouca dispersão dos dados em relação à média. As medidas de variabilidade confirmam que devido ao comportamento homogêneo dos dados, a média é representa a atuação dos pesquisadores em relação à referida prática.

2. Diretrizes que verifiquem, inspecionem e controlem a qualidade dos dados

Instituições especializadas estabelecem e fornecem à comunidade acadêmica manuais de boas práticas relacionados às fases que compõem a gestão de dados. Essas recomendações padronizadas permitem que os pesquisadores gerenciem seus dados com maior eficiência. Este indicador revela se os pesquisadores aplicam diretrizes, propostas por expertos, que orientam a inspeção, verificação e o controle da qualidade dos dados.

Pesquisadores de universidade particular e universidade pública, de acordo com seus respectivas médias, 3,35 e 3,43, em termos de categorização, indicam que a presença dessa prática é forte. A mediana para os dois grupos reforça o indicado pela média. A categorização forte revela que frequentemente os pesquisadores aplicam diretrizes estabelecidas em manuais de boas práticas das instituições expertas para verificar e monitorar a qualidade dos dados.

Para pesquisadores de universidade particular, observa-se uma variância de 1,59 e um desvio padrão de 1,26, indicando pouca dispersão dos dados em relação à média. As medidas

de variabilidade revelam que a atuação dos pesquisadores é próxima à característica da média, e, portanto, é considerada representativa.

Já em pesquisadores de universidade pública, observa-se uma variância de 2,03 e um desvio padrão de 1,43, indicando uma maior dispersão dos dados em comparação com as medidas de dispersão de universidades particulares. Os valores obtidos expressam que embora a média seja representativa, os pesquisadores atuam de uma forma relativamente heterogênea em relação à referida prática.

3. Assistência para garantir a qualidade dos dados

Além de estabelecer manuais de boas práticas, as instituições oferecem serviços que garantem a qualidade do conjunto de dados. Este indicador revela se os pesquisadores têm alguma assistência por parte de especialistas em dados para garantir a qualidade dos seus dados de pesquisa.

Em relação a receber assistência para garantir a qualidade dos dados, pesquisadores de universidades particulares observam uma média de 2,22, o que em termos de categorização indica uma presença neutra dessa prática. A mediana o classifica como fraco. Nesse sentido, a média indica que ocasionalmente os pesquisadores procuram ou recebem esse tipo de assistência. Já a mediana expressa que 50% dos pesquisadores raramente e nunca recebem assistência, enquanto que 50% recebem igual ou maior a raramente, ocasional, frequente e sempre.

No grupo de pesquisadores particulares observa-se uma variância de 1,65 e desvio padrão de 1,29, podem ser considerados relativamente pequenos, indicando pouca dispersão dos dados em relação à média. Os valores obtidos expressam que os pesquisadores atuam de forma similar à característica indicada pela média. A homogeneidade da amostra indica que a média é confiável em termos de representatividade.

Nos pesquisadores das universidades públicas, observa-se uma média de 2,35, o que significa presença neutra da prática na amostra estudada. A mediana o classifica como extremamente fraco. A categoria neutra evidencia que, em média, os pesquisadores ocasionalmente possuem assistência para assegurar a qualidade dos dados. Já a mediana indica que 50% dos pesquisadores nunca recebem assistência, enquanto que 50% recebem raramente ou com maior ocorrência sendo ocasional, frequente e sempre.

Observa-se uma variação de 2,62 e desvio padrão de 1,62, indicando uma maior dispersão dos dados em relação à média. Os valores obtidos evidenciam que, embora a média seja representativa, os pesquisadores atuam de uma forma relativamente heterogênea em relação à referida prática.

4. Considerações da dimensão garantia e controle de qualidade dos dados

Em termos gerais, pesquisadores de universidade particular e pública atuam de maneira similar em relação às práticas que garantem a qualidade dos dados. Os resultados revelam que frequentemente os pesquisadores monitoram e garantem a qualidade do conjunto de dados.

Tabela 29- Caracterização dimensão garantia e controle de qualidade dos dados

ITENS	Pesquisadores de universidade particular	Pesquisadores de universidade pública
Dimensão		
Verificação, controle e inspeção de dados, de forma a garantir sua qualidade.	Frequentemente	Frequentemente
Indicadores		
Inspeciona e monitora a confiabilidade, consistência e legibilidade dos dados	Sempre	Sempre
Aplica diretrizes, proposta por instituições especialistas, que orientam a verificação e o controle da qualidade dos dados.	Frequentemente	Frequentemente
Recebem assistência para garantir a qualidade dos dados	Ocasionalmente	Ocasionalmente

Fonte: Elaboração própria, 2020.

Inspecionar e verificar a confiabilidade, consistência e legibilidade dos dados é uma prática inerente ao processo de investigação, pois o monitoramento periódico dos dados evita que erros entrem ou permaneçam no conjunto de dados e afetem os resultados da investigação. Por ser uma prática implícita na pesquisa, os pesquisadores indicam que sempre verificam possíveis erros nos dados ao longo de seu ciclo de vida, de forma a garantir a qualidade dos dados. Esse tipo de prática também é reconhecido como uma fase da gestão de dados, porque, graças a isso, a qualidade dos dados é aprimorada. Deve ser mencionado que os mecanismos pelos quais a qualidade é inspecionada e verificada são determinados pelas características e particularidades dos dados.

Já em relação às diretrizes sobre métodos de inspeção, verificação e controle da qualidade dos dados, propostas por instituições especializadas em gestão de dados, os pesquisadores indicam que tais recomendações são frequentemente implementadas. Este indicador é considerado positivo porque a existência de um procedimento de controle de qualidade permite a fácil identificação de potenciais problemas no conjunto de dados e a rápida correção dos erros encontrados.

Finalmente, os pesquisadores indicam que ocasionalmente recebem assistência, de grupos especializados em dados, para orientá-los ou apoiá-los nas referidas atividades. Cabe ressaltar que os pesquisadores são, por si só, competentes para realizar essa tarefa, devido ao seu conhecimento em pesquisa e no campo de estudo.

De maneira geral, os pesquisadores indicam que inspecionam e monitoram constantemente a qualidade dos dados através de seus próprios mecanismos e daqueles

estabelecidos pelas instituições, porém ocasionalmente recebem assistência de especialistas para verificar e avaliar sua qualidade.

Tabela 30 - Dimensão garantia e controle de qualidade dos dados – pesquisadores de universidade particular

ITENS	Média	Mediana	Moda	Desvio padrão	Variância	Curtose	Coefficiente de assimetria	Mínimo	Máximo
Dimensão controle e garantia da qualidade dos dados	3,23	3,33	3,33	0,85	0,73	-0,19	-0,09	1,17	5
Inspeção e monitoramento da qualidade do conjunto de dados	4,11	4	4	0,66	0,44	3,56	-1,3	1,5	5
P.27: verifica a confiabilidade dos dados	4,29	4	4	0,88	0,77	6,16	-2,1	1	5
P.28: verifica a consistência dos dados	4,1	4	4	0,94	0,89	2,74	-1,39	1	5
P.29: O conteúdo dos dados é claro e compreensível	4,22	4	4	0,64	0,41	-0,59	-0,22	3	5
P.30: realiza metadados e verifica sua fácil compreensão	3,84	4	4	1,05	1,09	0,69	-0,98	1	5
Diretrizes que verifiquem, inspecionem e controlem o conjunto de dados	3,35	4	4	1,26	1,59	-0,58	-0,65	1	5
P.31: segue diretrizes, proposto por instituições, que orientem a verificação, inspeção e controle de dados	3,35	4	4	1,26	1,59	-0,58	-0,65	1	5
Assistência para garantir a qualidade dos dados	2,22	2	1	1,29	1,65	-0,76	0,64	1	5
P.32: tem algum tipo de assistência de um grupo de especialista em dados para garantir sua qualidade.	2,22	2	1	1,29	1,65	-0,76	0,64	1	5

Fonte: Elaboração própria, com base em dados processados SPSS 22®, 2020.

Tabela 31- Dimensão garantia e controle de qualidade dos dados – pesquisadores de universidade pública

ITENS	Média	Mediana	Moda	Desvio padrão	Variância	Curtose	Coefficiente de assimetria	Mínimo	Máximo
Dimensão controle e garantia da qualidade dos dados	3,40	3,33	5,00	1,00	1,00	-1,16	0,39	2,08	5,00
Inspeção e monitoramento da qualidade do conjunto de dados	4,43	4,50	5,00	0,57	0,33	0,24	-0,75	2,75	5,00
P.27: verifica a fiabilidade dos dados	4,68	5,00	5,00	0,53	0,28	1,00	-1,36	3	5
P.28: verifica a consistência dos dados	4,43	5,00	5,00	0,80	0,64	8,36	-2,34	1	5
P.29: O conteúdo dos dados é claro e compreensível	4,54	5,00	5,00	0,65	0,42	0,21	-1,12	3	5
P.30: realiza metadados e verifica sua fácil compreensão	4,05	4,00	5,00	1,08	1,16	-0,31	-0,95	2	5
Diretrizes que verifiquem, inspecionem e controlem o conjunto de dados	3,43	4,00	5,00	1,42	2,03	-1,30	-0,34	1	5
P.31: segue diretrizes, proposto por instituições, que orientem a verificação, inspeção e controle de dados	3,43	4,00	5,00	1,42	2,03	-1,30	-0,34	1	5
Assistência para garantir a qualidade dos dados	2,35	1,00	1,00	1,62	2,62	-1,34	0,63	1	5
P.32: tem algum tipo de assistência de um grupo de especialista em dados para garantir sua qualidade.	2,35	1,00	1,00	1,62	2,62	-1,34	0,63	1	5

Fonte: Elaboração própria, com base em dados processados SPSS 22 ®, 2020.

7.3.5 Dimensão preservação de dados

A preservação de dados envolve aspectos de backups e arquivamento. De acordo com o modelo proposto por *DataONE* (2016), o arquivamento está associado a processos de conversão de dados, reformatação e resgate como seleção de dados a serem preservados; criação de versões digitais dos dados coletados em físico; seleção de formatos legíveis; verificação da obsolescência dos dados, e sua migração quando é necessário. Já as atividades relacionadas aos backups consistem em fazer backups em vários formatos, armazenados em vários locais e inspeções periódicas de sua integridade.

A dimensão é composta por oito indicadores que medem as práticas que abrangem a preservação de dados de pesquisa.

- ✓ “Selecionam os dados que devem ser preservados a médio e longo prazo”.
- ✓ “Criam versões digitais dos dados coletados em físico”
- ✓ “Arquivam em formatos legíveis a médio e longo prazo”
- ✓ “Realizam migração de dados quando for necessário e verificam a obsolescência do formato”
- ✓ “Backups em vários formatos e armazenados em diversos locais”.
- ✓ “Inspeção periodicamente os backups para confirmar suas boas condições”.
- ✓ “Orienta a preservação dos dados a través de diretrizes propostas por instituições especialistas”.
- ✓ “Recebe assistência, por parte de especialistas, para preservar os dados.

Em geral, a dimensão preservação de dados, tanto para pesquisadores de universidade particular, quanto pesquisadores de universidade pública, de acordo com suas respectivas médias, 2,97 e 3,07, em termos de categorização, indicam que a presença dessa prática é neutra.

A categoria neutra para essa dimensão, revela que os pesquisadores ocasionalmente realizam atividades mencionadas anteriormente, de forma que os dados se preservem a médio e longo prazo.

Por outro lado, a respectiva mediana de cada grupo reforça o indicado pela média ao ser localizada no mesmo intervalo. Além disso, revela que 50% dos pesquisadores realizam essa prática com um grau de ocorrência igual ou maior que ocasionalmente, sendo frequente e sempre, e os outros 50% o fazem igual que ocasionalmente, ou menor, sendo raramente e nunca.

Em relação à variância e desvio padrão da dimensão, os dois grupos de pesquisadores, indicam pouca dispersão dos dados em relação à média. Os resultados das medidas de

variabilidade revelam que os pesquisadores atuam de forma homogênea, e, portanto, a característica traçada pela média é considerada representativa.

1. Seleção de dados para preservar

A seleção dos dados a serem preservados é feita por razões técnicas, como limites de espaço de armazenamento, e por razões de pesquisa, como a percepção dos pesquisadores sobre o valor e a relevância dos dados. Este indicador revela se os pesquisadores selecionam os dados que devem necessariamente ser preservados a médio e longo prazo.

Em pesquisadores de universidade particular, observa-se que a média é de 3,49, enquanto para pesquisadores de universidade pública é de 3,30, significando uma presença forte da referida prática. Em relação à mediana, para pesquisadores de universidade particular, é 4, reforçando o que é indicado pela média. Em pesquisadores de universidade pública, a mediana é 3, localizado na categoria neutra.

A categoria forte, indica que frequentemente os pesquisadores selecionam os dados, que pelas suas particularidades e valor informativo ou pelo caráter e relevância da pesquisa, devem ser preservados a médio e longo prazo.

No que refere à mediana do grupo de pesquisadores de universidade pública, essa medida revela que a metade dos pesquisadores realizam essa prática em um grau de ocorrência maior, isto é ocasional, frequente e sempre, enquanto a outra metade o realiza de forma ocasional, raramente e nunca.

As medidas de dispersão indicam que nos pesquisadores da universidade particular, observa-se uma variação de 1,33 e um desvio padrão de 1,16, indicando pouca dispersão dos dados em relação à média. Em pesquisadores de universidades públicas, observa-se uma variação de 1,88 e um desvio padrão de 1,37, o que implica uma relativa dispersão dos dados em relação à média. Os valores obtidos revelam que nos dois grupos de pesquisadores a atuação deles é próximo, similar à característica traçada pela média, portanto, devido ao seu comportamento homogêneo, considera-se confiável que a média é representativa.

2. Preservação de dados físicos em versão digital

O arquivamento dos dados em formatos físicos é considerado um dos meios mais vulneráveis e obsoletos, portanto, é essencial migrar esses dados para formatos e armazenamento digitais. Este indicador revela se os pesquisadores criam versões digitais dos dados coletados em forma física.

A média 3,45 para pesquisadores de universidade particular, e 3,43 para pesquisadores de universidade pública, indicam que, em termos de categorização, esta prática tem forte presença. Nos dois grupos de pesquisadores, a mediana reforça o que é indicado pela média. A

categoria forte, significa que frequentemente os pesquisadores migram seus dados em físico a formatos digitais.

Nos pesquisadores de universidade particular, observa-se uma variação de 1,73 e um desvio padrão de 1,32, considerado relativamente pequeno, com pouca dispersão dos dados em relação à média. Devido ao comportamento homogêneo dos dados, confirma-se que a média representa a característica do grupo em relação ao nível de ocorrência com o qual os formatos físicos são convertidos em digitais.

Em pesquisadores de universidade pública, observa-se uma variância de 2,36 e um desvio padrão de 1,54, indicando uma maior dispersão dos dados em relação à média. Os resultados expressam um comportamento um pouco mais heterogêneo nos dados, o que implica que é relativamente diverso o nível de ocorrência em que se realiza a referida prática entre os pesquisadores.

3. Arquivamento em formatos legíveis a médio e longo prazo

Um formato legível refere-se a formatos padrão não proprietários que podem ser abertos em diferentes sistemas operacionais e com diferentes programas ou aplicativos. O indicador mede a ocorrência em que os pesquisadores armazenam seus dados em formatos legíveis a médio e longo prazo.

Em relação ao armazenamento em formatos legíveis, tanto pesquisadores de universidade particular quanto pública, em termos de categorização, a prática tem uma presença forte. A mediana dos dois grupos reforça o indicado pela média.

A categoria forte, indica que os pesquisadores frequentemente armazenam os dados em formatos digitais legíveis a médio e longo prazo. Esta característica se reforça nos resultados obtidos na caracterização dos dados, em que se observa que os formatos de arquivo comumente usados são de caráter padrão possibilitando sua leitura em diferentes sistemas operacionais.

A variância e o desvio padrão para os dois grupos de pesquisadores são considerados relativamente pequenos, indicando pouca dispersão dos dados em relação à média. Devido ao comportamento homogêneo dos dados, confirma-se que a média representa a característica do grupo em relação ao nível de ocorrência com o qual os pesquisadores arquivam seus dados em formatos legíveis.

4. Realiza migração de dados quando for necessário e verifica a obsolescência do formato periodicamente.

A seleção dos formatos utilizados durante a pesquisa dependerá do processo metodológico da investigação. No final da pesquisa, os formatos devem ser avaliados com base no uso que terão os dados após encerrar a pesquisa, decidindo a necessidade de manter seu

formato inicial ou migrar para um formato mais apropriado. Nesse contexto, os formatos padrão e abertos são considerados adequados ao facilitar sua leitura em diversos programas.

Por outro lado, a periodicidade com que as condições dos formatos devem ser verificadas ocorre como resultado dos avanços tecnológicos nos programas que abrem e leem os formatos. É assim que a revisão periódica dos formatos é considerada necessário para evitar a obsolescência e, portanto, a perda de informações. Este indicador revela se os pesquisadores realizam migração de seus dados de um formato para outro, bem como se verificam periodicamente que os formatos de arquivo não estejam obsoletos.

Em geral, tanto o pesquisador de universidade particular quanto de universidade pública, de acordo com seus respectivas médias, 2,93 e 3,07, indicam que se dá de forma ocasional referidas práticas. A mediana para os dois grupos reforça o que é indicado pela média.

Embora o indicador seja caracterizado pelo grau de ocorrência ocasional, observa-se em maior profundidade que, em relação à migração de dados de um formato para outro, os pesquisadores frequentemente o fazem quando consideram necessário. No entanto, eles indicam que ocasionalmente verificam periodicamente se os formatos estão em boas condições.

Em relação à variância e desvio padrão dos dois grupos, eles revelam pouca dispersão dos dados em relação à média. As medidas de variabilidade, devido ao comportamento homogêneo dos dados, confirmam a representatividade da característica traçada pela média.

5. Backups de dados

O termo backup refere-se a uma cópia que é feita do arquivo original a fim de evitar a perda parcial ou total dos dados. Estas cópias de segurança são geralmente realizadas em diversos formatos e em locais de armazenamento diferentes, evitando danos ou perda de informação. Nesse sentido o indicador tem como propósito revelar se os pesquisadores efetuam backups em vários formatos, e armazenam esses backups em vários locais.

Em pesquisadores de universidade particular, observa-se que a média é de 3,34, enquanto para pesquisadores de uma universidade pública é de 3,54, significa que, em termos de categorização, há uma forte presença dessa prática. A mediana para os dois grupos reforça o que é indicado pelas médias.

A categoria forte indica que os pesquisadores frequentemente fazem cópias de segurança em vários formatos e as armazenam em locais diferentes. Essa característica é reforçada nos resultados obtidos na caracterização dos dados, nos quais se observa que os pesquisadores geralmente depositam seus dados em vários locais, evitando possíveis perdas.

Em relação à variância e desvio padrão dos dois grupos, eles revelam pouca dispersão dos dados em relação à média. As medidas de variabilidade indicam que os pesquisadores

atuam de forma similar ou próxima à média. Devido ao comportamento homogêneo dos dados a média é representativa.

6. Inspeção periódica de backups

Fazer vários backups não é suficiente para garantir a preservação dos dados. Além de criar vários backups, é necessário verificar periodicamente sua integridade para garantir que, em uma possível situação de perda do arquivo ou armazenamento original, as cópias estejam acessíveis. O indicador determina se os pesquisadores inspecionam periodicamente os backups para confirmar suas boas condições.

Em relação à periodicidade com que revisam os backups, pesquisadores de universidade particular e pública indicam em termos de categorização que esta prática é neutra. As medianas dos dois grupos reforçam o que é indicado pela média. A categoria neutra revela que os pesquisadores ocasionalmente inspecionam a condição dos backups.

No referente aos pesquisadores de universidade particular, observa-se uma variância de 1,33 e um desvio padrão de 1,16, quanto que para os pesquisadores de universidade pública se registra uma variância de 1,40 e desvio padrão de 1,18, o que indica pouca dispersão dos dados em relação à média. Os valores obtidos nas medidas de variabilidade revelam que devido ao comportamento homogêneo dos dados, a média é um valor confiável, e, portanto, representativa. Igualmente esses valores afirmam que os pesquisadores atuam de forma similar à característica da média.

7. Diretrizes que orientem a preservação de dados

As organizações especializadas em gestão de dados criam protocolos de preservação com o objetivo de expandir boas práticas, entre os pesquisadores, que garantam a máxima longevidade dos dados da pesquisa. O indicador revela se os pesquisadores orientam a preservação dos seus dados a través de diretrizes propostas por instituições, observatórios, associações científicas, data centers, repositórios, etc.

A média 2,23 para pesquisadores de universidade particular, e 2,54 para pesquisadores de universidade pública, indicam que, em termos de categorização, esta prática tem presença neutra. A categoria neutra aponta que os pesquisadores ocasionalmente orientam a preservação de seus dados a partir das diretrizes estabelecidas por instituições especializadas.

Já as medianas classificam a prática como fraca, revelando que a metade dos pesquisadores realizam tais práticas em um grau de ocorrência igual ou menor a raramente sendo nunca, e a outra metade o realiza igual ou maior a raramente, sendo ocasional, frequente e sempre.

Em pesquisadores de universidade particular, observa-se uma variância de 1,23 e um desvio padrão de 1,11. Pode ser considerado relativamente pequeno, indicando pouca dispersão dos dados em relação à média. Os valores obtidos reforçam a representatividade da média e revelam que os pesquisadores atuam de forma homogênea e similar à caracterizada traçada pela média.

Em pesquisadores de universidade pública, observa-se uma variância de 2,26 e um desvio padrão de 1,50, indicando uma maior dispersão dos dados em relação à média. Os valores obtidos indicam que os pesquisadores atuam relativamente de uma forma heterogênea em relação ao grau de ocorrência em que aplicam diretrizes de preservação.

8. Assistência para preservar dados

Devido à complexidade de preservar eficientemente dados de natureza diversa, atualmente as organizações fornecem serviços de preservação nos quais o autor entrega os dados a um custodiante qualificado que tomará os cuidados necessários para conservá-los. O indicador mede se os pesquisadores têm recebido algum tipo de assistência de um grupo de especialistas para preservar seus dados de pesquisa a médio e longo prazo.

A média 1,8 para pesquisadores de universidade particular, e 1,86 para pesquisadores de universidade pública, indicam que, em termos de categorização, esta prática tem uma presença fraca. Nos dois grupos de pesquisadores, a mediana reforça o que a média indica.

A categorização fraca indica que os pesquisadores raramente possuem alguma forma de assistência de um grupo de especialistas em dados para preservá-los a médio e longo prazo. Em relação à variância e desvio padrão os dois grupos de pesquisadores, indicam pouca dispersão dos dados em relação à média. As medidas de variabilidade indicam que os pesquisadores atuam, em relação à referida prática, de forma homogênea, próximo à característica traçada pela média. A homogeneidade da amostra indica que a média é confiável em termos de representatividade.

9. Considerações da dimensão preservação.

O resultado da dimensão, revela que a preservação entre pesquisadores de universidade particular e pública é igual. Observando os resultados de cada indicador, estes indicam que os pesquisadores atuam de forma similar, e apenas diferem na seleção dos dados a preservar devido a suas condições especiais.

Tabela 32- Caracterização dimensão preservação de dados

ITENS	Pesquisadores de universidade particular	Pesquisadores de universidade pública
Dimensão		
Preservação de dados	Ocasionalmente	Ocasionalmente
Indicadores		

ITENS	Pesquisadores de universidade particular	Pesquisadores de universidade pública
Dimensão		
Seleciona os dados que devem ser preservados a médio e longo prazo.	Frequentemente	Ocasionalmente
Cria versões digitais dos dados coletados em físico	Frequentemente	Frequentemente
Arquivamento em formatos legíveis a médio e longo prazo	Frequentemente	Frequentemente
Verifica a obsolescência do formato e realiza migração de dados quando for necessário	Ocasionalmente	Ocasionalmente
Backups em vários formatos e armazenados em diversos locais.	Frequentemente	Frequentemente
Inspeção periodicamente os backups para confirmar suas boas condições.	Ocasionalmente	Ocasionalmente
Orienta a preservação dos dados a través de diretrizes propostas por instituições especialistas.	Ocasionalmente	Ocasionalmente
Recebe assistência, por parte de especialistas, para preservar os dados.	Raramente	Raramente

Fonte: Elaboração própria, 2020

Os resultados das práticas de preservação revelam que os pesquisadores protegem sua informação criando frequentemente backups em vários formatos e armazenados em vários locais; no entanto, eles indicam que ocasionalmente revisam periodicamente as condições desses backups. Fazer backups em vários formatos e locais leva a minimizar o risco de perda acidental ou pela obsolescência tecnológica. A revisão ocasional das condições dos backups pode conduzir a divergências no conteúdo entre a cópia e o original, ou pode fazer com que os backups fiquem inacessíveis.

Em relação ao arquivamento de dados, é revelado que os pesquisadores frequentemente migram seus dados de um formato para outro quando consideram apropriado, especialmente criam versões digitais quando os dados estão em formato físico. Igualmente, frequentemente selecionam formatos padrão e abertos que garantem a leitura dos dados a médio e longo prazo; porém, os pesquisadores ocasionalmente verificam periodicamente se os formatos estão em boas condições. Como evidenciado na caracterização dos dados, os pesquisadores expressam que os formatos mais utilizados são do tipo Doc, Txt, pdf, XLS, XML, CSV, SPSS, SAS, SAV, sendo os formatos considerados padrão. No entanto, esses formatos não estão isentos de não serem comumente usados no futuro, de modo que questões de migração e revisão constante se tornam relevantes.

