

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

RENATA OLIVEIRA DE ARAUJO

**MAPEAMENTO DA ATUAÇÃO DO ARQUIVISTA NA GESTÃO DE DADOS DE
PESQUISA**

RIO DE JANEIRO

2022

RENATA OLIVEIRA DE ARAUJO

**MAPEAMENTO DA ATUAÇÃO DO ARQUIVISTA NA GESTÃO DE DADOS DE
PESQUISA**

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

Orientador: Prof. Dr. Luis Fernando Sayão.

Rio de Janeiro

2022

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada na fonte.

Renata Oliveira de Araujo

Data

RENATA OLIVEIRA DE ARAUJO

**MAPEAMENTO DA ATUAÇÃO DO ARQUIVISTA NA GESTÃO DE DADOS DE
PESQUISA**

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

Aprovado em ____ de _____ de 2022.

BANCA EXAMINADORA

Prof. Dr. Luis Fernando Sayão (Orientador) – Presidente
Universidade Federal do Estado do Rio de Janeiro – UNIRIO

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

Profa. Dra. Ivone Pereira de Sá – Titula Externo
Fundação Oswaldo Cruz – FIOCRUZ

Profa. Dra. Simone da rocha Weitzel – Suplente Externo
Universidade Federal do Estado do Rio de Janeiro – UNIRIO

Profa. Dra. Viviane Santos de Oliveira Veiga – Suplente Externo
Fundação Oswaldo Cruz- FIOCRUZ

Dedico este trabalho a minha família.

AGRADECIMENTOS

Agradeço a minha família pelo apoio.

Ao orientador Luís Fernando Sayão por ter acreditado na minha pesquisa.

Às profas. Luana e Ivone, que fizeram parte da banca avaliadora, pelas contribuições.

RESUMO

A pesquisa apresenta a atuação do arquivista na gestão de dados de pesquisa, a partir da análise dos ciclos de vida dos dados de pesquisa Data Documentation Initiative (DDI) e Data Observation Network For Earth (DataONE) e do ciclo da curadoria Digital Curation Centre (DCC), relacionando as atividades arquivísticas às etapas da gestão de dados de pesquisa. O objetivo geral do estudo propõe mapear as áreas de atuação do arquivista de dados de pesquisa. Por meio da análise dos ciclos de vida dos dados de pesquisa e da curadoria. O percurso metodológico traçado é de abordagem qualitativa, exploratória, fundamentado no desenvolvimento e no esclarecimento de ideias para a construção do mapeamento da atuação do arquivista de dados de pesquisa. A coleta de dados é baseada em buscas por periódicos, análise no repositório de trabalho do IASSIST e no LinkedIn. Os resultados obtidos revelam que a teoria e a prática arquivística inserem o arquivista na gestão de dados de pesquisa, possibilitando a esse profissional a aplicação dos processos arquivísticos de gestão à preservação a longo prazo no tratamento dos dados de pesquisa, com destaque para o conhecimento e prática em avaliação, seleção, descrição, gestão e preservação, o uso secundário dos dados de pesquisa; além de deixar claro que a atuação do arquivista deve começar nas primeiras discussões e no planejamento da gestão de dados de pesquisa. O perfil do arquivista de dados construído a partir da análise das vagas de emprego evidencia como são poucas as vagas ofertadas no mercado de trabalho. Nas bases do mercado de trabalho alguns conhecimentos arquivísticos são solicitados, tais como padrão do OAIS, metadados de preservação PREMIS e questões relacionadas a repositórios. A continuidade dos estudos deve ser desenvolvida para todas as questões passíveis de interlocução com a atuação do arquivista na gestão de dados de pesquisa, com o intuito de orientar melhor o profissional que irá atuar nesse contexto e fazer com que o conhecimento e a prática arquivista sejam aderentes ao universo dos dados de pesquisa, promovendo a atuação e a visibilidade do profissional no tratamento desses dados.

Palavras-chave: dados de pesquisa; arquivista de dados; gestão de dados; arquivística.

ABSTRACT

The research presents the role of the archivist in the management of research data, from the analysis of the data life cycles of the Data Documentation Initiative (DDI) and Data Observation Network For Earth (DataONE) research data and the cycle of the Digital Curation Center curatorship (DCC), relating archival activities to research data management steps. The general objective of the study proposes to map the areas of action of the research data archivist. By analyzing the lifecycles of research and curation data. The methodological course traced is a qualitative, exploratory approach based on the development and clarification of ideas for building the mapping of the research data archivist's performance. Data collection is based on searches for journals, analysis in the IASSIST work repository and on LinkedIn. The results obtained reveal that archival theory and practice include the archivist in the management of research data, enabling this professional to apply archival management processes to long-term preservation in the treatment of research data, with emphasis on knowledge and practice in evaluation, selection, description, management and preservation, the secondary use of research data; in addition to making it clear that the archivist's role should begin with the first discussions and planning of research data management. The profile of the data archivist built from the analysis of job vacancies shows how few vacancies are offered in the job market. On the basis of the job market some archival knowledge is requested such as the OAIS standard, PREMIS preservation metadata, issues related to repositories. The continuity of studies must be developed for all the issues that can be discussed with the archivist's performance in the management of research data, in order to better guide the professional who will work in this context and make the archival knowledge and practice more relevant. adherent to the universe of research data, promoting the performance and visibility of the professional in the treatment of this data.

Keywords: research data; data archivist; data management; data management, archival

LISTA DE ILUSTRAÇÕES

Figura 1 – Modelo do ciclo de vida da DDI.....	33
Figura 2 – Ciclo de vida da DataONE.....	35
Figura 3 – Modelo do ciclo de vida dos dados do DCC.....	38
Figura 4 – Modelo OAIS	56
Figura 5 – Pacote de informação modelo OAIS.....	57

LISTA DE QUADROS

Quadro 1 – Estruturação metodológica.	19
Quadro 2 – Os princípios orientadores do FAIR.	26
Quadro 3 – Semelhanças entre os princípios FAIR e a gestão de documentos arquivísticos digitais.	68
Quadro 4 – Planejamento.	71
Quadro 5 – Produção.	73
Quadro 6 – Utilização.	75
Quadro 7 – Avaliação e destinação.	78
Quadro 8 – Cargo de Arquivista de dados de pesquisa.	82
Quadro 9 – Cargo de Arquivista Assistente.	83
Quadro 10 – Cargo de Arquivista Associado.	84
Quadro 11 – Cargo de Arquivista de dados Sênior.	85
Quadro 12 – Cargo de Arquivista Chefe.	86

LISTA DE ABREVIATURAS E SIGLAS

ABC	Academia Brasileira de Ciências
APIS	Arquivo Português de Informação Social
Brapci	Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação
CCSDS	<i>Consultative Committee for Space Data System</i>
CHA	Conhecimentos, habilidades e atitudes
CONARQ	Conselho Nacional de Arquivos
DataONE	<i>Data Observation Network for Earth</i>
DCC	<i>Digital Curation Centre</i>
DDI	<i>Data Documentation Initiative</i>
DOI	<i>Digital Object Identifier</i>
EAD	Descrição arquivística codificada
e-ARQ Brasil	Modelo de Requisitos para sistemas informatizados de gestão arquivística de documentos
FAIR	<i>Findable, Accessible, Interoperable and Reusable</i>
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
IASSIST	<i>International Association for Social Science Information Service & Technology</i>
ICA	<i>International Council on Archives</i>
ICPSR	<i>Inter-university Consortium for Political and Social Research</i>
ISAD(G)	Norma Geral Internacional de Descrição Arquivística
LISA	<i>Library and Information Science Abstracts</i>
METS	Padrão de codificação e transmissão de metadados
NAA	<i>National Archives of Australia</i>
NARA	<i>National Archives and Records Administration</i>
NOBRADE	Norma Brasileira de Descrição Arquivística
NSF	<i>National Science Foundation</i>
OAI-PMH	Protocolo para coleta de metadados
OAIS	<i>Open Archival Information System</i>
OCLC	<i>Online Computer Library Center</i>
OECD	<i>Organization for Economic Cooperation and Development</i>

PREMIS	<i>Preservation Metadata: Implementation Strategies</i>
RDC-Arq	Repositório arquivístico digital confiável
RLG	<i>Research Libraries Group</i>
SAAI	Sistema Aberto de Arquivamento de Informação
SESRI	<i>Social and Economic Survey Research Institute</i>
SIGAD	Sistema informatizado de gestão arquivística de documentos
TICs	Tecnologias de Informação e da Comunicação
USP	Universidade de São Paulo
XML	<i>Extensible Markup Language</i>

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Objetivos	16
1.2	Justificativa	17
2	PROCEDIMENTOS METODOLÓGICOS	18
3	DESAFIO DOS DADOS DE PESQUISA	21
3.1	Entendendo o que é dado de pesquisa	22
3.2	Tipologia dos dados de pesquisa	22
3.3	Fair Data	24
3.4	Plano de gestão de dados de pesquisa: iniciando o registro da pesquisa	27
3.5	Repositório de dados de pesquisa	29
4	O CICLO DE VIDA DOS DADOS DE PESQUISA E DA CURADORIA: A VISÃO DETALHADA DA GESTÃO	32
4.1	Data Documentation Initiative (DDI).....	32
4.2	Data Observation Network for Earth (DataONE)	34
4.3	Digital Curation Centre (DCC).....	37
5	CONHECIMENTO TEÓRICO E PRÁTICO DO ARQUIVISTA	41
5.1	Do convencional ao digital: compreendendo o documento arquivístico	41
5.2	Gestão de documentos arquivísticos dos convencionais aos digitais	43
5.3	Modelo de Requisitos para sistemas informatizados de gestão arquivística de documentos (e-ARQ Brasil)	47
5.4	Sistema informatizado de gestão arquivística de documentos (SIGAD)	48
5.5	Preservação a longo prazo	51
5.6	Metadados de preservação (PREMIS)	53
5.7	Modelo Open Archival Information System (OAIS)	55
5.8	Repositório arquivístico digital confiável (RDC-Arq)	57
5.9	O profissional arquivista	61
5.10	A arquivística e os dados de pesquisa.....	64
6	MAPEAMENTO DA ATUAÇÃO DO ARQUIVISTA NA GESTÃO DE DADOS DE PESQUISA	70
6.1	Planejamento	70
6.2	Produção	72
6.3	Utilização.....	74

6.4	Avaliação e destinação	75
6.5	Perfil do Arquivista no contexto da gestão de dados de pesquisa	80
7	CONCLUSÃO	90
	REFERÊNCIAS	92

1 INTRODUÇÃO

O termo *e-Science* foi primeiramente enunciado em estudos do conselho Nacional de Pesquisa do Reino Unido para designar o modo então emergente de produção do conhecimento científico, o qual se baseava no uso compartilhado de recursos e tecnologias da informação e da comunicação e permearia diferentes disciplinas e domínios científicos (RESEARCH COUNCILS UK, 2010).

A *e-Science* tem como característica explorar, no próprio fazer científico, ferramentas computacionais avançadas, que possibilitam amplo acesso a recursos geograficamente dispersos, incluindo coleções de dados, instrumentos científicos e mecanismos de visualização de alto desempenho, sendo um movimento internacional de crescente interesse e debate. A meta essencial da *e-Science* é utilizar-se das Tecnologias de Informação e da Comunicação (TICs) como fundamento e infraestrutura para uma transformação profunda no método científico de modo a tirar o maior proveito possível dos resultados de pesquisa (GRAY, 2009).

Com os avanços em tecnologias da informação, a capacidade das organizações, grupos e indivíduos de gerar dados vem se tornando cada vez maior, trazendo oportunidades ímpares e, ao mesmo tempo, impondo dificuldades à organização e preservação desses grandes volumes de dados — fenômeno amplamente discutido na atualidade sob a designação de *Big Data* (CHEIN *et al.*, 2014).

O desenvolvimento científico sempre foi fundamentado na obtenção e análise dos dados produzidos das mais variadas formas e por meio de instrumentos diversos. Com o avanço tecnológico recente na área de instrumentação científica, pesquisadores têm produzido uma quantidade de dados sem precedentes, muitos dos quais são subutilizados ou pouco explorados em seu potencial para o avanço científico e tecnológico, e mesmo nas organizações de pesquisa mais modernas, os dados de pesquisa estão dispersos, mal documentados ou inacessíveis, o que impossibilita sua reutilização em novas análises e a obtenção de respostas a novas perguntas (DRUCKER, 2012).

A Gestão de dados de pesquisa tem assumido crescente relevância no cenário científico internacional, na medida em que proliferam estudos que demonstram os benefícios associados, como o aumento de citações e a reutilização dos dados arquivados (PIWOWAR; DAY; FRIDSMA, 2007). A publicação *Data's Shameful Neglect* da revista *Nature*, em 2009, já destacava a gestão de dados de pesquisa como uma das principais bases da produção científica, refletindo a necessidade de integrá-la aos currículos acadêmicos em todas as áreas do conhecimento (DATA'S..., 2009).

Este foco maior nos dados é relativamente recente e ainda não é possível ter uma visão completa sobre todos os aspectos envolvidos, mas neste processo de desenvolvimento é crucial uma compreensão comum de pontos-chave independentemente de tecnologias, linguagens ou sistemas específicos (MÖLLER, 2013).

No contexto do trabalho orientado a dados de pesquisa, Madeiro e Dias (2020) destacam que quem pode oferecer serviço de tratamento de dados são profissionais que atendem a nova tendência mundial de trabalho orientado a dados. Aliás, profissões focadas em dados têm crescido cada vez mais na atualidade. No cenário internacional, o mercado de trabalho apresenta demanda de profissionais qualificados e capacitados para trabalhar com gestão e curadoria de dados.

O foco crescente em ciência de dados, serviços de gerenciamento de dados de pesquisa e estratégias associadas de curadoria e preservação de dados constitui a evidência do crescente impacto operacional do dilúvio de dados, da necessidade de desenvolvimento de infraestrutura de dados e da percepção de que a capacidade significativa da força de trabalho e desafios de capacidade estão surgindo. Empregadores em todos os setores estão buscando graduados para preencher uma mistura diversificada de funções relacionadas a dados, caracterizada por uma ampla gama de habilidades e competências em alfabetização de dados, combinadas com conhecimento disciplinar e experiência prática. Neste cenário complexo, as Escolas de Informação estão revisando os requisitos do currículo e desenvolvendo novos cursos centrados em dados para construir capacidade no local de trabalho e apoiar carreiras baseadas em dados no século XXI (LYON *et al.*, 2015, p. 1, tradução nossa).

Lyon *et al.* (2015) apresentam o resultado do mapeamento de currículo e a análise dos requisitos de trabalho especificados pelo empregador para as funções selecionadas de bibliotecário de dados, administrador/curador de dados e arquivista de dados. O último, segundo Madeiro e Dias (2020), é quase desconhecido no Brasil e merece ser investigado.

Com base no que foi exposto, esta pesquisa investiga o papel do arquivista na gestão de dados de pesquisa, no sentido de verificar como o profissional, a partir do uso da teoria e prática arquivística, pode contribuir e atuar na gestão de dados de pesquisa, por se tratar de um profissional com o olhar acurado, devido ao seu conhecimento teórico e prático em gestão, passando pelo ciclo de vida do documento arquivístico com o foco na preservação a longo prazo de modo a garantir a memória institucional da sociedade, possibilitando o reuso dos documentos por gerações futuras. Assim se configura a temática dos dados de pesquisa na perspectiva arquivística, com o foco na atuação do arquivista como um tema emergente.

Esse foco da atuação do arquivista no tratamento dos dados de pesquisa é relativamente recente, e ainda não é possível ter uma visão completa sobre todos os aspectos que envolvem

sua atuação, mas essa ação de identificar a atuação do profissional é essencial para determinar as atividades pertinentes à atuação do arquivista no âmbito da gestão de dados de pesquisa.

Com relação à estruturação da dissertação, o capítulo 1, *Introdução*, contextualiza a pesquisa e apresenta os objetivos geral e específico e a justificativa necessária à compreensão da questão da pesquisa: Qual o papel do arquivista na gestão de dados de pesquisa? O capítulo 2, *Procedimentos metodológicos*, descreve o tipo da pesquisa, a natureza da pesquisa e o procedimento de coleta de dados.

A seguir, o capítulo 3, *Desafio dos dados de pesquisa*, apresenta a teoria, estabelecendo o que é dado de pesquisa, tipologia, *Fair Data*, plano de gestão de dados de pesquisa e repositório de dados de pesquisa.

O capítulo 4, *O ciclo de vida dos dados de pesquisa e da curadoria: a visão detalhada da gestão*, traz os modelos do ciclo de vida dos dados de pesquisa: *Data Documentation Initiative* (DDI), *Data Observation Network for Earth* (DataONE), e o *Digital Curation Centre* (DCC), ciclo da curadoria.

O capítulo 5, *Conhecimento teórico e prático do arquivista*, trata do profissional arquivista, do documento arquivístico, da gestão de documentos, do Modelo de Requisitos para sistemas informatizados de gestão arquivística de documentos (e-ARQ Brasil), do Sistema informatizado de gestão arquivística de documentos (SIGAD), da preservação a longo prazo, do metadado de preservação (*Preservation Metadata: Implementation Strategies – PREMIS*), do modelo *Open Archival Information System* (OAIS), do repositório arquivístico digital confiável (RDC-Arq) e das instituições arquivísticas e dados de pesquisa.

O capítulo 6, *Mapeamento da atuação do arquivista na gestão de dados de pesquisa*, apresenta os resultados da pesquisa e o produto da pesquisa: o mapeamento realizado a partir da análise do ciclo de vida do dado de pesquisa DDI, DataONE e do ciclo da curadoria DCC.

1.1 Objetivos

No que concerne ao objetivo geral, esta pesquisa busca mapear as áreas de atuação do arquivista de dados de pesquisa. Analisar os ciclos de vida dos dados de pesquisa DataONE, DDI, e do ciclo da curadoria DCC. Como resultado esperado, o mapeamento da atuação do arquivista de dados de pesquisa relacionando as atividades arquivísticas e as etapas da gestão de dados. Quanto aos objetivos específicos, propõe:

- a) compreender as etapas da gestão de dados de pesquisa e do ciclo de vida dos dados de pesquisa;

- b) reexaminar a teoria e prática arquivística;
- c) apresentar as características e o perfil do arquivista no âmbito da gestão de dados de Pesquisa.

1.2 Justificativa

Esta pesquisa justifica-se pela necessidade de capacitar e instruir arquivistas, bibliotecários, pesquisadores e demais profissionais a atuar na gestão de dados de pesquisa. A percepção da necessidade de capacitar todos que lidam com dados de pesquisa se deu a partir da publicação do artigo: “Curadoria digital: papéis e responsabilidade do arquivista” (ARAÚJO; FINAMOR, 2017), no qual foi identificada a possibilidade de atuação do arquivista na curadoria digital de dados de pesquisa.

Aprofundando-se a investigação, foram percebidas outras questões análogas às práticas arquivísticas e ao contexto dos dados de pesquisa, tais como as semelhanças entre documento arquivístico e dado de pesquisa, a proveniência arquivística e a rastreabilidade do dado de pesquisa.

O dado de pesquisa necessita de ações de gestão e preservação, tendo em vista o reuso. A presente pesquisa tentará identificar os pontos-chave da atuação do arquivista no tratamento dos dados de pesquisa.