Selecionar corretamente os formatos e executar a migração de dados garante seu acesso a longo prazo, minimizando os problemas de compatibilidade e desatualização entre o formato e os sistemas operacionais. O uso de formatos padrão e abertos permite a leitura eficiente dos

dados por vários programas ou aplicativos. Da mesma forma, migrar formatos quando necessário, especialmente quando são utilizados formatos tradicionais de coleta como o papel, são estratégias para combater a obsolescência de determinados formatos. Cabe ressaltar que entre as atividades necessárias à preservação está a verificação periódica das condições dos formatos, nos quais, neste caso, é apresentado ocasionalmente o que poderia causar problemas como os já mencionados.

Por outro lado, pesquisadores de universidade particular frequentemente selecionam, sob certos critérios, os dados que consideram deveriam ser preservados, enquanto pesquisadores de universidade pública o fazem ocasionalmente. Selecionar os dados para preservar implica identificar os dados de produção difícil, alto custo de reprodução, alto valor informativo, de eventos únicos ou experimentais, e os necessários para validar os resultados da investigação. Nem todos os dados de pesquisa têm características de valor relevantes para preservá-los; além disso, devido ao espaço de armazenamento, é necessário, com base em determinados critérios, selecionar e verificar os dados que se consideram devem ser preservados em um determinado tempo.

No referente ao tempo de armazenamento, a caracterização dos dados revela que os pesquisadores de universidade particular armazenam indefinidamente os dados, enquanto os pesquisadores de universidade pública o armazenamento de dados é entre 1 a 5 anos. A necessidade de estimar o tempo de preservação se determina a partir de diversos critérios estabelecidos pelos envolvidos na pesquisa, tendo como finalidade evitar o desperdício de espaço de armazenamento ou o abandono dos dados.

Finalmente, os pesquisadores indicam que ocasionalmente orientam a preservação de seus dados através das melhores práticas recomendadas por instituições especializadas, e raramente recebem qualquer tipo de assistência especializada para a referida prática. Receber raramente assistência especializada e, aplicar ocasionalmente boas práticas, pode resultar na preservação inadequada dos dados, expondo-o a danos ou perdas. Desde a perspectiva da abertura dos dados, sua preservação parcial ou ineficiente pode afetar seu acesso a médio e longo prazo, prejudicando sua potencial reutilização.

Por último, outro aspecto relevante na preservação dos dados é o local de armazenamento já que protege os dados de possíveis perdas e determina se potencial acesso ou abertura. Os resultados da caracterização dos dados demonstram que os pesquisadores diversificam os espaços onde armazenam os dados, nos quais costumam fazê-lo em computadores, discos rígidos externos e na nuvem. Desde o ponto de vista do acesso aberto aos dados, os repositórios são ocasionalmente usados, no qual os repositórios institucionais e de

grupos de pesquisa prevalecem, no entanto, sua abertura e visibilidade são limitadas. Em termos de preservação, os repositórios são considerados os locais mais adequados e mais seguros.

Tabela 33- Dimensão preservação– pesquisadores de universidade particular

ITENS	Média	Mediana	Moda	Desvio padrão	Variância	Curto se	Coefficiente de assimetria	Mínimo	Máximo
Dimensão preservação	2,97	2,94	2,25	0,75	0,56	-0,24	0,36	1,63	5
Seleção de dados para preservar	3,49	4	4	1,16	1,33	-0,85	-0,34	1	5
P. 33: seleção de dados para preservar a médio e longo prazo	3,49	4	4	1,16	1,33	-0,85	-0,34	1	5
Preservação de dados físicos em versão digital	3,45	4	4	1,32	1,73	-0,62	-0,63	1	5
P. 34: cria versões digitais dos dados coletados em físico	3,45	4	4	1,32	1,73	-0,62	-0,63	1	5
Arquivamento em formatos legíveis	3,92	4	5	1,04	1,07	-0,07	-0,74	1	5
P. 35: arquiva os dados em formatos legíveis a médio e longo prazo	3,92	4	5	1,04	1,07	-0,07	-0,74	1	5
Verifica a obsolescência do formato e realiza migração de dados quando for necessário	2,93	3	2	1,1	1,21	-0,79	0,25	1	5
P. 36 verifica periodicamente a obsolescência dos formatos que armazenam os dados	2,76	2	2	1,31	1,7	-1,01	0,4	1	5
P.37 realiza migração de dados de um formato para outro, conservando suas características, quando for necessário.	3,1	3	4	1,19	1,41	-0,89	-0,12	1	5
Backups de dados	3,34	3,5	4	1,03	1,05	-0,48	-0,31	1	5
P. 38 realiza backups em vários formatos.	3,39	4	4	1,18	1,4	-0,84	-0,3	1	5
P.39 realiza backups em diversos locais.	3,29	3	4	1,2	1,45	-0,85	-0,24	1	5
Inspeção periódica de backups	2,51	2	2	1,16	1,33	-0,84	0,26	1	5
P.40 inspeção periodicamente os backups para confirmar suas boas condições.	2,51	2	2	1,16	1,33	-0,84	0,26	1	5
Diretrizes que orientem a preservação de dados	2,33	2	2	1,11	1,23	-0,74	0,49	1	5
P. 41 orienta a preservação dos dados a través de diretrizes propostas por instituições especialistas.	2,33	2	2	1,11	1,23	-0,74	0,49	1	5
Assistência para preservar dados	1,8	1	1	1,11	1,24	0,65	1,31	1	5
P.42 recebe assistência, por parte de especialistas, para preservar os dados.	1,8	1	1	1,11	1,24	0,65	1,31	1	5

Tabela 34- Dimensão preservação – pesquisadores de universidade pública

	Média	Mediana	Moda	Desvio padrão	Variância	Curto se	Coefficiente de assimetria	Mínimo	Máximo
Dimensão preservação	3,07	3,00	3,19	0,93	0,86	0,10	0,28	1	5
Seleção de dados para preservar a médio e longo prazo	3,30	3,00	4,00	1,37	1,88	-1,10	-0,30	1	5
P. 33: seleção de dados para preservar a médio e longo prazo	3,30	3,00	4,00	1,37	1,88	-1,10	-0,30	1	5
Preservação de dados físicos em versão digital	3,43	4,00	5,00	1,54	2,36	-1,26	-0,50	1	5
P. 34: cria versões digitais dos dados coletados em físico	3,43	4,00	5,00	1,54	2,36	-1,26	-0,50	1	5
Arquivamento em formatos legíveis a médio e longo prazo	3,92	4,00	4,00	1,09	1,19	1,21	-1,19	1	5
P. 35: arquiva os dados em formatos legíveis a médio e longo prazo	3,92	4,00	4,00	1,09	1,19	1,21	-1,19	1	5
Verifica a obsolescência do formato e realiza migração de dados quando for necessário	3,07	3,00	3,00	1,12	1,25	-0,72	0,07	1	5
P. 36 verifica periodicamente a obsolescência dos formatos que armazenam os dados	3,00	3,00	2,00	1,31	1,72	-1,25	0,16	1	5
P.37 realiza migração de dados de um formato para outro, conservando suas características, quando for necessário.	3,14	3,00	4,00	1,25	1,56	-0,85	-0,27	1	5
Backups de dados	3,54	4,00	4,00	1,15	1,32	-0,77	-0,45	1	5
P. 38 realiza backups em vários formatos.	3,43	4,00	4,00	1,21	1,47	-0,95	-0,33	1	5
P.39 realiza backups em diversos locais.	3,65	4,00	4,00	1,23	1,51	-0,62	-0,70	1	5
Inspeção periódica de backups	2,86	3,00	2,00	1,18	1,40	-1,01	0,17	1	5
P.40 inspeção periodicamente os backups para confirmar suas boas condições.	2,86	3,00	2,00	1,18	1,40	-1,01	0,17	1	5
Diretrizes que orientem a preservação de dados	2,54	2,00	1,00	1,50	2,26	-1,37	0,39	1	5
P. 41 orienta a preservação dos dados a través de diretrizes propostas por instituições especialistas.	2,54	2,00	1,00	1,50	2,26	-1,37	0,39	1	5
Assistência para preservar dados	1,86	1,00	1,00	1,32	1,73	0,76	1,42	1	5
P.42 recebe assistência, por parte de especialistas, para preservar os dados.	1,86	1,00	1,00	1,32	1,73	0,76	1,42	1	5

7.4 PRÁTICAS DE COMPARTILHAMENTO DE DADOS DE PESQUISA

De acordo com Borgman (2012), o compartilhamento de dados refere-se à liberação de dados de pesquisa com a finalidade de conceder acesso e reuso desse recurso a terceiros. Anteriormente à disponibilização dos dados, é considerado que submeter os dados a um processo de gestão garantirá que os dados estejam adequadamente estruturados e curados e, portanto, em condições ideais para colocá-los no sistema de acesso aberto. Entretanto, para que os dados sejam disponibilizados ao público, não é condição que esse recurso seja submetido a um processo de gerenciamento. Nesse sentido, a liberação dos dados independe das condições de sua qualidade, encontrando publicamente dados minimamente curados ou não curados de forma alguma.

Entre outros aspectos a serem considerados no compartilhamento de dados, está o estabelecimento de políticas de propriedade intelectual que determinam as condições sob as quais será realizada uma prática justa. Nesse sentido, a política de acesso e compartilhamento de dados, bem como as licenças de direitos autorais, determinará o nível de restrição ou abertura e sob quais condições os dados podem ser reusados por potenciais usuários.

Finalmente, em relação aos meios pelos quais o acesso aberto é concedido, Borgman (2012) contempla que o acesso pode ser outorgado de forma privada por solicitação direta ou publicamente por meio de depósitos na *Web*, repositórios, centros ou revistas científicas.

Nesse contexto, para revelar as práticas associadas ao compartilhamento de dados de pesquisa são levadas em consideração as questões 43 a 57 do questionário aplicado, tendo em vista que as questões extraídas são de dois tipos: escala nominal e escala Likert.

7.4.1 Dimensão disponibilidade de dados de pesquisa

A disponibilidade dos dados é considerada uma consequência do cenário atual da ciência aberta ao reconhecer esse recurso como uma fonte inestimável de informações, o qual se for preservado eficientemente pode ser reutilizada para diversos fins científicos. Borgman (2012) menciona que, devido ao valor dos dados para transmitir informações ao longo do tempo, os pesquisadores decidem compartilhar todos ou parte de seus dados, seja por exigência dos setores que financiam as pesquisas, pelas revistas científicas ou por interesses individuais dos pesquisadores.

Quando os pesquisadores decidem disponibilizar abertamente os dados, surgem questões sobre em que estágio os dados serão compartilhados (brutos, processados, publicados), quais usuários poderão acessar os dados? Por quais meios de comunicação os usuários em potencial poderiam acessar? A dimensão, é, portanto, medida a partir de três indicadores:

- ✓ “Quantidade de dados disponíveis”
- ✓ “Tipos de dados disponíveis”
- ✓ “Usuários que têm acesso aos dados de pesquisa do pesquisador”
- ✓ “Meios de acesso aos dados de pesquisa”

Os referidos indicadores são de caráter nominal, devido a isso são apresentados em termos de frequência estatística.

1. Quantidade de dados disponíveis

As investigações são realizadas sob condições de privacidade, confidencialidade, ética, propriedade intelectual, políticas, considerações legais, envolvendo questões financeiras e possivelmente comerciais. Nessa direção, os aspectos mencionados, em certos casos particulares, podem restringir a liberação de dados, entendendo que, sob essas barreiras, nem todos os dados podem ou devem ser compartilhados abertamente.

A Tabela 35 exibe a quantidade de dados de pesquisa que os professores compartilham com outros pesquisadores, comunidade acadêmica, instituições, centros, associações, redes, entre outros. 70,6% dos pesquisadores de universidade particular, e 48,64% dos pesquisadores de universidade pública declararam compartilhar parte de seus dados. 27,02% dos pesquisadores de universidade pública e 11,8% dos pesquisadores de universidade particular indicaram não compartilhar seus dados de pesquisa. 24,32% dos pesquisadores de universidade pública e 17,6% dos pesquisadores de universidade particular indicaram compartilhar todos os dados da pesquisa.

No final da tabela 35, são encontradas informações sobre o teste do qui-quadrado de Pearson. Como o nível de significância selecionado foi de 5% (0,05) e o Pvalue associado à estatística do teste é de 0,086, a hipótese nula não é rejeitada, significando que não há associação estatisticamente significativa entre o tipo de universidade e a quantidade de dados que eles compartilham.

Tabela 35 – Quantidade de dados disponíveis

Itens	Estatística	Privada	Público	Total
Nenhum dos meus dados	Frequência	6	10	16
	Porcentagem	11,8%	27,02%	18,1%
Parte dos meus dados	Frequência	36	18	54
	Porcentagem	70,6%	48,64%	61,4%

Itens	Estatística	Privada	Público	Total
Todos os meus dados	Frequência	9	9	18
	Porcentagem	17,6%	24,32%	20,5%
Total	Frequência	51	37	88
	% do total	100%	100%	100%
Qui-quadrado de Pearson		Valor	df	Significância Bilateral
		4,897	2	0,086

Fonte: dados de pesquisa, SPSS 22 ®, 2020

Tende a predominar o compartilhamento de parte dos dados de pesquisa pelos pesquisadores de universidade particular e pública. Os resultados do teste do qui-quadrado, indicam que não há associação ou relação entre a universidade e o valor que os pesquisadores compartilham. Nesse sentido, infere-se que a quantidade de dados compartilhados não depende da universidade à qual o pesquisador está vinculado, ou vice-versa, ou seja, não há relação nas duas direções.

Os motivos, argumentos e incentivos para liberar os dados variam e dependem das partes envolvidas nas pesquisas, desde os pesquisadores, instituições educativas, organizações pesquisadoras, entidades financiadoras até o governo. A complexidade na abertura dos dados se dá por inúmeras razões, principalmente porque para que um sistema de acesso aberto se dê em um contexto científico é necessária uma base jurídica que o sustente, bem como a disposição dos pesquisadores para que aconteça.

2. Tipos de dados disponíveis

A biblioteca da Cepal caracteriza os dados de pesquisa a partir do seu estado. Segundo a biblioteca (*online*) os dados brutos ou primários referem-se aos dados originais, coletados na aplicação do instrumento científico, que não foram processados ou analisados. Os dados processados referem-se a dados que foram organizados, traduzidos, validados, verificados. Os dados publicados ou finais se referem aos dados consolidados que foram submetidos a determinados processos, levando-os ao seu estado final de publicação.

A Tabela 36 expõe o tipo de dados de pesquisa que os professores compartilham com outros pesquisadores. 17,6% dos pesquisadores de universidade particular afirmaram compartilhar dados brutos ou não processados, 8,1% dos pesquisadores de universidade pública compartilham esse tipo de dados. 25,5% dos pesquisadores de universidade particular e 21,6% dos pesquisadores de universidade pública compartilham dados processados. 45,1% dos pesquisadores universidade particular, e 43,3% dos pesquisadores de universidade pública declaram compartilhar dados publicados.

Os pesquisadores que na questão de quantidade de dados responderam não compartilhar nenhum dos seus dados de pesquisa, ignoraram as perguntas 45 a 58 e continuaram na seção do questionário reuso de dados de pesquisa. É assim que pelo motivo exposto, 27,0% dos pesquisadores de universidade público e o 11,8% dos pesquisadores de universidade particular não responderam a esta questão.

No final da Tabela 36, há informações sobre o teste do qui-quadrado de Pearson. Como o nível de significância selecionado foi de 5% (0,05) e o Pvalue associado à estatística do teste é de 0,924, a hipótese nula não é rejeitada. Em outras palavras, é evidente que não há associação estatisticamente significativa entre o tipo de universidade e o tipo de dados que os pesquisadores compartilham.

Tabela 36-Tipos de dados disponíveis

Itens	Estatística	Particulares	Público	Total
Dados brutos ou não processados	Frequência	9	3	12
	Porcentagem	17,6%	8,1%	13,7%
Dados processados	Frequência	13	8	21
	Porcentagem	25,5%	21,6%	23,9%
Dados publicados	Frequência	23	16	39
	Porcentagem	45,1%	43,3%	44,3%
Não responde	Frequência	6	10	16
	Porcentagem	11,8%	27,0%	18,1%
Total	Frequência	51	37	88
	% do total	100%	100%	100%
Qui-quadrado de Pearson		Valor	df	Significância Bilateral
		,475	3	,924

Fonte: dados de pesquisa, SPSS 22 ®, 2020

Em geral, os pesquisadores de universidade particular e pública tendem a compartilhar os dados em fase publicados, isto é, dados consolidados no estado final de publicação. Os dados publicados são aqueles que tendem a ser expostos resumidamente em gráficos, tabelas, textos, etc., de modo que sustente e justifique as descobertas da pesquisa. A depender da forma em que os dados sejam apresentados, serão considerados ricos ou não em termos da capacidade de transferir informação em outros contextos científicos.

Os resultados do teste do qui-quadrado, indica que não há associação ou relação entre a universidade e o tipo de dados que disponibilizam. Infere-se que a o estado dos dados compartilhados não depende da universidade à qual o pesquisador está vinculado, ou seja, não há relação entre tais variáveis.

3. Usuários que têm acesso aos dados científicos do pesquisador

Este indicador revela a quem o pesquisador dá acesso aos seus dados de pesquisa. A tabela 37 expõe as informações sobre o acesso a dados de pesquisa e o tipo de instituição. 19,6% dos

professores de universidade particular afirmaram que seus dados estão disponíveis para a instituição em que trabalham, enquanto isso ocorre apenas para 8,1% dos professores de instituição pública. 15,7% dos professores de universidade particular, e 18,9% dos professores de universidade pública declararam que seus dados estão disponíveis para todos. 31,4% dos pesquisadores de universidade particular, e 35,2% dos professores de universidade pública disponibilizam seus dados através de solicitação direta. 19,6% dos professores de universidade particular disponibilizam seus dados ao grupo de pesquisa ao qual pertencem. Por outro lado, 8,1% dos professores de universidade pública disponibilizam seus dados para a equipe docente ou departamento em que trabalham. Os professores que não responderam, devido a que não compartilham nenhum dos seus dados, são 27,0% dos professores que pertencem a uma instituição pública, e 11,8% dos professores de universidade particular.

No final da Tabela 37, há informações sobre o teste do qui-quadrado de Pearson. O nível de significância selecionado foi de 5% (0,05) e o Pvalue associado à estatística do teste é 0,030. A hipótese nula é rejeitada, em outras palavras, significa que existe uma associação estatisticamente significativa entre o tipo de universidade e o acesso aos dados da pesquisa. Infere-se que os usuários que acessam aos dados dependem da universidade à qual o pesquisador está vinculado, ou vice-versa, ou seja, há relação entre tais variáveis.

Tabela 37- Acesso a dados de pesquisa

Itens	Estatística	Particular	Público	Total
Não responde	Frequência	6	10	16
	Porcentagem	11,8%	27,0%	18,2%
Disponível para todos	Frequência	8	7	15
	Porcentagem	15,7%	18,9%	17,0%
Disponível para a instituição onde trabalho	Frequência	10	3	13
	Porcentagem	19,6%	8,1%	14,8%
Disponível para o meu grupo de pesquisa	Frequência	10	0	10
	Porcentagem	19,6%	0%	11,4%
Disponível sob solicitação direta	Frequência	16	13	29
	Porcentagem	31,4%	35,2%	32,9%
Disponível para a faculdade ou departamento onde trabalho	Frequência	0	3	3
	Porcentagem	0%	8,1%	3,4%
Disponível para a rede nacional/internacional a que pertença	Frequência	1	1	2
	Porcentagem	1,9%	2,7%	2,3%
Total	Frequência	51	37	88
	% do total	100,0%	100,0%	100,0%
Qui-Quadrado de Pearson		Valor	df	Significância bilateral
		13,962	6	,030

Fonte: dados de pesquisa, SPSS 22 ®, 2020

Em geral, os pesquisadores de universidade particular tendem a colaborar compartilhando seus dados, principalmente entre os científicos que os solicitem diretamente, ou entre colegas do grupo de pesquisa, e instituição ao qual pertencem como resultado de sua relação profissional. Já os pesquisadores de universidade pública disponibilizam seus dados, em geral, a todos e aos interessados que os solicitem diretamente.

A relação de dependência ou associação entre as variáveis faz sentido, uma vez que os processos ou políticas de pesquisa que as universidades determinam podem ou não contemplar aspectos sobre intercambio de dados, em que independentemente de serem concebidos, influem na decisão do pesquisador sobre a quem conceder o acesso a seus dados de pesquisa.

4. Meios para acessar os dados de pesquisa

Conforme afirmado por Borgman (2012) e *The University of Edinburgh* (2017, online), as partes interessadas podem acessar os dados por meio de canais informais de comunicação, como solicitação direta ao autor, ou por canais formais, como publicações em revistas, sites, centros ou repositórios de dados.

A tabela 38 indica os canais pelos quais os pesquisadores disponibilizam seus dados e o tipo de instituição a que pertencem. 39,2% dos pesquisadores de universidade particular, e 32,5% dos pesquisadores de universidade pública permitem o acesso a seus dados sob solicitação direta. 15,7% dos pesquisadores de universidade particular e 13,5% dos professores de universidade pública, compartilham seus dados através de repositórios de dados de revistas científicas. Aqueles que concedem acesso a seus dados por meio de repositórios de dados temáticos são 7,9% dos pesquisadores de universidade particular, e 18,9% dos pesquisadores de universidade pública. 15,7% dos professores que concedem acesso aos seus dados através de repositórios de dados institucionais são de universidade particular, enquanto apenas 2,7% pertencem a universidade pública. 5,9% dos pesquisadores de universidade particular, e 5,4% dos pesquisadores de universidade pública, compartilham seus dados através de repositórios pertencentes aos seus grupos de pesquisa. O 3,8% dos professores que concedem acesso aos seus dados através de repositórios de dados de associações científicas e repositórios de dados internacionais pertencem a instituições particulares, enquanto que professores de universidade pública não compartilham seus dados dessa maneira.

No final da Tabela 38, há informações sobre o teste do qui-quadrado de Pearson. O nível de significância selecionado foi de 5% (0,05) e o Pvalue associado à estatística do teste é 0,325. A hipótese nula não é rejeitada, o que indica que não há associação estatisticamente significativa entre o tipo de universidade e os canais de acesso aos dados da pesquisa.

Tabela 38- Canais de acesso a dados

Itens	Estatística	Particular	Público	Total
Não responde	Frequência	6	10	16
	Porcentagem	11,8%	27,0%	18,2%
Solicitação direta	Frequência	20	12	32
	Porcentagem	39,2%	32,5%	36,4%
Repositório de dados de revistas científicas	Frequência	8	5	13
	Porcentagem	15,7%	13,5%	14,8%
Repositório de dados temáticas	Frequência	4	7	11
	Porcentagem	7,9%	18,9%	12,5%
Repositório de dados institucional	Frequência	8	1	9
	Porcentagem	15,7%	2,7%	10,2%
Repositório do meu grupo de pesquisa	Frequência	3	2	5
	Porcentagem	5,9%	5,4%	5,7%
Repositório de dados de associações científicas	Frequência	1	0	1
	Porcentagem	1,9%	0	1,1%
Repositório de dados internacional	Frequência	1	0	1
	Porcentagem	1,9%	0	1,1%
Total	Frequência	51	37	88
	% do total	100,0%	100,0%	100,0%
Qui-quadrado		Valor	df	Significância bilateral
		8,088	7	,325

Fonte: Dados de pesquisa, SPSS 22®, 2020

Em geral o meio que predomina para obter acesso aos dados é por solicitação direta ao pesquisador. Além desse meio, os pesquisadores de universidade particular declaram usar repositórios de revistas científicas e repositório institucional, enquanto os pesquisadores de universidade pública usam repositórios de dados temáticos e repositórios de revistas científicas. Atualmente as revistas solicitam a vinculação das bases de dados aos artigos científicos.

Os resultados do teste qui-quadrado revelam que não há dependência entre a universidade e o canal de comunicação científica escolhido pelo pesquisador para acessar os dados. Nesse sentido, os processos investigativos ou política de pesquisa não se relacionam à decisão do pesquisador de compartilhar seus dados através de um meio específico.

5. Considerações da dimensão disponibilidade de dados de pesquisa

Com base no exposto, os pesquisadores são perfilados de acordo com o tipo de universidade a que pertencem e as características predominantes em relação à disponibilidade de seus dados em acesso aberto.

Tabela 39- Caracterização disponibilidade de dados de pesquisa

ITENS	Pesquisadores de Universidade Particular	Pesquisadores de Universidade Pública
Quantidade de dados disponíveis	Parte dos meus dados	Parte dos meus dados
Tipos de dados disponíveis	Dados publicados	Dados publicados
Usuários que têm acesso aos dados	Disponível aos usuários que os solicitem diretamente	Disponível aos usuários que os solicitem diretamente

ITENS	Pesquisadores de Universidade Particular	Pesquisadores de Universidade Pública
Meios de acesso aos dados de pesquisa	Solicitação direta	Solicitação direta

Fonte: Elaboração própria, 2020.

A partir das características predominantes, assume-se que:

Em relação à disponibilidade dos dados os pesquisadores de universidade particular e pública atuam de forma semelhante. Os pesquisadores tendem a disponibilizar parte de seus dados de pesquisa. Esses dados compartilhados em geral estão na fase publicada, ou seja, dados consolidados e em seu estado final de publicação. Os usuários a quem os pesquisadores permitem acesso a seus dados são os interessados que os solicitam diretamente a eles. O meio comum de acesso é por solicitação direta.

Nesse contexto em que os dados são comumente compartilhados em parte e em seu estado final de publicação, possíveis problemas de interpretação e reuso podem surgir uma vez que se elimina parte de seu contexto inicial. Adicionado à possibilidade de não atribuir metadados que os descrevam em sua totalidade, aumenta a probabilidade de que os dados sejam inúteis.

Por outro lado, os usuários que acessam os dados e os canais pelos quais são transferidos, predomina os usuários que solicitam os dados de forma privada, ou seja, por solicitação direta ao autor. Do ponto de vista da ciência aberta, embora os pesquisadores tendam a compartilhar seus dados, reconhece-se que a prática de intercambio tem barreiras associadas à restrição e limitação da abertura dos dados devido a que seu compartilhamento não se dá através de plataformas que incluam um maior público científico.

Finalmente, conclui-se que a disponibilidade de dados abertos é caracterizada por pesquisadores que compartilham todos ou parte de seus dados em seu estado final de publicação. O compartilhamento de seus dados tende a ocorrer em privado por solicitação direta ao autor. Os usuários que comumente acedem aos dados disponibilizados pelos pesquisadores são pesquisadores que os solicitam diretamente e colegas de seu grupo de pesquisa, como resultado da proximidade de suas relações profissionais e em apoio a suas ocupações como pesquisadores.

7.4.2 Dimensão qualidade dos dados de pesquisa compartilhados

Ante o panorama de tornar os dados disponíveis abertamente, um dos desafios para os pesquisadores que os compartilham é convertê-los em dados facilmente interpretáveis e, portanto, potencialmente utilizáveis ao longo do tempo. Nesse sentido, criar dados interpretáveis supõe que eles passem pelas fases de gestão de dados, o que garantirá a qualidade dos dados através de uma adequada organização, descrição e preservação. Portanto, dados

ricamente estruturados e curados são considerados em excelentes condições para serem compartilhados, evitando usos indevidos ou interpretações errôneas. A dimensão, é, portanto, medida a partir de dois indicadores:

- ✓ “Os dados compartilhados são gerenciados sob diretrizes de organizações especializadas em gestão de dados”.
- ✓ “Os dados compartilhados não são estruturados ou documentados”.

Os referidos indicadores são de caráter escala Likert, devido a isso são apresentados em termos de indicadores estatísticos básicos de forma a caracterizar as práticas.

1. Dados de pesquisa gerenciados sob diretrizes institucionais.

Pesquisadores de universidade particular e universidade pública, de acordo com seus respectivas médias, 2,43 e 2,38, em termos de categorização, indicam que a presença dessa prática é neutra. A categorização neutra revela que os dados compartilhados ocasionalmente são gerenciados e estruturados sob modelos ou diretrizes recomendadas por organizações especializadas em gestão de dados.

As medidas de variabilidade para os dois grupos de pesquisa, conforme se observa nas tabelas 41 e 42, indicam uma relativa dispersão dos dados em relação à média. Os valores obtidos expressam que embora a média seja representativa, os pesquisadores atuam de uma forma relativamente heterogênea em relação à referida prática.

A mediana em pesquisadores de universidade particular se localiza na categorização neutra, reforçando a representação da média, além de indicar que 50% dos pesquisadores gerenciam seus dados compartilhados sob diretrizes em um grau de ocorrência igual ou maior a ocasionalmente, frequente, e sempre, enquanto que o outro 50% o realiza igual ou inferior a ocasionalmente, raramente e nunca.

Já a mediana em pesquisadores de universidade pública se localiza na categoria fraca, revelando que 50% dos pesquisadores realizam tais práticas em um grau de ocorrência igual ou inferior a raramente e nunca, e a outro 50% o realiza igual ou maior raramente, ocasional, frequente e sempre.

2. Dados de pesquisa não estruturados ou documentados.

Entre os pesquisadores de universidade particular, a média 2,37, indica que, em termos de categorização, esta prática é neutra. A categorização neutra revela que os dados compartilhados ocasionalmente não são estruturados ou documentados. A mediana 3, reforça o afirmado pela média, ao ser localizado no mesmo intervalo.

Em pesquisadores de universidade pública, observa-se que a média é de 2,00, indica que, em termos de categorização, esta prática é fraca. A categoria fraca indica que os dados

compartilhados raramente não são estruturados ou documentados. A mediana 2 reforça o indicado pela média, ao ser localizado no mesmo intervalo.

As medidas de variabilidade para os dois grupos de pesquisa, conforme se observa nas tabelas 41 e 42, indicam uma relativa dispersão dos dados em relação à média. Os valores obtidos expressam que embora a média seja representativa, os pesquisadores atuam de uma forma heterogênea em relação à referida prática.

3. Considerações da dimensão qualidade dos dados de pesquisa compartilhados

Os pesquisadores de universidade pública ocasionalmente gerenciam seus dados de pesquisa seguindo recomendações estabelecidas pelas instituições especializadas e, raramente os dados compartilhados não são estruturados ou documentados. Nesse sentido, infere-se que os dados compartilhados em certo grau são curados e preservados pelos pesquisadores seja aplicando modelos de gestão de dados ou de maneira informal.

Por outro lado, pesquisadores de universidade particular indicam que ocasionalmente gerenciam seus dados sob diretrizes institucionais e que também ocasionalmente compartilham dados não estruturados ou sem documentação anexada. Infere-se que os pesquisadores às vezes compartilham dados gerenciados de acordo com diretrizes institucionais ou informalmente, e outras vezes compartilham dados não estruturados ou sem metadados.

Tabela 40- Qualidade dos dados de pesquisa compartilhados

ITENS	Pesquisadores de universidade particular	Pesquisadores de universidade pública
Indicadores		
Gerencia, sob diretrizes propostas por instituições, os dados de pesquisa compartilhados.	Ocasionalmente	Ocasionalmente
Os dados da pesquisa são compartilhados sem estruturação, metadados ou documentação anexa.	Ocasionalmente	Raramente

Fonte: Dados de pesquisa, 2020.

Um dos principais objetivos da gestão de dados é garantir sua preservação, integridade e acesso ao longo do tempo, de tal forma que os dados estejam em ótimas condições para serem compartilhados e possam ser reusados em outros contextos. Dados devidamente gerenciados implica sua fácil interpretação e adequado reúso. Assim, a abertura dos dados é significativa quando os dados são acessíveis, reproduzíveis e reutilizáveis, sendo atingido através de uma eficiente curadoria e preservação.

Por outro lado, o não gerenciamento dos dados não os impede de serem compartilhados; no entanto, os dados que não contêm metadados e não são estruturados, implica que o conjunto de dados está em esquemas que não são compreensíveis para um público amplo e não possuem documentos que descrevem os atributos dos dados, ou seja, que forneçam o contexto de sua composição e seu significado. Nesse contexto, os dados que não são curados dificilmente

podem ser reutilizados com eficiência e, portanto, tornam-se poucos significativos para um contexto de dados abertos.

Tabela 41- Qualidade dos dados de pesquisa compartilhados – Universidade particular

ITENS	Média	Mediana	Moda	Desvio padrão	Variância	Curtose	Coefficiente de assimetria	Mínimo	Máximo
Dados de pesquisa estruturados e gerenciados	2,43	3,00	3,00	1,58	2,49	-1,05	-0,09	1	5
P.47: os dados de pesquisa compartilhados são gerenciados a partir de diretrizes propostas por instituições especializadas em dados.	2,43	3,00	3,00	1,58	2,49	-1,05	-0,09	1	5
Dados de pesquisa não estruturados ou documentados	2,37	3,00	3,00	1,44	2,08	-0,74	-0,23	1	5
P.48: Os dados da pesquisa são compartilhados brutos, sem estruturação, metadados ou documentação anexa.	2,37	3,00	3,00	1,44	2,08	-0,74	-0,23	1	5

Fonte: dados de pesquisa, SPSS 22 ®, 2020

Tabela 42- Qualidade dos dados de pesquisa compartilhados– Universidade pública

ITENS	Média	Mediana	Moda	Desvio padrão	Variância	Curtose	Coefficiente de assimetria	Mínimo	Máximo
Dados de pesquisa estruturados e gerenciados	2,38	2,00	4,00	1,77	3,13	-1,37	0,02	1	5
P.47: os dados de pesquisa compartilhados são gerenciados a partir de diretrizes propostas por instituições especializadas em dados.	2,38	2,00	4,00	1,77	3,13	-1,37	0,02	1	5
Dados de pesquisa não estruturados ou documentados	2,00	2,00	2,00	1,45	2,11	-0,94	0,11	1	5
P.48: Os dados da pesquisa são compartilhados brutos, sem estruturação, metadados ou documentação anexa.	2,00	2,00	2,00	1,45	2,11	-0,94	0,11	1	5

Fonte: dados de pesquisa, SPSS 22 ®, 2020.

7.4.3 Dimensão condições de compartilhamento e reuso dos dados de pesquisa

Ao compartilhar dados, é relevante estabelecer claramente sob quais condições o pesquisador compartilhará os dados para reuso por terceiros. Determinar as condições que irão reger o compartilhamento e a reutilização de dados implica procurar e aplicar mecanismos adequados que concretizem como o pesquisador tem previsto que seus dados serão utilizados no futuro, eliminando qualquer ambiguidade que possa surgir. Pesquisadores ou instituições podem criar suas próprias condições instituídas em acordos de cooperação, formais ou informais, ou da mesma forma, usar licenças amplamente usadas em termos de abertura de dados, como as licenças Creative Commons (CC).

1. Tipos de licenças

O licenciamento ou acordos de cooperação estabelecem de forma clara quais direitos cede o autor; sob que condições; e que podem ou não fazer os usuários com os dados. Conforme discutido anteriormente, as condições podem ser estabelecidas por meio de acordos de cooperação formais ou informais, nos quais as partes envolvidas estabelecem como os dados serão abertos. Esses acordos podem ser estabelecidos de forma padronizada pelos pesquisadores ou instituições ou, pelo contrário, podem ser determinados caso a caso.

Por outro lado, organizações especializadas no assunto oferecem modelos de licenciamento ou licenças de direitos autorais padronizados com o objetivo de facilitar ao autor a forma pela qual ele irá gerenciar os direitos de seus dados, de tal forma que possam compartilhá-los nos termos e condições de sua escolha.

A Tabela 43 indica o tipo de licença que usam para compartilhar dados de pesquisa. 17,64% dos pesquisadores de universidade particular e 5,40% dos pesquisadores de universidade pública utilizam acordos formais de cooperação. Pesquisadores que usam licenças Creative Commons (CC) são 15,68% de instituições particulares e 13,51% são de instituições públicas. Por meio de acordos de cooperação informal, 13,72% dos pesquisadores de universidade particular e 13,51% de universidade pública, compartilham seus dados. Aqueles que declararam não usar nenhum tipo de licença ou acordo para compartilhar dados, 24,49% dos pesquisadores são de universidade particular e 13,51% de universidade pública. 9,80% dos pesquisadores de universidade particular, e 10,81 de universidade pública indicam o uso de licenças Open Data Commons (ODC).

Igualmente a tabela mostra os professores que usam dois ou mais tipos de licenças, no entanto, sua porcentagem é significativamente menor. Dos professores que não responderam

devido a que não compartilham nenhum dos seus dados são 11,76% pertencem a universidade particular e 27,02% de universidade pública.

Em relação ao teste do qui-quadrado de Pearson, com nível de significância de 5% (0,05), e o Pvalue associado à estatística do teste é 0,659, a hipótese nula não é rejeitada, ou seja, não há associação estatisticamente significativa entre os tipos de universidade e o tipo de licença que os pesquisadores usam para compartilhar dados de pesquisa.

Tabela 43- Tipo de licenças

Itens	Estatística	Particulares	Público	Total
Acordos de cooperação formal	Frequência	9	2	11
	Porcentagem	17,64%	5,40%	12,5%
Acordos de cooperação informal Acordos de cooperação formal	Frequência	2	2	4
	Porcentagem	3,92%	5,40%	4,54%
Licenças Creative Commons (CC)	Frequência	8	5	13
	Porcentagem	15,68%	13,51%	14,77%
Acordos de cooperação informal	Frequência	7	5	12
	Porcentagem	13,72%	13,51%	13,63%
Licenças do Open Data Commons (ODC)	Frequência	5	4	9
	Porcentagem	9,80%	10,81%	10,22%
Licenças Creative Commons (CC) Acordos de cooperação formal	Frequência	0	1	1
	Porcentagem	0	2,70%	1,13%
Licenças do Open Data Commons (ODC); Acordos de cooperação formal	Frequência	1	1	2
	Porcentagem	1,96%	2,70%	2,27%
Nenhuma	Frequência	13	5	18
	Porcentagem	24,49%	13,51%	20,45%
Não responde	Frequência	6	10	16
	Porcentagem	11,76%	27,02%	18,18%
Licenças Creative Commons (CC) Acordos de cooperação informal Acordos de cooperação formal	Frequência	0	2	2
	Porcentagem	0	5,40%	2,27%
Total	Frequência	51	37	88
	% do total	100,0%	100,0%	100,0%
Chi-quadrado de Pearson		Valor	df	Significância (bilateral)
		7,696	10	,659

Fonte: Dados de pesquisa, SPSS 22®, 2020

Em geral, predomina nos pesquisadores de universidade particular não usar nenhum tipo de licenças ou acordos, como segunda opção usam acordos de cooperação formal, seguido de Licenças Creative Commons (CC). Os pesquisadores de universidade pública tendem a usar acordos de cooperação informal, Licenças Creative Commons (CC), e não usar nenhuma licença ou acordo quando compartilham seus dados com terceiros.

Por outra parte, o teste do qui-quadrado de Pearson indica que o tipo de licença escolhido pelo pesquisador não depende ou não está associado ao tipo de universidade em que labora o pesquisador.

2. Condições de uso dos dados disponibilizados pelo pesquisador

Estabelecer condições para o compartilhamento e reúso de dados de pesquisa implica determinar a compensação que o autor terá em troca de colocar os dados em acesso aberto para seu reúso. Nesse sentido, as perguntas 49 a 56 do questionário aplicado visam revelar o grau de ocorrência com o qual os pesquisadores aplicam determinadas condições para compartilhar seus dados de pesquisa. As condições que são medidas são as seguintes:

- ✓ “Citado em toda a produção acadêmica que usa seus dados”
- ✓ “Coautoria em toda a produção acadêmica que usa seus dados”
- ✓ “Oportunidade de se envolver no projeto de pesquisa que usa seus dados”
- ✓ “Revisar e fazer sugestões aos resultados da pesquisa que usa seus dados”
- ✓ “Receber toda a produção acadêmica (artigos, livros, apresentações, materiais etc.) que usa seus dados”
- ✓ “Reciprocidade na troca de dados”
- ✓ “Condições de acordo com as especificidades de cada caso”
- ✓ “Não há condições específicas”

De acordo com a média obtida 2,52, pesquisadores de universidade particular ocasionalmente compartilham seus dados nas condições mencionadas anteriormente. A mediana 2,88 reforça o que é indicado pela média.

Os resultados obtidos nas medidas de dispersão, variância de 2,20 e desvio padrão de 1,48 indicam uma maior dispersão em relação à média, significando que os pesquisadores em relação as condições para compartilhar seus dados atuam de maneira relativamente heterogênea.

Especificamente, os resultados por questão revelam que os pesquisadores de universidade particular, quando compartilham seus dados para reúso em outras pesquisas, ocasionalmente exigem: a) ser citados, b) ser co-autor da pesquisa, c) oportunidade de se envolver no projeto de pesquisa, d) revisar e fazer sugestões nos resultados da pesquisa; e) receber a produção acadêmica; f) reciprocidade na troca de dados; g) definem as condições de acordo com as especificidades de cada caso; e h) não estabelecem as condições. As medidas de dispersão para cada questão indicam uma dispersão relativa. Embora a média seja representativa, os pesquisadores se comportam de maneira relativamente heterogênea, ou seja, o grau de ocorrência com o qual as condições mencionadas são aplicadas tendem a ser diversas entre os pesquisadores de universidade particular.

Em relação aos pesquisadores de universidade pública, a média de 1,93 indica que, em geral, os pesquisadores raramente compartilham seus dados nas condições estabelecidas neste indicador. A mediana 2.0 reforça o que é indicado pela média.

As medidas de dispersão, variância de 1,80 e desvio padrão de 1,34, indicam uma maior dispersão em relação à média, o que significa que, embora a média seja representativa, os pesquisadores tendem a se comportar de maneira heterogênea.

Especificamente, as perguntas revelam que pesquisadores de universidade pública, ao compartilhar seus dados para reuso em outras pesquisas, raramente exigem: a) serem coautores da pesquisa, b) oportunidade de se envolver no projeto de pesquisa, c) revisam e fazer sugestões em resultados da investigação; d) recebem a produção acadêmica; e) reciprocidade na troca de dados; f) definem as condições de acordo com as especificidades de cada caso; g) não estabelecem condições e, ocasionalmente, exigem ser citados nas investigações que faça uso dos dados.

As medidas de dispersão para cada questão indicam uma dispersão relativa. Embora a média seja representativa, os pesquisadores tendem a agir de forma heterogênea, o que significa que o grau de ocorrência é relativamente diverso em relação às condições de reutilização dos dados.

3. Considerações dimensão condições de compartilhamento e reuso de dados de pesquisa

Os pesquisadores de universidade particular e publica tendem a usar acordos de cooperação informal e licenças Creative Commons para compartilhar e outorgar o reuso de seus dados a terceiros. Igualmente os pesquisadores tendem a não usar nenhum tipo de licenças e acordos.

O uso de licenças ou acordos explícitos possibilita eliminar ambiguidades relacionadas ao que o usuário pode ou não fazer com os dados do autor. Especialmente as licenças são consideradas o mecanismo mais fácil e simples para gerenciar os direitos dos dados de pesquisa e para compartilhá-los, devido a que defendem um uso mais livre e com menos restrições dos dados, além de permitir que os autores estabeleçam suas próprias condições sob um padrão recomendado.

Embora a licença seja uma boa opção para gerenciar os direitos dos dados, não existe um único mecanismo. A critério do autor ou das instituições, podem ser criados acordos de cooperação entre as partes envolvidas, estabelecendo, a partir das suas necessidades e condições, como os dados podem ser compartilhados e reusados, desde que a autoria seja reconhecida.

Igualmente, os pesquisadores declaram que não usam licenças ou acordos quando compartilham seus dados. Nesse caso, gera-se lacunas sobre como usar de forma correta os dados, criando assim a possibilidade de uso indevido do recurso.

Tabela 44- Caracterização tipos de licença

ITENS	Pesquisadores de universidade particular	Pesquisadores de universidade pública
	Indicadores	
Tipos de licença	Nenhuma Acordo de cooperação formal Licenças Creative Commons (CC)	Nenhuma Acordo de cooperação informal Licenças Creative Commons (CC)

Fonte: dados de pesquisa, 2020.

Já investigando as condições para o uso dos dados disponíveis, os pesquisadores declaram o que é apresentado na tabela 45.

Tabela 45- Caracterização condições de uso dos dados disponibilizados

ITENS	Pesquisadores de universidade particular	Pesquisadores de universidade pública
	Indicador	
Condições de uso dos dados disponibilizados pelos pesquisadores	Ocasionalmente	Raramente
Questões		
Exige ser citado em toda a produção acadêmica que usa seus dados	Ocasionalmente	Ocasionalmente
Exige coautoria em toda a produção acadêmica que usa seus dados	Ocasionalmente	Raramente
Exige a oportunidade de se envolver no projeto de pesquisa que usa seus dados	Ocasionalmente	Raramente
Exige revisar e fazer sugestões aos resultados da pesquisa que usa seus dados	Ocasionalmente	Raramente
Exige receber toda a produção acadêmica (artigos, livros, apresentações, materiais etc.) que usa seus dados	Ocasionalmente	Raramente
Exige reciprocidade na troca de dados	Ocasionalmente	Raramente
Define as condições de acordo com as especificidades de cada caso.	Ocasionalmente	Raramente
Não há condições	Ocasionalmente	Raramente

Fonte: Dados de pesquisa, 2020.

Uma das condições básicas do uso de dados é o reconhecimento do autor através da citação. Os pesquisadores reconhecem que ocasionalmente solicitam que os usuários os citem em toda a produção científica que faz uso dos dados. Como é comumente feito ao usar uma publicação científica em que a citação é obrigatória, é necessário que na utilização dos dados por terceiros também seja outorgado o reconhecimento ao autor. A relevância da citação é dada por ser um dos mecanismos em que o criador dos dados recebe o devido crédito, dá maior visibilidade à

pesquisa que os gerou, mede o valor dos dados através do impacto no seu reuso, e possibilita o aumento de seu uso ao ser facilmente localizados.

Por outro lado, pesquisadores de universidade pública reconhecem que seus dados raramente são compartilhados sem nenhuma condição, enquanto pesquisadores de universidade particular indicam que ocasionalmente não são estabelecidas condições de uso. Em termos de abertura de dados, é essencial que eles sejam disponibilizados sob certas condições de uso, de modo que o autor seja recompensado pela sua prática, e as ambiguidades em seu uso sejam evitadas.

Tabela 46- Condições de uso dos dados disponibilizados – Universidade particular

ITENS	Média	Mediana	Moda	Desvio padrão	Variância	Curtose	Coeficiente de assimetria	Mínimo	Máximo
Dimensão condições de compartilhamento e reúso de dados de pesquisa	2,52	2,88	1,00	1,48	2,20	-0,88	-0,39	1	5
Condições de uso dos dados disponibilizados pelo pesquisador	2,52	2,88	1,00	1,48	2,20	-0,88	-0,39	1	5
P.49 exige ser citado em toda a produção acadêmica que usa seus dados	2,94	3,00	5,00	1,82	3,30	-1,22	-0,41	1	5
P.50 exige coautoria em toda a produção acadêmica que usa seus dados	2,53	2,00	4,00	1,76	3,09	-1,39	0,01	1	5
P.51 exige a oportunidade de se envolver no projeto de pesquisa que usa seus dados	2,35	2,00	1,00	1,69	2,87	-1,24	0,11	1	5
P.52 exige revisar e fazer sugestões aos resultados da pesquisa que usa seus dados	2,51	3,00	3,00	1,68	2,81	-1,24	-0,07	1	5
P.53 exige receber toda a produção acadêmica (artigos, livros, apresentações, materiais etc.) que usa seus dados	2,45	2,00	1,00	1,74	3,01	-1,37	0,07	1	5
P.54 exige reciprocidade na troca de dados	2,59	3,00	3,00	1,71	2,93	-1,22	-0,12	1	5
P.55 define as condições de acordo com as especificidades de cada caso.	2,31	2,00	1,00	1,71	2,94	-1,34	0,21	1	5
P.56 não há condições específicas	2,51	3,00	1,00	1,76	3,09	-1,44	-0,03	1	5

Fonte: Dados de pesquisa, SPSS 22 ®, 2020

Tabela 47- Condições de uso dos dados disponibilizados – Universidade pública

ITENS	Média	Mediana	Moda	Desvio padrão	Variância	Curtose	Coefficiente de assimetria	Mínimo	Máximo
Dimensão condições de compartilhamento e reúso de dados de pesquisa	1,93	2,00	1,00	1,34	1,80	-0,32	0,15	1	5
Condições de uso dos dados disponibilizados pelo pesquisador	1,93	2,00	1,00	1,34	1,80	-0,32	0,15	1	5
P.49 exige ser citado em toda a produção acadêmica que usa seus dados	2,19	2,00	2,00	1,84	3,38	-1,08	0,50	1	5
P.50 exige coautoria em toda a produção acadêmica que usa seus dados	1,89	2,00	1,00	1,65	2,71	-0,50	0,74	1	5
P.51 exige a oportunidade de se envolver no projeto de pesquisa que usa seus dados	1,92	2,00	2,00	1,53	2,35	-0,53	0,53	1	5
P.52 exige revisar e fazer sugestões aos resultados da pesquisa que usa seus dados	1,84	2,00	1,00	1,54	2,36	-0,50	0,63	1	5
P.53 exige receber toda a produção acadêmica (artigos, livros, apresentações, materiais etc.) que usa seus dados	1,73	1,00	1,00	1,48	2,20	-0,01	0,82	1	5
P.54 exige reciprocidade na troca de dados	1,95	2,00	2,00	1,65	2,72	-0,79	0,52	1	5
P.55 define as condições de acordo com as especificidades de cada caso.	1,89	2,00	2,00	1,49	2,21	-0,70	0,41	1	5
P.56 não há condições específicas	2,00	2,00	1,00	1,65	2,72	-0,92	0,47	1	5

Fonte: Dados de pesquisa, SPSS 22 @, 2020

7.5 PRÁTICAS DE USO E REÚSO DE DADOS DE PESQUISA

Um entorno científico aberto significa disponibilizar os materiais científicos gerados durante a pesquisa de modo a conceder acesso a tais recursos entre a comunidade acadêmica. A abertura dos materiais científicos, especificamente a disponibilização dos dados, é uma forma de promover o progresso da ciência, ao ter como finalidade: a) reproduzir pesquisas para verificar sua confiabilidade; b) destacar a transparência da pesquisa; c) reusar os dados para fazer descobertas adicionais ou complementares, e para desenvolver novas investigações.

Em termos de uso e reúso dos dados de pesquisa, sua efetiva reutilização dependerá em parte dos métodos de curadoria e preservação pelos quais os dados foram submetidos. Atribuir aos dados metadados de qualidade é considerado um determinante no uso adequado dos dados disponibilizados. Do mesmo modo, os canais de comunicação, ao conceder acesso aos dados às comunidades acadêmicas, são considerados outro aspecto relevante que influenciará na visibilidade destes e no aumento ou não de sua possível reutilização.