A gestão e curadoria de dados de pesquisa deve ser desempenhada por profissionais qualificados. Já que tal meio carece ainda de metodologia e prática especialmente no ambiente digital, por se tratar de um ambiente que está sendo constantemente atualizado e substituído por tecnologias mais modernas, isso nos leva a repensar as atividades de produção, uso, acesso, compartilhamento, preservação e reuso dos dados de pesquisa, além da atuação do arquivista na gestão desses dados.

Considerando a necessidade de atuação de profissionais capacitados no tratamento dos dados de pesquisa, a demonstração da possibilidade de atuação do arquivista neste contexto amplia e amadurece as concepções teóricas e práticas acerca dos dados de pesquisa, contribuindo para a manutenção da memória científica, o acesso e promovendo o reuso dos dados de pesquisa.

O objetivo é conduzir a atuação do arquivista no gerenciamento e preservação dos dados de pesquisa com vistas ao reuso e, sobretudo, que o arquivista se veja desenvolvendo sua prática e aplicando a teoria arquivística no âmbito dos dados de pesquisa, apropriando-se deste contexto com toda sua *expertise* e se tornando um gestor de conhecimento.

2 PROCEDIMENTOS METODOLÓGICOS

A pesquisa, para Minayo (1999), reflete a:

[...] atividade básica das ciências na sua indagação e descoberta da realidade. É uma atitude e uma prática de constante busca que define um processo intrinsecamente inacabado e permanente. É uma atividade de aproximação sucessiva da realidade que nunca se esgota, fazendo uma combinação particular entre teoria e dados (MINAYO, 1999, p. 23).

Desse modo, Menezes *et al.* (2019, p. 13) esclarecem que “para se realizar qualquer pesquisa, é necessário partir de uma questão que norteará todo o percurso investigativo a ser empreendido; por isso, tal problema pode ser apresentado em forma de pergunta”. A presente pesquisa questiona: qual o papel do arquivista na gestão de dados de pesquisa?

A investigação é realizada de modo a apresentar estudos sobre a temática de gestão de dados de pesquisa que vêm mostrando a necessidade de atuação do arquivista neste contexto, devido ao seu conhecimento em gestão e preservação de longo prazo, e ao pouco e/ou nenhum conhecimento da atuação do profissional no tratamento dos dados de pesquisa, promovendo discussões sobre sua atuação nos cursos de graduação e pós-graduação em Arquivologia e Biblioteconomia. Cabe ressaltar que é em tais áreas do conhecimento que estão se iniciando os estudos sobre a possibilidade de atuação do profissional para além da instituição e setor de arquivo.

O percurso metodológico utilizado na condução desta pesquisa baseou-se numa abordagem bibliográfica na perspectiva de Vergara (2006, p. 48): o procedimento de coleta “utiliza fontes bibliográficas ou material elaborado, como livros, publicações periódicas, artigos científicos, impressos diversos ou, ainda, textos extraídos da internet”. Foram adotadas estratégias de busca na Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (Brapci) e *Library and Information Science Abstracts* (LISA), uma base de dados sobre ciência da informação com cobertura internacional. Com a intenção de identificar a produção científica sobre a atuação do arquivista na gestão de documentos arquivísticos, foram utilizados autores que fazem parte da fundamentação da Arquivística, tais como: Bellotto (2005), Schellenberg (2006, 2015), Rousseau e Couture (1998), Indolfo (2007), entre outros.

Trata-se de uma pesquisa exploratória, a qual, segundo Gonsalves (2003), se caracteriza pelo desenvolvimento e esclarecimento de ideias, com o objetivo de fornecer uma visão panorâmica, uma primeira aproximação a um determinado fenômeno que é pouco explorado. Esse tipo de pesquisa também é denominado “pesquisa de base”, pois oferece dados

elementares que dão suporte à realização de estudos mais aprofundados sobre o tema. Gil (2002, p. 41) acrescenta que “na maioria dos casos essa pesquisa envolve: levantamento bibliográfico, análise de exemplos que estimulem a compreensão [...]”.

Dessa forma, torna-se necessário identificar a atuação do arquivista e como o profissional pode contribuir com seu conhecimento e prática na gestão de dados de pesquisa. No que concerne ao objetivo, esta pesquisa busca determinar a possibilidade de atuação profissional do arquivista e de inserção de processos arquivísticos no ciclo de gestão de dados de pesquisa científica, mapeando as áreas de atuação do arquivista de dados de pesquisa a partir da análise do ciclo de vida dos dados de pesquisa DDI¹, DataONE² e do ciclo da curadoria DCC.³

Quadro 1 – Estruturação metodológica.

Objetivo específico	Etapa metodológica	Resultado esperado
Compreender as etapas da gestão de dados de pesquisa e do ciclo de vida dos dados de pesquisa.	Pesquisa bibliográfica Análise descritiva dos modelos do ciclo de vida dos dados de pesquisa DDI, DataONE, e do ciclo da curadoria DCC.	Capítulo teórico.
Reexaminar a teoria e prática arquivística.	Pesquisa bibliográfica Análise descritiva da teoria e prática arquivística	Capítulo teórico.
Apresentar as características e o perfil do arquivista no âmbito da gestão de dados de Pesquisa.	Busca e análise no repositório de trabalho do IASSIST e no LinkedIn.	Perfil do arquivista de dados de pesquisa, relacionando o conhecimento do arquivista com as demandas do mercado de trabalho (quadro sistemático).

Fonte: elaborado pela autora.

Os modelos escolhidos para subsidiar esta análise possuem maior destaque na literatura. A DataONE é um exemplo da nova ciência ambiental inovadora, visando atender a demanda crescente de produção de dados através de estruturas distribuídas e sustentáveis que atendam às necessidades da ciência e da sociedade de acesso. Configurando-se como o mais conhecido na literatura, o DCC também é considerado o mais completo. Já o modelo DDI tem sua importância por ter servido de base para o modelo da *MRC Centre of Epidemiology for Child Health* da *University College London*, de Humphrey.

Para montar o modelo do perfil do arquivista que irá atuar na gestão de dados de pesquisa, e tomando como base a literatura analisada e os modelos, foram realizadas buscas em

¹ Disponível em: <https://ddialliance.org/>.

² Disponível em: <https://www.dataone.org/>.

³ Disponível em: <https://dcc.ac.uk/>.

duas bases de bancos internacionais de trabalho com o objetivo de delinear o perfil do arquivista de dados de pesquisa: *The IASSIST Jobs Repository* da *International Association for Social Science Information Service & Technology* (IASSIST) — e LinkedIn.

A busca foi efetuada no site de trabalho da *IASSIST Jobs Repository* (repositório de trabalho da IASSIST), originando um conjunto de dados com descrições de trabalho postadas desde o dia 11 de janeiro de 2005. IASSIST é a abreviação da *International Association for Social Science Information Service and Technology*. Trata-se de uma organização internacional de trabalhadores da tecnologia da informação e serviços de dados que apoiam as ciências sociais e promovem o seu ensino (IASSIST, 2021a).

LinkedIn é a maior rede profissional do mundo, com mais de 774 milhões de usuários em mais de 200 países e territórios em todo o mundo. Tem como visão criar oportunidades econômicas para cada integrante da força de trabalho global, com a missão de conectar profissionais do mundo todo, tornando-os mais produtivos e bem-sucedidos. É uma rede social voltada para relacionamentos profissionais que pode ser usada para buscar emprego, anunciar vagas, fazer parcerias e *networking* (LINKEDIN, 2021).

Esta pesquisa é de natureza qualitativa, que, segundo Minayo (2001), responde a perguntas muito particulares. Ela se preocupa, nas ciências sociais, com um nível de realidade que não pode ser quantificado.

3 DESAFIO DOS DADOS DE PESQUISA

Instituições e pesquisadores devem garantir o gerenciamento e a preservação dos dados de pesquisa de modo a possibilitar a reutilização dos dados originais. Particularmente, estudos têm apontado que dados de pesquisa bem organizados, documentos preservados, acessíveis e verificados quanto a sua acurácia e validade são mais facilmente compartilháveis e reutilizáveis (APPEL; MACIEL; ALBAGLI, 2006; NATIONAL RESEARCH COUNCIL UK, 1999).

No decorrer dos últimos anos, diversas agências internacionais de fomento, como a National Science Foundation (NSF), solicitam desde janeiro de 2011 que as propostas de projetos incluam um documento complementar designado: o “Plano de Gestão de Dados de Pesquisa”, seguindo a iniciativa do *Natural Environment Research Council* de 2010. Agências de fomento inglesas, como o *Medical Research Council* e o *Wellcome Trust* (UK), determinam, além do plano de gestão de dados, que o pesquisador compartilhe seus dados com a comunidade científica “no momento certo e de modo responsável, garantindo, assim, que esses dados possam ser verificados e usados como base para alcançar o conhecimento científico e suas aplicações na área da saúde” (WELLCOME TRUST, 2017, *online*, tradução nossa).

No âmbito nacional, alinhada às agências de fomento internacional, temos a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), que:

[...] reconhece a importância da gestão adequada dos dados de pesquisa como parte essencial das boas práticas de pesquisa. Considera necessário que os dados resultantes de projetos financiados pela Fundação sejam gerenciados e compartilhados de forma a garantir o maior benefício possível para o avanço científico e tecnológico. (FAPESP, 2020, *online*).

No contexto recente, a Academia Brasileira de Ciências (ABC) criou um grupo de estudo sobre a abertura e a gestão de dados com o objetivo de discutir e apresentar iniciativas que possam ajudar o Brasil a lidar com a produção, abertura e gestão de grandes volumes de dados ao mais variado campo científico nacional. Em setembro de 2020, ela lançou o documento: “Abertura e gestão de dados: desafios para a Ciência brasileira” (ABC, 2020).

Nesse sentido, os dados de pesquisa devem possuir estrutura e organização para que sejam considerados úteis. Assim, fica evidente a necessidade de aprimoramento dos mecanismos e das práticas de gestão de dados de pesquisa em universidades e organizações de pesquisa e desenvolvimento. Na próxima seção desse trabalho será apresentado o que é dado de pesquisa.

3.1 Entendendo o que é dado de pesquisa

O Relatório da *Organization for Economic Cooperation and Development* (OECD) descreve o termo “dados de pesquisa” como “registros factuais usados como fonte primária para a pesquisa científica e que são comumente aceitos pelos pesquisadores como necessários para validar os resultados do trabalho científico” (OECD, 2007, p. 7, tradução nossa).

O *National Science Board* da NSF adota a seguinte definição para se referir a dados de pesquisa: “qualquer informação que possa ser armazenada em formato digital, incluindo texto, números, imagens, vídeo ou filmes, áudio, software, algoritmos, equações, animações, modelos, simulações etc.” (NATIONAL SCIENCE BOARD, 2005, p. 10, tradução nossa).

Corrêa (2016) esclarece que o dado de pesquisa pode ser do tipo numérico, descritivo, visual e reproduzir-se em formato de papel (incluindo notas de pesquisa em cadernos, fotografias etc.) ou digital. O conceito de dados científicos também faz referência às distintas ferramentas como protocolos, códigos numéricos, gráficos e tabelas, que são necessárias para recolher e organizar os dados, tanto em trabalho de campo quanto em laboratório. Os dados compreendem não somente os materiais e amostras biológicas e/ou ambientais extraídas mas também os resumos gerados durante o transcurso da realização de uma pesquisa: “Os dados científicos são todas as evidências que um investigador necessita para validar suas conclusões após uma pesquisa. São fonte de prova do exercício do processo da pesquisa científica” (CORRÊA, 2016, p. 21, tradução nossa).

O dado de pesquisa é a matéria-prima, a parte mais palpável da pesquisa e o insumo para o desenvolvimento de novas pesquisas. Ele reflete o contexto no qual é produzido por meio das abordagens utilizadas por pesquisadores, instituições e pela própria área do conhecimento na qual está inserido e é criado. Trata-se de informações das mais variadas e registradas em suporte convencional e/ou digital no exercício da atividade de pesquisa, ou seja, é a prova do processo de pesquisa desempenhado pelo pesquisador. Tais dados podem ser gerados por vários meios, incluindo observação, computação ou experimento conforme veremos a seguir.

3.2 Tipologia dos dados de pesquisa

A distinção dos dados de pesquisa, segundo a NSF (2022), ocorre por meio da classificação segundo o procedimento de coleta. Os dados podem ser classificados em: computacionais, observacionais e experimentais.

Os dados computacionais são o resultado da execução de um modelo de computador ou simulação. Se informações abrangentes sobre o modelo (incluindo uma descrição completa do hardware, software e dados de entrada) estiverem disponíveis, a preservação em um repositório de longo prazo pode não ser necessária, porque os dados podem ser reproduzidos. Assim, embora as saídas de um modelo não precisem ser preservadas, o arquivamento do próprio modelo e de um conjunto de metadados robusto pode ser essencial (NSF, 2022).

Na perspectiva de Corrêa (2016), mesmo que sejam utilizadas como parte do processo de pesquisa em diversos campos, a bioinformática e a genômica têm sido precursoras no uso de dados computacionais com a posse de padrões e previsões de comportamentos. Seu uso é mais comum nas ciências físicas e biológicas, mesmo que também sejam utilizadas para o desenvolvimento de pesquisas em ciências sociais e humanas. Podemos citar, como exemplo, os oceanógrafos, que utilizam dados computacionais para estudar os mares e realizar análises sobre as possibilidades de aparição ou previsão de tsunâmis em determinadas regiões, ou os economistas, que estudam as interações políticas e os mercados.

Dados experimentais como mediações de padrões de expressão gênica, taxas de reações químicas ou desempenho de motor apresentam um quadro mais complexo. Em princípio, os dados de experimentos que podem ser reproduzidos com precisão não precisam ser armazenados indefinidamente. Na prática, entretanto, pode não ser possível reproduzir precisamente todas as condições experimentais, particularmente nos casos em que algumas condições e variáveis experimentais talvez não sejam conhecidas e quando os custos de reprodução do experimento são proibitivos (NSF, 2022).

Dados observacionais, para Corrêa (2016), são dados provenientes de observações científicas, nas quais os pesquisadores medem tantas variáveis quanto lhes for possível, a fim de elucidar as possíveis relações de causa e efeito. São um tipo de registro que tem caráter único e que, com a diferença dos experimentos, não se pode voltar a reproduzir, e, portanto, são insubstituíveis. Definitivamente, são dados capturados em tempo real, geralmente fora do laboratório.

Os dados observacionais podem ser, por exemplo, observações climatológicas ou de flora e fauna, por satélite, redes de sensores ou mediante anotações. Mas também podem ser, em ciências sociais, observações do comportamento de distintos grupos ou pessoas. Além disso, estas observações vêm associadas a lugares e momentos específicos, ou a uma multiplicidade de lugares e nos estudos transversais e longitudinais (CORRÊA, 2016, p. 28, tradução nossa).

Esses dados são provenientes de registros dos fatos ou das evidências de fenômenos, usualmente com instrumentos, como as observações diretas da temperatura do oceano em uma data específica, a atitude dos eleitores antes de uma eleição ou fatos de uma supernova, ou até de registros históricos. Assim, esses dados observacionais são geralmente arquivados indefinidamente (NSF, 2022).

Nessa perspectiva, dados computacionais, observacionais e experimentais, segundo Sayão e Sales (2020), são criados ou coletados em ambientes, contextos ou situações condicionadas ou controladas por rígidos protocolos científicos, como laboratórios, observatórios, expedições, ambientes científicos virtualizados e outros. São dados de pesquisa na sua essência. Entretanto, os dados governamentais, registros médicos — como os prontuários — e arquivísticos, dados provenientes de censo e dados extraídos das redes sociais, por exemplo, são dados coletados para outros propósitos que são frequentemente utilizados para a pesquisa científica. Trata-se de dados usados para a pesquisa que têm sua origem fora do mundo científico. Borgman (2015), citada por Sayão e Sales (2020), caracteriza esses dados como uma quarta categoria em relação à origem, denominando-os registros, que consistem em formas de dados que não se enquadram facilmente nas categorias observacionais, experimentais ou computacionais ou que não resultam de nenhuma dessas categorias. Nessa perspectiva, “Registros de quase todos os fenômenos ou de qualquer atividade humana podem ser tratados como dados para a pesquisa”, conclui Borgman (2015, p. 24 *apud* SAYÃO; SALES, 2020, p. 22).

Com intensificação da produção do dado de pesquisa de forma contínua e dinâmica por meio da tecnologia. Faz-se necessário a adoção de instrumento para a descoberta, o acesso, interoperabilidade, compartilhamento e a reutilização dos dados de pesquisa. Nesse sentido será apresentado na próxima seção o Fair Data.

3.3 Fair Data

O FORCE 11 é uma comunidade de acadêmicos, bibliotecários, arquivistas, editores e financiadores de pesquisas que surgiu organicamente para ajudar a facilitar a mudança em direção à criação e compartilhamento de conhecimentos aprimorados. Individualmente e coletivamente, eles pretendem realizar uma mudança nas comunicações acadêmicas modernas por meio do uso eficaz da tecnologia da informação (FORCE 11, 2021).

Eles acreditam que um dos grandes desafios da ciência com uso intensivo de dados é facilitar a descoberta de conhecimentos, auxiliando humanos e máquinas em sua descoberta,

acesso, integração e análise de dados científicos apropriados para tarefas e seus algoritmos e fluxos de trabalho (FORCE 11, 2021).

A criação dos princípios *Findable, Accessible, Interoperable and Reusable* (FAIR) é datada do início de 2014, fruto de uma conferência internacional denominada “Jointly designing a data FAIRPORT”, que reuniu especialistas de diversas áreas do conhecimento envolvidos na utilização, no tratamento e no reuso dos dados de pesquisa no âmbito da *e-Science* (HENNING *et al.*, 2019, p. 177).

O resultado dessa conferência foram os princípios FAIR, um acrônimo para *Findable, Accessible, Interoperable e Reusable*, figurando como elementos norteadores do processo de gestão de dados de pesquisa, que vêm sendo adotados por instituições de pesquisa, universidades, editores, gestores de dados, infraestruturas de dados e agências financiadoras de pesquisa. Esses princípios são considerados o instrumento norteador para a descoberta, o acesso, a interoperabilidade global, o compartilhamento e a reutilização dos dados de pesquisa, pois se propõem estabelecer orientações estruturais sobre a publicação de recursos digitais, conjuntos de dados, códigos e objetos de pesquisa de maneira FAIR (HENNING *et al.*, 2019; WILKINSON *et al.*, 2016).

Wilkinson *et al.* (2016) enumeram os princípios FAIR: localizável, acessível, interoperável e reutilizável como princípios básicos que servem para orientar os produtores e editores de dados, além de maximizar o valor agregado obtido pela publicação digital acadêmica. Os princípios devem ser aplicados não apenas aos dados no sentido convencional mas também aos algoritmos, ferramentas e fluxos de trabalho que levaram a esses dados. Todos os objetos de pesquisa digital acadêmica se beneficiam da aplicação desses princípios, uma vez que todos os componentes do processo de pesquisa devem estar disponíveis para garantir transparência, reprodutibilidade e reutilização.

Wilkinson *et al.* (2016) expõem os princípios abaixo:

Quadro 2 – Os princípios orientadores do FAIR.