Nesse contexto, procura-se revelar se os pesquisadores usam dados de pesquisa em acesso aberto de forma a integrá-los em seus projetos de pesquisa; bem como a qualidade desses dados em relação a eles terem ou não metadados, e os meios pelos quais acessam a eles. A variável está associada às perguntas 58 a 61 do questionário aplicado, considerando que as questões extraídas são de dois tipos: escala nominal e escala Likert.

7.5.1 Dimensão uso e reúso de dados de pesquisa

No âmbito da ciência moderna, em que se percebe o valor informativo dos dados e há uma maior abertura e acesso a esse recurso, cada vez mais, a prática de usar e reusar dados, para fins diferentes daqueles criados, está presente nos processos de pesquisa. A dimensão revela o grau de ocorrência com que os pesquisadores usam seus dados para gerar novos projetos de pesquisa, além de reutilizar dados de terceiros. A dimensão é, portanto, medida a partir dos seguintes indicadores:

- ✓ “Reúso, em uma nova pesquisa, de dados coletados em pesquisas anteriores”.
- ✓ “Uso, em uma nova pesquisa, de dados sob autoria de terceiros”.

Em geral, a dimensão, tanto para pesquisadores de universidade particular, quanto pesquisadores de universidade pública, de acordo com suas respectivas médias, 2,83 e 2,76, indicam uma presença neutra dessa prática. A categoria neutra revela que, em média, os pesquisadores ocasionalmente reusam dados em suas pesquisas, sejam de autoria própria ou de terceiros. No entanto, se for explorado os indicadores que o compõem, na tabela 50 se observa

uma exceção em pesquisadores de universidade pública ao expressar que frequentemente reusam seus próprios dados para gerar novos projetos de pesquisa.

Por outro lado, as respectivas medianas da dimensão para cada grupo reforçam o que suas médias indicam, ao ser localizada no mesmo intervalo. Além disso, indica que 50% dos pesquisadores reusam dados com um grau de ocorrência igual ou maior a ocasionalmente, sendo frequente e sempre; e 50%, o fazem igual ou menor que ocasionalmente, raramente e nunca.

Em relação à variância e desvio padrão da dimensão, tanto em pesquisadores de universidade particular e pública, identifica-se pouca dispersão dos dados em relação à média. Os valores das medidas de variabilidade revelam que os pesquisadores atuam de forma próxima à característica traçada pela média. Devido ao comportamento homogêneo dos dados, a média é representativa.

1. Reúso, em uma nova pesquisa, de dados coletados em pesquisas anteriores.

Em relação ao indicador sobre se os pesquisadores reusam seus dados de pesquisas anteriores para gerar outros projetos de investigação, observa-se que entre os pesquisadores de universidade particular, a média 2,96, esta pratica tem uma presença neutra. A mediana 3, reforça o afirmado pela média, ao ser localizado no mesmo intervalo. A categorização neutra significa que os pesquisadores ocasionalmente reusam seus próprios dados para desenvolver novas linhas de pesquisa.

Em relação aos pesquisadores de universidade pública têm uma média de 3,11, indica que, em termos de categorização, esta pratica tem uma presença forte. A categoria forte significa que os pesquisadores, em média, frequentemente reusam seus dados em novas pesquisas. A mediana 3, localiza-se na categoria neutra.

No referente às medidas de variabilidade se observa em pesquisadores de universidade particular uma variância de 0,92 e um desvio padrão de 0,96, quanto que para os pesquisadores de universidade pública se registra uma variância de 0,77 e desvio padrão de 0,88. Os valores obtidos indicam pouca dispersão dos dados em relação à média, significando que os pesquisadores atuam de forma similar, próxima à média. Considera-se que devido ao comportamento homogêneo dos dados, a característica traçada pela média é considerada representativa.

2. Uso, em uma nova pesquisa, de dados sob autoria de terceiros.

Em relação a usar dados de pesquisa sob a autoria de terceiros, os pesquisadores de universidade particular, obtiveram uma média de 2,71, indicando que a prática tem uma presença neutra. A categoria neutra significa que, em média, os pesquisadores ocasionalmente

usam dados sob autoria de terceiros para suas pesquisas. A mediana reforça o indicado pela média, ao ser localizado no mesmo intervalo.

Em pesquisadores de universidade pública, a média 2,41, revela que os pesquisadores ocasionalmente reutilizam dados de terceiros em suas pesquisas. A mediana, localiza-se na categoria fraca, significando que 50% dos pesquisadores reusam dados de terceiros em um grau de ocorrência igual ou inferior a raramente e nunca, e o outro 50% os reusam igual ou maior a raramente, ocasional, frequente e sempre.

Em relação às medidas de variabilidade se observa em pesquisadores de universidade particular uma variância de 1,13 e um desvio padrão de 1,06, quanto que para os pesquisadores de universidade pública se registra uma variância de 0,64 e desvio padrão de 0,80. Os valores obtidos indicam pouca dispersão dos dados em relação à média, significando que os pesquisadores atuam de forma próxima à média. Considera-se que devido ao comportamento homogêneo dos dados, a característica traçada pela média é considerada representativa para os dois grupos de pesquisadores.

3. Considerações da dimensão uso e reúso de dados próprios ou de terceiros.

Os resultados gerais da dimensão, indicam que entre os pesquisadores de universidade particular e pública, em média, ocasionalmente reusam dados de pesquisa em novos projetos. Observando o resultado de cada indicador, confirma-se que pesquisadores de universidade particular ocasionalmente reutilizam dados de suas pesquisas anteriores, ou dados de terceiros para gerar novas pesquisas. Pesquisadores de universidade pública já afirmam que frequentemente reutilizam seus dados em outros projetos, e ocasionalmente usam dados de pesquisas de terceiros.

Tabela 48- Caracterização dimensão uso e reúso de dados próprios ou de terceiros

ITENS	Pesquisadores de universidade particular	Pesquisadores de universidade pública
Dimensão		
Reusa dados de pesquisa próprios e de terceiros	Ocasionalmente	Ocasionalmente
Indicadores		
Reusa os dados de sua pesquisa anterior para gerar outros projetos de investigação	Ocasionalmente	Frequentemente
Usa os dados de pesquisa gerados por terceiros em sua investigação. (Pesquisadores, instituições, observatórios, centros, associações, etc.)	Ocasionalmente	Ocasionalmente

Fonte: Elaboração própria, 2020.

As tecnologias da informação e comunicação inseridas nos processos de pesquisa e o fortalecimento dos movimentos de ciência aberta tem favorecido cada vez mais a criação de espaços de pesquisa colaborativos, dispostos a compartilhar e reutilizar, sob reconhecimento

autoral e uso adequado, materiais científicos, especialmente os dados que se caracterizam por terem um valor informativo que, se bem explorados, levarão a novas linhas de pesquisa.

Assim, as características gerais da ciência moderna permitem condições ideais e seguras para que se desenvolvam ecossistemas de acesso e reúso a dados de pesquisa. Entretanto, cada ambiente científico tem suas peculiaridades que podem se opor e entrar em conflito com a abertura da ciência, colocando assim barreiras que impedem o uso e a reúso dos dados de pesquisa disponíveis para essa prática.

No contexto acadêmico estudado, é evidente que o reúso de dados se dá de forma ocasional, apontando que esta prática não é comum entre os pesquisadores. Nesse sentido, reconhece-se que esse contexto apresenta desvantagens quando comparado aos ambientes de pesquisa em países com maior presença na ciência aberta. Os possíveis benefícios que os pesquisadores não estão aproveitando são, desde a minimização de tempo e recursos econômicos investidos em pesquisa; o possível aumento na produção de pesquisas; possíveis alianças criando comunidades acadêmicas mais colaborativas e abertas; e a possibilidade de criar novos conjuntos de dados, a partir de diversas bases, para realizar e aumentar a pesquisa interdisciplinar.

Cabe mencionar que uma das áreas de conhecimento que possuem maior número de bancos de dados a seu serviço é o campo de ação de economia e negócios, portanto, nesse caso, não estaria sendo constantemente aproveitados.

Tabela 49- Dimensão reúso de dados de pesquisa próprios e de terceiros – pesquisadores de universidade particular

ITENS	Média	Mediana	Moda	Desvio padrão	Variância	Curtose	Coefficiente de assimetria	Mínimo	Máximo
Dimensão reúso de dados de pesquisa próprios e de terceiros	2,83	3,00	3,00	0,89	0,80	0,29	0,15	1,00	5,00
Reúso, em uma nova pesquisa, de dados coletados em pesquisas anteriores	2,96	3,00	3,00	0,96	0,92	0,26	-0,06	1,00	5,00
P.58 Reusa os dados de sua pesquisa anterior para gerar outros projetos de investigação	2,96	3,00	3,00	0,96	0,92	0,26	-0,06	1,00	5,00
Uso, em uma nova pesquisa, de dados sob autoria de terceiros.	2,71	3,00	3,00	1,06	1,13	-0,61	0,11	1,00	5,00
P. 59 Usa os dados de pesquisa gerados por terceiros em sua investigação. (Pesquisadores, instituições, observatórios, centros, associações, etc.)	2,71	3,00	3,00	1,06	1,13	-0,61	0,11	1,00	5,00

Fonte: Elaboração própria, com base em dados processados SPSS 22 ®, 2020.

Tabela 50- Dimensão reúso de dados de pesquisa próprios e de terceiros – pesquisadores de universidade público

ITENS	Média	Mediana	Moda	Desvio padrão	Variância	Curtose	Coefficiente de assimetria	Mínimo	Máximo
Dimensão reúso de dados de pesquisa próprios e de terceiros	2,76	3,00	3,00	0,58	0,34	1,09	-0,70	1,00	4,00
Reúso, em uma nova pesquisa, de dados coletados em pesquisas anteriores	3,11	3,00	3,00	0,88	0,77	-0,34	-0,22	1,00	5,00
P. 58 Reusa os dados de sua pesquisa anterior para gerar outros projetos de investigação	3,11	3,00	3,00	0,88	0,77	-0,34	-0,22	1,00	5,00
Uso, em uma nova pesquisa, de dados sob autoria de terceiros.	2,41	2,00	2,00	0,80	0,64	-0,26	0,15	1,00	4,00
P. 59 Usa os dados de pesquisa gerados por terceiros em sua investigação. (Pesquisadores, instituições, observatórios, centros, associações, etc.)	2,41	2,00	2,00	0,80	0,64	-0,26	0,15	1,00	4,00

Fonte: Elaboração própria, com base em dados processados SPSS 22 ®, 2020.

7.5.2 Dimensão qualidade dos dados de pesquisa

A qualidade dos dados, refere-se às informações adicionais - metadados - atribuídos aos dados que serão reutilizados. Ter uma descrição associada ao contexto da pesquisa; nível de banco de dados; nível de variáveis e administrativo, contribui para que o reúso de dados se dê de forma assertiva. A dimensão, é, portanto, medida a partir do indicador: Os dados de pesquisa reusados possuem metadados, o qual está associado à questão 60 aplicada no questionário.

Em geral, a dimensão, tanto para pesquisadores de universidade particular quanto pesquisadores de universidade pública, de acordo com suas respectivas médias, 2,57 e 2,57, indicam uma presença neutra dessa prática. A categoria neutra para essa dimensão revela que, em média, os pesquisadores ocasionalmente reusam datasets com metadados.

Já em relação às respectivas medianas de cada grupo, estas se encontram na categoria fraca, indicando que 50% dos pesquisadores realizam essa prática de forma igual ou inferior a raramente e nunca, enquanto que 50% o fazem igual ou maior a raramente, ocasional, frequente e sempre.

Em relação às medidas de variabilidade se observa em pesquisadores de universidade particular uma variância de 1,21 e um desvio padrão de 1,10, enquanto que para os pesquisadores de universidade pública se registra uma variância de 1,31 e desvio padrão de 1,14. Os valores obtidos indicam pouca dispersão em relação à média, significando que os pesquisadores atuam, em relação à referida prática, de forma homogênea, próximo à característica traçada pela média. A homogeneidade da amostra indica que a média é confiável em termos de representatividade.

Tabela 51- Atribuição de metadados a dados usados.

ITENS	Pesquisadores de universidade particular	Pesquisadores de universidade pública
Dimensão – Indicador-Pergunta		
Reusa datasets com metadados	Ocasionalmente	Ocasionalmente

Fonte: Elaboração própria, 2020.

Um dos aspectos relevantes na prática de reutilizar dados de pesquisa é garantir que os dados possam ser entendidos e interpretados corretamente pelos usuários. Para uma boa interpretação, é essencial que os dados sejam descritos de forma clara e detalhada em vários níveis, tanto em sua contextualização investigativa, técnica, quanto administrativamente. É por esse motivo que, em termos de gestão de dados, uma das fases mais relevantes é a documentação dos dados, especialmente quando o autor pretende colocá-los em acesso aberto.

A atribuição de metadados de qualidade aos dados geralmente concede aos usuários o uso e a reuso adequado desse recurso, além de aumentar a probabilidade de serem citados com mais frequência e, conseqüentemente, seu autor receber uma maior visibilidade na comunidade acadêmica. Ao contrário do cenário ideal, neste caso, os pesquisadores declaram que ocasionalmente usam dados documentados, o que poderia implicar um uso incorreto deles, gerando um viés na pesquisa na qual os dados foram integrados.

Tabela 52-. Dimensão qualidade dos dados de pesquisa– pesquisadores de universidade particular

	Média	Mediana	Moda	Desvio padrão	Variância	Curtose	Coefficiente de assimetria	Mínimo	Máximo
Dimensão qualidade dos dados de pesquisa	2,57	2,00	2,00	1,10	1,21	-0,34	0,48	1,00	5,00
Reusa datasets com metadados	2,57	2,00	2,00	1,10	1,21	-0,34	0,48	1,00	5,00
P. 60 Os dados de pesquisa reusados possuem metadados.	2,57	2,00	2,00	1,10	1,21	-0,34	0,48	1,00	5,00

Fonte: Elaboração própria, com base em dados processados SPSS 22 ®, 2020.

Tabela 53- Dimensão qualidade dos dados de pesquisa– pesquisadores de universidade público

	Média	Mediana	Moda	Desvio padrão	Variância	Curtose	Coefficiente de assimetria	Mínimo	Máximo
Dimensão qualidade dos dados de pesquisa	2,57	2,00	2,00	1,14	1,31	-0,60	0,35	1,00	5,00
Reusa datasets com metadados	2,57	2,00	2,00	1,14	1,31	-0,60	0,35	1,00	5,00
P. 60 Os dados de pesquisa reusados possuem metadados.	2,57	2,00	2,00	1,14	1,31	-0,60	0,35	1,00	5,00

Fonte: Elaboração própria, com base em dados processados SPSS 22 ®, 2020.

7.5.3 Dimensão meios de acesso a conjuntos de dados de terceiros

A prática de compartilhar e reusar dados de pesquisa leva ao surgimento de novos canais de comunicação científica, como repositórios e data centers. Esses meios formais de comunicação fornecem aos dados maior preservação a médio e longo prazo, além de facilitar a disponibilidade dos dados em acesso aberto. Embora atualmente cada vez mais repositórios estejam surgindo, existem outras modalidades não formais de acesso a dados, como por solicitação direta ao autor. A dimensão revela os meios pelos quais os pesquisadores acessaram aos dados de pesquisa de terceiros que eles têm reutilizado em seus projetos de investigação. A dimensão, é, portanto, medida a partir do indicador: canais de comunicação para acessar a dados de terceiros, o qual está associado à questão 61 aplicada no questionário.

A Tabela 54 mostra a relação entre o tipo de instituição e os canais de acesso a dados de terceiros. O canal de acesso predominante é mediante solicitação direta ao autor, em que 54,1% do total de pesquisadores de universidade pública afirmaram utilizá-lo, enquanto 37,3% do total de pesquisadores de universidade particular acessam os dados dessa maneira.

O 13,7% do total de pesquisadores de universidade particular, enquanto 10,8% de pesquisadores de universidade pública, acessam aos dados através de repositórios de dados temático. Dos que indicaram acessar a dados de terceiros por meio de repositórios de dados de revistas, são 11,8% do total de professores de universidade particular e 8,1% de universidade pública.

Em relação aos repositórios de dados nacionais apenas os pesquisadores de universidade particular afirmaram utilizar esse canal, enquanto aqueles que declararam acesso através de repositórios de dados internacionais 9,8% são pesquisadores de universidade particular e 10,8% de universidade pública. O acesso a dados através do repositório do seu grupo de pesquisa, 5,9% do total de professores de universidade particular afirmaram utilizá-lo, enquanto que 2,7% dos pesquisadores de universidade pública o usam.

Finalmente, a pergunta é respondida só se o pesquisador afirma ter reusado dados da pesquisa sob autoria de terceiros. Nesse aspecto, não responderam 11,8% dos pesquisadores de universidade particular e 8,1% dos pesquisadores de universidade pública.

Tabela 54 - Meios de acesso a dados de terceiros

Itens	Estatística	Particular	Pública	Total
Solicitação direta ao autor	Frequência	19	20	39
	Porcentagem	37,3%	54,1%	44,3%
Não responde	Frequência	6	3	9
	Porcentagem	11,8%	8,1%	10,2%
	Frequência	3	1	4

Itens	Estatística	Particular	Pública	Total
Repositórios / centros / base de dados do meu grupo de pesquisa	Porcentagem	5,9%	2,7%	4,6%
Repositórios / centros / base de dados da minha instituição	Frequência	1	2	3
	Porcentagem	1,9%	5,4%	3,4%
Repositórios / centros / base de dados de revistas científicas	Frequência	6	3	9
	Porcentagem	11,8%	8,1%	10,2%
Repositórios / centros / base de dados internacionais	Frequência	5	4	9
	Porcentagem	9,8%	10,8%	10,2%
Repositórios / centros / base de dados nacionais	Frequência	4	0	4
	Porcentagem	7,8%	0	4,6%
Repositórios / centros / base de dados temática	Frequência	7	4	11
	Porcentagem	13,7%	10,8%	12,5%
Total	Frequência	51	37	88
	% do total	100%	100%	100%
Chi-quadrado de Pearson		Valor	df	Significância bilateral
		6,218	7	0,514

Fonte: dados de pesquisa, SPSS 22 ®, 2020

Os repositórios são concebidos como os canais de comunicação com uma relevante pertinência devido a seus propósitos estarem alinhados com os fins da ciência aberta. Entre os objetivos dos repositórios encontram-se a proteção dos dados de possíveis ameaças, e garantir um maior fluxo informacional mediante o acesso aberto. Do ponto de vista da ciência aberta, o uso constante de canais informais, como solicitar diretamente ao autor, leva a que os sistemas de acesso aberto sejam desenvolvidos de maneira informal e não estruturada, prejudicando a disponibilidade permanente de dados; restringindo as oportunidades para uma comunidade científica de reutilizar os dados; reduzindo as possibilidades de ser uma referência entre os cientistas pela transparência dos seus processos de investigação; limitando sua visibilidade como pesquisador e afetando o progresso científico ao não conceder maior acesso aos seus dados.

Em geral, a dimensão sobre o reúso de dados revelou que, ocasionalmente os pesquisadores fazem uso de dados de terceiros em suas pesquisas. Pode-se apontar que sua reutilização esporádica se deve a que os meios de acesso predominantes terem mais barreiras e limitações ao apresentar aspectos mais ambíguos na troca e reutilização de dados. Assinala-se que se os repositórios fossem o meio mais usado para acessar os dados, possivelmente sua reutilização seria mais frequente.

8 CONSIDERAÇÕES FINAIS

As mudanças na maneira de fazer ciência evoluíram nas últimas décadas devido aos fenômenos ligados à e-Science e à reivindicação dos movimentos de ciência aberta. A ciência moderna concebe a pesquisa como um processo que deve ser aberto para permitir que os recursos gerados sejam distribuídos, reproduzidos e reutilizados pela comunidade acadêmica. Nesse sentido, a abertura do processo de pesquisa é, por si só, uma maneira de cumprir com os princípios de colaboração, contribuição, transparência e abertura, nos quais a ciência se baseia atualmente.

No atual contexto de abertura do processo científico, a comunidade acadêmica começa a reconhecer o valor dos dados de pesquisa por sua capacidade de continuar transmitindo informações ao longo do tempo, se forem devidamente preservados e curados (SAYÃO; SALES, 2015). Os dados apresentam-se como um recurso com alto potencial para ser compartilhado e reutilizado para diversas finalidades, gerando uma série de benefícios tanto para o pesquisador que o disponibiliza, para o usuário que o coleta, como para o progresso da ciência.

A pesquisa realizada por Eynden *et al* (2011, p. 3) menciona que, dentre os benefícios do compartilhamento de dados de pesquisa, destaca-se a possibilidade de dar um novo uso aos dados; promover a colaboração entre usuários e autores dos dados; maximizar a transparência da pesquisa, reduzir esforços e recursos nas investigações; aumentar a visibilidade e o impacto da pesquisa; e reconhecer publicamente a credibilidade do autor como pesquisador.

Nessa direção, diante da dinâmica do acesso aberto aos dados de pesquisa, um dos desafios cruciais para os pesquisadores que os compartilham é convertê-los em dados de qualidade, isto é, dados facilmente acessíveis, interpretáveis e reutilizáveis. A criação de dados de pesquisa de qualidade é determinada por uma gestão de dados eficiente, implicando submetê-los a atividades de organização, descrição, preservação, garantia e controle de qualidade, de forma que se garanta seu acesso, eficiente interpretação e reúso a longo prazo, cumprindo assim o propósito da ciência aberta.

A gestão de dados, além de garantir o acesso, a correta interpretação e reutilização dos dados, também implica, reduzir ou eliminar problemas de perda, integridade e validade dos dados durante e após o encerramento da investigação. Assim, em termos de pesquisa, independentemente de estarem disponíveis ou não abertamente, o gerenciamento de dados

fornece ao pesquisador uma série de atividades que permitem que a coleta e processamento dos dados fique livre de erros que possam prejudicar a fiabilidade e validade da investigação.

Garantir que os dados de pesquisa sejam de qualidade é em parte responsabilidade do pesquisador, pois ele é o autor e benfeitor das vantagens que fornece uma eficiente gestão de dados. Nessa direção, instituições especializadas em curadoria e preservação de dados têm fornecido modelos, diretrizes, ferramentas e equipe de suporte para facilitar o processo ao pesquisador.

Os modelos de ciclo de vida propostos pelas instituições DCC, UKOLN, e especialmente *DataONE*, proporcionam uma estrutura que padroniza as diversas práticas de curadoria e preservação que serão executadas sistematicamente, economizando tempo e recursos. No entanto, a gestão dos dados, como parte intrínseca das atividades de pesquisa, pode ser desenvolvida com base nas experiências e necessidades de pesquisa do cientista, sem a aplicação de um modelo ou diretrizes instituídos por organizações especializadas.

Diante desse contexto, a pesquisa teve como objetivo revelar as práticas de gestão, compartilhamento e reúso de dados de pesquisa na Colômbia, especificamente na cidade de Bucaramanga e sua região metropolitana. A ausência de estudos científicos que revelem o estado atual de tais práticas no contexto acadêmico colombiano gerou a seguinte questão:

Como os pesquisadores, vinculados aos grupos de pesquisa classificados pela Colciencias, na área de economia e negócios, praticam a gestão, compartilhamento e reúso de dados de pesquisa?

Para responder à pergunta, foi elaborado e aplicado um questionário composto por I. Perfil do investigador, II. Características dos dados de pesquisa, III. Prática de gestão de dados, IV. Prática de compartilhamento de dados, V. Prática de uso e reúso de dados.

Os pesquisadores participantes atenderam aos critérios de inclusão estabelecidos na pesquisa, isto é, pesquisadores vinculados a grupos de pesquisa na área de economia e negócios, classificados pela Colciencias, pertencentes a universidade particular e pública na cidade de Bucaramanga e região metropolitana. Participaram 88 pesquisadores, dos quais 51 são de universidades particulares e 37 de uma universidade pública.

Com base nos dados coletados, processados, analisados e interpretados, a questão foi respondida e as conclusões apresentadas a seguir foram alcançadas:

✓ Práticas de gestão de dados de pesquisa:

Segundo Ball (2010), os ciclos de vida dos dados construídos pelas diferentes instituições especializadas padronizaram os processos de curadoria. *DataONE*, como rede de observação de dados, propôs um modelo de ciclo de vida que explicita graficamente as fases de

gerenciamento. Entretanto, embora o modelo represente um ciclo lógico de gerenciamento de dados, na prática científica as fases podem não seguir o ciclo linear proposto, ocorrendo em diferentes sequências (*DataONE*, 2013).

Na realidade científica estudada, a afirmação de *DataONE* é confirmada ao revelar que os pesquisadores usam apenas parte de seu ciclo, aplicando as fases que consideram apropriadas para o seu contexto, assim como parte das atividades que compõem as fases implementadas. Cabe destacar que o comportamento dos pesquisadores de universidade particular e pública é semelhante em relação às fases de gestão de dados implementadas por eles.

As fases do ciclo de vida dos dados são compostas por atividades exclusivas da curadoria, bem como outras que fazem parte intrinsecamente do processo de pesquisa. Tais atividades inerentes ao processo de investigação são integradas à fase de gestão, pois possuem um componente de curadoria e preservação.

Nessa direção, observou-se que as práticas de gestão de dados que os pesquisadores realizam com mais frequência fazem parte do processo de pesquisa. Práticas como o planejamento da coleta, processamento e análise dos dados, e a garantia e controle da qualidade dos dados, os pesquisadores relatam que o fazem com frequência ou sempre, pois garantem o adequado procedimento investigativo, a qualidade dos dados e a validação dos resultados da pesquisa.

Já as práticas exclusivas de curadoria, como a construção de um plano formal de gestão de dados e atividades associadas à preservação, tendem a ser realizadas pelos pesquisadores ocasionalmente. O grau de ocorrência com que essas práticas são implementadas, revela que as fases consideradas fundamentais para gerar dados de pesquisa de qualidade, em termos de curadoria, preservação e acesso aberto, não são constantes no cotidiano investigativo do professor.

Em particular, os pesquisadores ocasionalmente criam um plano de gestão de dados que lhes permite traçar o método pelo qual os dados serão gerenciados, e explicita o que será feito com eles uma vez finalizada a pesquisa. Os planos de gestão independente da abertura ou não dos dados, são uma ferramenta que facilita, em tempo e esforço, o ciclo de vida dos dados na pesquisa. Igualmente, este é comumente solicitado pelas entidades que financiam as investigações ou pelos repositórios onde os pesquisadores decidem armazenar seus dados, caso que, no contexto colombiano, ocorre esporadicamente.

Da mesma forma, as atividades de preservação são caracterizadas por serem realizadas ocasionalmente, o que implica que os dados podem estar em condições vulneráveis de perda ou danos. Em relação à construção de metadados, tende a ser realizada com frequência, mas sem

descrevê-los a partir de metadados padrão que permitam a descrição integral dos dados e com uma linguagem padronizada fácil de interpretar por um público maior. A descrição e atribuição de metadados é imprescindível na criação de dados de qualidade, pois documenta o contexto, condições, significados, questões técnicas, semânticas, e administrativas dos dados, informação que é necessária para preservar os dados.

Finalmente, como resultado do auge das práticas de gestão de dados, as instituições especializadas em curadoria e preservação produziram um grande número de ferramentas tecnológicas, diretrizes de boas práticas e suporte especializado, a fim de sistematizar e facilitar esse processo aos pesquisadores.

Nesse sentido, os pesquisadores declararam ocasionalmente ter recebido assistência ou ter usado ferramentas digitais para realizar um plano de gestão de dados, criar metadados padrão, inspecionar, verificar e monitorar a qualidade dos dados. Igualmente, indicaram que ocasionalmente orientavam a preservação dos seus dados através de boas práticas, e raramente recebiam assistência para uma adequada preservação.

A implementação de boas práticas, receber assistência por expertos, e o uso de ferramentas digitais é um mecanismo que garante que em todas as etapas do ciclo de vida dos dados as práticas sejam realizadas de forma eficiente. Uma das vantagens de usar esse tipo de ferramenta é a facilidade de executar um processo de gestão. Assim, os serviços oferecidos por essas instituições são uma ferramenta de navegação que facilita ao pesquisador as múltiplas atividades que devem ser desenvolvidas para garantir uma curadoria e preservação eficaz de seus dados.

✓ **Compartilhamento, uso e reúso de dados de pesquisa:**

Um dos objetivos do gerenciamento de dados é gerar dados de qualidade que estejam em perfeitas condições de modo que sejam úteis em um sistema de acesso aberto. Em um cenário científico ideal, os pesquisadores ao final de seus projetos e depois de verificar a qualidade dos dados selecionados para compartilhamento, os tornam públicos através de canais formais de comunicação científica, como repositórios ou data centers, e em condições claras que permitem a uso deles por terceiros.

Diante do cenário de compartilhamento de dados de pesquisa, Borgman (2012) declara que, embora essa prática seja desenvolvida de maneira complexa e difícil, alguns dados são compartilhados com algumas pessoas, em certas ocasiões. Com base no que Borgman afirmou, foi revelado que o contexto científico estudado está longe do cenário ideal e tende a ocorrer de forma intermitente e informal. Em geral, o comportamento dos pesquisadores de universidade pública e particular em relação à prática de compartilhar dados tende a ser divergente.

88,2% dos pesquisadores de universidade particular compartilham todos ou parte de seus dados e 11,8% não disponibilizam nenhum de seus dados. Esses pesquisadores se caracterizam por conceder acesso a seus dados a usuários que os solicitem diretamente, ao seu grupo de pesquisa e à instituição à qual estão vinculados. Os meios que usam para compartilhar seus dados são por solicitação direta, repositórios de revistas científicas e repositórios institucionais. Quanto à qualidade de seus dados, eles declaram que, ocasionalmente, os dados compartilhados são gerenciados sob a orientação de instituições especializadas e outras vezes, compartilhados sem serem estruturados e sem metadados. Por último declaram que tendem a não usar nenhum tipo de licença, a realizar acordos formais de cooperação, ou a usar as licenças Creative Commons.

Dos pesquisadores de universidade pública, 72,96% compartilham todos ou parte de seus dados e 27,02%, nenhum deles. Esses pesquisadores caracterizam-se por uma tendência a dar acesso a dados a todos e àqueles que os solicitam diretamente. Os meios que eles usam para compartilhar seus dados são de forma privada por solicitação direta, e publicamente por repositórios temáticos e revistas científicas. Quanto a qualidade dos dados compartilhados, eles tendem a ser curados informalmente ou sob diretrizes especialistas, e raramente compartilham sem metadados. As licenças comumente usadas para proteger os dados são licenças Creative Commons, acordos de cooperação informal ou não usam nenhum tipo de licença.

O panorama revelado pelo estudo mostra que a prática de compartilhar dados pode abranger inúmeras formas, tão variadas no tipo de dados compartilhados, nos usuários que podem acessar os dados, nas maneiras como os dados são gerenciados, nos canais de comunicação pelo qual eles são compartilhados e o tipo de licenças que os protegem. Entretanto, o principal desafio para o pesquisador que compartilha seus dados é convertê-los em dados interpretáveis e reutilizáveis. Nesse aspecto Borgman (2012) menciona que a liberação dos dados é mais benéfica quando os dados são gerenciados de uma maneira que os torne úteis ao longo do tempo. Nesse aspecto, utilidade dos dados será determinado fortemente pela apropriada gestão dos dados.

Em termos de estabelecer um sistema aberto de acesso a dados, as barreiras identificadas são restrição de acesso a um público maior; baixo uso de canais formais de comunicação científica; compartilhamento de dados não estruturados e sem metadados; não usar nenhuma licença ou acordo de cooperação que os proteja; não especificar as condições em que os dados são compartilhados, e não exigir ser citado na produção acadêmica que os utiliza.

Contrário à posição do pesquisador como autor dos dados que compartilha, e desde a postura de usuário, pesquisadores de universidade particular declaram que ocasionalmente

reusam seus dados ou dados sob a autoria de terceiros. Eles também indicam que esporadicamente os dados usados possuem metadados. Os meios pelos quais os pesquisadores acessam aos dados reusados são por solicitação direta ao autor, ou através de repositórios temáticos ou revistas científicas.

Já os pesquisadores de universidade pública expressam que frequentemente usam seus dados de pesquisas anteriores para realizar novos projetos e, ocasionalmente, reusam dados de terceiros. Eles também mencionam que, ocasionalmente, os dados reusados têm metadados e os meios de acesso são principalmente por solicitação direta ou por repositórios temáticos.

Assim, o uso e reuso de dados não é uma prática comum entre os pesquisadores. Além disso, os dados tendem a não ter metadados, o que prejudica uma adequada reutilização. Outro aspecto relevante é o meio pelo qual os pesquisadores acessam os dados, prevalecendo a solicitação direta ao autor, sendo um canal privado que prejudica a visibilidade e o impacto causado pelo reuso dos dados.

Diante disso, destaca-se que os pesquisadores não aproveitam os dados de pesquisa que são disponibilizados para eles, principalmente nesta área de conhecimento que possui um número significativo de bancos de dados abertos.

De maneira geral, conclui-se que as práticas de gestão de dados incorporadas ao cotidiano investigativo do cientista são emergentes na medida que são realizadas ocasionalmente por eles. Por outro lado, concebe-se que o acesso aberto aos dados é desenvolvido de maneira informal e concentrado no contexto próximo e rede profissional do pesquisador. Por fim, a gestão, compartilhamento e reuso de dados é uma questão de produtividade e eficiente administração dos recursos científicos, em que se o contexto científico estudado incorporasse boas práticas de forma mais constante o progresso acadêmico e científico do pesquisador seria maior do que o atual.

Contribuição do estudo para a comunidade acadêmica:

O estudo contribuiu para decifrar o estado atual das práticas associadas à ciência aberta e *e-Science* no campo da gestão, compartilhamento e reuso de dados de pesquisa em uma cidade universitária da Colômbia. Por meio do estabelecimento de tais práticas, a pesquisa permitiu compreender a realidade científica da amostra estudada, para que, com base nos descobrimentos, a comunidade acadêmica possa propor estratégias para a construção e adoção de métodos padronizados de gestão de dados e boas práticas de compartilhamento e reuso.

Atualmente, investigações sobre gestão de dados de pesquisa é um dos tópicos mais relevantes e pertinentes no cenário científico colombiano. A gestão de dados é considerada um

método eficiente para preservar e curar os dados, a qual deve ser necessariamente incorporada nos processos de pesquisa para gerar dados de qualidade e úteis em ambientes científicos abertos. Revelar os mecanismos pelos quais os pesquisadores colombianos gerenciam seus dados implica em saber em que medida os dados da pesquisa são preservados e curados para garantir sua integridade e disponibilidade. Por sua parte, revelar os métodos de compartilhamento e reúso de dados de pesquisa, estabelece uma visão geral do uso desse recurso e o grau de abertura que a comunidade científica possui em relação ao acesso aberto a dados.

Sugestões para pesquisas futuras:

Com base nos resultados apresentados na presente investigação, são sugeridos possíveis estudos para complementar o tópico investigado:

- ✓ A revisão bibliográfica mostra que, na gestão, compartilhamento e reúso de dados de pesquisa, há uma série de atores que determinam sua implementação, e independe do conhecimento, interesse e vontade do pesquisador. Atores como instituições de ensino superior, agências de fomento, governo, entre outros, devem ser estudados para entender, desde uma perspectiva holística, a incorporação de tais práticas no cotidiano do pesquisador. Assim, para complementar os resultados desta pesquisa e abordar o assunto desde uma perspectiva sistêmica, é necessário revelar como o governo, agências de financiamento, instituições de ensino superior, associações acadêmicas, instituições de pesquisa abordam práticas de gestão, intercambio e reúso de dados.
- ✓ Para ampliar o escopo da investigação, propõe-se um estudo de natureza explicativa que permita revelar as percepções e razões pelas quais pesquisadores e outros atores incorporam práticas de gestão de dados e implementam seu compartilhamento e reúso.
- ✓ Com base no que foi revelado na pesquisa sobre as experiências dos pesquisadores na gestão, compartilhamento e reúso de dados de pesquisa, o estudo possibilita estabelecer diretrizes, boas práticas e modelos que atendam às necessidades da população estudada e permitam que as práticas sejam realizadas de forma sistematizada.

REFERÊNCIAS

- ABBOTT, D. (2008). **What is Digital Curation?**. DCC Briefing Papers: Introduction to Curation. Edinburgh: Digital Curation Centre. Handle: 1842/3362. Disponível em: <http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/what-digital-curation> Acesso em: 8 jun 2019.
- ALBAGLI, S.; APPEL, A.L.; MACIEL, M.L. E-science e ciência aberta: questões em debate. In: ENANCIB, 14., 2013, Florianópolis. **Anais...** Florianópolis: ANCIB, 2013. Disponível em: <http://ridi.ibict.br/handle/123456789/465>. Acesso em: 3 ago. 2019
- ALBUQUERQUE, M.R. **A curadoria em um repositório institucional: uma análise sob a ótica do ciclo de vida dos dados de Sant'Ana**. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade Federal do Ceará, Fortaleza, 2018.
- ALLARD, Suzie. DataONE: Facilitating eScience through Collaboration. **Journal of eScience Librarianship**, v. 1, n. 1, p. 4-17, 2014. Disponível em: <http://escholarship.umassmed.edu/cgi/viewcontent.cgi?article=1004&context=jeslib>. Acesso em: 25 out 2019.
- ANDRADE, M.M. **Introdução à metodologia do trabalho científico**. 5. ed. São Paulo: Atlas, 200.
- ARIAS, G.F. **El proyecto de Investigación: introducción a la metodología científica**. 5. ed. Caracas: Editorial Episteme, 2006.
- BABBIE, E. **Métodos de pesquisas de Survey**. Tradução: Guilherme Cezarino. 2.ed. Belo Horizonte: Editora UFMG, 2003. Título original: Survey Research Methods.
- BARRERA, M.A.M. Uso de la correlación de Spearman en un estudio de intervención en fisioterapia. **Mov.cient.** V. 8 , n.1, p. 98-104, 2014. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=5156978>. Acesso em: 2 Mar. 2020.
- BALL, A. **Review of the State of the Art of the Digital Curation of Research Data**. Other. Londres: University of Bath, Bath, 2010. Disponível em: https://www.researchgate.net/publication/242710555_Review_of_the_State_of_the_Art_of_the_Digital_Curation_of_Research_Data. Acesso em: 9 jun.2020
- BALL, A. **Review of Data Management Lifecycle Models**. Londres: University of Bath, 2012.
- BELL, G. Foreword. In: HEY, T.; TANSLEY, S.; TOLLE, K. (org.). **The Fourth Paradigm: Data-intensive Scientific Discovery**. Redmond, Washington: Microsoft Research, 2009.
- BELL, G.; HEY, T.; SZALAY, A. Beyond the data deluge (computer Science). **Science**, v. 323, n. 5919, p. 1297-1298, 6 Mar. 2009. Disponível em: https://www.researchgate.net/publication/24180847_Beyond_the_Data_Deluge_Computer_Science. Acesso em: 13 Out 2019.

BIBLIOTECA DA CEPAL. ¿Cómo identificar tipos de datos de investigación?. Disponível em: <https://biblioguias.cepal.org/gestion-de-datos-de-investigacion/tipos-datos>. Acesso em: 5 fev.2020.

BORGMAN, C.L. Advances in information science: The Conundrum of Sharing Research Data. **Journal of The American Society for Information Science and Technology**. Maryland, v. 63, n.6, p.1059–1078, junho de 2012. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1002/asi.22634>. Acesso em: 20 ago. 2019.

_____. **Big data, little data, no data: scholarship in the networked world**. Londres, Inglaterra: The MIT Press Cambridge, Massachusetts, 2015.

BOAI- Budapest Open Access Initiative. **Dez anos depois da Budapest Open Access Initiative: estabelecendo o Acesso Aberto como padrão**. Budapest, 2002. Disponível em: <https://www.budapestopenaccessinitiative.org/boai-10-translations/portuguese>. Acesso em: 15 Out 2019.

BOAI- Budapest Open Access Initiative. **Dez anos da Iniciativa de Budapeste em Acesso Aberto: a abertura como caminho a seguir**. Budapest, 2002. Disponível em: <https://www.budapestopenaccessinitiative.org/boai-10-translations/portuguese-brazilian-translation>. Acesso em: 15 Out 2019.

BRICEÑO, M.P.L.E. **Gestión de datos de investigación en universidades en base al ciclo de vida de los datos. Caso de estudio: Área de Ciencias de la Salud**. Dissertação (Mestrado Gestión de la información y conocimiento) – Facultad de Ingeniería de Sistemas e Informática, Universidad Nacional Mayor de San Marcos, Lima, 2017.

_____. **“Gestión de datos de investigación en universidades en base al ciclo de vida de los datos. Caso de estudio: Área de Ciencias de la Salud”**. Questionário. Facultad de Ingeniería de Sistemas e Informática, Universidad Nacional Mayor de San Marcos. Lima: 2017.

CAROL, T.E.D *et al.* **“Scientists and research data: Continuing to build an understanding of your data needs”**. Questionário. PLoS ONE 10(8): e0134826. 2015. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0134826>. Acesso em: 20.jan.2020.

CARRASCO, S.D. **Metodología de investigación científica: Pautas metodológicas para diseñar y elaborar el proyecto de investigación**. Lima: Editorial San Marcos, 2009.

CASTAÑO, G.A., CALDERÓN, G. & POSADA, R. Perfiles y tipologías del investigador en administración en Colombia y su producción científica. *Innovar*, V. 24, N. 52, p. 45-57, 2014. Disponível em: <https://revistas.unal.edu.co/index.php/innovar/article/view/42505>. Acesso em: 15 fev 2020.

CAVALCANTI, M.T.; SALES, L.F. Gestão de dados de pesquisa: um panorama da atuação da união europeia. **Biblos: Revista do Instituto de Ciências Humanas e da Informação**, v. 31, n. 1, p. 73-98, jan./jun. 2017. Disponível em: <https://periodicos.furg.br/biblos/article/view/5789/4616>. Acesso em: 19 sep 2019.

CENTRO DE CIÊNCIAS DA SAÚDE (CCS). **Comitê de Ética em Pesquisa**. João Pessoa. Disponível em: <http://www.ccs.ufpb.br/eticaccsufpb/>. Acesso em: 27 out. 2019

COLCIENCIAS - DEPARTAMENTO ADMINISTRATIVO DE CIENCIA, TECNOLOGÍA E INNOVACIÓN- (Colombia). **Lineamientos para una Política de Ciencia Abierta en Colombia**. Bogotá, 2018. Disponível em: https://www.colciencias.gov.co/sites/default/files/ckeditor_files/Lineamientos%20ciencia%20abierta%2017-dic-2018-doc.pdf. Acesso em: 5 jun. 2019.

_____. **Convocatoria Nacional para el Reconocimiento y Medición de Grupos de Investigación, Desarrollo Tecnológico o de Innovación y para el Reconocimiento de Investigadores del SNCITEI-2018**. Bogotá, 2018. Disponível em: <https://minciencias.gov.co/convocatorias/investigacion/convocatoria-nacional-para-el-reconocimiento-y-medicion-grupos-0>. Acesso em: 1 mai. 2020.

_____. **Modelo de medición de grupos de investigación, desarrollo tecnológico o de innovación y de reconocimiento de investigadores del sistema nacional de ciencia tecnológica e innovación, año 2018**. Bogotá, 2018. Disponível em: <https://minciencias.gov.co/sites/default/files/upload/convocatoria/4._anexo_1._documento_conceptual_del_modelo_de_reconocimiento_y_medicion_de_grupos_de_investigacion_2018.pdf>. Acesso em: 1 mai. 2020.

COLOMBIA. **Plan de Desarrollo Nacional 2018-2022**. Colombia: Departamento Nacional de Planeación. Bogotá, 2019. Disponível em: <https://colaboracion.dnp.gov.co/CDT/Prensa/BasesPND2018-2022n.pdf>. Acesso em: 5 jun. 2019.

COLLIS, J.; HUSSEY, R. **Pesquisa em Administração: um guia prático para alunos de graduação e pós-graduação**. 2. ed. Porto Alegre: Bookman, 2005

CURTY, R.G.. **Beyond “Data Thrifting”: an investigation of factors influencing research data reuse in the social sciences**. 2015. Dissertação. (Syracuse University)-New York, 2015.

DATAONE. **What is DataONE?**. Disponível em: <<https://www.dataone.org/what-dataone>> Acesso em: 20 out 2019.

_____. **Education modules**. 2016. Disponível em: https://www.dataone.org/sites/all/documents/education-modules/handouts/L06_DataProtection_Handout.pdf. Acesso em: 25 out. 2019.

_____. **Education modules - Lesson 1: Data Management**. 2016. Disponível em: https://www.dataone.org/sites/all/documents/education-modules/handouts/L01_DataManagement_Handout.pdf Acesso em: 20 out 2019.

_____. **Education modules- Lesson 2: Data Sharing**. 2016. Disponível em: https://www.dataone.org/sites/all/documents/education-modules/handouts/L02_DataSharing_Handout.pdf Acesso em: 20 oct 2019.

_____. **Education modules- Lesson 3:** Data Management Planning. 2016. Disponível em: https://www.dataone.org/sites/all/documents/education-modules/handouts/L03_DataManagement_Handout.pdf Acesso em: 20 oct 2019.

_____. **Education modules- Lesson 4:** Data Entry & Manipulation. 2016. Disponível em: https://www.dataone.org/sites/all/documents/education-modules/handouts/L04_DataEntry_Handout.pdf Acesso em: 20 oct 2019

_____. **Education modules- Lesson 5:** Data Quality Control and Assurance. 2016. Disponível em: https://www.dataone.org/sites/all/documents/education-modules/handouts/L05_DataQualityAssurance_Handout.pdf Acesso em: 20 out 2019

_____. **Education modules- Lesson 6:** Protecting Your Data. 2016. Disponível em: https://www.dataone.org/sites/all/documents/education-modules/handouts/L06_DataProtection_Handout.pdf Acesso em: 20 out 2019

_____. **Education modules- Lesson 7:** Metadata. 2016. Disponível em: https://www.dataone.org/sites/all/documents/education-modules/handouts/L07_DefiningMetadata_Handout.pdf Acesso em: 20 out 2019

_____. **Education modules- Lesson 10:** Legal and Policy Issues. 2016. Disponível em: https://www.dataone.org/sites/all/documents/education-modules/handouts/L07_DefiningMetadata_Handout.pdf Acesso em: 20 out 2019

_____. **Data Management Guide for Public Participation in Scientific Research.** Disponível em: <https://www.dataone.org/sites/all/documents/DataONE-PPSR-DataManagementGuide.pdf> Acesso em: 10 jun 2020.

DIAS, G.A; ANJOS, R. L.; ARAÚJO, D.G. 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 reúso de dados. **LiINC em revista**. v.15, n.2, p. 5-31. 2019.

DIAS, G.A; ANJOS, R. L.; ARAÚJO, D.G. “**A gestão de dados científicos: Desvendando as práticas dos pesquisadores brasileiros na área da Ciência da Informação**”. Questionário. Universidade Federal da Paraíba. Joao pessoa: 2017.

DIAS, G.A; ANJOS, R. L.; RODRIGUES, A. A. Os princípios FAIR: viabilizando o reúso de dados científicos. In: DIAS, G.A.; OLIVEIRA, B.M.J.F. **Dados científicos. Perspectivas e desafios**. Joao Pessoa: Editora da UFPB, 2019, p. 177-187

DIGITAL CURATION CENTRE (DCC). **History of the DCC**. Disponível em: <http://www.dcc.ac.uk/about-us/history-dcc/history-dcc>. Acesso em: 10 sep.2019

_____. **About the DCC**. [S.l.] [2019?]. Disponível em: <<http://www.dcc.ac.uk/about-us>>. Acesso em: 20 out. 2019.

_____. **DCC Curation Lifecycle Model**. Disponível em: <http://www.dcc.ac.uk/resources/curation-lifecycle-model> Acesso em: 25 out. 2019.

DIGITAL SCIENCE, *et al.* **The State of Open Data Report**. Figshare. 2018. Disponível em: https://figshare.com/articles/The_State_of_Open_Data_Report_2018/7195058 Acesso em: 20 ago. 2019

DUDZIAK, E. **Dados de pesquisa agora devem ser armazenados e citados**. 2016. Disponível em: <http://www.sibi.usp.br/noticias/dados-materiais-metodos-revistas-exigem-dados-pesquisa-estejam-disponiveis/> Acesso em: 19 out. 2019.

EYNDEN, Veerle *et al.* **Managing and data sharing: best practice for researchers**. Colchester: UK Data Archive, 2011. Disponível em: <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>. Acesso em: 8 jun.2020

FAPESP. **Open Science @ FAPESP**. Disponível em: <http://fapesp.br/openscience/>. Acesso em: 8 nov.2019.

FECHER, B.; FRIESIKE, S. Open Science: One Term, Five Schools of Thought. *In*: Bartling S., Friesike S. (org.) **Opening Science**. Springer, Cham, 2013, p.17-47.

GILLILAND-SWETLAND, J.A. La definición de los metadatos. *In*: TRUST, G.P.J. **Introducción a los Metadatos: vías a la información digital**. Estados Unidos de América: The J.Paul. Getty Trust, 1999, p.3-4. Disponível em: <http://d2aohiyo3d3idm.cloudfront.net/publications/virtuallibrary/0892365358.pdf>. Acesso em: 5 fev. 2020.

HERNÁNDEZ, R.S; FERNÁNDEZ, C.C; BAPTISTA, L.P. **Metodología de la investigación**. 5. ed. México, D.F: McGraw Hill INTERAMERICANA, 2010.

HEY, T.; HEY, J. E-Science and its implications for the library community. **Library Hi Tech**, v. 24, n. 4, out. 2006. Disponível em: https://www.researchgate.net/publication/28806348_E-science_and_Its_Implications_for_the_Library_Community. Acesso em: 4 jun. 2019.

HUSSON, F.; LÊ, S.; PAGÈS, J. **Análisis de datos com R**. 1. ed. Bogotá: Escuela Colombiana de Ingeniería, 2013.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4.ed. São Paulo: Atlas, 2002

_____. **Métodos e técnicas de pesquisa social**. 5. ed. São Paulo: Atlas, 2007.

_____. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GRAY, J. eScience: A Transformed Scientific Method. *In*: HEY, T.; TANSLEY, S.; TOLLE, K. (org.). **The Fourth Paradigm: Data-intensive Scientific Discovery**. Redmond, Washington. UNT Digital Library. Palestra de Jim Gray em 11 de janeiro de 2007, cedida a NRC-CSTB1, em Mountain View, CA.