Para ser localizável (F - <i>Findable</i>):
F 1 – (meta)dados são atribuídos a uma identidade globalmente única e persistente
F 2 – os dados são descritos com metadados ricos
F 3 – (meta)dados são registrados ou indexados em um recurso pesquisável
F 4 – metadados especificam o identificador persistente
Para ser acessível (A - <i>Accessible</i>):
A 1 – os (meta)dados são recuperáveis por seu identificador por meio de um protocolo de comunicação padronizado
A 1,1 – o protocolo é aberto, gratuito e universalmente implementável
A 1,2 – o protocolo permite um procedimento de autenticação e autorização, quando necessário
A 2 – metadados são acessíveis, mesmo quando os dados não estão mais disponíveis
Para ser interoperável (I - <i>Interoperable</i>):
I 1 – (meta)dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento
I 2 – (meta)dados usam vocabulários que seguem os princípios FAIR
I 3 – (meta)dados incluem referências qualificadas a outros (meta)dados
Para ser reutilizável (R- <i>Reusable</i>)
R 1 – (meta)dados têm uma pluralidade de atributos precisos e relevantes
R 1,1 – (meta)dados são liberados com uma licença de uso de dados clara e acessível
R 1,2 – (meta)dados estão associados à sua proveniência
R 1, 3 – (meta)dados atendem aos padrões de comunidade relevantes para o domínio

Fonte: Adaptado e traduzido do Force 11 (2021).

Os elementos dos princípios FAIR estão relacionados, mas são independentes e separáveis. Os princípios definem características que recursos de dados, ferramentas, vocabulários e infraestrutura devem possuir para auxiliar na descoberta e reutilização por diversas áreas. Para tanto, o bom gerenciamento de dados não é um objetivo em si mesmo, mas sim o principal canal que leva à descoberta e inovação do conhecimento e à subsequente integração e reutilização de dados e conhecimento pela comunidade após o processo de publicação de dados. Infelizmente, o ecossistema digital existente em torno da publicação de dados acadêmicos nos impede de extrair o máximo benefício de nossos investimentos em pesquisa (WILKINSON, 2016).

Cabe destacar o exemplo de aplicabilidade dos princípios FAIR no caso do Arquivo Português de Informação Social (APIS). O APIS é uma infraestrutura científica para as ciências sociais. Com sede no Instituto de Ciências Sociais da Universidade de Lisboa, o arquivo promove a angariação e a partilha de dados de investigação em formatos digitais. Do seu acervo faz parte um conjunto de bases de dados resultantes de projetos de investigação realizados pela comunidade científica (APIS, 2021).

A principal atividade do APIS consiste em assegurar a continuidade do acesso aos dados produzidos pelos projetos de investigação, através de um conjunto de atividades de curadoria, preservação e disseminação, promovendo ao mesmo tempo a visibilidade dos dados e a sua

utilização por terceiros para efeitos de análise secundária, consulta pública e uso pedagógico (APIS, 2021).

Com relação à aplicabilidade dos princípios FAIR, o APIS destaca que “tornar os dados FAIR é cada vez mais uma responsabilidade conjunta dos investigadores e dos repositórios” (APIS, 2021, *online*). O repositório ainda disponibiliza um link⁴ com mais informações sobre os princípios FAIR.

Em setembro de 2021, o APIS obteve a certificação *Core Trustworthy Data Repositories*. A certificação, lançada em 2017, é oferecida a quem atende a certos requisitos definidos para repositórios de dados confiáveis que armazenam dados para preservação de longo prazo (CORETRUSTSEAL, 2022).

O CoreTrustSeal está engajado pelos princípios FAIR e outros projetos e grupos de trabalho relacionados ao FAIR. Os repositórios de dados confiáveis certificados pelo CoreTrustSeal são componentes vitais para permitir a concretização dos princípios de dados localizáveis, acessíveis, interoperáveis e reutilizáveis (FAIR) (CORETRUSTSEAL, 2022).

À medida que os princípios FAIR são esclarecidos por meio de indicadores e avaliados por meio de testes (idealmente autorizados) em objetos digitais, o CoreTrustSeal continua a integrar as melhores práticas essenciais aos requisitos. O Conselho do CoreTrustSeal continuará a acompanhar o trabalho realizado pelo FAIR e outras iniciativas relacionadas ao FAIR em todo o mundo e se envolver nesse trabalho para garantir que a certificação CoreTrustSeal continue a atender as necessidades da comunidade relativas à certificação (CORETRUSTSEAL, 2022).

O exemplo de aplicabilidade dos princípios FAIR no repositório APIS e o uso do FAIR como requisito para obter o certificado CoreTrustSeal mostram que o FAIR está sendo empregado e que assimila as melhores práticas de requisitos de repositórios de dados. Outro instrumento, o plano de gestão de dados de pesquisa documenta todo o processo de pesquisa, veremos a seguir.

3.4 Plano de gestão de dados de pesquisa: iniciando o registro da pesquisa

O Plano de Gestão de Dados de Pesquisa é uma ferramenta útil para documentar metodologias, práticas de gerenciamento de dados e registros de pesquisa e qualquer responsabilidade associada a eles. Conforme enfatizado por Silva, Marinho e Santos (2007):

⁴ How FAIR are your data? Disponível em https://www.cessda.eu/content/download/3845/35038/file/20170707_How_FAIR_are_your_data_Jones.pdf.

[...] o plano de gestão de dados é um documento formal que descreve todo o ciclo de vida dos dados, desde sua coleta até a documentação completa do processo de pesquisa, e registra as decisões tomadas em relação aos padrões de metadados, formatos, bases de dados, métodos, segurança e período de armazenamento, assim como os custos associados com a gestão de dados. (SILVA; MARINHO; SANTOS, 2007, p. 55).

Corrêa (2016) assevera que o plano de gestão de dados requer uma sequência documentada de ações destinadas a identificar e assegurar recursos, coletar, manter, proteger e utilizar os arquivos de dados. Isso inclui a obtenção de financiamento e a identificação dos recursos técnicos e de pessoal para o complexo ciclo de gestão de dados. Uma vez que se determinem as necessidades de uso dos dados, o passo seguinte deve ser a adoção de um sistema para armazenar e manipular os que podem ser identificados e desenvolvidos, mensurando-se o alcance e a quantidade de detalhes em um plano de gestão de dados.

A Universidade de São Paulo (USP, 2019) define o Plano de Gestão de Dados como um documento formal relacionado ao projeto de pesquisa que deve responder a duas perguntas básicas. 1. Quais dados são gerados? O pesquisador deve informar o formato dos dados e uma breve descrição destes, de tal forma que aqueles que pretendem utilizá-los possam compreendê-los. Os dados gerados podem, por exemplo, estar em formato de planilha eletrônica, texto ou banco de dados digital. 2. Como e onde esses dados são armazenados e disponibilizados? O pesquisador deve informar o local de armazenamento e disponibilização dos dados, que pode ser um repositório público ou particular dele. A resposta a essa pergunta deve envolver aspectos éticos e legais que eventualmente estarão associados aos dados.

O plano de gestão de dados solicitado pela FAPESP (2020) varia conforme a disciplina, os tipos de dados considerados e a forma de disponibilização por parte dos responsáveis pelo projeto. Além disso, o plano de gestão de dados deve ser de até duas páginas, contendo as seguintes informações:

- a) descrição dos dados e metadados produzidos pelo projeto – por exemplo, amostras, registros de coleta, formulários, modelos, resultados experimentais, software, gráficos, mapas, vídeos, planilhas, gravações de dados, banco de dados, material didático e outros;
- b) quando aplicável, restrições legais ou óticas para compartilhamento de tais dados, práticas para garantir a privacidade, confidencialidade, segurança, propriedade intelectual e outros;
- c) política de preservação e compartilhamento (por exemplo, compartilhamento imediato ou apenas após a aceitação de publicação associada). Período de carência

(antes do compartilhamento) e período durante o qual os dados serão preservados e disponibilizados;

- d) descrição de mecanismos, formatos e padrões para armazenar tais itens de forma a torná-los acessíveis por terceiros. Essa descrição pode incluir o uso de repositórios e serviços de outras instituições (FAPESP, 2020).

O plano de gestão de dados de pesquisa é visto como uma ferramenta de planejamento, acompanhamento e registro da pesquisa. A elaboração do plano constitui parte do ciclo de vida dos dados de pesquisa. Além disso, a gestão de dados de pesquisa conta com uma infraestrutura para gerenciar, acessar, compartilhar e arquivar os conjuntos de dados de pesquisa. Tal infraestrutura denomina-se repositório de dados de pesquisa e será apresentado a seguir.

3.5 Repositório de dados de pesquisa

O re3data⁵ é um registro global de repositórios de dados de pesquisa que cobre repositórios de pesquisa de diferentes disciplinas acadêmicas, aponta que os dados de pesquisa são valiosos e onipresentes. O acesso permanente aos dados de pesquisa é um desafio para todas as partes interessadas na comunidade científica. A preservação a longo prazo e o princípio do acesso aberto aos dados de pesquisa oferecem amplas oportunidades à comunidade científica. Na última década, mais e mais universidades e centros de pesquisa estabeleceram repositórios de dados de pesquisa, permitindo o acesso permanente a conjuntos de dados em um ambiente confiável. Devido a requisitos disciplinares, o panorama de repositórios de dados é muito heterogêneo (RE3DATA.ORG, 2021).

Segundo Rodrigues e Saraiva (2010, p. 22), o termo “repositório” designa um sistema informático em que existe uma plataforma de armazenamento de objetos representados em arquivos capaz de incorporar novos objetos à medida que são produzidos ou submetidos. O repositório oferece serviços que são dirigidos a quem deposita, a quem pesquisa e aos administradores do sistema.

Rodrigues e Saraiva (2010) acreditam que nos repositórios de dados pode-se ir muito além, uma vez que cada conjunto de dados tem características próprias e por isso pode requerer um tratamento diferenciado. Como exemplo de um processamento específico de um conjunto de dados, pode-se citar a conversão de dados originalmente organizados numa base de dados em um formato textual (por exemplo, na linguagem de anotação — *Extensible Markup*

⁵ Disponível em: <https://www.re3data.org/>.

Language — XML). Por outro lado, para o usuário do repositório pode ser importante, por exemplo, em vez de baixar a totalidade de um conjunto de dados cujo volume seja difícil de gerir, selecionar apenas um subconjunto dos itens incluídos no conjunto, ou apenas alguns dos atributos desses itens.

Nos repositórios de dados atuais são utilizadas diversas soluções tecnológicas. Algumas destas soluções são plataformas desenvolvidas genericamente para repositórios, outras são soluções desenvolvidas sob medida para um caso específico (RODRIGUES; SARAIVA, 2010).

Uzwyszyn (2016) define os repositórios de dados de pesquisa como grandes infraestruturas de banco de dados configuradas para gerenciar, compartilhar, acessar e arquivar conjuntos de dados de pesquisadores. Os repositórios podem ser especializados disciplinares ou mais gerais, coletando dados sobre áreas de conhecimento mais amplas, como as ciências exatas ou ciências sociais. Repositórios on-line também podem agregar dados de especialistas, coletando dados de pesquisadores de uma universidade ou consórcio de universidades para benefício mútuo. A ideia é que o compartilhamento de dados melhora os resultados e impulsiona a pesquisa e a descoberta.

Um repositório permite exame, prova, revisão, transparência e validação dos resultados de um pesquisador por outros especialistas, além do artigo acadêmico referenciado publicado. A disponibilização de dados de pesquisa permite o acesso instantâneo por um grupo globalmente disperso de pesquisadores para compartilhar, compreender e sintetizar os resultados. Essa agregação e síntese fornecem uma oportunidade de percepção, progresso e de busca exclusivamente humana por uma compreensão maior. Os repositórios de dados também permitem a publicação de dados negativos anteriormente ocultos, sobretudo experimentos que não funcionaram. Isso permite que outros pesquisadores evitem becos sem saída para aqueles que tentaram um caminho antes deles e encontrem melhor o caminho em direção a um terreno mais fértil (UZWYSHYN, 2016).

Além de oferecer uma base tecnológica para a execução dos processos de contextualização dos dados, os repositórios têm um papel importante nas interações que envolvem a validação do trabalho de pesquisa e na própria dinâmica social da comunicação científica. A possibilidade de se ter os dados de pesquisa disponíveis on-line, indexados, documentados e anotados relativos a uma pesquisa publicada ou pré-publicada num artigo acadêmico redimensiona a revisão por pares, estendendo-a a uma comunidade mais ampla e conectada em rede (SAYÃO; SALES, 2016).

De acordo com Pampel (2013), os repositórios de dados de pesquisa são divididos em quatro tipos: Repositórios de dados de pesquisas institucionais; Repositórios de dados de

pesquisa disciplinar; Repositórios de dados de pesquisa multidisciplinar; Repositórios de dados de pesquisa específicos do projeto.

Sayão e Sales (2016) apresentam o elenco das funções dos repositórios de dados na qualidade de um sistema digital que integra diversas funções, tem como perspectiva oferecer um ambiente dinâmico e flexível — principalmente pela natureza heterogênea dos dados — para dar apoio à execução dos processos de gestão de dados de pesquisa. mas de forma ideal, (as funções) devem estar orientadas em uma política institucional explícita de gestão de dados de pesquisa, registrada em um documento disponível publicamente que deve refletir as opções e condicionantes das instituições, as demandas dos grupos de pesquisadores e as singularidades dos domínios disciplinares em que atuam.

Sales (2014) indica os benefícios dos repositórios de dados de pesquisa: visibilidade dos dados, compartilhamento de dados, crédito ao autor dos dados, preservação digital, memória científica e transferência, segurança dos dados, disponibilidade, curadoria digital, serviços inovadores, reuso dos dados, redes de repositórios e indicadores de qualidade e produtividade da instituição.

O contexto no qual os recursos são gerenciados e preservados visando manter os serviços de descoberta, acesso, compartilhamento e reuso dos dados de pesquisa, o ciclo de vida dos dados de pesquisa e o ciclo da curadoria, serão retratados na próxima seção.

4 O CICLO DE VIDA DOS DADOS DE PESQUISA E DA CURADORIA: A VISÃO DETALHADA DA GESTÃO

A importância do ciclo de vida para Green (2006) é que um ciclo implica em um ambiente no qual os recursos são gerenciados e preservados para descoberta e reaproveitamento. Os recursos são criados, disponibilizados e preservados para pesquisas, aprendizado e atividades de política subsequentes. O desafio do ciclo de vida dos dados é manter metadados e serviços que permitirão acesso, avaliação e reutilização ao longo do tempo. Além disso, ao diagramar diferentes visões dos ciclos de vida, temos uma maneira de reunir ideias sobre o ciclo de vida de programas de pesquisa a ideias sobre recursos e processos de dados (captura, gerenciamento e vinculação de recursos de dados). Visualizar o ciclo de vida completo da pesquisa nos permite identificar lacunas nos serviços.

Na sequência, são apresentados os modelos do ciclo de vida de maior destaque na literatura, tais como a DDI e a DataONE, que são ciclos de vida de dados de pesquisa, e o DCC, que se refere ao ciclo de vida da curadoria.

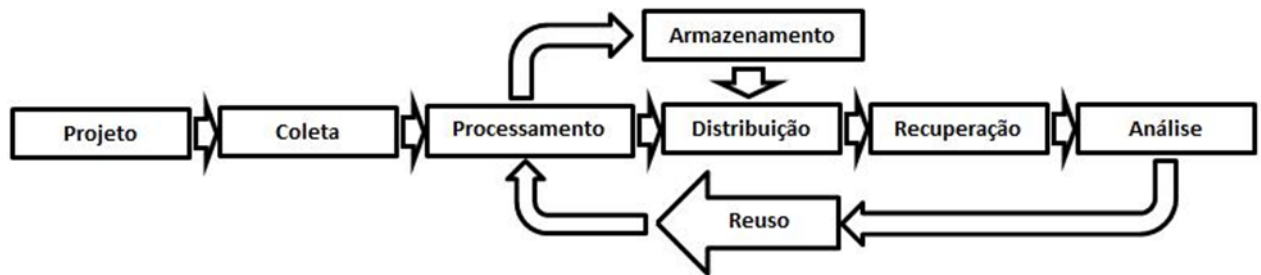
4.1 Data Documentation Initiative (DDI)

A DDI é uma iniciativa para definir e divulgar um padrão internacional de descrição de dados em ciências sociais. Esse padrão é definido em um formato XML que pode ser lido por humanos e máquinas. A DDI visa apoiar todo o ciclo de vida dos dados de pesquisa. Os metadados DDI estão relacionados ao desenho do estudo, coleta de dados, processamento e avaliação de dados, bem como ao uso secundário e arquivamento (ZENK-MÖLTGEN, 2012).

A DDI baseia-se no princípio de reutilização de metadados para eliminar redundâncias dispendiosas e apoiar a comparação explícita dentro dos estudos e entre eles. Como exemplo, categorias de resposta, conceitos e universos podem ser definidos uma vez e, em seguida, usados várias vezes por questões e variáveis. Grupos de pesquisa, como séries temporais, podem herdar informações comuns como questões ou variáveis centrais, alterando apenas os itens que mudam nas rodadas subsequentes de coleta de dados. A comparação implícita por meio da herança é complementada pelo suporte, o que permite descrever comparações explicitamente por meio de mapeamento (VARDIGAN; HEUS; THOMAS, 2008).

Vejamos a Figura 1, do ciclo de vida da DDI:

Figura 1 – Modelo do ciclo de vida da DDI



Fonte: Sant'Ana (2013).

Zenk-Möltgen (2012), da DDI Alliance⁶, esclarece que as oito fases não precisam ser executadas em uma sequência linear. A *fase de conceito* começa com um conceito de estudo, no qual a questão de pesquisa e a metodologia da investigação são determinadas.

Na *fase de coleta*, os dados são coletados e preparados. As fontes de coleta incluem pesquisas, censos, registros de votação ou de saúde, estatísticas de comércio ou coleções baseadas na web. As fontes de dados primárias e secundárias devem ser claramente distinguidas (BALL, 2012; MÖLLER, 2013).

Na fase de processamento, uma vez que os dados tenham sido coletados, eles são processados e analisados para produzir dados que respondam à questão da pesquisa. Esses dados podem ser registrados em uma forma legível por máquina ou por mãos humanas, como em um relatório técnico (BALL, 2012).

Na *fase de arquivamento*, a fim de garantir o acesso de longo prazo aos dados, eles devem ser passados para um arquivo em vez de apenas mantidos por pesquisadores. O arquivo não apenas preserva os dados (e metadados), mas também agrega valor a eles ao longo do tempo (BALL, 2012).

Na fase de distribuição subsequente, que trata da disseminação dos dados com o objetivo de acesso do usuário final aos metadados, dados e outras informações relacionadas ao ciclo de vida dos dados, os dados são disponibilizados aos usuários diretamente ou por meio de uma biblioteca ou arquivo de dados (BALL, 2012; MÖLLER, 2013).

Segue-se a *fase de Descoberta*, que torna os dados localizáveis, de modo que possam ser divulgados por meio de livros, publicações em periódicos, páginas da web ou outros serviços on-line (BALL, 2012; MÖLLER, 2013).

Na *fase de análise* dos dados da pesquisa, são feitas novas descobertas e adquiridos novos conhecimentos (ZENK-MÖLTGEN, 2012).

⁶ Disponível em: <https://ddialliance.org/>.

Na fase de reuso, os dados podem ser usados para outros fins, tais como uso secundário e realimentação para a fase de processamento (ZENK-MÖLTGEN, 2012).

Dessa maneira, Zenk-Möltgen (2012) considera que, neste modelo geral das fases, a ordem em que as fases individuais são usadas é completamente aberta e não é determinada na especificação DDI. Portanto, todos os outros caminhos através das fases são concebíveis e também podem ser documentados na DDI.