LIND, A.D.; MARCHAL, G.W; WATHEN, A.S. **Estadística Aplicada a los Negocios y la Economía**. 13.ed. México, D.F: McGraw Hill Interamericana, 2008.

LUZI, D; RUGGIERI, R; BIAGIONI, S; SCHIANO, E. **“Data sharing in environmental sciences: A survey of CNR researchers”**. Questionário. DANS. Itália: 2012. Disponível em: <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:53292/tab/1>. Acesso em: 1 fev. 2020.

MARCONI, M. A.; LAKATOS, E.M. **Técnicas de pesquisa**. 5 ed. São Paulo: Atlas, 2002.

_____. **Fundamentos da metodologia científica**. 7 ed. São Paulo: Atlas, 2010.

MCKEN, K. et al. Research360: Data in the Research Lifecycle. University of Bath: **JISC: UKOLN**, 2012. Disponível em: https://purehost.bath.ac.uk/ws/portalfiles/portal/9193618/Bath_Mcken_research_data.pdf. Acesso em: 10 out. 2019.

MEDEIROS, J. da S.; CAREGNATO, S. E. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **Liinc em Revista**, v. 8, n. 2, jul./dez. 2012. Disponível em: <http://revista.ibict.br/liinc/article/view/3368>. Acesso em: 1 sep 2019.

MIQUEL, S. (1997). **Investigación de mercados**. 1. ed. España: McGraw-Hill, 1997.

MUNAFÓ, M. R. *et al.* A manifesto for reproducible science. **Nature Human Behaviour**, v.1, n. 0021, p. 1-9. 2017. Disponível: <https://www.nature.com/articles/s41562-016-0021> em: Acesso em: 18 sep 2019.

NATIONAL RESEARCH COUNCIL, *et al.* **BITS OF POWER: Issues in Global Access to Scientific Data**. 1. ed. Washington, D.C: National Academies Press, 1997.

ROJAS, A.N.H. Notas sobre Psicometría. Universidad Nacional de Colombia. Bogotá: 1998.

OBSERVATORIO COLOMBIANO DE CIENCIA Y TECNOLOGÍA- OCyT (Colombia). **Estudio para identificar conocimientos, capacidades, percepciones y experiencias de los investigadores del país frente a la ciencia abierta**. Bogotá, 2017. Disponível em: https://www.colciencias.gov.co/sites/default/files/ckeditor_files/Estudio%20identificacion%20conocimientos%2C%20capacidades%2C%20percepciones%20y%20experiencias%20Ciencia%20Abierta%20Ocyt.pdf. Acesso em: 20 ago. 2019.

OECD- Organisation for Economic Co-operation and Development. **Principles and Guidelines for Access to Research Data from Public Funding**. 2007. Disponível em: <https://www.oecd-ilibrary.org/docserver/9789264034020-en-fr.pdf?expires=1556811511&id=id&accname=guest&checksum=2065D324FE488C42172940986C3719A4>. Acesso: 2 mai. 2019.

OPEN KNOWLEDGE FOUNDATION. **The Open Definition 2.1**. Disponível em: <https://opendefinition.org/>. Acesso em: 10 out. 2019.

OLIVEIRA, A.C.S. **Desvendando a autorialidade colaborativa na e-Science sob a ótica dos direitos de propriedade intelectual**. 2016. 297f.:il. Tese (Doutorado em Ciências da Informação) – Programa de Pós-graduação em Ciências da Informação, Universidade Federal da Paraíba, João pessoa, 2016.

PEREZ, J.E.; MARTÍNEZ, A.C. Validez de contenido y juicio de expertos: una aproximación a su utilización. **Revista avances en Mediciación**. v.9, n.2, p. 27-36. 2008. Disponível em: http://www.humanas.unal.edu.co/psicometria/files/7113/8574/5708/Articulo3_Juicio_de_expertos_27-36.pdf. Acesso em: 14 dec.2019

POWERDATA. Introducción a la Calidad de Datos: Definición, Control y Beneficios. Powerdata, 2014. Disponível em: <<https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/368784/introducci-n-a-la-calidad-de-datos-definici-n-control-y-beneficios>>. Acesso em: 5 fev 2020.

RAMÍREZ, P.A.; SAMOILOVICH, D. **Ciencia abierta**: Reporte para tomadores de decisiones. 2. ed. Montevideo: Ediciones Unesco, 2018. Disponível em: <https://www.columbus-web.org/es/nuestros-activos/publicaciones/item/250-ciencia-abierta-report-e-para-tomadores-de-decisiones.html>. Acesso em: 17 ago.2019

RIBES, D.; LEE, C. P. Sociotechnical studies of cyberinfrastructure and e-research: current themes and future trajectories. **Computer supported cooperative work**, v. 19, n. 3-4, p. 231-244, 2010. Disponível em: <http://www.davidribes.com/storage/Ribes%20Lee%20-%20Cyberinfrastructure%20Studies.pdf> Acesso em: 11 jun. 2019.

SALKIND, J. N. **Métodos de pesquisa**. 3. ed México D.F: Printice Hall, 1999.

SAYÃO, L.F.; SALES, L.F. Curadoria Digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**, 22(3), João Pessoa, v.22, n.3, p. 179-191, set/dez. 2012. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/12224/8586>. Acesso em: 19 ago. 2019

_____. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 8, n. 2. 2013. Disponível em: <<https://periodicos.ufpb.br/ojs/index.php/pbcib/article/view/18634>>. Acesso em: 13set 2019

_____. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **RECIIS – Rev. Eletron. de Comun. Inf. Inov. Saúde**, v.8, n.2, p.76-92, jun. Rio de Janeiro, 2014. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611>. Acesso em: 18 ago. 2019

_____. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. Disponível em: http://www.cnen.gov.br/images/CIN/PDFs/GUIA_DE_DADOS_DE_PESQUISA.pdf. Acesso em: 11 jun. 2019.

_____. Curadoria digital e dados de pesquisa. **AtoZ: novas práticas em informação e conhecimento**, 5(2), 67 – 71. 2016. Disponível em: <https://revistas.ufpr.br/atoz/article/view/49708/30161>. Acesso em: 19 out. 2019.

SANT’ANA, R.C.G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação & Informação**. V.21,n.2, 2016. p. 116-142. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940/20124> Acesso em: 8 nov. 2019

SEGUNDO, S. J. S.; ARAÚJO, W. J. Curadoria e Ciclo de Vida dos Dados. In: DIAS, G.A.; OLIVEIRA, B.M.J.F. **Dados científicos: Perspectivas e desafios**. Joao Pessoa: Editora da UFPB, 2019, p. 113-152.

THE UNIVERSITY OF EDINBURGH. Choose the best file formats. Edinburgh, 2020. Disponível em: <https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/choosing-file-formats>. Acesso em: 2 mai. 2020.

THE UNIVERSITY OF EDINBURGH. (s. f.-b). Sharing, preservation, and licensing. Edinburgh, 2017. Disponível em: <http://mantra.edina.ac.uk/sharingpreservationandlicensing/> Acesso em: 20 mai. 2020

THE UNIVERSITY OF MELBOURNE. Management of Research Data and Records Policy (MPF1242). Melbourne, 2013. Disponível em: <https://policy.unimelb.edu.au/MPF1242#section-8>. Acesso em: 16 out. 2019.

UHLIR P, SCHRÖDER P. Open Data for Global Science. **Data Science Journal** Jun 2007; 6 (Open Data Issue). Disponível em: https://www.researchgate.net/publication/220390291_Open_Data_for_Global_Science. Acesso em: 20 ago. 2019.

UKOLN. **About UKOLN**. [S.l.] [2013?]. Disponível em: <http://www.ukoln.ac.uk/about/>. Acesso em: 20 out. 2019.

VANZ, S.A. S; PASSOS, P.C. S. J; CAREGNATO, S.E.; PAVÃO, C. G; BORGES, N. B.; ROCHA, R. P.; GABRIEL, R.F.; AZAMBUJA, L.A.B. “**Acesso aberto a dados de pesquisa no Brasil: práticas e percepções dos pesquisadores: relatório 2018**”. Questionário. Rede de Dados de Pesquisa Brasileira. Porto Alegre: 2018. Disponível em: <https://lume.ufrgs.br/bitstream/handle/10183/185195/001082283.pdf?sequence=1&isAllowed=y>. Acesso em: 20 mai. 2019.

VERGARA, S.C. **Projetos e relatórios de pesquisa em administração**. 2.ed. São Paulo: Atlas, 1998.

VILLAVICENCIO-CAPARÓ. E. *et al.* ¿Cómo plantear las variables de una investigación?: Operacionalización de las variables. **Revista OACTIVA**. V.4, n.1, 2019. p. 9-14. Disponível em: https://www.researchgate.net/publication/332032600_OPERACIONALIZACION_DE_VARIABLES . Acesso em: 9 fev.2020

WIGGINS, *et al.* Data management guide for public participation in scientific research. **DataONE**. 2013. Disponível em: <https://www.dataone.org/sites/all/documents/DataONE-PPSR-DataManagementGuide.pdf>. Acesso em: 25 out. 2019.

YANG, X.; WANG, L.; VON LASZEWSKI, G. Recent research advances in e-science, **Cluster Computing**, v 12, n. 4, p. 353-356, 2009. Disponível em: https://www.researchgate.net/publication/220405889_Recent_Research_Advances_in_e-Science. Acesso em: 1 set. 2019.

VERGUGO, M.O. Percepción Sistemática del Clima Organizacional a través de la Práctica de Desarrollo Orgnizacional. **Working paper**. Sonora: Universidad de Sonora, 2005.

APÊNDICES

APÊNDICE A – QUESTIONARIO
QUESTIONÁRIO⁴⁷:
PRÁTICAS DE GESTÃO, COMPARTILHAMENTO E REÚSO DE DADOS DE
PESQUISA

I. PERFIL DO PESQUISADOR

A continuação, apresenta-se perguntas relacionadas à carreira profissional e de pesquisa do professor. Os critérios de inclusão determinados para participar da pesquisa são:

1. Estar vinculado como professor a uma universidade pública ou privada localizada na cidade de Bucaramanga e na região metropolitana

2. Estar vinculado a um grupo de pesquisa classificado por Colciencias e da área de administração, economia, negócios e áreas afins

1. Selecione o tipo de instituição à qual está vinculado (a):

Pública
Particular

2. Qual é o seu grau de formação acadêmica?

Graduação
Especialização
Mestrado em curso
Mestrado completo
Doutorado em curso
Doutorado completo
Pós-doutorado

3. Tempo de atuação como professor (a) universitário (a):

Menos de 5 anos
Entre 5 e 10 anos

⁴⁷ Instrumento de pesquisa baseado nos questionários intitulados “Acesso Aberto a Dados de Pesquisa (AADP)” desenvolvido pela Rede de Dados de Pesquisa Brasileira (RDP Brasil); “A gestão de dados científicos: Desvendando as práticas dos pesquisadores brasileiros na área da Ciência da Informação” desenvolvido pelos pesquisadores do Grupo de Pesquisa Web Representação do Conhecimento e Ontologias – WRCO, da Universidade Federal da Paraíba; “Scientists and research data: Continuing to build an understanding of your data needs” desenvolvido por Data Observation Network for Earth; “Data sharing in environmental sciences: A survey of CNR researchers” desenvolvido pelos pesquisadores Daniela Luzi, Roberta Ruggieri, Stefania Biagioni, Elisabetta Schiano; “Gestión de datos de Investigación em universidades em base al ciclo de vida de los datos. Caso de estudio: Área de Ciencias de la Salud” desenvolvido por María del Pilar Liliana Eléspuru Briceño; módulos de educação: “Data sharing”, “Data Management Planning”, “Data Entry and Manipulation”, “Data Quality Control and Assurance”, “Protecting your Data”, “Metadata”, “Legal and Policy Issues”, desenvolvido por Data Observation Network for Earth.

Entre 11 e 15 anos

Entre 16 e 20 anos

Mais de 20 anos

4. Selecione sua categoria atual como pesquisador (a) reconhecido (a) por Colciencias⁴⁸:

Emérito

Associado

Sênior

Junior

Membro com doutorado

Membro com mestrado

Membro vinculado com especialização

Sem categoria

Outra

5. Tempo de atuação como pesquisador:

Menos de 5 anos

Entre 5 e 10 anos

Entre 11 e 15 anos

Entre 16 e 20 anos

Mais de 20 anos

6. Qual é a área de conhecimento em que realiza sua pesquisa:

II. CARACTERÍSTICAS DOS DADOS DA PESQUISA

A continuação, enumera-se o termo " dados de pesquisa" com o propósito de que o participante conheça o significado que se assigna e responda a seguinte seção:

Dados da pesquisa: unidades de informação criadas ou coletadas no decorrer de pesquisas científicas (planilhas estatísticas, textos, imagens, sons, vídeos, números, algoritmos etc.) (Sayão & Sales, 2015, p.80)

7. Selecione a principal característica dos seus dados de pesquisa:

Dados quantitativos

Dados qualitativos

Dados quantitativos e qualitativos

8. Que tipo de dados você produz e manipula em suas pesquisas?

Selecione as opções que se aplicam.

Numérico (equações, algoritmos, dados estatísticos, etc.)

Textual (documentos, cadernos de laboratório ou de campo, etc.)

⁴⁸ Entidade administrativa de ciência, tecnologia e inovação a qual promove políticas públicas relacionadas ao progresso científico da Colômbia.

Imagens (imagens, filmes, fotografias, animações, etc.)
 Audição (entrevistas, sons, etc.)
 Software
 Outro, especifique:

9. Selecione os tipos de formatos em que armazena seus dados de pesquisa:

Documentos de texto (Doc., Txt, pdf, ect)
 Documentos digitalizados (pdf, etc)
 Planilhas (Excel, XML, CSV, XLS, etc)
 Dados: estatística / SPSS, SAS (sav, spv, etc)
 Imagens / gráficos (JPEG, GIF, etc.)
 Áudio (mp3, wav, etc.)
 Vídeos / filmes (wmv, etc.)
 Software
 Web (html, etc)
 Outro, especifique

10. Por quanto tempo armazena os dados da pesquisa:

Menos de 1 ano
 Entre 1 e 5 anos
 Mais de 5 anos
 Indefinidamente
 Não sei

11. Tipo de armazenamento usado para arquivar seus dados de pesquisa:

Itens	Nenhum dos meus dados	Parte dos meus dados	Todos os meus dados
Em físico- papel			
Computador pessoal			
Computador de trabalho			
Disco rígido externo (HD externo)			
USB/pendrive			
Nuvem (Dropbox, Google Drive)			
Repositório / observatório / centro / base de dados do meu grupo de pesquisa			
Repositório / observatório / centro / base de dados da minha instituição			
Repositório / observatório / centro / base de dados temática			
Repositório / observatório / centro / base de dados de redes científicas			
Repositório / observatório / centro / base de dados nacionais			
Repositório / observatório / centro / base de dados internacionais			

III. PRÁTICAS DE GESTÃO DE DADOS DE PESQUISA

A continuação, enumera-se os termos presentes nesta pesquisa com o propósito de que o participante conheça o significado que se assigna e responda a seguinte seção:

Gestão de dados de pesquisa: práticas relacionadas à manutenção, segurança e preservação de dados de pesquisa a fim de garantir seu uso, reúso e intercambio. (SAYÃO & SALES, 2015). Entre as diferentes fases que compõem a gestão estão:

Planejar: definição dos tipos de dados que serão produzidos, processados e analisados. Determina métodos de preservação e manutenção, políticas de propriedade intelectual, entre outros aspectos. Nesta fase, é construído o Plano de Gerenciamento de Dados, documento que contém os itens mencionados acima.

Coleta: estabelece padrões para evitar erros ou contaminação de dados na sua coleta, processamento e análise.

Descrever (metadados): documentação ou informação que possibilita compreender os dados de pesquisa coletados. Descreve sua autoria, contexto, propósito, métodos, organização, dentre outros aspectos.

Metadados de nível de pesquisa: contexto, metodologia, fonte de dados, processamento, garantia e controle de qualidade, etc.

Metadados ao nível da base de dados: tabelas de base de dados, formatos, etc.

Metadados ao nível da variável: nomes, etiquetas, significado, descrição das variáveis, categorias, etc.

Metadados a nível administrativo: técnico, preservação, direitos de autor, etc.

Preservar: Conjunto de métodos tecnológicos e gerenciais voltados para garantir que os dados permaneçam intactos, acessíveis e compreensíveis ao longo do tempo. (SAYÃO & SALES, 2015)

Garantia da qualidade dos dados: atividades que impedem que erros entrem ou permaneçam em um conjunto de dados. Estas atividades asseguram a qualidade dos dados antes da sua coleta, e monitorizam e mantêm a qualidade dos dados durante e após a conclusão da investigação (DATAONE, 2016).

As práticas de gestão de dados podem ser implementadas seguindo as diretrizes propostas pelas instituições, centros de dados, etc., ou podem estar implícitas no processo de pesquisa, sem que o pesquisador siga critérios ou associe essas atividades ao conceito de gestão de dados.

12. ¿Qual é a sua familiaridade com a gestão de dados de pesquisa?

- Eu tenho conhecimento completo
- Eu tenho um conhecimento parcial
- Eu tenho pouco conhecimento
- Desconheço

➤ Em relação ao planejamento dos seus dados de pesquisa, afirma que:

ITENS	Nunca (1)	Raramente (2)	Às Vezez (3)	Frequentemente (4)	Sempre (5)
13. Planeja antes de iniciar a pesquisa como os dados serão adquiridos, processados, analisados, armazenados, protegidos e assegurados.					
14. Planeja a descrição (metadados) dos dados					
15. Planeja a preservação dos dados (estabelece um método que protege os dados, tentando preservar seu estado e evitando danos)					
16. Estabelece critérios de propriedade intelectual sobre os dados					
17. Cria um documento formal - plano de gestão de dados -, a partir de diretrizes propostas pelas instituições, repositórios, centros de dados, etc., o qual descreve a coleta, processamento, armazenamento, preservação, garantia de qualidade, metadados, critérios de propriedade intelectual, dentre outros aspectos.					
18. Utiliza ferramentas digitais que assistem na criação de um plano de gestão de dados					
19. Tem algum tipo de assistência, por parte de um grupo de especialistas em dados, para construir seu plano de gestão de dados					

➤ Em relação aos metadados associados aos dados de pesquisa, afirma que:

ITENS	Nunca (1)	Raramente (2)	Às Vezez (3)	Frequentemente (4)	Sempre (5)
20. Descreve os dados a nível de pesquisa (contexto, metodologia, coleta de dados, processamento, garantia e controle de qualidade, etc.)					
21. Descreve os dados a nível de base de dados (tabelas de base de dados, formatos, etc.)					

ITENS	Nunca (1)	Raramente (2)	Às Vezes (3)	Frequentemente (4)	Sempre (5)
22. Descreve os dados a nível de variável (nomes, etiquetas, significado, descrição das variáveis, categorias, etc.)					
23. Descreve os dados a nível administrativo (técnico, preservação, direitos de autor, etc.)					
24. Orienta a descrição de seus dados com base em metadados padrões determinados por instituições, associações científicas ou data centers.					
25. Usa ferramentas digitais para criar metadados de seus dados.					
26. Tem algum tipo de assistência de um grupo de especialistas em dados para criar os metadados.					

- Em relação à verificação, inspeção e controle da qualidade dos dados de pesquisa, afirma que:

Critérios de qualidade dos dados:

Confiabilidade: Os dados fornecidos são precisos. A representação dos dados (ou valor) reflete bem o estado real da informação de origem.

Consistência: Após o processamento dos dados, seus conceitos, domínios de valor e formatos ainda coincidem como antes do processamento. Todos os dados são consistentes ou verificáveis.

Legibilidade: os dados (conteúdo) são claros e compreensíveis. A descrição dos dados, a classificação e o conteúdo da codificação satisfazem as especificações e são fáceis de compreender

ITENS	Nunca (1)	Raramente (2)	Às Vezes (3)	Frequentemente (4)	Sempre (5)
27. Verificar a confiabilidade dos dados					
28. Verificar a consistência dos dados					
29. O conteúdo dos dados é claro e compreensível					
30. A descrição dos dados (metadados), é fácil de entender					

ITENS	Nunca (1)	Raramente (2)	Às Vezez (3)	Frequentemente (4)	Sempre (5)
31. Segue diretrizes que orientem a verificação, inspeção e controle de dados proposto por instituições, repositórios, centro de dados, etc.					
32. Tem algum tipo de assistência de um grupo de especialistas em dados para garantir a qualidade dos dados da pesquisa.					

➤ Em relação à preservação dos dados de pesquisa, afirma que:

ITENS	Nunca (1)	Raramente (2)	Às Vezez (3)	Frequentemente (4)	Sempre (5)
33. Seleciona os dados que devem ser preservados a médio e longo prazo.					
34. Criar versões digitais dos dados coletados em suporte físico (papel)					
35. Armazena os dados em formatos considerados legíveis a médio e longo prazo.					
36. Verifica periodicamente a obsolescência dos formatos que armazenam os dados					
37. Realiza migração de dados de um formato para outro, preservando suas características, quando for necessário.					
38. Realiza backups em vários formatos.					
39. Realiza backups em diversos locais.					
40. Inspecciona os backups periodicamente para confirmar seu bom estado.					
41. Orienta a preservação de seus dados de pesquisa através de diretrizes propostas por instituições, repositórios, centros de dados, etc.					

ITENS	Nunca (1)	Raramente (2)	Às Vezez (3)	Frequentemente (4)	Sempre (5)
42. Tem algum tipo de assistência de um grupo de especialistas em dados para preservar a médio e longo prazo os dados de pesquisa.					

IV. COMPARTILHAMENTO DE DADOS DE PESQUISA

A continuação, enumera-se o termo “ compartilhamento de dados de pesquisa” com o propósito de que o participante conheça o significado que se assigna e responda a seguinte seção:

"O intercâmbio de dados é a publicação de dados de pesquisa para uso de outras pessoas. A liberação pode assumir muitas formas, desde troca privada sob pedido, até depósito em centros ou repositórios de dados pública. A publicação dos conjuntos de dados em um site público ou seu fornecimento a uma revista como material suplementar também é considerada uma troca. Contudo, o grau de utilidade, confiabilidade e valor dos dados compartilhados varia muito. Algumas podem ser ricamente estruturadas e curadas. Outros podem ser arquivos em bruto com o mínimo de documentação. (BORGMAN, 2011, P. 3)

43. Quanto dos seus dados de pesquisa disponibiliza (compartilha) para pesquisadores, comunidade acadêmica, instituições, centros, etc.?

Todos os meus dados
Parte dos meus dados
Nenhum dos meus dados

Se a sua resposta anterior é não compartilha nenhum dos seus dados de pesquisa, ignore as perguntas 44 a 57 e continue na seção reuso de dados de pesquisa. Se compartilhar todos ou parte dos seus dados de pesquisa, continue:

44. Que tipo de dados compartilha:

Dados brutos ou não processados
Dados processados
Dados publicados

45. A quem você concede acesso aos seus dados de pesquisa?

Selecione as opções que se aplicam.

Disponível para todos

Disponível para o meu grupo de pesquisa

Disponível para a faculdade ou departamento onde trabalho

Disponível para a instituição onde trabalho

Disponível para a rede nacional/internacional a que pertença

Disponível mediante solicitação direta

Outro, especifique

46. Através de que meios podem outros investigadores, comunidade acadêmica, etc., acessar os seus dados de pesquisa?

Selecione as opções que se aplicam.

Solicitação direta

Repositório/ base de dados do meu grupo de pesquisa

Repositório / centro / base de dados institucional

Repositório / centro / base de dados nacional

Repositório / centro / base de dados de revistas científicas

Repositório / centro / base de dados internacional

Repositório/ centro / base de dados de associações científicas

Repositório/ centro / base de dados temáticas

Outro, especifique:

➤ Em relação à qualidade dos dados de pesquisa compartilhados:

ITENS	Nunca (1)	Raramente (2)	Ocasionalmente (3)	Frequentemente (4)	Sempre (5)
47. Os dados de pesquisa compartilhados foram gerenciados com base em diretrizes propostas por instituições, associações científicas, repositórios, centros de dados, etc.					
48. Os dados da pesquisa são compartilhados em bruto, sem estruturação ou documentação anexada.					

➤ Selecione sob quais condições você compartilha seus dados de pesquisa:

ITENS	Nunca (1)	Raramente (2)	Às Vezes (3)	Frequentemente (4)	Sempre (5)
49. Exige ser citado em toda a produção acadêmica que usa seus dados					
50. Exige co-autoria em toda a produção acadêmica que usa seus dados					
51. Exige a oportunidade de se envolver no projeto de pesquisa que usa seus dados					
52. Exige revisar e fazer sugestões aos resultados da pesquisa que usa seus dados					
53. Exige receber toda a produção acadêmica (artigos, livros, apresentações, materiais etc.) que usa seus dados					
54. Exige reciprocidade na troca de dados					
55. Define as condições de acordo com as especificidades de cada caso.					
56. Não há condições específicas					

57. Selecione o tipo de licença que você usa ao compartilhar seus dados de pesquisa:

Licenças Creative Commons (CC)
 Licenças do Open Data Commons (ODC)
 Acordos de cooperação informal
 Acordos de cooperação formal
 Nenhuma
 Outro, especifique:

IV. REÚSO DE DADOS DE PESQUISA:

A continuação, enumera-se o termo “reúso de dados” com o propósito de que o participante conheça o significado que se assigna e responda a seguinte seção:

“Uso dos dados para propósitos diferentes do qual eles foram coletados, geralmente por outros pesquisadores que não os autores dos dados” (SAYÃO & SALES, 2015, p.82).

➤ Em relação ao uso e reúso de dados de pesquisa, afirma que:

ITENS	Nunca (1)	Raramente (2)	Às Vezes (3)	Frequentemente (4)	Sempre (5)
58. Reusa os dados da sua pesquisa anterior para gerar outros projetos de pesquisa.					
59. Reusa dados de pesquisa gerados por outros pesquisadores, instituições, centros, etc., em suas pesquisas.					
60. Os dados de pesquisa reusados possuem metadados.					

Se reusa os dados de pesquisa gerados por outros pesquisadores, instituições, observatórios, centros, associações científicas, etc., responda à seguinte pergunta. Se não reusa, evite responder e termine o questionário.

61. Selecione os meios de acesso aos dados de pesquisa de outros pesquisadores que você usa em suas pesquisas:

Solicito diretamente ao pesquisador
 Repositórios / centros / base de dados do meu grupo de pesquisa
 Repositórios / centros / base de dados institucionais
 Repositórios / centros / base de dados nacionais
 Repositórios / centros / base de dados de revistas científicas
 Repositórios / centros / base de dados internacionais
 Repositórios / centros / base de dados de associações científicas
 Repositórios / centros / base de dados temáticas
 Outro, especifique:

Agradecemos a sua participação!

APÊNDICE B – MATRIZ DE CONSISTÊNCIA METODOLÓGICA

Problema	Objetivos	Operacionalização			
		Variáveis	Dimensões	Indicadores	Metodologia
Como os pesquisadores, pertencentes a grupos de	Revelar as práticas relacionadas à gestão,	1 Perfil do pesquisador	1.1 Trajetória profissional.	42 indicadores	A pesquisa adota uma abordagem quantitativa, de

Problema	Objetivos	Operacionalização			
		Variáveis	Dimensões	Indicadores	Metodologia
pesquisa classificados pelas Colciencias na área de economia e negócios, praticam a questão da gestão, compartilhamento e reúso de dados de pesquisa?	compartilhamento e reúso de dados de pesquisa, de pesquisadores vinculados a grupos de pesquisa classificados por Colciencias na área de economia e negócios. Caracterizar o perfil dos pesquisadores da área de economia e negócios, localizados na cidade de Bucaramanga e sua região metropolitana; Descrever as particularidades dos dados produzidos pelos pesquisadores da área de economia e negócios, localizados na cidade de Bucaramanga e sua região metropolitana Caracterizar as práticas associadas à gestão, compartilhamento e reúso de dados de pesquisa por pesquisadores da área de economia e negócios, localizados na cidade de Bucaramanga e sua região metropolitana	2 Dados de pesquisa 3 Gestão de dados de pesquisa 4 Compartilhamento de dados de pesquisa 5 Reúso de dados de pesquisa	Trajetória investigativa. 2.1 Particularidades dos dados 3.1 Conhecimento sobre gestão de dados Planejamento da gestão de dados. Metadados. Controle e Garantia da qualidade de dados. Preservação dos dados. 4.1 Disponibilidade dos dados de pesquisa. Qualidade dos dados de pesquisa. Políticas de compartilhamento e reúso de dados de pesquisa 5.1 Reúso de dados próprios e de outros pesquisadores. Qualidade dos dados Meios de acesso ao conjunto de dados de outros pesquisadores.		natureza exploratória e descritiva, sob o procedimento de pesquisa tipo survey. A análise dos dados se realiza a partir do método estatístico descritivo, sendo interpretado sob o arcabouço teórico estabelecido.

APÊNDICE C – FORMATO MATRIZ DE OPERACIONALIZAÇÃO DE VARIÁVEIS

Variável	Definição		Dimensões	Indicadores	Instrumento Questões	Tipo de variáveis	Escala de medida
	Conceitual	Operacional					

APÊNDICE D- CARTA DE CONVITE A EXPERTOS

CARTA DE PRESENTACIÓN

Profesor (a):

Asunto: Validación de instrumentos a través de juicio de experto.

Nos es muy grato comunicarnos con usted para expresarle nuestros saludos y así mismo, hacer de su conocimiento que siendo estudiante del programa de Maestría en GESTIÓN DE LAS ORGANIZACIONES APRENDIENTES en el CENTRO DE EDUCACIÓN de la Universidad Federal de Paraíba, bajo la orientación del Prof. Dr. GUILHERME ATAÍDE DIAS, requiero validar los instrumentos con los cuales se recogerá la información necesaria para poder desarrollar nuestra investigación y con el cual se optará el grado de Magíster.

El título nombre del proyecto de investigación es: Prácticas de gestión, intercambio y reúso de datos científicos entre investigadores colombianos en el área de economía y negocios, y siendo imprescindible contar con la aprobación de especialistas para poder aplicar los instrumentos en mención, hemos considerado conveniente recurrir a usted, ante su connotada experiencia en temas de gestión de datos.

El expediente de validación, que le hacemos llegar contiene:

- Carta de presentación
- Resumen de la investigación.
- Matriz de consistencia metodológica.
- Definiciones conceptuales de las variables y dimensiones
- Matriz de operacionalización de las variables.
- Certificado de validez de contenido de los instrumentos.

Expresándole mis sentimientos de respeto y consideración me despido de usted, no sin antes agradecerle por la atención que dispense a la presente.

Atentamente,
Maria Paula Fernández Jiménez
Cedula 1.095.817.90 de Floridablanca, Santander

APÉNDICE E- FORMATO DE CERTIFICADO DE VALIDADE DE CONTEÚDO

NOMBRES Y APELLIDOS DEL JUEZ:

FORMACION ACADEMICA

AREAS DE EXPERIENCIA

TIEMPO

CARGO ACTUAL

INSTITUCION

De acuerdo con los siguientes indicadores califique cada uno de los ítems según corresponda.

CATEGORÍA	CALIFICACIÓN	INDICADOR
Suficiencia Los ítems que pertenecen a una misma dimensión bastan para obtener la medición de ésta	5. No cumple con el criterio 6. Bajo nivel 7. Moderado nivel 8. Alto nivel	5. Los ítems no son suficientes para medir la dimensión 6. Los ítems miden algún aspecto de la dimensión pero no corresponden con la dimensión total 7. Se deben incrementar algunos ítems para poder evaluar la dimensión complemente. 8. Los ítems son suficientes
Claridad El ítem se comprende fácilmente, es decir, su sintáctica y semántica son adecuadas	5. No cumple con el criterio 6. Bajo nivel 7. Moderado nivel 8. Alto nivel	5. El ítem no es claro 6. El ítem requiere bastantes modificaciones o una modificación muy grande en el uso de las palabras de acuerdo con su significado o por la ordenación de las mismas 7. Se requiere una modificación muy específica de algunos de los términos de ítem 8. El ítem es claro, tiene semántica y sintaxis adecuada.
Coherencia El ítem tiene relación lógica con la dimensión o indicador que está midiendo.	5. No cumple con el criterio 6. Bajo nivel 7. Moderado nivel 8. Alto nivel	5. El ítem no tiene relación lógica con la dimensión 6. El ítem tiene una relación tangencial con la dimensión 7. El ítem tiene una relación moderada con la dimensión que está midiendo. 8. El ítem se encuentra completamente relacionado con la dimensión que está midiendo.
Relevancia El ítem es esencial o importante, es decir debe ser incluido	5. No cumple con el criterio 6. Bajo nivel 7. Moderado nivel 8. Alto nivel	5. El ítem puede ser eliminado sin que se va afectada la medición de la dimensión. 6. El ítem tiene alguna relevancia, pero otro ítem puede estar incluyendo lo que mide éste. 7. El ítem es relativamente importante. 8. El ítem es muy relevante y debe ser incluido.

Fuente: (Escobar & Cuervo, 2008)

Variable:

Dimensión:

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBSERVACIONES

SUFICIENCIA				

Fuente: (Escobar & Cuervo, 2008)

ANEXOS

ANEXO A- AVALIAÇÃO DO INSTRUMENTO POR EXPERTOS

Nombres y apellidos del juez: Eduardo Carrillo Zambrano

Formación académica: posdoctorado

Áreas de experiencia: profesional tic y emprendimiento

Tiempo: 16 años

Cargo actual: profesor titular

Institución: UNAB

Variable: perfil de investigador

Dimensión: Trayectoria profesional

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
1 Seleccione la institución a la cual está vinculado (a)	4	4	4	
3 ¿Cuál es su nivel de formación académica?	4	4	4	
4 ¿Cuánto tiempo lleva ejerciendo como profesor (a) universitario (a)?	4	3	4	
SUFICIENCIA	4			

Dimensión: Trayectoria investigativa

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
2 Seleccione el grupo de investigación al que se encuentra vinculado (a)	4	4	4	
5 Seleccione su actual categoría como investigador reconocido por Colciencias	4	4	4	
6 ¿Cuánto tiempo lleva realizando actividades de investigación?	4	4	4	
7 Seleccione las líneas de investigación a las que se encuentran vinculados (as) sus investigaciones	4	4	4	
8 ¿Cuántas horas por semana dedica a sus investigaciones?	4	3	4	
SUFICIENCIA	4			

Variable: datos de investigación

Dimensión: Particularidades de los datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
9 ¿De dónde provienen los datos de investigación que produce?	4	4	4	
10 Seleccione la principal característica de sus datos de investigación	4	4	4	
11 ¿Qué tipo de datos produce en sus investigaciones?	4	4	4	
SUFICIENCIA	4			

Dimensión: Investigadores involucrados en la recolección de datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
12 Las investigaciones, específicamente la recolección o producción de datos, se llevan a cabo principalmente	4	3	3	
SUFICIENCIA	4			

Variable: gestión de datos de investigación

Dimensión: conocimiento sobre gestión de datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
13 ¿Cuál es su familiaridad con las prácticas de gestión de datos de investigación?	4	4	4	
14 ¿Ha realizado cursos o entrenamientos sobre gestión de datos de investigación?	4	4	4	
SUFICIENCIA	3			

Dimensión: asistencia para la gestión de datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
15 ¿En la institución donde labora existen servicios que asistan la gestión de datos de investigación?	4	3	4	
17 ¿Utiliza servicios que asistan la gestión de datos de investigación?	4	3	4	
SUFICIENCIA	4			

Dimensión: gestión de los datos de investigación

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
16 ¿Cómo investigador, gestiona los datos científicos recolectados en sus investigaciones?	4	4	4	
SUFICIENCIA	4			

Dimensión: planeación de la gestión de datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
18 Planea antes de iniciar la investigación cómo se adquirirá, procesará, analizará almacenará, asegurará, verificará y controlará los datos.	4	4	4	
19 Planea la descripción semántica y técnica (metadatos) de los datos	4	4	4	
20 Planea que datos serán preservados, como será su preservación, bajo qué tipo de formatos, y donde se preservaran a mediano y largo plazo	4	4	3	Posiblemente se podría subdividir la pregunta por contener diversas posibles respuestas
21 Establece cuestiones sobre propiedad intelectual, normatividad sobre acceso, intercambio, reúso, y privacidad de los datos.	4	4	3	Posiblemente se podría subdividir la pregunta por contener diversas posibles respuestas
22 Determina la disponibilidad de los datos. Cuáles y donde serán compartidos	4	4	3	Revisar signos de interrogación. Podría ser más clara
23 Establece roles y responsabilidades para la gestión de los datos	4	4	3	
24 Crea el documento plan de gestión de datos de forma estándar, a partir de directrices propuestas por instituciones, asociaciones científicas, repositorios, centros de datos, etc. Contiene cómo los datos serán recolectados, procesados, descritos, almacenados, autoría, políticas de intercambio e reúso, determinar roles y responsabilidades, entre otros aspectos.	4	4	4	

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
25 Utiliza herramientas digitales que asisten la creación de un plan de gestión de datos	4	4	4	
26 Posee algún tipo de asistencia, por parte de un grupo especialista en datos, para realizar su plan de gestión de datos	4	3	3	En otra pregunta anterior se preguntaba si se tenía apoyo institucional, revisar para evitar redundancia
27 Planea el presupuesto para la gestión de los datos	4	4	4	
SUFICIENCIA	4			

Dimensión: recolecta, procesamiento y almacenamiento de los datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
28 Recolecta los datos de acuerdo a lo planeado	4	4	4	
29 Prepara los datos de acuerdo a lo planeado, de forma a detectar errores	4	4	3	Sugiero revisar redacción
30 Organiza el conjunto de datos de manera que sean fáciles de comprender, procesar y usar.	4	4	4	
31 Usa programas de software para recolecta, exploración, visualización, análisis de sus datos.	4	4	4	
32 Los formatos que almacenan los datos son legibles a mediano y largo plazo.	4	4	4	Me queda la duda de si plantear que el formato sea legible a largo plazo
33 Almacena los datos en formatos que son accesibles a diversas aplicaciones.	4	4	4	
34 Crea versiones digitales de los datos recolectados en físico (papel)	4	4	4	
35 Mantiene una copia de los datos brutos (no procesados).	4	4	4	
36 Sigue un modelo de ciclo de vida de datos o directrices estándares para la recolecta, procesamiento y almacenamiento de datos, propuesto por instituciones, asociaciones científicas, centros de datos, repositorios, etc.	4	4	4	
37 Posee algún tipo de asistencia por parte de un grupo especialista en datos para recolectar, procesar y almacenar los datos de investigación	4	4	4	
38 Seleccione los tipos de formatos en que almacena sus datos de investigación	4	4	4	
39 Tipo de almacenamiento que utiliza para archivar sus datos de investigación	4	4	4	
SUFICIENCIA	4			

Dimensión: metadatos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
40 Realiza descripción de los datos a nivel de investigación (contexto, metodología, procedencia de los datos, procesamiento, aseguramiento y control de la calidad, modificaciones hechas en los datos, etc.), a nivel de base de datos (tablas de base de datos,	4	4	4	

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
formatos), a nivel de variables (nombres, etiquetas, significado, descripción de variables, categorías, etc.), administrativos (técnico, preservación, derechos de autor), entre otros aspectos –metadatos-.				
41 Orienta la descripción de sus datos a partir de metadatos estándares propuestos por instituciones, asociaciones científicas, centros de datos, repositorios, etc.	4	4	4	
42 Utiliza herramientas digitales para crear metadatos.	4	4	4	
43 Posee algún tipo de asistencia por parte de un grupo especialista en datos para crear los metadatos.	4	3	4	Podría ser redundante o bastante relacionada con la pregunta 37
SUFICIENCIA	4			

Dimensión: aseguramiento de la calidad de los datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
44 Valida el instrumento de recolecta de datos	4	4	4	Duda si el termino apropiado es recolecta o recolección, en preguntas anteriores también se usa recolecta
45 Periódicamente inspecciona la existencia de errores en el conjunto de datos, durante y después de finalizar la investigación	4	4	4	
46 Periódicamente monitorea y asegura la calidad (datos correctos, exactos, fácil lectura, procesamiento, uso) del conjunto de datos durante y después de finalizar la investigación.	4	4	4	
47 Verifica si existe duplicación de datos	4	4	4	
48 Documenta alteraciones realizadas al conjunto de datos	4	4	3	No me resulta suficientemente claro lo relacionado con la alteración al conjunto de datos
49 Describe los datos (metadatos) e inspecciona que la descripción sea correcta y exacta	4	4	4	
50 Sigue un modelo de ciclo de vida de datos o directrices estándares que orientan la verificación, inspección y control de datos propuesto por instituciones, asociaciones científicas, centros de datos, repositorios, etc.	4	4	4	
51 Posee algún tipo de asistencia por parte de un grupo especialista en datos para garantizar la calidad de los datos de Investigación	4	4		
SUFICIENCIA	4			

Dimensión: preservación de los datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
52 Selecciona los datos que pueden ser eliminados y cuáles deben ser preservados a mediano y largo plazo.	4	4	4	
53 Lleva a cabo copias de seguridad en múltiples versiones	4	4	4	

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
54 Lleva a cabo copias de seguridad en múltiples formatos	4	4	4	
55 Lleva a cabo copias de seguridad en múltiples lugares	4	4	4	
56 Inspecciona periódicamente las copias de seguridad para confirmar su buen estado	4	4	4	
57 Verifica periódicamente la obsolescencia de los formatos digitales que almacenan los datos	4	4	4	
58 Realiza migración de datos de un formato digital a otro, conservando sus características, cuando es necesario	4	4	4	
59 Dispone los datos de investigación a través de repositorios, banco o centros de datos, etc. para preservarlos a mediano y largo plazo	4	4	4	
60 Orienta la preservación de los datos de investigación a través de un modelo de ciclo de vida de datos o de directrices estándares propuestas por instituciones, asociaciones científicas, centros de datos, repositorios, etc.	4	4	4	
61 Posee algún tipo de asistencia por parte de un grupo especialista en datos para preservar a mediano y largo plazo los datos de Investigación	4	4	4	
SUFICIENCIA	4			

Variables: intercambio de datos de investigación

Dimensiones: disponibilidad de datos de investigación

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
62 ¿Qué cantidad de datos de investigación comparte con investigadores, comunidad académica, instituciones, centros, asociaciones, redes etc.?	4	4	4	
63 ¿A quién concede el acceso a sus datos de investigación?	4	4	4	
64 ¿A través de que medio pueden acceder otros investigadores, comunidad académica, etc., a sus datos de investigación?	4	4	4	Se podría explorar la relación del encuestado con iniciativas de ciencia abierta
SUFICIENCIA	4			

Dimensión: calidad de los datos de investigación compartidos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
65 Los datos de investigación compartidos están estructurados y han sido gestionados de forma estándar a partir de un modelo de ciclo de vida de datos o directrices propuestas por instituciones, asociaciones científicas, repositorios, centros de datos, etc.	4	4	4	
66 Los datos de investigación compartidos están estructurados y han sido gestionados sin seguir un modelo de ciclo de vida de datos o directrices propuestas por instituciones, asociaciones científicas, repositorios, centros de datos, etc.	4	4	4	
67 Los datos de investigación son compartidos en bruto, sin estructuración o documentación adjunta.	4	3	3	Posiblemente se pueda inferir la

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
				respuesta en relación con las preguntas 65 y 66
SUFICIENCIA	3			

Dimensión: políticas de acceso e intercambio de datos de investigación

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
68 Reconocimiento como autor de los datos de investigación mediante citación de referencia en toda la producción académica que haga uso de ellos	4	4	4	
69 Reconocimiento como autor de los datos de investigación mediante coautoría en toda la producción académica que haga uso de ellos	4	4	4	
70 Oportunidad de involucrarse en un proyecto de investigación que utilice los datos de investigación que provee	4	4	4	
71 Retribución económica para cubrir parte de los costos de recolecta, procesamiento y preservación de los datos de investigación	4	4	4	
72 Aprobación de los resultados que estén sustentados en los datos de investigación que provee.	4	4	4	
73 Oportunidad de revisar y realizar sugerencias a los resultados de la investigación sustentados en los datos que provee.	4	4	4	
74 Recibir toda producción académica (artículos, libros, presentaciones, materiales, etc.) que haga uso de los datos de investigación que provee	4	4	4	
75 Acuerdo mutuo solamente de intercambio de datos de investigación entre el autor y el usuario.	4	4	4	
76 Establecimiento de las condiciones para acceso y reúso de datos de acuerdo a la consideración del autor	4	4	4	
77 No hay condiciones específicas para el acceso y reúso de datos	4	4	4	
78 Seleccione que tipo de licencia utiliza cuando comparte sus datos de investigación	4	4	4	
SUFICIENCIA	4			

Variable: reúso de datos de investigación

Dimensiones: reúso de datos propios y de otros investigadores

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
79 Reúsa los datos de sus anteriores investigaciones para generar otros proyectos de investigación	4	4	4	
80 Reúsa los datos de investigación generados por otros investigadores, instituciones, centros, asociaciones, etc., en sus investigaciones	4	4	4	
SUFICIENCIA	4			

Dimensión: conjuntos de datos formados a partir de datos propios y de otros investigadores

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
81 Fusiona múltiples conjuntos de datos de otros investigadores y de su propia autoría para sustentar los resultados de su investigación	4	4	4	

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
SUFICIENCIA	4			

Dimensión: medios para acceder a conjuntos de datos de otros investigadores

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
82 Seleccione los medios de acceso a datos de investigación de otros de investigadores que utiliza en sus investigaciones ⁴	4	4	4	
SUFICIENCIA	4			

Nombres y apellidos del juez: Paola Andrea Ramírez Pérez

Formación académica bibliotecóloga

Áreas de experiencia profesional desarrollo y evaluación de colecciones, documentación para el área biomédica, ciencia abierta, gestión de datos.

Tiempo 22 años

Cargo actual asesora y consultora independiente, documentalista de la unidad de evidencia y deliberación para la toma de decisiones de la facultad de medicina de la Universidad de Antioquia.

Institución Universidad de Antioquia

De acuerdo con los siguientes indicadores califique cada uno de los ítems según corresponda.

Variable: perfil de investigador

Dimensión: Trayectoria profesional

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
1 Seleccione la institución a la cual está vinculado (a)	4	4	4	
3 ¿Cuál es su nivel de formación académica?	4	4	4	
4 ¿Cuánto tiempo lleva ejerciendo como profesor (a) universitario (a)?	4	4	4	
SUFICIENCIA	4			

Dimensión: Trayectoria investigativa

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
2 Seleccione el grupo de investigación al que se encuentra vinculado (a)	4	4	4	
5 Seleccione su actual categoría como investigador reconocido por Colciencias	4	4	4	
6 ¿Cuánto tiempo lleva realizando actividades de investigación?	4	4	4	
7 Seleccione las líneas de investigación a las que se encuentran vinculados (as) sus investigaciones	4	4	4	
8 ¿Cuántas horas por semana dedica a sus investigaciones?	4	4	4	
SUFICIENCIA	4			

Variable: datos de investigación

Dimensión: Particularidades de los datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
------	------------	------------	----------	-----

9 ¿De dónde provienen los datos de investigación que produce?	4	4	4	Es necesario definir las categorías y opciones dadas para aplicar este ítem de forma clara. (Como en la pregunta 11)
10 Seleccione la principal característica de sus datos de investigación	4	4	3	
11 ¿Qué tipo de datos produce en sus investigaciones?	4	4	4	
SUFICIENCIA	4			Falta preguntar por la cantidad o volumen de los datos producidos en un periodo de tiempo definido.

Dimensión: Investigadores involucrados en la recolecta de datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
12 Las investigaciones, específicamente la recolección o producción de datos, se llevan a cabo principalmente	2	3	2	La redacción de la pregunta no es clara y las opciones ofrecidas de cantidad de personas, no tiene relevancia directa con la dimensión. Sugiero preguntar directamente ¿Quién recolecta los datos de investigación en los procesos...? Y ofrecer opciones de los roles en la investigación (Inv principal, co-investigadores, auxiliares, otros.)
SUFICIENCIA	4			Considerar que además de la cantidad es importante el perfil de quienes recolectan o analizan.

Variable: gestión de datos de investigación

Dimensión: conocimiento sobre gestión de datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
13 ¿Cuál es su familiaridad con las prácticas de gestión de datos de investigación?	3	4	3	
14 ¿Ha realizado cursos o entrenamientos sobre gestión de datos de investigación?	4	4	4	
SUFICIENCIA	4			

Dimensión: asistencia para la gestión de datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
15 ¿En la institución donde labora existen servicios que asistan la gestión de datos de investigación?	4	4	4	
17 ¿Utiliza servicios que asistan la gestión de datos de investigación?	4	4	4	
SUFICIENCIA	4			

Dimensión: gestión de los datos de investigación

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
16 ¿Cómo investigador, gestiona los datos científicos recolectados en sus investigaciones?	4	4	4	
SUFICIENCIA	4			

Dimensión: planeación de la gestión de datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
18 Planea antes de iniciar la investigación cómo se adquirirá, procesará, analizará almacenará, asegurará, verificará y controlará los datos.	4	4	4	
19 Planea la descripción semántica y técnica (metadatos) de los datos	4	4	4	
20 Planea que datos serán preservados, como será su preservación, bajo qué tipo de formatos, y donde se preservaran a mediano y largo plazo	4	4	4	
21 Establece cuestiones sobre propiedad intelectual, normatividad sobre acceso, intercambio, reúso, y privacidad de los datos.	4	4	2	Establecer cuestiones no es claro. Sugiero: Establece criterios de propiedad intelectual sobre...
22 Determina la disponibilidad de los datos. Cuáles y donde serán compartidos	4	4	4	
23 Establece roles y responsabilidades para la gestión de los datos	4	4	4	
24 Crea el documento plan de gestión de datos de forma estándar, a partir de directrices propuestas por instituciones, asociaciones científicas, repositorios, centros de datos, etc. Contiene cómo los datos serán recolectados, procesados, descritos, almacenados, autoría, políticas de intercambio e reúso, determinar roles y responsabilidades, entre otros aspectos.	4	4	4	
25 Utiliza herramientas digitales que asisten la creación de un plan de gestión de datos	4	4	4	
26 Posee algún tipo de asistencia, por parte de un grupo especialista en datos, para realizar su plan de gestión de datos	4	4	4	
27 Planea el presupuesto para la gestión de los datos	4	4	4	
SUFICIENCIA	4			

Dimensión: recolecta, procesamiento y almacenamiento de los datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
28 Recolecta los datos de acuerdo a lo planeado	4	4	4	
29 Prepara los datos de acuerdo a lo planeado, de forma a detectar errores	3	4	2	Sugiero: Con la posibilidad de identificar y corregir errores
30 Organiza el conjunto de datos de manera que sean fáciles de comprender, procesar y usar.	4	4	4	
31 Usa programas de software para recolecta, exploración, visualización, análisis de sus datos.	4	4	4	Aquí y en muchas otras sugiero cambiar recolecta por recolección ya que la encuesta está dirigida a hispanohablantes.