Vardigan, Heus e Thomas (2008) esclarecem que no ciclo de vida da DDI, quantidades extensas de metadados são geradas ao longo da vida de uma pesquisa típica de ciências sociais, começando com uma articulação dos conceitos a serem estudados e continuando com a proposta e financiamento para conduzir o estudo. O próprio esforço de coleta de dados e disseminação, descoberta de dados e finalmente análise de dados leva a novas descobertas e conhecimentos que enriquecem a literatura das ciências sociais.

Na concepção de Vardigan, Heus e Thomas (2008), a DDI permite descrever o ciclo de vida dos dados e aumentar a quantidade e os tipos de metadados que junto com os dados ampliam o contexto, contribuindo assim para a avaliação da qualidade dos dados.

4.2 Data Observation Network for Earth (DataONE)

A área ambiental tem crescido em relevância e em foco e representa uma fronteira de pesquisa sobre acesso a dados, principalmente em função do grande volume de dados gerados, volume este que tende a crescer de forma exponencial em função da disseminação de sensores e coletores de dados cada vez mais acessíveis (SANT'ANA, 2013, p. 11).

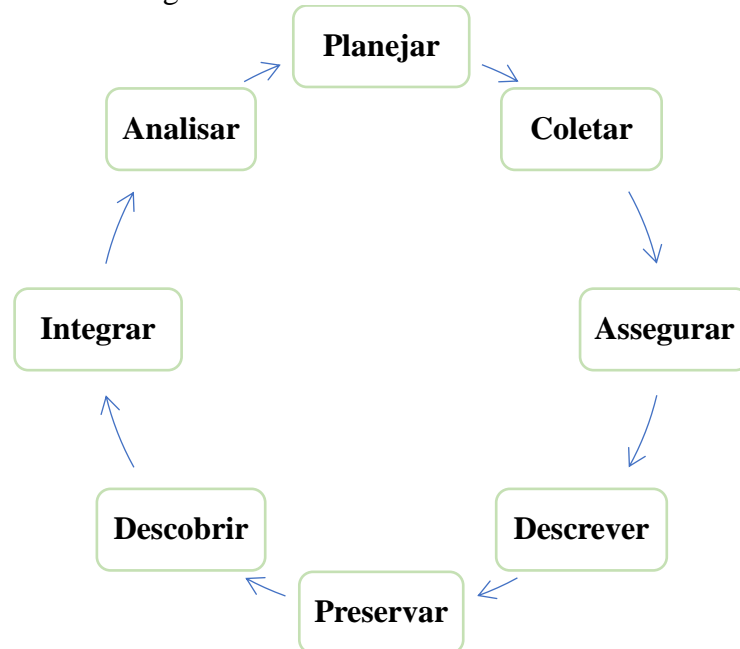
DataONE⁷ é um programa administrado pela comunidade que fornece acesso a dados em vários repositórios de membros, apoiando a pesquisa e descoberta aprimoradas de dados da Terra e ambientais. A DataONE promove as melhores práticas em gerenciamento de dados por meio de recursos e materiais educacionais (DATAONE, 2021).

Sant'Ana (2013) refere-se à DataONE como um exemplo desta nova ciência ambiental inovadora, que busca atuar através de estruturas distribuídas e sustentáveis que atendam às necessidades da ciência e da sociedade de acesso aberto. Para atender essa demanda, a DataONE propõe um modelo de ciclo de vida de dados que possa nortear o planejamento e realização da gestão desses dados.

⁷ Disponível em: <https://www.dataone.org/>.

O ciclo de vida desenvolvido pela DataONE nos mostra uma visão geral com oito componentes. Na Figura 2, podemos visualizar o ciclo de vida dos dados de pesquisa DataONE.

Figura 2 – Ciclo de vida da DataONE.



Fonte: Adaptado de DataONE (2021).

- a) **Planejar:** o planejamento se refere ao mapeamento dos processos e à ciberinfraestrutura (metodológica e tecnológica) necessária para a execução de todos os estágios do ciclo de vida. Recomenda-se como uma boa prática adotar uma ferramenta de planejamento para detalhar o projeto de dados. Está relacionado ao plano de pesquisa que será desenvolvido pelo pesquisador que optou pelo gerenciamento dos seus dados de pesquisa. Todas as etapas e procedimentos metodológicos relativos ao projeto de pesquisa devem ser definidos neste estágio (OLIVEIRA *et al.*, 2015; OLIVEIRA; SILVA, 2016);
- b) **coletar:** coleta dos dados primários. Apoia todo o processo do ciclo dos dados desde a sua origem. Os dados primários são oriundos de investigações de campo, observacionais ou experimentais, desenvolvidas por pesquisadores ou grupos de pesquisa. Os dados de pesquisa são coletados de acordo com as metodologias de cada pesquisa. A natureza dos dados se relaciona com o projeto de pesquisa que os originou. Dessa forma, além da coleta, o pesquisador deve se preocupar com a curadoria em curto prazo dos dados de pesquisa. Para auxiliar nessa curadoria, é recomendado que se adote uma infraestrutura inicial com ferramentas de comunicação geral, *drive*, software específico e equipamentos laboratoriais que

condicionem e assegurem a proteção dos dados primários (OLIVEIRA *et al.*, 2015; OLIVEIRA; SILVA, 2016);

- c) **assegurar**: realizados os procedimentos preliminares de qualidade e validação, transferem-se os dados de pesquisa para uma ferramenta compatível que favoreça o compartilhamento, curadoria e preservação a longo prazo em uma única interface. Estimula-se neste estágio a adoção de identificadores para os dados digitais, a exemplo do *Digital Object Identifier* (DOI), um sistema de identificação de objetos digitais on-line que, além de assegurar a qualidade e identificação, permite o gerenciamento dos direitos de propriedade intelectual de dados e objetos digitais e a adoção de critérios e procedimentos que garantam a qualidade e validação dos dados em todos os estágios e o gerenciamento a longo prazo (OLIVEIRA *et al.*, 2015; OLIVEIRA; SILVA, 2016);
- d) **descrever**: é de extrema relevância, e a escolha de um padrão de metadados adequado para a descrição dos dados de pesquisa é essencial. Na literatura, já estão disponíveis inúmeros padrões de metadados específicos que podem ser adotados para recursos informacionais (impressos e eletrônicos) e para diferentes tipos de dados de pesquisa (OLIVEIRA *et al.*, 2015; OLIVEIRA; SILVA, 2016);
- e) **preservar**: a adoção de um plano de preservação dos dados primários a curto e longo prazos é recomendada neste estágio. A preservação de dados a longo prazo tem a finalidade de proporcionar *backups* e arquivamento, por meio de processos, tais como: conversão, reformatação e salvamento de dados para reuso no futuro, além de pensar no descarte digital a longo prazo. A preservação realizada em um repositório especializado permite o armazenamento a longo prazo dos dados de pesquisa (OLIVEIRA; SILVA, 2016);
- f) **descobrir**: estimula-se a elaboração de parcerias institucionais para garantir que os dados tenham uma maior visibilidade na pesquisa e permitir o compartilhamento e a colaboração científica. Recomenda-se a adoção de repositórios de dados, *datasets* e registros em diretórios relacionados à curadoria e preservação de dados de pesquisa. Reforça-se a importância de os metadados estarem disponíveis para serem localizados e serem potencialmente passíveis de uso. A partir da integração com o repositório especializado, é também possível acessar outros diretórios de dados (OLIVEIRA; SILVA, 2016);
- g) **integrar**: a adoção dos protocolos de interoperabilidade para a integração dos metadados tem a finalidade de prover visibilidade, novas análises e investigações

neste estágio. Podem-se adotar as ferramentas disponíveis no OAIS com o intuito de preservar e disponibilizar, por meio de seus protocolos de interoperabilidade, dados, informações e documentos relacionados ao projeto de pesquisa. Além disso, a integração do projeto de dados com outros recursos de dados que sejam capazes de proporcionar novas análises e investigações cumpre dois propósitos: 1) agregação interna dos diferentes dados ou conjuntos de dados gerados na pesquisa científica em um mesmo ambiente; 2) integração do projeto de dados com outras fontes externas (OLIVEIRA *et al.*, 2015; OLIVEIRA; SILVA, 2016);

- h) **analisar**: este estágio está voltado para a análise e as avaliações que podem ser realizadas entre os dados de pesquisa por meio de software, câmaras de compensação de dados e ferramentas analíticas e métricas especializadas. Todo o processo de pesquisa já está completo, e o pesquisador pode fazer derivações dos seus dados através de publicações em jornais, repositórios, *datasets* e portais de dados. Sugere-se a adoção de ferramentas de alta performance para a análise métrica e visualização de dados (OLIVEIRA; SILVA, 2016).

4.3 Digital Curation Centre (DCC)

O DCC consiste num modelo para alinhar as tarefas de curadoria com os estágios do ciclo de vida de um objeto digital, sendo uma ferramenta de planejamento para criadores de dados, curadores e usuários (BALL, 2012).

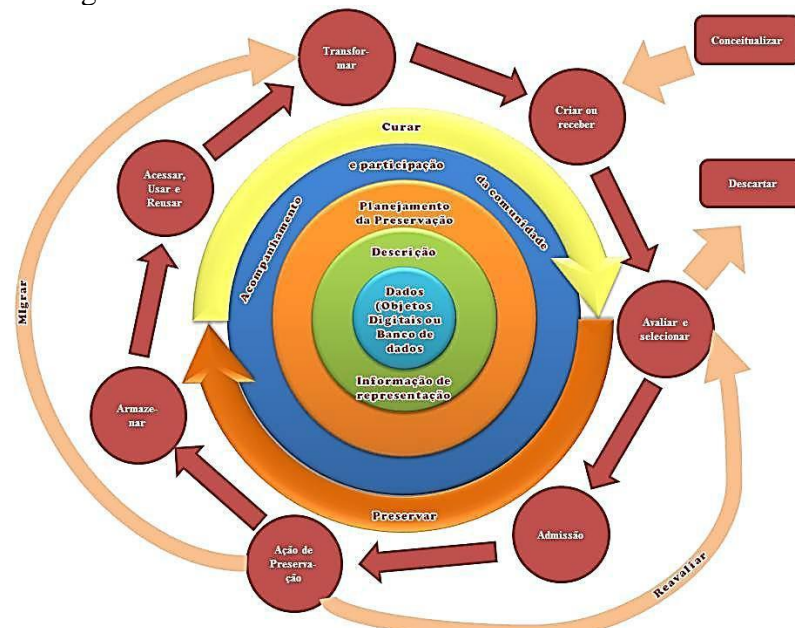
O modelo DCC, segundo Higgins (2008), identifica entre as ações de curadoria, que são aplicáveis a todo o ciclo de vida digital, aquelas que precisam ser realizadas sequencialmente para que a curadoria seja bem-sucedida e aquelas que são realizadas ocasionalmente, conforme as circunstâncias o exigirem.

No núcleo do modelo, encontram-se os dados em formato digital, compreendidos como objetos digitais como arquivos textuais, de imagem, de som, em conjunto com identificadores e metadados relacionados. Já os objetos digitais complexos são uma combinação de objetos digitais. O modelo conta também com coleções estruturadas de dados armazenados em um sistema de computador — neste caso, o banco de dados (HIGGINS, 2008).

No centro do modelo estão os dados digitais, que aqui são identificados com objetos ou bancos de dados digitais simples e complexos. O modelo considera três níveis de ações do ciclo de vida completo (BALL, 2012):

- a) **descrição da informação de representação:** definir os metadados administrativos, descritivos, técnicos, estruturais e de preservação, utilizando padrões apropriados, para garantir descrição e controle a longo prazo. Definir a forma de atribuição de informações de representação pertinente para compreender e contextualizar o material digital;
- b) **planejamento da preservação:** planejar as estratégias de preservação a serem aplicadas no decorrer de todo o ciclo de vida da curadoria de material digital;
- c) **acompanhamento e participação da comunidade:** motivar e envolver a comunidade interna (produtores do conteúdo/dados, desenvolvedores, pessoas da equipe) e externa (comunidade em geral, usuários da informação) no ciclo de curadoria para ajudar na representação da informação, definir a melhor forma de acesso aos dados, colaborar na definição de padrões, ferramentas e produtos de software apropriados, entre outros;
- d) **curar e preservar:** estar consciente de ações administrativas e gerenciais planejadas e intencionais para contribuir com a curadoria e a preservação no decorrer de todo o ciclo de vida da curadoria e realizá-las. Tratar questões de direitos autorais e permissões de acesso (ARAÚJO *et al.*, 2019).

Figura 3 – Modelo do ciclo de vida dos dados do DCC



Fonte: Yamaoka (2012).

As ações sequenciais não se preocupam exclusivamente com a curadoria, mas representam estágios do ciclo de vida dos dados que devem ter um componente de curadoria (BALL, 2012):

- a) **conceitualização**: conceber e planejar a criação do dado, considerando também os métodos de captura e as opções de armazenamento. Questões tais como propriedade intelectual, embargos e restrições, financiamento, responsabilidades, objetivos específicos da pesquisa, ferramentas de captura e calibração devem ser registradas;
- b) **criação e/ou recebimento**: compreende a criação do dado, incluindo o elenco de metadados necessários à sua gestão e compreensão, ou seja, metadados administrativos, descritivos, estruturais e técnicos (os metadados de preservação também podem ser incluídos no momento da criação do dado). Nem sempre os dados são arquivados por quem os gerou, e, portanto, este estágio compreende também a recepção dos dados segundo políticas bem documentadas, seja dos seus criadores, de outros arquivos, de repositórios ou centros de dados. Quando necessário, assinalar metadados apropriados para a curadoria e a preservação dos dados recepcionados;
- c) **avaliação e seleção**: avaliar o dado e selecionar o que será objeto dos processos de curadoria e de preservação por longo prazo. Manter-se aderente tanto às boas práticas quanto às políticas pertinentes e também às exigências legais;
- d) **arquivamento**: transferir o dado para um arquivo, repositório, centro de dados ou outro custodiante apropriado;
- e) **ações de preservação**: promover ações para assegurar a preservação de longo prazo e a retenção do dado de natureza oficial. As ações de preservação devem garantir que o dado permaneça autêntico, confiável e capaz de ser usado enquanto mantém sua integridade. Essas ações de preservação incluem: a limpeza do dado e a sua validação, a adição de metadados de preservação e de informação de representação e a garantia de estruturas de dados ou formatos de arquivos aceitáveis;
- f) **armazenamento**: armazenar o dado de forma segura, mantendo a aderência aos padrões relevantes;
- g) **acesso, uso e reuso**: garantir que o dado possa ser cotidianamente acessado tanto pela sua comunidade-alvo quanto pelos demais usuários interessados no reuso do dado. Isso pode ser realizado por meio de publicação disponível às

comunidades interessadas. Porém, controle de acesso e procedimentos de autenticação também podem ser aplicados;

- h) **transformação**: compreende a criação de novos dados a partir do original, por exemplo, pelo processo de migração para diferentes formatos ou pela criação de subconjuntos — realizada por meio de seleção ou formulação de consultas —, derivando novos resultados que podem ser publicados (SAYÃO; SALES, 2012, p. 186).

As ações derradeiras são as ocasionais, começando pelo descarte dos dados que não foram selecionados para curadoria e preservação de longo prazo de acordo com as políticas documentadas, orientações ou requisitos legais. Normalmente, os dados podem ser transferidos para outro arquivo, repositório, centro de dados ou outro custodiante. Em alguns casos, os dados são destruídos. A natureza dos dados pode, por razões legais, exigir sua destruição segura. Já as ações de reavaliar os dados e migrá-los para um formato diferente podem ser feitas de acordo com o ambiente de armazenamento ou para garantir a imunidade dos dados contra a obsolescência do hardware ou software (HIGGINS, 2008):

- a) **eliminação**: eliminar os dados que não foram selecionados para curadoria e preservação de longo prazo de acordo com políticas documentadas, diretrizes ou exigências legais;
- b) **reavaliação**: retornar ao dado cujos procedimentos de avaliação foram falhos para nova avaliação e possível seleção para curadoria;
- c) **migração**: migrar os dados para um formato diferente. Isso pode ser feito no sentido de compatibilizá-lo com o ambiente de armazenamento ou para assegurar a imunidade do dado contra a obsolescência de hardware e de software (SAYÃO; SALES, 2012).

A curadoria digital de dados é um conjunto de ações com um ciclo próprio que visa preservar e documentar a origem e a autenticidade dos dados, descrevendo, registrando e catalogando seus conteúdos, providenciando armazenamento e preservação a longo prazo.

Por outro lado, o o conhecimento teórico e prático do arquivista em gestão e preservação de longo prazo. Atividades inerentes ao profissional podem ser aplicadas na gestão de dados de pesquisa. A seguir veremos a teoria e prática arquivística.

5 CONHECIMENTO TEÓRICO E PRÁTICO DO ARQUIVISTA

Conservar os arquivos, segundo Delmas (2010), é um ato indispensável. Eles são o produto necessário do funcionamento de toda sociedade organizada. Quanto mais uma sociedade se desenvolve, mais as atividades humanas são numerosas, diversificadas e interdependentes. Quanto mais documentos são usados pelos homens para registrar seus atos, assegurar a sua continuidade e estabelecer relacionamentos duráveis entre si mesmos, mais eles produzem e conservam arquivos.

“A gestão dos arquivos, a profissão de arquivista e a disciplina arquivística desenvolveram-se consideravelmente ao longo das últimas décadas, e a quantidade de serviços de arquivos e de arquivistas aumenta constantemente” (ROUSSEAU; COUTURE, 1998, p. 23).

5.1 Do convencional ao digital: compreendendo o documento arquivístico

Na perspectiva de Delmas (2010), os arquivos servem para provar, lembrar-se, compreender e identificar-se. Provar seus direitos é uma utilidade jurídica e judiciária. Lembrar-se é uma utilidade de gestão. Compreender é uma utilidade científica de conhecimento. Identificar-se pela transmissão da memória é uma utilidade social. Essas utilidades desenvolveram-se em paralelo, porém cada uma delas sofreu transformações diversas dependendo da época, em razão da própria evolução da sociedade. Na Idade Média, a função de prova dos direitos era o principal motivo da conservação dos documentos, enquanto a da memória ia se desenvolvendo. No Renascimento, surgiu a função de compreender, que iria se desenvolver continuamente, enquanto a de identificar-se foi impulsionada, como sabemos, só na época contemporânea.

Dito isso, o documento arquivístico nasce de forma involuntária para atender e cumprir uma necessidade, isto é, “a circunstância do documento implica tanto a presença de um fato e da vontade de manifestá-lo quanto a vontade de dar origem ao ato” (BELLOTTO, 2005, p. 59). Neste sentido, “o documento de arquivo é produzido de forma involuntária, criado no decurso de uma atividade. É o resíduo material da ação que lhe dá origem” (RODRIGUES, 2006, p. 34).

Com base em Dirks, citado por Rondinelli (2002), os documentos arquivísticos são o resultado de ações específicas em momentos específicos, e, para servir de prova, eles precisam ser mantidos inalterados. Com base nessa definição, para constituir prova, o documento arquivístico tem que ser:

- a) **autêntico**: o documento é exatamente o que diz ser. Para tanto, precisa ser protegido de intervenções não autorizadas e ter sua tramitação controlada;
- b) **fidedigno**: o conteúdo do documento representa verdadeiramente a transação a que se refere. Nesse sentido, deve ser criado e capturado na hora certa, pela pessoa que conhece a transação, ou gerado automaticamente por processos consagrados pela instituição;
- c) **completo e inalterado**: o documento tem que ser protegido de intervenções não autorizadas, sendo que acréscimos, anotações e eliminações permitidos devem ser monitorados e rastreados;
- d) **usável**: o documento tem que ser passível de localização, recuperação e interpretação pelo tempo em que seu valor de prova persistir.

Para Delmas (2010), o documento de arquivo é o resultado de uma ação passada que se pretende guardar, ao longo do tempo, por um prazo mais ou menos extenso, para necessidades futuras. A guarda dos arquivos se insere na duração, no tempo longo da vida dos homens e das instituições, ao contrário das necessidades da comunicação, da informação e, em geral, da documentação, que são orientadas para a atualidade e o imediato.

Os documentos arquivísticos possuem características que os distinguem dos demais documentos. Destacam-se a imparcialidade, a autenticidade, a naturalidade, o inter-relacionamento ou organicidade e a unicidade.

A característica da imparcialidade diz respeito à veracidade do documento, tendo em vista que ele não tem a intenção de ser produzido, mas nasce no decorrer de uma atividade e da necessidade de registrar. Quem produz o documento é livre de ideias pré-concebidas (DURANTI, 1994).

A autenticidade se refere à credibilidade de um documento enquanto documento, isto é, a qualidade de um documento ser o que diz ser e estar livre de adulteração ou qualquer outro tipo de corrupção. A autenticidade está vinculada ao *continuum* da criação, manutenção e custódia. Os documentos são autênticos porque são criados tendo em vista a necessidade de agir através deles, são garantias para futuras ações ou para informação e “são definitivamente separados para preservação, tacitamente julgados dignos de serem conservados” por seu criador ou legítimo sucessor como “testemunhos escritos de suas atividades no passado”. Assim, os documentos são autênticos porque são criados, mantidos e conservados sob custódia de acordo com procedimentos regulares que podem ser comprovados (CONARQ, 2009; DURANTI, 1994).

A terceira característica dos documentos é a naturalidade. A naturalidade diz respeito à maneira como eles se acumulam no decurso das transações de acordo com as necessidades da matéria em pauta: eles não são “coletados artificialmente, [...], mas acumulados naturalmente em função dos objetivos práticos da administração [...], isto é, de se acumularem de maneira contínua e progressiva” (DURANTI, 1994, p. 52).

O inter-relacionamento é também conhecido como organicidade. Esse inter-relacionamento se deve ao fato de que os documentos estabelecem relações no decorrer do andamento das transações e de acordo com as necessidades destas. Cada documento está intimamente relacionado com outro, tanto dentro quanto fora do grupo no qual está preservado, e seu significado depende dessas relações que os documentos arquivísticos guardam entre si e que expressam as funções e atividades da pessoa ou organização que os produziu. A organicidade se constitui em um atributo essencial para que um determinado conjunto de documentos seja considerado um arquivo (CONARQ, 2009; DURANTI, 1994).

A quinta e última característica é a unicidade, que advém do fato de que cada registro documental ocupa um lugar único na estrutura documental do grupo ao qual pertence e no universo documental (DURANTI, 1994).

Conduzindo a temática para o contexto digital, Rondinelli (2002) afirma que no mundo digital tudo é codificado em linguagem binária e, para se tornar acessível aos olhos humanos, precisa da intermediação de programas computacionais igualmente codificados em *bits*, numa sofisticação tecnológica que passa despercebida à maioria dos usuários. Dito isso, o documento arquivístico constitui o registro de ações humanas, independentemente da forma como se apresenta e da base em que se encontra afixado. No caso específico da base, o documento arquivístico pode se apresentar em qualquer suporte. Por essas assertivas, infere-se que documentos digitais, gerados no decurso de atividades desempenhadas por pessoas físicas e jurídicas e em suportes tão diferentes como os magnéticos e ópticos, também podem ser documentos arquivísticos.

Na seção seguinte, será apresentada a gestão de documentos, que cobre o ciclo de vida do documento desde a sua produção até a destinação final (eliminação ou guarda permanente).

5.2 Gestão de documentos arquivísticos dos convencionais aos digitais

Em meados do século XIX, começa a desabrochar um crescente interesse pelo valor histórico dos arquivos e os documentos ganham o status de testemunho da história. O trabalho dos arquivistas da época se concentra, basicamente, na organização e utilização, mas, a partir

da Segunda Guerra Mundial, em decorrência do progresso científico e tecnológico alcançado pela humanidade, a produção de documentos cresceu a níveis tão elevados que superou a capacidade de controle e organização das instituições, as quais se viram forçadas a buscar novas soluções para gerir as grandes massas documentais acumuladas nos arquivos (PAES, 2004).

Os Estados Unidos, entre outros países anglo-saxônicos, são considerados pioneiros, desde os anos 1940, na elaboração do conceito de gestão de documentos (*records management*) cuja ótica, inicialmente, era nitidamente mais administrativa e econômica do que arquivística, uma vez que se tratava, essencialmente, de otimizar o funcionamento da administração, limitando a quantidade de documentos produzidos e o prazo de guarda, o que se confirma na própria definição do conceito na legislação americana (INDOLFO, 2007).

A propósito, Jardim (1987) esclarece que a legislação norte-americana estabelece a seguinte definição: o planejamento, o controle, a direção, a organização, a capacitação, a promoção e outras atividades gerenciais relacionadas com a criação de documentos, sua manutenção, uso e eliminação, incluindo o manejo de correspondência, formulários, diretrizes, informes, documentos informáticos, microformas, recuperação de informação, fichários, correios, documentos vitais, equipamentos e materiais, máquinas reprográficas, técnicas de automação e elaboração de dados, preservação e centros de arquivamento intermediários ou outras instalações para armazenagem.

A gestão de documentos é um processo arquivístico que busca intervir nas fases que compõem o ciclo vital dos documentos, determinando o tempo necessário de permanência nas fases de arquivamento (corrente e intermediária) até que seja dada a destinação final (eliminação ou guarda permanente), para que se possa alcançar a eficácia, eficiência e o menor custo na gestão de documentos e arquivos. Tal intervenção é impulsionada por motivos ligados a praticidade, agilidade, imediatismo ou economia (SILVA *et al.*, 1999).

Considera-se gestão de documentos o conjunto de procedimentos e operações técnicas referentes a sua produção, tramitação, uso, avaliação e arquivamento em fase corrente e intermediária, visando a sua eliminação ou recolhimento para guarda permanente (BRASIL, 1991).

Na concepção de Bernardes e Delatorre (2008), a gestão documental deverá definir normas e procedimentos técnicos referentes a produção, tramitação, classificação, avaliação, uso e arquivamento dos documentos durante todo o seu ciclo de vida (idade corrente, idade intermediária e idade permanente), com a definição de seus prazos de guarda e de sua destinação final, requisitos necessários, inclusive, para o desenvolvimento de sistemas informatizados de gestão de informações.

Paes (2004) destaca três fases básicas da gestão de documentos:

- a) **produção de documentos:** refere-se à elaboração dos documentos em decorrência das atividades de um órgão ou setor. Nessa fase, o arquivista deve contribuir para que sejam criados apenas documentos essenciais à administração da instituição e evitadas duplicação e emissão de vias desnecessárias, propor consolidação de atos normativos alterados ou atualizados com certa frequência, visando à perfeita compreensão e interpretação dos textos, sugerir criação ou extinção de modelos e formulários, apresentar estudos sobre a adequação e o melhor aproveitamento de recursos reprográficos e informáticos, contribuir para a difusão de normas e informações necessárias ao bom desempenho institucional, opinar sobre escolha de materiais e equipamentos, participar da seleção de recursos humanos que deverão desempenhar tarefas arquivísticas e afins;
- b) **utilização de documentos:** esta fase abrange as atividades de protocolo (recebimento, classificação, registro, distribuição, tramitação), de expedição, de organização e arquivamento de documentos em fase corrente e intermediária, bem como a elaboração de normas de acesso à documentação (empréstimo, consulta) e à recuperação das informações, indispensáveis ao desenvolvimento de funções administrativas, técnicas ou científicas das instituições;
- c) **avaliação e destinação de documentos:** talvez seja a mais complexa das três fases de gestão de documentos, desenvolvendo-se mediante a análise e avaliação dos documentos acumulados nos arquivos, com vistas a estabelecer seus prazos de guarda, determinando quais serão objeto de arquivamento permanente e quais deverão ser eliminados pela mensuração do seu valor de prova e de informação para a instituição (PAES, 2004).

Na perspectiva de Jardim (1987), a gestão cobre todo o ciclo de existência dos documentos, e, para Indolfo (2007), a gestão de documentos pressupõe essa intervenção no ciclo de vida dos documentos por intermédio de um conjunto das operações técnicas e processos que governam todas as atividades dos arquivos correntes e intermediários e que são capazes de controlar e racionalizar as atividades desde a produção e uso até a destinação final dos documentos.

O ciclo vital e a teoria das três idades são complementares. A teoria das três idades consiste em uma interpretação do conceito do ciclo vital, sobretudo no que concerne a uma tentativa de melhor aplicação prática. Com base em Bellotto (2005), a primeira idade, dos arquivos correntes, é constituída de documentos que estão em uso funcional, administrativo,

jurídico, e sua tramitação legal e utilização estão ligadas às razões pelas quais foram criados. A permanência de tais documentos nesse tipo de arquivo depende de sua tipologia/função.

A segunda idade, o arquivo intermediário, é constituída de documentos que deixam de ser frequentemente consultados, mas cujos órgãos que os receberam e produziram podem ainda solicitá-los. É nessa fase que os documentos são submetidos às tabelas de temporalidade, que determinam seus prazos de vigência e de vida, ou seja, eliminação ou guarda permanente (BELLOTTO, 2005; PAES, 2004).

A terceira idade, arquivo permanente, é constituída dos documentos que perderam todo valor de natureza administrativa, que se conservam em razão de seu valor histórico ou documental e que constituem os meios de conhecer o passado e sua evolução, iniciando os usos científico, social e cultural dos documentos (BELLOTTO, 2005; PAES, 2004).

Na concepção de Indolfo (2007), os documentos arquivísticos possuem valores a eles atribuídos, primário ou secundário, segundo o seu potencial de uso. O valor primário refere-se ao uso administrativo, razão primeira da criação do documento, o que pressupõe o estabelecimento de prazos de guarda ou retenção anteriores à eliminação ou ao recolhimento para guarda permanente — portanto, ao período de uso do documento para o cumprimento dos fins administrativos, legais ou fiscais. O valor secundário refere-se ao uso dos documentos para outros fins que não aqueles para os quais eles foram, inicialmente, criados, passando a ser considerados fontes de pesquisa e informação para terceiros e para a própria administração, pois contêm informações essenciais sobre matérias com as quais a organização lida para fins de estudo.

Rondinelli (2005) esclarece que o documento arquivístico deverá estar inserido num sistema capaz de executar as seguintes operações:

- a) **captura** determina formalmente que os documentos devem ser capturados pelo sistema;
- b) **registro**: formaliza a captura por meio de um identificador e uma breve descrição do documento, tal como data, hora, título, assunto;
- c) **classificação e indexação**: identifica as atividades relacionadas ao documento e o liga a outros documentos de maneira a facilitar a descrição, o controle, a destinação e o acesso;
- d) **acesso e segurança**: estipula direitos ou restrições para uso de determinados documentos;
- e) **avaliação**: identifica o período de retenção do documento no momento de sua captura e registro;

- f) **armazenamento**: mantém os documentos pelo tempo em que possam ser exigidos legalmente;
- g) **uso e rastreamento**: assegura que apenas determinadas funcionárias possam ter acesso a determinados documentos e que tal acesso possa ser rastreado como medida de segurança;
- h) **destinação**: identifica documentos com temporalidade semelhante, revê todo o uso desses documentos para confirmar ou alterar sua temporalidade e mantém um termo de eliminação para fins de auditoria.

O Conselho Nacional de Arquivos (CONARQ, 2011) considera que a gestão arquivística compreende a responsabilidade de um órgão ou entidade produtora em assegurar que a documentação produzida, independentemente do suporte, seja o registro fiel das suas atividades e que os documentos permanentes sejam preservados, garantindo que os documentos arquivísticos sejam confiáveis, autênticos, acessíveis e compreensíveis. Ademais, o conselho define a especificação de requisitos funcionais e de metadados para sistemas de gestão arquivística de documentos, no caso de documentos arquivísticos digitais.

Por fim, os procedimentos e operações da gestão arquivística de documentos são apoiados pelo plano de classificação, tabela de temporalidade e destinação, manual de gestão arquivística de documentos e esquema de classificação, sendo estes os instrumentos principais da gestão. Há também os instrumentos adicionais: glossário, vocabulário controlado, tesouro, entre outros. Por sua vez, na seção seguinte será exposto o Modelo de requisitos para sistemas informatizados de gestão arquivística de documentos – e ARQ Brasil.

5.3 Modelo de Requisitos para sistemas informatizados de gestão arquivística de documentos (e-ARQ Brasil)

O e-ARQ Brasil é uma especificação de requisitos a serem cumpridos pela organização produtora recebedora de documentos, pelo sistema de gestão arquivística e pelos próprios documentos, a fim de garantir sua confiabilidade e autenticidade, assim como sua acessibilidade. Além disso, ele pode ser usado para orientar a identificação de documentos arquivísticos digitais, estabelecendo requisitos mínimos para um SIGAD, independentemente da plataforma tecnológica em que for desenvolvido e implementado (CONARQ, 2011).

O e-ARQ Brasil especifica todas as atividades e operações técnicas da gestão arquivística de documentos, desde a produção, tramitação, utilização e arquivamento até a sua destinação final. Todas essas atividades poderão ser desempenhadas pelo SIGAD, o qual, tendo

sido desenvolvido em conformidade com os requisitos aqui apresentados, conferirá credibilidade à produção e à manutenção de documentos arquivísticos (CONARQ, 2011).

O e-ARQ Brasil (CONARQ, 2011) estabelece que o SIGAD deve ser capaz de gerenciar, simultaneamente, os documentos digitais e os convencionais. No caso dos documentos convencionais, o sistema registra apenas as referências sobre os documentos, e, para os documentos digitais, a captura, o armazenamento e o acesso são feitos por meio do SIGAD.

O conjunto de requisitos propostos no e-ARQ Brasil fornece a base para a implementação de um Sistema informatizado de gestão arquivística de documentos (isto é, SIGAD), comportando, assim, os requisitos para desenvolver um ambiente confiável para a gestão de documentos arquivísticos. Dessa forma, as fases corrente e intermediária serão contempladas pelo SIGAD e, após o cumprimento dos prazos estabelecidos, este deverá exportar os documentos de valor permanente para um RDC-Arq (FLORES; ROCCO; SANTOS, 2016). Ao passo que, na seção seguinte, será apresentado o sistema informatizado de gestão arquivística de documentos – SIGAD.

5.4 Sistema informatizado de gestão arquivística de documentos (SIGAD)

O SIGAD é um conjunto de procedimentos e operações técnicas, característico do sistema de gestão arquivística de documentos processados por computador. Pode compreender um software particular, um determinado número de softwares integrados, adquiridos ou desenvolvidos por encomenda, ou uma combinação dos dois. Para mais, o sucesso do SIGAD dependerá da implementação prévia de um programa de gestão arquivística de documentos (CONARQ, 2011).

O CONARQ (2011) estabelece que o SIGAD deve contar com procedimentos e operações técnicas do sistema de gestão arquivística de documentos digitais e convencionais, a saber:

- a) **captura:** consiste em declarar um documento como um documento arquivístico, incorporando-o ao sistema de gestão arquivística por meio das seguintes ações: registro, classificação, indexação, atribuição de restrição de acesso e arquivamento;
- b) **avaliação, temporalidade e destinação:** é uma atividade vital em um programa de gestão arquivística de documentos, pois permite racionalizar o acúmulo de documentos nas fases corrente e intermediária, facilitando a constituição dos arquivos permanentes;

- c) **pesquisa, localização e apresentação dos documentos:** o sistema de gestão arquivística deve prever funções de recuperação dos documentos e das informações neles contidas e acesso a eles, de forma a facilitar a condução das atividades e satisfazer os requisitos relativos à transparência do órgão ou entidade. A recuperação engloba pesquisa, localização e apresentação dos documentos;
- d) **segurança:** controle de acesso, trilhas de auditoria e cópias de segurança: o sistema de gestão arquivística deve prever controle de acesso e procedimentos de segurança que garantam a integridade dos documentos. Entre esses procedimentos, podem ser destacados: o uso de controles técnicos e programáticos, distinguindo entre diferentes tipos de documentos; perfis de usuários e características de acesso aos dados; e a manutenção de trilhas de auditoria e de rotinas de cópia de segurança;
- e) **armazenamento:** as considerações e ações relativas ao armazenamento dos documentos arquivísticos convencionais e digitais permeiam todo o seu ciclo de vida. Esse armazenamento deve garantir a autenticidade dos documentos e o acesso a estes pelo tempo estipulado na tabela de temporalidade e destinação;
- f) **preservação:** os documentos arquivísticos têm de se manter acessíveis e utilizáveis pelo tempo que for necessário, garantindo-se sua longevidade, funcionalidade e acesso. Devem ser asseguradas as características dos documentos, tais como autenticidade e acessibilidade, pela adoção de estratégias institucionais e técnicas proativas de produção e preservação que garantam sua perenidade. Essas estratégias são estabelecidas por uma política de preservação (CONARQ, 2011).

O SIGAD conta com especificações de requisitos, ou seja, aspectos e funcionalidades, que são: organização dos documentos arquivísticos; plano de manutenção dos documentos (configuração e plano de classificação no SIGAD, classificação e metadados das atividades de arquivamento, gerenciamento dos dossiês/processos, requisitos adicionais para o gerenciamento de processos, volumes: abertura, encerramento e metadados; gerenciamento de documentos e processos/dossiês arquivísticos convencionais e híbridos); tramitação e fluxo de trabalho (controle de fluxo de trabalho, controle de versões e status do documento); captura (procedimentos gerais, captura em lote, captura de mensagens de correio eletrônico, captura de documentos convencionais ou híbridos, formato de arquivo e estrutura dos documentos a serem capturados e estrutura dos procedimentos de gestão); avaliação e destinação (configuração da tabela de temporalidade e destinação de documentos, aplicações da tabela de temporalidade e destinação de documentos, exportação de documentos, eliminação e avaliação e destinação de documentos arquivísticos convencionais e híbridos); pesquisa, localização e apresentação dos

documentos (aspectos gerais, pesquisa e localização, apresentação: visualização, impressão e emissão de som); segurança (cópias de segurança, controle de acesso, classificação da informação quanto ao grau de sigilo e restrição de acesso à informação sensível, trilha de auditoria, assinaturas digitais, criptografia, marcas d'água digitais, acompanhamento de transferência, autoproteção e alteração, exclusão e truncamento de documentos arquivísticos digitais); armazenamento (durabilidade, capacidade e efetividade de armazenamento); preservação (aspectos físicos, aspectos lógicos e aspectos gerais); funções administrativas; conformidade com a legislação e regulamentações; usabilidade; interoperabilidade; disponibilidade; desempenho e escalabilidade (CONARQ, 2011).

A identificação dos metadados referidos no e-ARQ Brasil teve como uma das premissas a complementação dos requisitos do SIGAD, compreendendo a identificação de: documentos simples, processos ou dossiês, que podem se apresentar em formato convencional, híbrido ou digital, e as ações de gerenciamento do seu ciclo de vida (CONARQ, 2011).