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
32 Los formatos que almacenan los datos son legibles a mediano y largo plazo.	4	4	4	
33 Almacena los datos en formatos que son accesibles a diversas aplicaciones.	4	4	4	
34 Crea versiones digitales de los datos recolectados en físico (papel)	4	4	4	
35 Mantiene una copia de los datos brutos (no procesados).	4	4	4	
36 Sigue un modelo de ciclo de vida de datos o directrices estándares para la recolecta, procesamiento y almacenamiento de datos, propuesto por instituciones, asociaciones científicas, centros de datos, repositorios, etc.	4	4	4	
37 Posee algún tipo de asistencia por parte de un grupo especialista en datos para recolectar, procesar y almacenar los datos de investigación	4	4	4	
38 Seleccione los tipos de formatos en que almacena sus datos de investigación	4	4	4	
39 Tipo de almacenamiento que utiliza para archivar sus datos de investigación	4	4	4	Sugiero cambiar archivar por almacenar.
SUFICIENCIA	4			

Dimensión: metadatos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
40 Realiza descripción de los datos a nivel de investigación (contexto, metodología, procedencia de los datos, procesamiento, aseguramiento y control de la calidad, modificaciones hechas en los datos, etc.), a nivel de base de datos (tablas de base de datos, formatos), a nivel de variables (nombres, etiquetas, significado, descripción de variables, categorías, etc.), administrativos (técnico, preservación, derechos de autor), entre otros aspectos –metadatos-.	2	4	2	Ameritaría dividirse en cuatro preguntas dirigidas a: investigación, base de datos, variables y administración. O en una sola pregunta con esas opciones de respuesta.
41 Orienta la descripción de sus datos a partir de metadatos estándares propuestos por instituciones, asociaciones científicas, centros de datos, repositorios, etc.	4	4	4	
42 Utiliza herramientas digitales para crear metadatos.	4	4	4	
43 Posee algún tipo de asistencia por parte de un grupo especialista en datos para crear los metadatos.	4	4	4	
SUFICIENCIA	4			

Dimensión: aseguramiento de la calidad de los datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
44 Valida el instrumento de recolecta de datos	4	4	4	Sugiero: incluir o usa uno que ya está validado.
45 Periódicamente inspecciona la existencia de errores en el conjunto de datos, durante y después de finalizar la investigación	4	4	4	En esta pregunta y la siguiente ¿Cómo diferencias la existencia de errores de la calidad?
46 Periódicamente monitorea y asegura la calidad (datos correctos, exactos, fácil lectura, procesamiento, uso) del conjunto de datos durante y después de finalizar la investigación.	4	4	4	¿Cómo diferencias la existencia de errores de la calidad? Si dejas las dos opciones debes indicar qué es calidad y qué es la existencia de errores.
47 Verifica si existe duplicación de datos	4	4	4	También estaría incluido en la calidad.
48 Documenta alteraciones realizadas al conjunto de datos	4	4	4	También estaría incluido en la calidad.
49 Describe los datos (metadatos) e inspecciona que la descripción sea correcta y exacta	4	4	4	También estaría incluido en la calidad.
50 Sigue un modelo de ciclo de vida de datos o directrices estándares que orientan la verificación, inspección y control de datos propuesto por instituciones, asociaciones científicas, centros de datos, repositorios, etc.	4	4	4	
51 Posee algún tipo de asistencia por parte de un grupo especialista en datos para garantizar la calidad de los datos de investigación	4	4	4	
SUFICIENCIA	4			

Dimensión: preservación de los datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
52 Selecciona los datos que pueden ser eliminados y cuáles deben ser preservados a mediano y largo plazo.	4	4	4	
53 Lleva a cabo copias de seguridad en múltiples versiones	4	4	4	Sugiero una sola pregunta con múltiples opciones para este conjunto de ítems.
54 Lleva a cabo copias de seguridad en múltiples formatos	4	4	4	
55 Lleva a cabo copias de seguridad en múltiples lugares	4	4	4	
56 Inspecciona periódicamente las copias de seguridad para confirmar su buen estado	4	4	4	
57 Verifica periódicamente la obsolescencia de los formatos digitales que almacenan los datos	4	4	4	
58 Realiza migración de datos de un formato digital a otro, conservando sus características, cuando es necesario	4	4	4	
59 Dispone los datos de investigación a través de repositorios, banco o centros de datos, etc. para preservarlos a mediano y largo plazo	4	4	4	
60 Orienta la preservación de los datos de investigación a través de un modelo de ciclo de vida de datos o de directrices estándares propuestas por instituciones, asociaciones científicas, centros de datos, repositorios, etc.	4	4	4	

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
61 Posee algún tipo de asistencia por parte de un grupo especialista en datos para preservar a mediano y largo plazo los datos de investigación	4	4	4	
SUFICIENCIA	3			Sugiero agregar un ítem sobre dónde está almacenando los datos de investigación (nube, discos, servidor, etc.).

Variables: intercambio de datos de investigación

Dimensiones: disponibilidad de datos de investigación

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
62 ¿Qué cantidad de datos de investigación comparte con investigadores, comunidad académica, instituciones, centros, asociaciones, redes etc.?	4	4	4	
63 ¿A quién concede el acceso a sus datos de investigación?	4	4	4	
64 ¿A través de que medio pueden acceder otros investigadores, comunidad académica, etc., a sus datos de investigación?	4	4	4	
SUFICIENCIA	3			Falta preguntar por qué tipo de datos comparte (resultados, raw data, etc.

Dimensión: calidad de los datos de investigación compartidos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
65 Los datos de investigación compartidos están estructurados y han sido gestionados de forma estándar a partir de un modelo de ciclo de vida de datos o directrices propuestas por instituciones, asociaciones científicas, repositorios, centros de datos, etc.	4	4	4	
66 Los datos de investigación compartidos están estructurados y han sido gestionados sin seguir un modelo de ciclo de vida de datos o directrices propuestas por instituciones, asociaciones científicas, repositorios, centros de datos, etc.	2	2	2	Esta pregunta es innecesaria considerando lo que se preguntó en la categoría gestión.
67 Los datos de investigación son compartidos en bruto, sin estructuración o documentación adjunta.	4	4	4	
SUFICIENCIA	4			

Dimensión: políticas de acceso e intercambio de datos de investigación

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
68 Reconocimiento como autor de los datos de investigación mediante citación de referencia en toda la producción académica que haga uso de ellos	4	4	2	Sugiero mejorar la redacción e iniciar la redacción de todas
69 Reconocimiento como autor de los datos de investigación mediante coautoría en toda la producción académica que haga uso de ellos	4	4	2	
70 Oportunidad de involucrarse en un proyecto de investigación que utilice los datos de investigación que provee	4	4	2	

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
71 Retribución económica para cubrir parte de los costos de recolecta, procesamiento y preservación de los datos de investigación	4	4	2	estas preguntas con ¿Exige la...?
72 Aprobación de los resultados que estén sustentados en los datos de investigación que provee.	4	4	2	
73 Oportunidad de revisar y realizar sugerencias a los resultados de la investigación sustentados en los datos que provee.	4	4	2	
74 Recibir toda producción académica (artículos, libros, presentaciones, materiales, etc.) que haga uso de los datos de investigación que provee	4	4	2	
75 Acuerdo mutuo solamente de intercambio de datos de investigación entre el autor y el usuario.	4	4	2	
76 Establecimiento de las condiciones para acceso y reúso de datos de acuerdo a la consideración del autor	4	4	2	
77 No hay condiciones específicas para el acceso y reúso de datos	4	4	2	
78 Seleccione que tipo de licencia utiliza cuando comparte sus datos de investigación	4	4	2	
SUFICIENCIA	4			

Variable: reúso de datos de investigación

Dimensiones: reúso de datos propios y de otros investigadores

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBSERVACIONES
79 Reúsa los datos de sus anteriores investigaciones para generar otros proyectos de investigación	4	4	4	
80 Reúsa los datos de investigación generados por otros investigadores, instituciones, centros, asociaciones, etc., en sus investigaciones	4	4	4	
SUFICIENCIA	3			Sugiero indagar por dónde (red de colegas, repositorios, etc.) y cómo (intercambio personal, búsqueda selectiva, etc.) obtiene los datos que reúsa.

Dimensión: conjuntos de datos formados a partir de datos propios y de otros investigadores

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
81 Fusiona múltiples conjuntos de datos de otros investigadores y de su propia autoría para sustentar los resultados de su investigación	2	2	2	Esta pregunta no agrega mucho al instrumento. Preguntar esto implicaría un desconocimiento del proceso mismo de investigación.
SUFICIENCIA	2			

Dimensión: medios para acceder a conjuntos de datos de otros investigadores

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
82 Seleccione los medios de acceso a datos de investigación de otros de investigadores que utiliza en sus investigaciones	4	4	4	Leer observación de pregunta 80
SUFICIENCIA	4			

Nombres y apellidos del juez: _william humberto zamudio peña

Formación académica economista especialista en finanzas privadas y maestría en administración de empresas con énfasis en formulación de proyectos.

Áreas de experiencia profesional:

Tiempo: 30 años

Cargo actual director del centro de formación en tecnologías

Institución Universidad de Santander UDES

Variable: perfil de investigador

Dimensión: trayectoria profesional

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
1 Seleccione la institución a la cual está vinculado (a)	4	4	4	
3 ¿Cuál es su nivel de formación académica?	4	4	4	
4 ¿Cuánto tiempo lleva ejerciendo como profesor (a) universitario (a)?	4	4	4	
SUFICIENCIA	4			

Dimensión: Trayectoria investigativa

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
2 Seleccione el grupo de investigación al que se encuentra vinculado (a)	4	4	4	
5 Seleccione su actual categoría como investigador reconocido por Colciencias	4	4	4	
6 ¿Cuánto tiempo lleva realizando actividades de investigación?	4	4	4	
7 Seleccione las líneas de investigación a las que se encuentran vinculados (as) sus investigaciones	3	3	3	Incluye demasiados ítems para seleccionar, debería agregar en ítems más gruesos y mantener la opción de otro cual.
8 ¿Cuántas horas por semana dedica a sus investigaciones?	4	4	4	
SUFICIENCIA	4			

Variable: datos de investigación

Dimensión: Particularidades de los datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
9 ¿De dónde provienen los datos de investigación que produce?	4	4	4	
10 Seleccione la principal característica de sus datos de investigación	4	4	4	
11 ¿Qué tipo de datos produce en sus investigaciones?	4	4	4	
12 Seleccione los tipos de formatos en que almacena sus datos de investigación	4	4	4	
SUFICIENCIA	4			

Dimensión: Investigadores involucrados en la recolecta de datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
13 Las investigaciones, específicamente la recolección o producción de datos, se llevan a cabo principalmente	4	4	4	
SUFICIENCIA	4			

Variable: gestión de datos de investigación

Dimensión: conocimiento sobre gestión de datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
14 ¿Cuál es su familiaridad con las prácticas de gestión de datos de investigación?	4	4	4	
15 ¿Ha realizado cursos o entrenamientos sobre gestión de datos de investigación?	4	4	4	
SUFICIENCIA	4			

Dimensión: asistencia para la gestión de datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
16 ¿En la institución donde labora existen servicios que asistan la gestión de datos de investigación?	4	4	4	
18 ¿Utiliza servicios que asistan la gestión de datos de investigación?	4	4	4	
SUFICIENCIA	4			

Dimensión: gestión de los datos de investigación

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
17 ¿Cómo investigador, gestiona los datos científicos recolectados en sus investigaciones?	4	4	4	
SUFICIENCIA	4			

Dimensión: planeación de la gestión de datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
19 Planea antes de iniciar la investigación cómo se adquirirá, procesará, analizará almacenará, asegurará, verificará y controlará los datos.	4	4	4	
20 Planea la descripción semántica y técnica (metadatos) de los datos	4	4	4	
21 Planea que datos serán preservados, como será su preservación, bajo qué tipo de formatos, y donde se preservaran a mediano y largo plazo	4	4	4	
22 Establece cuestiones sobre propiedad intelectual, normatividad sobre acceso, intercambio, reúso, y privacidad de los datos.	4	4	4	
23 Determina la disponibilidad de los datos. Cuáles y donde serán compartidos	4	4	4	
24 Establece roles y responsabilidades para la gestión de los datos	4	4	4	
25 Crea el documento plan de gestión de datos de forma estándar, a partir de directrices propuestas por instituciones, asociaciones científicas, repositorios, centros de datos, etc. Contiene cómo los datos serán recolectados, procesados, descritos, almacenados, autoría, políticas de intercambio e reúso, determinar roles y responsabilidades, entre otros aspectos.	4	4	4	
26 Utiliza herramientas digitales que asisten la creación de un plan de gestión de datos	4	4	4	

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
27 Posee algún tipo de asistencia, por parte de un grupo especialista en datos, para realizar su plan de gestión de datos	4	4	4	
28 Planea el presupuesto para la gestión de los datos	4	4	4	
SUFICIENCIA	4			

Dimensión: recolecta, procesamiento y almacenamiento de los datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
29 Recolecta los datos de acuerdo a lo planeado	4	4	4	
30 Prepara los datos de acuerdo a lo planeado, de forma a detectar errores	4	4	4	
31 Organiza el conjunto de datos de manera que sean fáciles de comprender, procesar y usar.	4	4	4	
32 Usa programas de software para recolecta, exploración, visualización, análisis de sus datos.	4	4	4	
33 Los formatos que almacenan los datos son legibles a mediano y largo plazo.	4	4	4	
34 Almacena los datos en formatos que son accesibles a diversas aplicaciones.	4	4	4	
35 Crea versiones digitales de los datos recolectados en físico (papel)	4	4	4	
36 Mantiene una copia de los datos brutos (no procesados).	4	4	4	
37 Sigue un modelo de ciclo de vida de datos o directrices estándares para la recolecta, procesamiento y almacenamiento de datos, propuesto por instituciones, asociaciones científicas, centros de datos, repositorios, etc.	4	4	4	
38 Posee algún tipo de asistencia por parte de un grupo especialista en datos para recolectar, procesar y almacenar los datos de investigación	4	4	4	
39 Seleccione el tipo de almacenamiento que utiliza para archivar sus datos de investigación	4	4	4	
SUFICIENCIA	4			

Dimensión: metadatos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
40 Realiza descripción de los datos a nivel de investigación (contexto, metodología, procedencia de los datos, procesamiento, aseguramiento y control de la calidad, modificaciones hechas en los datos, etc.), a nivel de base de datos (tablas de base de datos, formatos), a nivel de variables (nombres, etiquetas, significado, descripción de variables, categorías, etc.), administrativos (técnico, preservación, derechos de autor), entre otros aspectos –metadatos-.	4	4	4	
41 Orienta la descripción de sus datos a partir de metadatos estándares propuestos por instituciones, asociaciones científicas, centros de datos, repositorios, etc.	4	4	4	
42 Utiliza herramientas digitales para crear metadatos.	4	4	4	
43 Posee algún tipo de asistencia por parte de un grupo especialista en datos para crear los metadatos.	4	4	4	
SUFICIENCIA	4			

Dimensión: aseguramiento de la calidad de los datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
44 Valida el instrumento de recolecta de datos	4	4	4	
45 Periódicamente inspecciona la existencia de errores en el conjunto de datos, durante y después de finalizar la investigación	4	4	4	
46 Periódicamente monitorea y asegura la calidad (datos correctos, exactos, fácil lectura, procesamiento, uso) del conjunto de datos durante y después de finalizar la investigación.	4	4	4	
47 Verifica si existe duplicación de datos	4	4	4	
48 Documenta alteraciones realizadas al conjunto de datos	4	4	4	
49 Describe los datos (metadatos) e inspecciona que la descripción sea correcta y exacta	4	4	4	
50 Sigue un modelo de ciclo de vida de datos o directrices estándares que orientan la verificación, inspección y control de datos propuesto por instituciones, asociaciones científicas, centros de datos, repositorios, etc.	4	4	4	
51 Posee algún tipo de asistencia por parte de un grupo especialista en datos para garantizar la calidad de los datos de investigación	4	4	4	
SUFICIENCIA	4			

Dimensión: preservación de los datos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
52 Selecciona los datos que pueden ser eliminados y cuáles deben ser preservados a mediano y largo plazo.	4	4	4	
53 Lleva a cabo copias de seguridad en múltiples versiones	4	4	4	
54 Lleva a cabo copias de seguridad en múltiples formatos	4	4	4	
55 Lleva a cabo copias de seguridad en múltiples lugares	4	4	4	
56 Inspecciona periódicamente las copias de seguridad para confirmar su buen estado	4	4	4	
57 Verifica periódicamente la obsolescencia de los formatos digitales que almacenan los datos	4	4	4	
58 Realiza migración de datos de un formato digital a otro, conservando sus características, cuando es necesario	4	4	4	
59 Dispone los datos de investigación a través de repositorios, banco o centros de datos, etc. para preservarlos a mediano y largo plazo	4	4	4	
60 Orienta la preservación de los datos de investigación a través de un modelo de ciclo de vida de datos o de directrices estándares propuestas por instituciones, asociaciones científicas, centros de datos, repositorios, etc.	4	4	4	
61 Posee algún tipo de asistencia por parte de un grupo especialista en datos para preservar a mediano y largo plazo los datos de investigación	4	4	4	
SUFICIENCIA	4			

Variables: intercambio de datos de investigación

Dimensiones: disponibilidad de datos de investigación

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
62 ¿Qué cantidad de datos de investigación comparte con investigadores, comunidad académica, instituciones, centros, asociaciones, redes etc.?	4	4	4	
63 ¿A quién concede el acceso a sus datos de investigación?	4	4	4	

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
64 ¿A través de que medio pueden acceder otros investigadores, comunidad académica, etc., a sus datos de investigación?	4	4	4	
SUFICIENCIA	4			

Dimensión: calidad de los datos de investigación compartidos

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
65 Los datos de investigación compartidos están estructurados y han sido gestionados de forma estándar a partir de un modelo de ciclo de vida de datos o directrices propuestas por instituciones, asociaciones científicas, repositorios, centros de datos, etc.	4	4	4	
66 Los datos de investigación compartidos están estructurados y han sido gestionados sin seguir un modelo de ciclo de vida de datos o directrices propuestas por instituciones, asociaciones científicas, repositorios, centros de datos, etc.	4	4	4	
67 Los datos de investigación son compartidos en bruto, sin estructuración o documentación adjunta.	4	4	4	
SUFICIENCIA	4			

Dimensión: políticas de acceso e intercambio de datos de investigación

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
68 Reconocimiento como autor de los datos de investigación mediante citación de referencia en toda la producción académica que haga uso de ellos	4	4	4	
69 Reconocimiento como autor de los datos de investigación mediante coautoría en toda la producción académica que haga uso de ellos	4	4	4	
70 Oportunidad de involucrarse en un proyecto de investigación que utilice los datos de investigación que provee	4	4	4	
71 Retribución económica para cubrir parte de los costos de recolecta, procesamiento y preservación de los datos de investigación	4	4	4	
72 Aprobación de los resultados que estén sustentados en los datos de investigación que provee.	4	4	4	
73 Oportunidad de revisar y realizar sugerencias a los resultados de la investigación sustentados en los datos que provee.	4	4	4	
74 Recibir toda producción académica (artículos, libros, presentaciones, materiales, etc.) que haga uso de los datos de investigación que provee	4	4	4	
75 Acuerdo mutuo solamente de intercambio de datos de investigación entre el autor y el usuario.	4	4	4	
76 Establecimiento de las condiciones para acceso y reúso de datos de acuerdo a la consideración del autor	4	4	4	
77 No hay condiciones específicas para el acceso y reúso de datos	4	4	4	
78 Seleccione que tipo de licencia utiliza cuando comparte sus datos de investigación	4	4	4	
SUFICIENCIA	4			

Variable: reúso de datos de investigación

Dimensiones: reúso de datos propios y de otros investigadores

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
79 Reúsa los datos de sus anteriores investigaciones para generar otros proyectos de investigación	4	4	4	
80 Reúsa los datos de investigación generados por otros investigadores, instituciones, centros, asociaciones, etc., en sus investigaciones	4	4	4	
SUFICIENCIA	4			

Dimensión: conjuntos de datos formados a partir de datos propios y de otros investigadores

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
81 Fusiona múltiples conjuntos de datos de otros investigadores y de su propia autoría para sustentar los resultados de su investigación	4	4	4	
SUFICIENCIA	4			

Dimensión: medios para acceder a conjuntos de datos de otros investigadores

ÍTEM	COHERENCIA	RELEVANCIA	CLARIDAD	OBS
82 Seleccione los medios de acceso a datos de investigación de otros de investigadores que utiliza en sus investigaciones	4	4	4	
SUFICIENCIA	4			

ANEXO B- TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (Elaborado de acordo com a Resolução 466/13 e 510/17 do CNS)

O (A) Sr.(a) está sendo convidado (a) a participar da pesquisa intitulada: **“Práticas de gestão, compartilhamento e reúso de dados de pesquisa entre pesquisadores colombianos na área de economia e negócios”**, desenvolvida por **MARIA PAULA FERNÁNDEZ JIMÉNEZ**, aluno regularmente matriculado no **PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO NAS ORGANIZAÇÕES APRENDENTES** do **CENTRO DE EDUCAÇÃO** da Universidade Federal da Paraíba, sob a orientação do **Prof. Dr. GUILHERME ATAÍDE DIAS**, nesta instituição.

Os objetivos da pesquisa são:

Objetivo geral: revelar as práticas relacionadas à gestão, compartilhamento e reúso de dados de pesquisa, de pesquisadores vinculados a grupos de pesquisa classificados por Colciencias na área de economia e negócios.

Objetivos Específicos:

Caracterizar o perfil dos pesquisadores da área de economia e negócios, localizados na cidade de Bucaramanga e sua região metropolitana.

Descrever as particularidades dos dados produzidos pelos pesquisadores da área de economia e negócios, localizados na cidade de Bucaramanga e sua região metropolitana.

Caracterizar as práticas associadas à gestão, compartilhamento e reúso de dados de pesquisa por pesquisadores da área de economia e negócios, localizados na cidade de Bucaramanga e sua região metropolitana.

Justifica-se o presente estudo por se tratar de uma pesquisa de caráter exploratória e descritiva. A literatura sobre o tema é escassa e pouco divulgada, fato que despertou real interesse em estudá-lo e divulgá-lo.

A participação do (a) sr.(a) na presente pesquisa é de fundamental importância, mas será voluntária, não lhe cabendo qualquer obrigação de fornecer as informações e/ou colaborar com as atividades solicitadas pelos pesquisadores se não concordar com isso, bem como, participando ou não, nenhum valor lhe será cobrado, como também não lhe será devido qualquer valor.

Caso decida não participar do estudo ou resolver a qualquer momento dele desistir, nenhum prejuízo lhe será atribuído, sendo importante o esclarecimento de que os riscos da sua participação são considerados mínimos, limitados à possibilidade de eventual desconforto psicológico ao responder o questionário que lhe será apresentado, para que isso não venha a ocorrer, será escolhido um local privado sem a interferência de pessoas alheias ao estudo, enquanto que, em contrapartida, os benefícios obtidos com este trabalho serão importantíssimos e traduzidos em esclarecimentos para a população estudada.

Em todas as etapas da pesquisa serão fielmente obedecidos os Critérios da Ética em Pesquisa com Seres Humanos, conforme Resoluções nºs. 466/2012 e 510/17, ambas do Conselho Nacional de Saúde, que disciplina as pesquisas envolvendo seres humanos no Brasil.

Solicita-se, ainda, a sua autorização para apresentar os resultados deste estudo em eventos científicos ou divulgá-los em revistas científicas, assegurando-se que o seu nome será mantido no mais absoluto sigilo por ocasião da publicação dos resultados.

Caso a participação de vossa senhoria implique em algum tipo de despesas, as mesmas serão ressarcidas pelo pesquisador responsável, o mesmo ocorrendo caso ocorra algum dano.

Os pesquisadores estarão a sua disposição para qualquer esclarecimento que considere necessário em qualquer etapa da pesquisa.

Eu, _____, declaro que fui devidamente esclarecido (a) quanto aos objetivos, justificativa, riscos e benefícios da pesquisa, e dou o meu consentimento para dela participar e para a publicação dos resultados, assim como o uso de minha imagem nos slides destinados à apresentação do trabalho final. Estou ciente de que receberei uma cópia deste documento, assinada por mim e pelo pesquisador responsável. Como trata-se de um documento em duas páginas, a primeira deverá ser rubricada tanto pelo pesquisador responsável quanto por mim, assim como a última assinada por ambos.

Bucaramanga, Santander, ____ de _____ de 2020.

Maria Paula Fernández Jiménez
Pesquisador responsável

Participante da Pesquisa

Pesquisador Responsável: Maria Paula Fernández Jiménez
Endereço do Pesquisador Responsável: carrera 17 A # 3 37 Jardin del Limoncito Telefono: 3005790844 – E-mail: mariapaula.fernandezjmail.com
E-mail do Comitê de Ética em Pesquisa do Centro de Ciências da Saúde da Universidade Federal da Paraíba: eticaccs@ccs.ufpb.br – fone: (83) 3216-7791 – Fax: (83) 3216-7791
Endereço: Cidade Universitária – Campus I – Conj. Castelo Branco – CCS/UFPB – João Pessoa-PB - CEP 58.051-900