Metadados, segundo o Glossário da Câmara Técnica de documentos eletrônicos, são definidos como: “dados estruturados que descrevem e permitem encontrar, gerenciar, compreender e/ou preservar documentos arquivísticos ao longo do tempo” (CONARQ, 2009, p. 17). O e-ARQ Brasil definiu metadados para as entidades: documento (refere-se aos documentos arquivísticos que são gerenciados pelo SIGAD); evento de gestão (refere-se às ações de gestão que ocorrem com os documentos arquivísticos ao longo de seu ciclo de vida); classe (refere-se aos diversos níveis de agregação do plano de classificação: classes, subclasses, grupos e subgrupos, tendo informações associadas a respeito da temporalidade e da destinação para os documentos assim classificados); agente (refere-se aos usuários que acessam o SIGAD); componente digital (refere-se aos objetos digitais que compõem o documento arquivístico); e o evento de preservação (refere-se às ações de preservação realizadas nos documentos arquivísticos digitais, tais como migração, atualização, conversão, compreensão, validação, decifração) (CONARQ, 2011).

Na percepção de Indolfo (2013), o tratamento adequado dos documentos arquivísticos digitais levou à necessidade da criação de sistemas informatizados de gestão arquivística de documentos. Nesses sistemas informatizados, os requisitos implementados para a manutenção da autenticidade dos documentos arquivísticos digitais devem estar bem definidos, e sua importância deve estar clara. Indolfo (2013) destaca que:

[...] para assegurar a longevidade, é essencial a inclusão de requisitos de natureza arquivística desde o momento da concepção desses sistemas, tornando-se imprescindível o cuidadoso controle ao longo de todo o seu ciclo vital, de modo a garantir a autenticidade, a fidedignidade, a integridade e a acessibilidade. (INDOLFO, 2013, p. 53).

Em linha com esse pensamento, Flores, Rocco e Santos (2016, p. 5) destacam que o SIGAD é o “sistema de gestão que trata da captura, armazenamento, indexação e recuperação de todos os componentes digitais do documento arquivístico como uma unidade complexa, até os sistemas de preservação e acesso”. Na próxima seção desse trabalho será apresentada a preservação de longo prazo com mais detalhes.

5.5 Preservação a longo prazo

A preservação de acervos arquivísticos requer o envolvimento da estrutura organizacional, na qual as partes envolvidas devem auxiliar no processo como um todo. Inicialmente, devem-se definir questões basilares de forma a criar uma “cultura organizacional” que vislumbre a preservação de longo prazo. No entanto, é preciso ter claro que documentos em suportes tradicionais e documentos digitais possuem a mesma relevância, não devendo assim haver distinção entre eles no tocante à prioridade de tratamento. A partir desse ponto, observam-se questões relacionadas à preservação dos documentos, no caso dos digitais, e definem-se as estratégias a serem utilizadas (FLORES; SANTOS, 2016).

A preservação envolve princípios, estratégias e políticas que norteiam as atividades para assegurar a estabilidade física e tecnológica dos documentos. No âmbito digital, trata-se do conjunto de ações destinadas a manter a autenticidade e acessibilidade dos documentos arquivísticos digitais ao longo do tempo, garantindo as funções e abarcando todo o ciclo de vida dos documentos.

A preservação de documentos arquivísticos tem por objetivo garantir a autenticidade e a integridade da informação, enquanto o acesso depende de os documentos estarem em condições de serem utilizados e compreendidos. O desafio da preservação dos documentos arquivísticos digitais está em garantir o acesso contínuo a seus conteúdos e funcionalidades, por meio de recursos tecnológicos disponíveis à época em que ocorrer a sua utilização (CONARQ, 2005).

Com relação às influências tecnológicas na Arquivística, “a informática está incorporada na gestão e na disseminação, seja dos documentos tradicionais, ou digitais”

(BELLOTTO, 2005, p. 53). A preservação digital deve garantir o acesso de longo prazo a documentos arquivísticos autênticos, o que envolve a adoção dos seguintes princípios:

- a) focar especificamente documentos arquivísticos, e não em objetos digitais de forma genérica;
- b) focar documentos arquivísticos digitais autênticos;
- c) pressupor que a autenticidade dos documentos arquivísticos digitais esteja sob ameaça, principalmente no momento da transmissão no espaço (entre pessoas e sistemas) e no tempo (atualização/substituição de hardware e software usados para armazenar, processar e transmitir os documentos);
- d) reconhecer que a preservação digital é um processo contínuo, que começa na concepção do documento;
- e) reconhecer que a autenticidade dos documentos arquivísticos digitais tem por base os procedimentos de gestão e preservação e a confiança tanto no repositório como no órgão responsável pela guarda desses documentos;
- f) decidir o que se considera como documento original, uma vez que a preservação digital pressupõe a necessidade de conversão de formatos e atualização de suportes;
- g) reconhecer que a elaboração de manuais e os procedimentos de preservação desempenhados pelo repositório digital apoiam a presunção de autenticidade desses documentos;
- h) reconhecer que o registro, em metadados, das intervenções de preservação em cada documento apoia a presunção de autenticidade desses documentos;
- i) reconhecer que a autenticidade dos documentos digitais deve ser avaliada e presumida no momento de sua submissão ao repositório;
- j) reconhecer que o repositório digital é responsável pela manutenção permanente da autenticidade dos documentos a ele submetidos; e
- k) distinguir claramente entre autenticidade e autenticação de documentos, considerando que a primeira é a qualidade de o documento ser verdadeiro, e a segunda é uma declaração dessa qualidade, feita, em um dado momento, por uma pessoa autorizada para tal (CONARQ, 2011).

As ações de preservação devem contemplar os diferentes níveis de abstração do documento digital, na seguinte ordem: físico (mídia, registro físico), lógico (software e formato digital) e conceitual (estrutura, conteúdo exibido). Além disso, deve considerar os elementos para produção, manutenção dos documentos digitais e acesso a estes: software que o criou e pode apresentá-lo; formato digital em que é codificado; hardware para o funcionamento do

software e para permitir a apresentação do formato digital. Essas são questões que podem implicar na mudança nos tipos de mídia de armazenamento, podendo chegar ao ponto da conversão em outros formatos digitais, atualização do ambiente tecnológico, ou de hardware e software.

A propósito, segundo o InterPARES 2 Project (2007), os documentos digitais devem ser geridos cuidadosamente durante toda a sua existência, a fim de garantir sua acessibilidade e legibilidade ao longo do tempo, mantendo-se intactos sua forma, seu conteúdo e suas relações até quando for necessário, para a continuidade de sua credibilidade como documentos de arquivo. Também já se sabe bem que a gestão de documentos arquivísticos digitais deve ser realizada a partir de uma abrangente compreensão de todas as fases ou estágios da existência desses documentos, desde quando são gerados, passando pela sua manutenção por parte do produtor e chegando à sua avaliação, destinação e preservação a longo prazo. Todas as atividades para gerenciar os documentos, no decurso de sua existência, estão ligadas, como em uma cadeia, e são independentes. Se um elo se rompe, a cadeia não pode executar sua função.

Em seguida, serão enfatizados aspectos importantes na preservação de longo prazo, como os metadados de preservação PREMIS, o Modelo OAIS e o RDC-Arq.

5.6 Metadados de preservação (PREMIS)

Os metadados, para Alves (2010), são atributos que representam uma entidade (objeto do mundo real) em um sistema de informação. Em outras palavras, são elementos descritivos ou atributos referenciais codificados que representam características próprias ou atribuídas às entidades; são, ainda, dados que descrevem outros dados em um sistema de informação, com o intuito de identificar de forma única uma entidade (recurso informacional) para posterior recuperação.

Na concepção de Caplan (2009) e Riley (2017), os metadados são categorizados de acordo com o que se destinam a realizar. Metadados descritivos ajudam na descoberta e identificação de recursos. Metadados administrativos são divididos em: metadados técnicos, utilizados para decodificar e renderizar arquivos; metadados de preservação, que auxiliam nas atividades destinadas a garantir a usabilidade a longo prazo de um recurso digital, ou seja, o gerenciamento a longo prazo; metadados de direitos, que estão relacionados à propriedade intelectual do conteúdo; metadados estruturais, que estabelecem as relações de partes de recursos entre si; e linguagens de marcação, que contribuem com metadados e sinalizadores para outros recursos estruturais ou semânticos dentro do conteúdo.

Nesse sentido, o PREMIS é uma norma internacional que apresenta um conjunto básico (*core*) de elementos de metadados de preservação para apoiar sistemas que gerenciam objetos digitais. O grupo de trabalho PREMIS tem ampla abrangência no âmbito da comunidade dedicada à preservação digital, e seu principal documento de referência é o *PREMIS Data Dictionary for Preservation Metadata*⁸. Os metadados definidos no *PREMIS Data Dictionary* (CONARQ, 2011):

- a) contribuem para a viabilidade, disponibilidade, clareza, autenticidade e identidade de objetos no contexto da preservação digital;
- b) representam as informações sobre os documentos digitais que a maioria dos repositórios precisa saber para preservar esses documentos ao longo do tempo;
- c) prestam especial atenção aos metadados rigorosamente definidos, com base em diretrizes para a criação, gestão e uso voltadas para fluxos de trabalho automatizados, e são tecnicamente neutros, ou seja, não pressupõem o uso, em particular, de quaisquer tecnologias de preservação, estratégias, sistemas de armazenamento, gerenciamento de metadados etc. (CONARQ, 2011).

O *PREMIS Data Dictionary* também conta com um modelo de esquema em XML, que permite incorporar o dicionário de dados em sistemas de gestão de objetos digitais (CONARQ, 2011).

Entretanto, para Caplan (2009), o Dicionário de Dados PREMIS não se destina a definir todos os elementos de metadados de preservação possíveis, mas apenas aqueles que a maior parte dos repositórios precisará saber na maioria das vezes. Várias categorias de metadados são excluídas por estarem fora do escopo, incluindo: metadados específicos de formato, ou seja, metadados que pertencem a apenas um formato de arquivo ou classe de formatos, como áudio, vídeo ou gráficos vetoriais; metadados específicos de implementação e regras de negócio, ou seja, metadados que descrevem as políticas ou práticas de um repositório individual (por exemplo, como ele fornece acesso aos materiais); metadados descritivos, pois embora a descrição dos recursos seja obviamente relevante para a preservação, muitos padrões independentes podem ser usados para esse propósito, como MARC21, MODS e Dublin Core, além de informações detalhadas sobre mídia ou hardware. Mais uma vez, embora claramente relevantes para a preservação, esses metadados foram deixados a outras comunidades para que definissem as informações sobre agentes (pessoas, organizações ou software) além do mínimo necessário à identificação, e as informações sobre direitos e permissões, com exceção daqueles

⁸ Disponível em: <https://www.loc.gov/standards/premis/>.

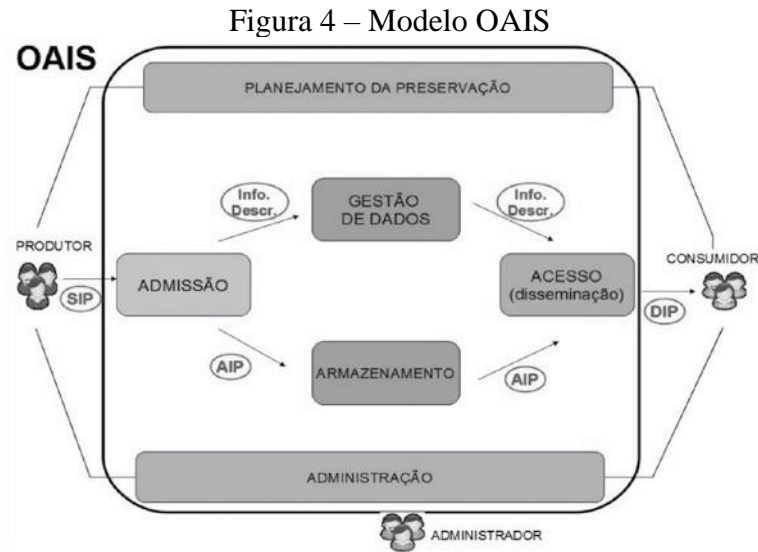
que afetam diretamente as funções de preservação. A seguir será apresentado o modelo Open Archival Information System – OAIS que define um repositório digital.

5.7 Modelo Open Archival Information System (OAIS)

O modelo de referência OAIS é uma recomendação internacional desde 2003 (ISO 14721). Trata-se de um modelo conceitual que define um repositório digital, identificando o ambiente, os componentes funcionais, suas interfaces internas e externas, os objetos de dados e informações. No Brasil, foi adaptado e publicado como norma ABNT NBR 15472: 2007, com o título “Sistema Aberto de Arquivamento de Informação (SAAI)” (CONARQ, 2011).

Rocha (2015) destaca a importância do modelo OAIS no que diz respeito à preservação digital e a repositórios digitais. O modelo OAIS foi desenvolvido sob a coordenação do Comitê Consultivo para Sistemas de Dados Espaciais (CCSDS) da Nasa, que contou com a colaboração da comunidade científica internacional. Sua elaboração levou dez anos. Uma primeira versão foi publicada em 1999, outra em 2002, e em 2003 transformou-se na norma ISO 14721:2003. O OAIS é um sistema conceitual que descreve um modelo funcional e um modelo de informação para a preservação dos materiais digitais administrados pelo repositório e o acesso a estes.

De acordo com Márdero Arellano (2008), as entidades funcionais gerenciam o fluxo de informação entre as entidades que formam o ambiente OAIS e identificam os componentes funcionais dos arquivos relacionados com a preservação dos objetos digitais. No ambiente OAIS, cada entidade é responsável por um dos processos do sistema: *produtor*: é a pessoa ou sistema cliente que fornece a informação; *consumidor*: é a pessoa ou sistema cliente interessado no material arquivado ou em uma versão transformada deste; *comunidade-alvo*: são os usuários em potencial, a serem reconhecidos para a realização da meta maior da preservação de longo prazo.

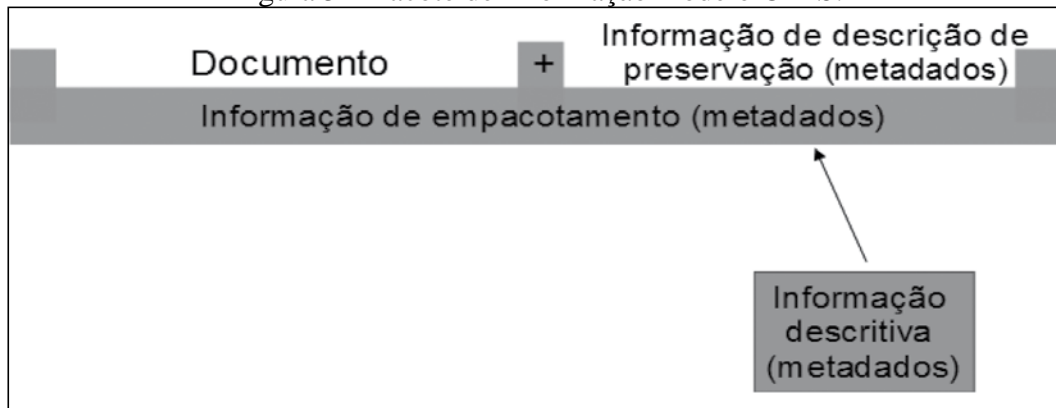


Fonte: Rocha (2015).

De acordo com o modelo OAIS, as funcionalidades de um repositório devem estar organizadas em seis grandes grupos: a admissão (*ingest*), o armazenamento, a gestão de dados, o planejamento da preservação, a administração e o acesso. As funcionalidades de cada grupo são detalhadas no modelo funcional e podem ser acessadas por três tipos de agentes: produtores (pessoas ou sistemas que depositam os objetos digitais no repositório), consumidores (pessoas ou sistemas que interagem com o OAIS para acessar os objetos digitais) e administradores (responsáveis pelo estabelecimento das políticas e pela gestão dos objetos digitais preservados) (ROCHA, 2015).

No modelo de informação são elaborados pacotes de informação, que atuam como um recipiente conceitual. Um pacote de informação é composto de informação de conteúdo (o documento) e informação de descrição de preservação (metadados necessários para apoiar a preservação e acesso do documento a longo prazo), que são encapsuladas e identificadas pela informação de empacotamento. Ao pacote de informação são, ainda, associadas as informações descritivas do pacote (metadados descritivos que possibilitam a localização do pacote no repositório) (ROCHA, 2015).

Figura 5 – Pacote de informação modelo OAIS.



Fonte: Rocha (2015).

O Consultative Committee for Space Data System (CCSDS, 2012) argumenta que o modelo OAIS define requisitos para o arquivamento digital, isto é, o armazenamento em um repositório digital com a responsabilidade de preservar informações e disponibilizá-las a uma comunidade designada. Portanto, os documentos mantidos no OAIS devem ser de preservação de longo prazo, independentemente da continuidade do modelo. A próxima seção tem como tema o RDC-Arq.

5.8 Repositório arquivístico digital confiável (RDC-Arq)

No entendimento de Rocha (2015), o repositório, ou seja, todo esse ambiente, pode ser utilizado em diversas situações, como na área de arquivos e outras áreas de gestão de informação: arquivo de documentação corrente (associado a um sistema de gestão documental), arquivo permanente, biblioteca digital, coleção de obras de arte digital, coleção de áudio e vídeo digital, curadoria de dados de pesquisa digitais etc.

Na concepção do CONARQ (2015), o repositório digital é um ambiente de armazenamento e gerenciamento de materiais digitais. Esse ambiente constitui-se em uma solução informatizada em que os materiais são capturados, armazenados, preservados e acessados. Um repositório digital é, então, um complexo que apoia o gerenciamento dos materiais digitais, pelo tempo que for necessário, e é formado por elementos de hardware, software e metadados, bem como por uma infraestrutura organizacional e procedimentos normativos e técnicos. Tal ambiente tem sido empregado em diversas situações, tais como: arquivos corrente e intermediário (em associação com um SIGAD); arquivo permanente; biblioteca digital; acervo de obras de arte digitais; depósito legal de material digital; e curadoria

de dados digitais de pesquisa. Um repositório digital não se resume a uma solução informatizada para armazenamento (*storage*), que é apenas um componente do repositório.

O conceito de repositório digital confiável surgiu em um documento-base, preparado por um grupo de trabalho liderado pelo *Research Libraries Group* (RLG) e o *Online Computer Library Center* (OCLC), intitulado *Trusted Digital Repositories: Attributes and Responsibilities* (Repositórios digitais confiáveis: atributos e responsabilidades). Esse documento propôs as bases conceituais para repositórios digitais confiáveis e estabeleceu os atributos e as responsabilidades que estes devem assumir. Também apresentou um importante debate sobre como criar novamente, em um ambiente instável como a internet, a ideia tradicional de *confiança*, que é um princípio fundamental nas instituições arquivísticas (ROCHA, 2015).

Um repositório digital confiável é um repositório digital capaz de manter autênticos os materiais digitais, de preservá-los e prover acesso a eles pelo tempo necessário. Para cumprir essa missão, segundo o relatório *Trusted Digital Repositories: Attributes and Responsibilities* (RLG, 2002), os repositórios digitais confiáveis devem:

- a) aceitar, em nome de seus depositantes, a responsabilidade pela manutenção dos materiais digitais;
- b) dispor de uma estrutura organizacional que apoie não somente a viabilidade de longo prazo dos próprios repositórios mas também dos materiais digitais sob sua responsabilidade;
- c) demonstrar sustentabilidade econômica e transparência administrativa;
- d) projetar seus sistemas de acordo com convenções e padrões comumente aceitos, no sentido de assegurar, de forma contínua, a gestão, o acesso e a segurança dos materiais depositados;
- e) estabelecer metodologias para avaliação dos sistemas que considerem as expectativas de confiabilidade esperadas pela comunidade;
- f) considerar, para desempenhar suas responsabilidades de longo prazo, os depositários e os usuários de forma aberta e explícita;
- g) dispor de políticas, práticas e desempenho que possam ser auditáveis e mensuráveis; e
- h) observar os seguintes fatores relativos às responsabilidades organizacionais e de curadoria dos repositórios: escopo dos materiais depositados, gerenciamento do ciclo de vida e preservação, atuação em conjunto com uma ampla gama de

parceiros, questões legais relacionadas com a propriedade dos materiais armazenados e implicações financeiras (CONARQ, 2015).

Rocha (2015) ressalta que um repositório digital confiável deve dispor de uma série de funções para verificar a integridade dos pacotes de informação na entrada, no armazenamento e no acesso aos documentos. Além desses controles, o repositório deve registrar, em metadados, informação descritiva e de apoio à preservação, em conformidade com as principais normas e padrões internacionais.

Além disso, Rocha (2015) assinala que a implementação de diretrizes para repositórios digitais confiáveis significa que é necessário satisfazer as necessidades de seus usuários e as especificidades do acervo sob sua custódia. Assim, os repositórios destinados à manutenção de documentos arquivísticos necessitam de requisitos adicionais para proteger as características desses documentos e de previsão de procedimentos em conformidade com as normas da área de arquivo.

O CONARQ (2015) apresenta as diretrizes para a implementação de RDC-Arq para documentos arquivísticos digitais, nas fases corrente, intermediária e permanente. O repositório arquivístico digital confiável deve ser capaz de atender aos procedimentos arquivísticos em suas diferentes fases e aos requisitos de um repositório digital confiável.

Um repositório arquivístico digital é um repositório digital que armazena e gerencia esses documentos, seja nas fases corrente e intermediária, seja na fase permanente. Como tal, esse repositório deve:

- gerenciar os documentos e metadados de acordo com as práticas e normas da Arquivologia, especificamente relacionadas à gestão documental, descrição arquivística multinível e preservação; e
- proteger as características do documento arquivístico, em especial a autenticidade (identidade e integridade) e a relação orgânica entre os documentos (CONARQ, 2015, p. 5).

Nos cenários do ciclo de vida da implementação do RDC-Arq, Flores, Rocco e Santos (2016) destacam que haverá um RDC-Arq para as fases corrente e intermediária e outro para a fase permanente. Esse cenário poderá envolver a mudança da cadeia de custódia para o RDC-Arq da fase permanente, com as fases corrente e intermediária sendo associadas a um SIGAD. Dessa maneira, o repositório irá auxiliar na preservação de documentos que ficam por longos períodos em fase corrente/intermediária. Isso se justifica pelos acelerados ciclos de obsolescência tecnológica. Logo, haverá uma preocupação com a preservação dos documentos arquivísticos digitais, que chegarão à fase permanente, devendo assim ser preservados permanentemente. Portanto, o RDC-Arq irá atuar na manutenção da autenticidade e da relação orgânica, além de auxiliar no arranjo e na descrição, visando ao acesso.

Quando abordamos a característica confiável do repositório arquivístico digital, estamos falando que esses repositórios precisam ser auditados e certificados, e, nesse cenário, a Resolução n. 43 do CONARQ (2015) deixa clara a sua posição quanto à adoção da ISO 16.363:2012⁹ - *Space data and information transfer systems — Audit and certification of trustworthy digital repositories*. Isso significa que a autenticidade dos documentos arquivísticos recebe mais uma camada para reforçar a cadeia de custódia digital, ou seja, da abordagem de Jenkinson (1922 apud FLORES; ROCCO; SANTOS, 2016) da custódia focada no produtor à instituição arquivística. Agora temos, além dos ambientes SIGAD e RDC-Arq e da instituição arquivística, a instituição certificadora, que vai garantir que a autenticidade se manteve de acordo com os requisitos estabelecidos, indo então além da instituição arquivística e reforçando essa linha ininterrupta digital.

O CONARQ (2015) apresenta algumas considerações sobre os repositórios arquivísticos digitais, como: a reponsabilidade pelo repositório (a responsabilidade pelo projeto, implantação e manutenção de um repositório digital de documentos arquivísticos deve ser compartilhada por profissionais de arquivo e de tecnologia da informação, de forma a se cumprirem os requisitos tecnológicos e os procedimentos do tratamento arquivístico); o tratamento arquivístico (o repositório digital tem que ser capaz de organizar e recuperar os documentos, de forma a manter a relação orgânica entre eles; deve apoiar a organização hierárquica dos documentos digitais, a partir de um plano de classificação de documentos, e a descrição multinível, de acordo com a norma internacional para descrição arquivística, a “Norma Geral Internacional de Descrição Arquivística (ISAD(G))”, e a “Norma Brasileira de Descrição Arquivística (NOBRADE)”; princípios de preservação digital (a preservação digital implica em adotar princípios de preservação digital, visando garantir o acesso de longo prazo a documentos arquivísticos autênticos); independência dos repositórios (o repositório digital deve ter independência; isso significa que seu funcionamento e o acesso aos documentos não podem depender das aplicações que funcionam em conjunto com ele, isto é, deve ser possível acessar os documentos independentemente da aplicação, diretamente no repositório, desde que isso seja feito de forma controlada, para não ameaçar a autenticidade dos documentos no repositório) e, por último, a interoperabilidade (o repositório digital deve estar em conformidade com as normas e padrões estabelecidos, de forma a possibilitar níveis de interoperabilidade com outros repositórios digitais e sistemas informatizados que tratem de documentos arquivísticos).

⁹ Disponível em: <https://www.iso.org/standard/56510.html>.

Para a construção de RDC-Arq são definidos padrões e normas de referência, tais como: Modelo OAIS, proposta de implementação do *Relatório do RLG e do Online Computer Library Center (OCLC) – Repositórios digitais confiáveis: atributos e responsabilidades*, Certificação e auditoria de repositórios confiáveis: critérios e *checklist* – TRAC, Requisitos técnicos para entidades de auditoria e certificação de organizações candidatas a serem repositórios digitais confiáveis, PREMIS, ISAD(G), NOBRADE, Metadados do e-ARQ Brasil, Protocolo para coleta de metadados (OAI-PMH), Padrão de codificação e transmissão de metadados (METS), Descrição arquivística codificada (EAD) (CONARQ, 2015).

Concluindo, segundo Flores, Rocco e Santos (2016), a implementação do RDC-Arq irá depender do cenário em questão. Mas fica expressa a necessidade de um repositório com tais características para os documentos de valor permanente e a possibilidade de se implementar outro para os documentos de valor corrente e intermediário, sendo possível reunir os documentos das três idades em um único RDC-Arq, porém de forma que permaneçam independentes, já que é o SIGAD que interopera os documentos de valor corrente e intermediário, enquanto os de valor permanente ficarão a cargo da instituição arquivística, pois o SIGAD somente interopera para o recolhimento, curadoria digital, auditoria e certificação. Na próxima seção, será apresentado, o profissional arquivista detentor de todo conhecimento e prática mencionado anteriormente.

5.9 O profissional arquivista

A data do surgimento da profissão de arquivista não é consenso entre os autores. A ideia compartilhada por Duchein (1991) é que a profissão de arquivista não é talvez a mais antiga do mundo, mas, inequivocamente, é tão antiga como a invenção da escrita, o que lhe assegura, no mínimo, uma experiência de três ou quatro mil anos.

Conhecer a origem do profissional arquivista nos permite traçar o caminho percorrido e vislumbrar perspectivas de atuação, permitindo o desenvolvimento e acolhimento de possibilidades futuras de atuação do profissional. Na concepção de Rousseau e Couture (1998), ainda que o termo “arquivista” seja uma designação contemporânea, desde a Alta Antiguidade, em diferentes regimes, nota-se a presença de pessoas exercendo atividades relativas à preparação, ao tratamento, à recuperação e à conservação dos documentos produzidos por essas administrações, semelhantes às funções dos arquivistas no século XXI.

Na sua origem, a função de arquivista se confunde com a do escriba do antigo Egito. Pessoas de confiança, nos recônditos do poder ou dos poderosos, redigem, transcrevem e

conservam os documentos normativos, as contas e a correspondência. O trabalho do arquivista surge quando essas duas atribuições — a escrita e sua guarda — se separam (DELMAS, 2010).

Na Grécia, eram os pítons, principais magistrados da cidade, os responsáveis pelos documentos administrativos, com o auxílio de escravos públicos qualificados cuja função era relativamente invejável. Desse modo, concluímos que no período clássico a técnica arquivística atingiu um alto grau de maturidade, superando as expectativas para a época. Já em Roma, a tarefa de conservar os documentos era confiada aos gestores, magistrados que prestavam assistência aos cônsules em questões financeiras. Havia também os tabulários, responsáveis por arquivar e conservar os documentos (ROUSSEAU; COUTURE, 1998).

De acordo com Rousseau e Couture (1998), na França (século XII), a partir da criação das comunas e das cidades, a responsabilidade pelas tarefas do arquivo é confiada ao secretário-escrivão do município ou a um almotacé, denominado de arquivista, secretário, ou guarda dos arquivos. Esse hábito de confiar a um dos altos funcionários da administração a guarda dos documentos é mantido até os dias atuais nas cidades e vilas do país.

Os arquivistas, tal como são hoje conhecidos, apareceram no final da Idade Média, quando eram chamados de arquivários (*arquivaires*), tornando-se arquivistas em 1701. Trata-se, portanto, de fenômeno relativamente recente, particularmente visível com o desenvolvimento da administração moderna no Antigo Regime (Secretarias de Estado da Guerra, da Marinha etc.) (DELMAS, 2010).

Ainda no século XIX, surge na Europa uma preocupação com a formação dos arquivistas. Diante disso, são criadas escolas de Arquivologia: a *École Nationale des Chartes* (França), com o ensino voltado para a erudição histórica, além de outras escolas que priorizavam a administração dos arquivos estatais, como a Escola de Florença (Itália). Contudo, a história dominava os currículos dessas escolas, e esse modelo de ensino se expandiu para outros países, assim como os Estados Unidos (RODRIGUES, 2006).

Os Estados Unidos passaram a se interessar efetivamente pelos arquivos no século XIX, quando foi criada a *American Historical Association*. No século XX, foram introduzidos nos arquivos do Estado alguns princípios arquivísticos em vigor na Europa, especialmente na França, voltados para o conhecimento histórico. Todavia, durante e após a Primeira Guerra Mundial, surgiu a preocupação com a avaliação de documentos, decorrente do aumento significativo da produção documental, e, na segunda metade dos anos 1940, a evolução tecnológica influenciou em diversos aspectos a vida dos arquivos e o trabalho dos arquivistas (SANTOS, 2008).

Para Souza (2011), desde que o homem passou a registrar a história da humanidade até os dias de hoje, há a preocupação em conservar tais registros para posterior utilização. Assim, ao longo dos anos, os arquivistas têm manejado diversas tipologias documentais registradas em diferentes suportes e que dão lugar a novos desafios para esses profissionais.

Diante disso, as novas tecnologias exercem influência, segundo Duchein (1991), como um aspecto da evolução da profissão arquivística que impacta o trabalho na área de arquivo, seja porque produzem novos tipos de documentos (audiovisual, informática), seja porque modificam as condições de trabalho dos serviços de arquivo. Os novos tipos de documentos (ou, mais resumidamente, os “novos arquivos”) suscitam problemas muito complexos do ponto de vista do seu acervo, da sua preservação, da sua descrição e da sua utilização.

A profissão de arquivista vem ao longo dos séculos se adaptando e evoluindo. Segundo Bellotto (2005):

[...] deve haver um diálogo constante entre a concepção que o arquivista tem do arquivo e a forma pela qual a sociedade expressa suas necessidades. Isso demonstra a capacidade de interagir com os novos desafios da globalização, da era da informação e as novas tecnologias da informação e comunicação, que surgem no contexto social, principalmente relacionados à preservação da memória para a sociedade. Gestão e preservação são pontos latentes da atuação do profissional e desenvolvem também as atividades de disseminação, acesso e uso de documentos arquivísticos, em qualquer suporte que for, elaborados por pessoa física ou jurídica (pública e/ou privada) no desenvolvimento de atividades de área meio ou área fim (BELLOTTO, 2005, p. 301-302).

Nessas circunstâncias, segundo Moro-Cabero (2021), o arquivista do século XXI administra documentos eletrônicos compostos de *bytes* — objeto digital —, assinatura(s) eletrônica(s), metadados e aparência delimitados por um determinado formato de criação/armazenamento/visualização que permeia as propriedades significativas do objeto, no melhor dos casos. Sua identificação e complexidade de curadoria aumentam devido à diversidade de categorias de objetos eletrônicos reconhecíveis, como som, audiovisual, imagem fixa, aeroespacial, páginas da web, bancos de dados, materiais de escritório, redes sociais etc., bem como à variedade de tecnologias que podem ser aplicadas para análise, processamento, armazenamento e descoberta: tecnologias *big data*, tecnologias de nuvem e *blockchain*, tecnologias para link aberto de dados, entre outras. Na seção derradeira do capítulo será apresentada a relação das instituições arquivísticas e os dados de pesquisa.

5.10 A arquivística e os dados de pesquisa

A partir da evolução científica do século XVII, seguida pela Era Industrial, a utilidade de compreender os arquivos passou a ser evidente. Os arquivos são fontes de conhecimento. Voltada de início para a religião e o direito, essa utilidade estendeu-se rapidamente à erudição e, sobretudo, às ciências (DELMAS, 2010).

Os arquivos preservam documentos científicos e técnicos, notadamente do século XVIII, e seu uso foi progressivamente ampliado. É muitas vezes difícil imaginar que possa haver um acúmulo de experiências e de saberes científicos e técnicos não explorados nas instituições arquivísticas (DELMAS, 2010).

Em harmonia de pensamento, os *National Archives and Records Administration* (NARA) dos Estados Unidos constituíram a política de avaliação do Arquivo Nacional, que faz parte da gestão de registros federais. Tal política estabelece a estrutura estratégica, os objetivos e as diretrizes que a NARA usa para determinar se os registros federais têm valor arquivístico, e fornece diretrizes mais específicas para avaliar certas categorias de registros. Como os dados pessoais, dados observacionais das ciências físicas, registros de saúde e segurança ambiental, com destaque para os *dados irreprodutíveis* que apoiam estudos epidemiológicos ou outras análises significativas de exposição de humanos e do meio ambiente a produtos químicos, pesticidas e resíduos perigosos que podem ter valor suficiente para garantir a retenção permanente, a depender da amplitude e escopo da análise e da reutilização dos dados. *Os dados que rastreiam ou modelam* as condições e respostas de exposição de longo prazo em locais de resíduos abandonados, inativos ou não controlados, e as liberações de resíduos perigosos também podem ter valor suficiente para retenção permanente, dependendo do interesse do público em geral ou da comunidade de pesquisa no local ou liberação e a finalidade para a qual os dados foram coletados. Além de *registros e dados* relacionados à produção, uso, descarte e monitoramento rotineiro de substâncias perigosas. No caso dos Registros de Pesquisa e Desenvolvimento com consideração aos cadernos de laboratório e os dados criados por projetos de pesquisa e desenvolvimento (NARA, 2007).

A *Section on University and Research Institution Archives Committee on Scientific and Research Data* do *International Council on Archives* (ICA) lançou, em 2010, o *Management and Preservation of Scientific Records and Data*. O manual visa melhorar a comunicação entre as comunidades de pesquisa e arquivística e fornecer orientações abrangentes sobre as melhores práticas para preservação e acesso a dados de pesquisa. A função arquivística das universidades e instituições de pesquisa, por muitas razões, não tem sido suficientemente visível e instrutiva

para criar um diálogo com a comunidade de pesquisa e cooperar no apoio à preservação e acesso aos dados de pesquisa. Uma dessas razões pode ser a falta de um manual adequado e diretrizes sobre documentação de pesquisa e gerenciamento de dados de pesquisa. As demandas atuais de compartilhamento e reutilização de dados, demandas de acesso a eles e pesquisa sobre dados existentes, no entanto, exigem uma abordagem diferente para documentar e uma cooperação mais estreita entre cientistas e arquivistas. Estas diretrizes não pretendem cobrir e resolver todos os problemas de preservação e acesso aos dados, mas pretendem apontar algumas maneiras possíveis de trabalhar com eles e, esperançosamente, dar alguns conselhos à comunidade de pesquisa (AROVILIUS *et al.*, 2010).

A estrutura do manual apresenta na seção 2: Escopo e definições de Ciência; Pesquisar; Registros e dados da ciência; Uso de diferentes fontes de dados de pesquisa. Na seção 3 aborda os Antecedentes legais para gestão de registros de ciência; Estudos de caso da Áustria, Israel, Suécia, Reino Unido e semelhanças e diferenças. A seção 4 discorre sobre: o Processo de pesquisa e fluxo de dados; Atividades do processo de pesquisa; Processo de pesquisa – atividades e dados; Fase de planejamento; Coleta de dados; Análise; Avaliação, auditoria; Relato dos resultados; Relatórios financeiros e Arquivamento. Já na seção 5 o foco é na Estratégia de preservação, curadoria digital e avaliação; Critérios de preservação; Preservação e curadoria digital. Na seção 6 descreve os benefícios do arquivamento (AROVILIUS *et al.*, 2010).

Os *National Archives of Australia* (NAA) estabelecem orientações na página institucional sobre a gestão da informação. As informações criadas, enviadas ou recebidas como parte do trabalho para o governo australiano são um registro. Ele fornece evidências do que a agência fez e o porquê. No contexto do governo australiano, usa-se “informações” como um termo abrangente para dados, informações e registros em qualquer formato, seja analógico ou digital, além de orientações específicas, como gerenciar, reter e descartar dados e conjuntos de dados (identificação de dados e conjuntos de dados, descrição, armazenamento, preservação digital, controle de acesso, segurança, temporalidade dos dados e conjuntos de dados, compartilhamento de dados) (NAA, 2021).

O APIS incentiva e promove o depósito de dados de investigação com origem em inquéritos produzidos nas áreas da Sociologia, Ciência Política e Psicologia Social. Contudo, qualquer investigador na área das ciências sociais poderá enviar dados para o APIS. O Arquivo menciona também os dados FAIR como uma responsabilidade conjunta dos investigadores e dos repositórios. Neste contexto, e de forma a otimizar o arquivo, a disseminação e reutilização

de dados e o desenvolvimento de metadados ricos são um requisito dos princípios FAIR (APIS, 2021).

Em 18 de outubro de 2016, a comissão do *European Archives Group* adotou uma estratégia corporativa em matéria de gestão de dados, informação e conhecimento. Essa nova abordagem visa melhorar a maneira como dados, informações e conhecimento são coletados, gerenciados, compartilhados e preservados, desenvolvendo novas oportunidades de trabalho colaborativo (EUROPEAN COMMISSION, 2016).

Conforme apontam Rousseau e Couture (1998), arquivos são condutores de dados e informação, e, na percepção da pesquisa, isso fica mais evidente por meio da necessidade de preservá-los ao longo do tempo para as gerações futuras.

Questões como gestão e preservação a longo prazo foram apuradas pela arquivística no desenvolvimento da sua teoria e prática. Os estudos aplicados a essas questões configuram o trabalho do arquivista. Temos estudos que conseguem responder várias questões que são colocadas hoje, como questões da gestão de dados de pesquisa. Sendo possível apresentar o lugar da arquivística na gestão de dados de pesquisa, no momento atual de produção intensa de dados. Este lugar autêntico, organizado, é tratado como um recurso tão importante quanto os recursos financeiros.

São muitos os conceitos, termos e expressões que valem um olhar dialógico. A própria definição de dados de pesquisa ainda não é consenso e merece exame, assim como o termo associado, “gestão de dados de pesquisa”. Para alguns autores do campo da arquivologia, os dados de pesquisa podem apresentar extensões, classificações e formatos próprios, além de “tipologias e padrões documentais” tradicionais já estabelecidos. Parece evidente o emprego de termos muito familiares à teoria, aos princípios, métodos e práticas da arquivologia (SANTOS, 2021, p. 16).

A iniciativa parte da perspectiva de que os dados de pesquisa são produzidos em contextos institucionais e podem ser utilizados para validar as evidências e comprovar os resultados obtidos; portanto, destaca-se que os princípios e qualidades arquivísticas, como proveniência, organicidade, confiabilidade e autenticidade, podem servir como uma base teórico-metodológica segura para preservação efetiva, disponibilização qualificada e reuso desses dados (SANTOS, 2021).

É possível ainda perceber outras similaridades no manejo dos termos. Um caso exemplar é a afirmação de que os “dados são criados para atender a uma finalidade específica”, conforme Jorge e Albagli (2018, p. 417 *apud* SANTOS, 2021, p. 17), administrativa, por exemplo, e tornam-se dados de pesquisa científica de acordo com seu uso. De nítida inspiração nas ideias

do arquivista norte-americano Theodore Schellenberg, a finalidade administrativa nos remete ao valor primário, “qualidade inerente às razões de criação de todo documento, típica das fases iniciais de seu ciclo vital”, de acordo com Camargo e Bellotto (1996, p. 78 *apud* SANTOS, 2021, p. 17), e seu uso para outros fins, como a pesquisa, nos conduz ao seu valor secundário.

Landim e Jorente (2016), tecendo uma comparação entre a lista de procedimentos aplicados em todo o ciclo de vida dos objetos digitais elencados por Higgins e os procedimentos de Gestão Documental realizados na Arquivologia, permitem constatar diversos procedimentos realizados tanto na abordagem da Curadoria Digital quanto na Gestão de Documentos, como a Avaliação, a Conceitualização, a Criação e o Descarte.

Os exemplos se multiplicam. Segundo Rocha (2019 *apud* SANTOS, 2021), as práticas de gestão e a curadoria adequada dos dados de pesquisa servem para evitar as constantes perdas de dados relatadas na literatura. Na perspectiva do projeto Febre Amarela, desenvolvido no Instituto René Rachou (Fiocruz Minas), a modelagem dos “ciclos de vida de dados de pesquisa” e sua “tipificação” cumprem a função de orientar a promoção de boas práticas de gestão e curadoria nos laboratórios, conforme Rocha (2019 *apud* SANTOS, 2021, p. 17). No intuito de dialogar com a autora, cumpre mencionar que dados de pesquisa são registrados em documentos de diferentes suportes e formatos, sobre os quais impõe-se uma análise para determinar suas possíveis qualidades de natureza arquivística. Dados de pesquisa são encontrados em cadernos de laboratório e bases de dados, entre outros tantos tipos documentais, o que nos aponta a necessidade do trabalho de caracterização de seus elementos arquivísticos. Ao mesmo tempo, o ciclo vital dos documentos, o *records life-cycle*, é um conceito anunciado na década de 1940 nos Estados Unidos que gerou impactos significativos na estruturação posterior do sistema de gestão de documentos norte-americano e forte influência em diferentes países, como o Brasil (SANTOS, 2021).

Diniz, Dias e Silva (2018) traçam pontos em comum dos princípios FAIR e a gestão de documentos arquivísticos. Os princípios FAIR trazem como pilares serem encontráveis, acessíveis, interoperáveis e reutilizáveis como preocupações da gestão de dados de pesquisa. Estes princípios focam potencialmente em administração para o uso e reuso dos dados, não abordando a questão da preservação, que também é uma preocupação da gestão dos dados de pesquisa. A gestão de documentos arquivísticos digitais tem como prerrogativas não apenas a administração, mas assuntos que envolvem a política institucional e preservação dos documentos.

Quadro 3 – Semelhanças entre os princípios FAIR e a gestão de documentos arquivísticos digitais.

Princípios FAIR	Gestão de documentos arquivísticos digitais	Pontos em comum
Dados encontráveis	Encontrabilidade de documentos.	Pontos em comum sobre encontrabilidade (serem encontráveis)
Encontrável a partir dos metadados, uso de identificador persistente, uso de palavras-chave, nomenclatura apropriada.	Uso de identificador, descrição informativa = metadados, recursos de indexação, uso de palavras-chave, documentos encontrados a partir dos metadados.	Uso de identificador, encontrável a partir de metadados, uso de palavras-chave.
Dados acessíveis	Acessibilidade de documentos.	Pontos em comum sobre acessibilidade
Mediante autorização apropriada, protocolo bem definido, controle de acesso, especificar onde se depositam os dados, metadados, documentação e códigos.	Controle de acesso dos documentos, atributos de segurança, administrador do sistema determina quem pode ter acesso aos documentos.	Acesso mediante a autorização apropriada, controle de acesso.
Dados interoperáveis	Interoperabilidade de documentos.	Pontos em comum sobre interoperabilidade
Metadados acionados por máquinas, metaformatos utilizam vocabulários controlados e ontologias, metadados dentro dos objetos de dados devem ser reproduzíveis e acessíveis, especificar dados, vocabulários, padrões e metodologias a seguir.	Consulta, recuperação, importação e exportação de documentos e seus metadados, operações de interoperabilidade devem respeitar a legislação vigente e a política de segurança, capaz de interoperar com outros sistemas por meio de padrões abertos.	Não foram encontrados pontos em comum neste aspecto.
Dados reusáveis	Reuso de documentos.	Pontos em comum sobre reuso
Devem estar em conformidade com os princípios de encontrabilidade (encontráveis) e interoperabilidade (interoperáveis) bem descritos para que possam ser integrados a outros dados, referência a fonte e metadados, processo de garantia de qualidade de dados, cumprimento de tempo da validade de dados.	Os documentos arquivísticos devem possuir valores de autenticidade, confiabilidade, organicidade, acessibilidade e unicidade. No plano de classificação e destinação deve estar atribuído o tempo de guarda para fins administrativos ou históricos.	Deve ser estabelecido o tempo de guarda e validade dos dados/documentos. Garantia de qualidade, confiabilidade da informação.

Fonte: adaptado de Diniz, Dias e Silva (2018).

Os princípios FAIR aliados aos princípios da gestão de documentos arquivísticos digitais demonstram qualidades essenciais na administração e preservação de dados de pesquisa. Quem tende a ganhar com isto é a ciência, pois terá acesso e preservação para o desenvolvimento presente e futuro (DINIZ; DIAS; SILVA, 2018).

Além das semelhanças entre a arquivologia e a gestão de dados de pesquisa, alguns conhecimentos pertinentes ao tratamento dos dados de pesquisa fazem parte do conhecimento

do arquivista, tais como: em sistema informatizado em gestão, padrões PREMIS, OAIS e conhecimento em repositórios, demonstrado nas vagas selecionadas para delinear o perfil do arquivista que irá trabalhar com os dados de pesquisa. No capítulo a seguir, veremos com detalhes as práticas e conhecimentos do arquivista na gestão de dados de pesquisa.

6 MAPEAMENTO DA ATUAÇÃO DO ARQUIVISTA NA GESTÃO DE DADOS DE PESQUISA

Como resultado, chegou-se ao mapeamento da atuação do arquivista na gestão de dados de pesquisa, construído por meio da análise dos modelos do ciclo de vida dos dados de pesquisa DDI e DataONE e do modelo do ciclo da curadoria do DCC e da aplicação da teoria e prática arquivística, o que permitiu identificar as atividades desempenhadas pelo profissional na gestão de dados de pesquisa.

A atuação do arquivista de dados de pesquisa identificada nos ciclos foi distribuída de acordo com a teoria e prática arquivística. Segundo a concepção de Paes (2004), a organização de arquivos pressupõe o desenvolvimento de etapas de trabalho entre as quais temos o planejamento. A autora destaca as três fases básicas da gestão de documentos: a produção, a utilização e a avaliação e destinação de documentos.

Os quadros elaborados descrevem as etapas da DDI, DataONE e DCC que se correlacionam com a gestão arquivística e a competência do arquivista na gestão de dados de pesquisa descritas por autores que destacam a necessidade, o papel e a importância da atuação do arquivista no tratamento dos dados de pesquisa.

6.1 Planejamento

Qual é o papel do arquivista em um projeto? Primeiramente, e certamente no início, o papel do arquivista é educar a equipe de pesquisa. Os pesquisadores precisam vislumbrar o uso mais amplo dos dados, especialmente o valor destes para fins de pesquisa secundária (HUMPHREY *et al.*, 2000).

A fase de planejamento é composta da concepção e das primeiras discussões sobre a gestão, envolvendo: a definição dos tipos de documentos que devem ser produzidos e mantidos; a definição dos metadados a serem criados no momento da produção do documento e ao longo do seu ciclo de vida; a definição dos procedimentos de arquivamento, captura, acesso, uso, transmissão de documentos, armazenamento, entre outras decisões compreendidas no planejamento de gestão arquivística. Para que um arquivo, em todos os estágios de sua evolução, possa cumprir seus objetivos, torna-se indispensável a elaboração de um plano arquivístico que tenha em conta tanto as disposições legais quanto as necessidades das instituições (CONARQ, 2011; PAES, 2004).

O contato com o arquivista deve se dar no início do projeto de pesquisa e durante todo o processo de pesquisa. Contar com a experiência do profissional em gestão de documentos é uma forma de auxiliar na estruturação do fluxo de trabalho, na preparação de um plano de gestão de dados de pesquisa, em questões de legislação e ética e sobre curadoria de dados, e ajuda a determinar os tipos de recursos que serão necessários para preparar os dados para arquivar, o programa de computador que deve ser usado e em que formato os dados originais devem ser preparados, assim como a documentação que deve ser estruturada e a terminologia a ser utilizada (CESSDA, 2022).

Quadro 4 – Planejamento.

GESTÃO ARQUIVÍSTICA	CICLO DE VIDA DOS DADOS DE PESQUISA		CICLO DA CURADORIA	COMPETÊNCIAS DO ARQUIVISTA	AUTORES
	DDI	DataONE	DCC		
Planejamento	Conceito	Planejar	Conceitualização Planejamento da preservação	<ul style="list-style-type: none"> Estabelecer e implementar padrões de coleta, práticas de arquivamento 	Humphrey <i>et al.</i> (2000)
				<ul style="list-style-type: none"> Planejar e implementar a curadoria 	Noonan e Chute (2014)
				<ul style="list-style-type: none"> Planejar, selecionar e implementar sistemas de preservação 	Dooley (2015)
				<ul style="list-style-type: none"> Desenvolver sistemas de gestão de dados de pesquisa 	Rutz e Collins (2005)

Fonte: elaborado pela autora.

Nessa fase são definidas as metodologias e tecnologias para o desenvolvimento de todo o ciclo de vida dos dados de pesquisa. Planeja-se a criação do dado, o método de captura, armazenamento; abordam-se questões tais como propriedade intelectual, embargos e restrições, financiamento, objetivos específicos da pesquisa, preservação de longo prazo. Na perspectiva de Humphrey *et al.* (2000), ao se envolver no início de um projeto de pesquisa, o arquivista de dados tem a oportunidade de introduzir padrões para a preparação de dados e documentação. Um dos desafios que os arquivistas enfrentam é a grande variedade de formatos em que os dados e a documentação são preparados.

6.2 Produção

“A realidade é que os arquivistas provavelmente coletaram dados de pesquisa ao longo dos anos, sejam eles realmente identificados como tal ou não. Os arquivistas devem coletar, preservar e fornecer acesso aos dados de pesquisa” (NOONAN; CHUTE, 2014, p. 211, tradução nossa).

As etapas dos ciclos relacionadas à produção de documentos referem-se à elaboração dos documentos em decorrência das atividades de um órgão. Nessa fase, o arquivista deve contribuir para que sejam criados apenas documentos essenciais à administração da instituição e evitadas duplicação e emissão de vias desnecessárias, propor consolidação de atos normativos alterados ou atualizados com certa frequência, visando à perfeita compreensão e interpretação dos textos, sugerir criação ou extinção de modelos, apresentar estudos sobre a adequação e o melhor aproveitamento de recursos reprográficos e informáticos, introduzir os metadados capazes de assegurar a conservação dos documentos e a perpetuidade dos dados e opinar sobre escolha de materiais e equipamentos (DELMAS, 2010; PAES, 2004).

As circunstâncias em torno de sua criação, propriedade, custódia e uso pretendido informam o entendimento do arquivista sobre o material, de modo que seu significado completo possa ser determinado e o próprio contexto possa ser compreendido e preservado. A compreensão do contexto é crítica em quase todos os estágios da gestão de dados, desde a investigação preliminar de uma aquisição até o processamento dos materiais e a criação de metadados (DOOLEY, 2015). Huvila (2016) afirma que do ponto de vista do arquivista, os dados são essencialmente uma caixa preta que precisa ser descrita, gerenciada e disponibilizada de acordo com os padrões relevantes e as necessidades dos usuários dos dados.

Com relação à descrição no contexto dos dados de pesquisa, Ramírez (2011, tradução nossa) assevera que conhecimento especializado, principalmente do autor dos dados, é necessário para a criação e aplicação de ontologias e metadados.

Neste sentido, a descrição arquivística identifica e explica o contexto e o conteúdo de documentos de arquivo, a fim de promover o acesso a estes, por meio de representações precisas e adequadas e pela organização dessas representações de acordo com modelos predeterminados. Processos relacionados à descrição podem começar na ou antes da produção dos documentos e continuam durante sua vida. Esses processos permitem instituir controles intelectuais necessários para tornar confiáveis, autênticas, significativas e acessíveis descrições que serão mantidas ao longo do tempo. O processo de descrição consiste na elaboração de instrumentos de pesquisa que possibilitem a identificação, o rastreamento, a localização e a utilização de

dados. Além disso, a descrição deve ser feita em relação à sua substância (indicando-se unidade de organização, funções, atividades, operações, assuntos) e estrutura. Por outro lado, a EAD desenvolvida e utilizada para a descrição de metadados arquivísticos baseados na linguagem de marcação XML atualiza e incorpora metadados relacionados aos padrões de metadados MARC, ISAD(G) e Dublin Core. A EAD permite a descrição, estruturação e interoperabilidade dos metadados arquivísticos referenciais, que, quando associados ao XML, possibilitam a decodificação e a apresentação das informações referenciais de forma estruturada aos usuários (BELLOTTO, 2005; CONARQ, 2011; PAES, 2004).

Segundo Ramírez (2011), a metodologia da arquivística está imersa em uma rica tradição de atividades curatoriais, com foco particular na avaliação, seleção e descrição. Essas práticas têm valor no domínio digital e devem ser consultadas ao se lidar com questões complicadas de gestão de dados. Os estudos arquivísticos, no seu cerne, são de natureza interdisciplinar, incorporando ideias das ciências e das humanidades, que se adaptam às necessidades da profissão arquivística. Portanto, os arquivistas estão bem posicionados para aconselhar sobre o alinhamento das práticas curatoriais entre as disciplinas e dentro delas. Eles não são estranhos à navegação nas complexas águas jurídicas, culturais e políticas e podem fornecer uma perspectiva bem-vinda sobre como lidar com questões de conformidade para dados digitais, principalmente atuando como consultores quando questões de ética e legalidade estiverem em jogo.

Quadro 5 – Produção.

Gestão Arquivística	Ciclo de vida dos dados de pesquisa		Ciclo da curadoria	Competências do Arquivista	Autores
	DDI	DataONE	DCC		
Produção	Coleta	Coletar Descrever	Criação e/ou recebimento	• Descrever os dados de pesquisa	Huvila (2016)
				• Auxiliar na coleta dos dados de pesquisa • Garantir a integridade dos dados de pesquisa	Humphrey <i>et al.</i> (2000)
			Descrição e representação	• Coletar (identificar tipos de dados de pesquisa; desenvolver processo de coleta)	Noonan e Chute (2014)
				• Criar metadados contextuais e descritivos para descoberta dos dados de pesquisa e acesso a estes	Dooley (2015)

(continua)

Quadro 5 – Produção (continuação).

Gestão Arquivística	Ciclo de vida dos dados de pesquisa		Ciclo da curadoria	Competências do Arquivista	Autores
	DDI	DataONE	DCC		
Produção	Coleta	Coletar Descrever	Criação e/ou recebimento	<ul style="list-style-type: none"> • Coletar dados, metadados • Descoberta de dados de pesquisa • Registrar os dados de pesquisa na base de dados • Descrever formato de arquivo 	Rutz e Collins (2005)
			Descrição e representação	<ul style="list-style-type: none"> • Questões éticas e legais • Garantir da proveniência e autenticidade 	Ramírez (2011)

Fonte: elaborado pela autora.

Rutz e Collins (2005) consideram que o envolvimento do arquivista motiva a equipe de pesquisa a trabalhar de forma mais eficaz para garantir que os dados e metadados sejam devidamente coletados, organizados e arquivados a longo prazo.

6.3 Utilização

O envolvimento precoce permite ao arquivista promover padrões que apoiem o acesso e a usabilidade, sem criar mais trabalho para o pesquisador (HUMPHREY *et al.*, 2000). Esses padrões alinham-se à operação de utilização dos documentos, que engloba atividades de protocolo (recebimento, classificação, registro, distribuição, tramitação) de expedição, de organização e arquivamento de documentos em fase corrente e intermediária, bem como a elaboração de normas de acesso à documentação, “empréstimo, consulta e a recuperação de informações, indispensáveis ao desenvolvimento de funções administrativas, técnicas ou científicas das instituições” (PAES, 2004, p. 54).

A usabilidade e a documentação adequada de dados são uma questão de preocupação para um arquivista de dados (HUVILA, 2016). Sob a custódia do arquivista, o documento torna-se acessível, de modo que possa ser usado pelas gerações presentes, bem como pelas futuras. O arquivista deve, pois, formular normas e métodos relativos ao uso dos documentos que sejam de interesse de todos. Ele deve fazer com que se possa dispor do material num nível máximo compatível com uma estimativa razoável quanto à preservação deste, pesando, de um lado, as solicitações dos consulentes atuais, e do outro as solicitações da posterioridade e de preservação dos documentos (SCHELLENBERG, 2006). Podemos visualizar o que foi descrito acima no Quadro 6:

Quadro 6 – Utilização.

Gestão Arquivística	Ciclo de vida dos dados de pesquisa		Ciclo da curadoria	Competências do Arquivista	Autores
	DDI	DataONE	DCC		
Utilização	Recuperação, análise e reuso	Descobrir, integrar e analisar	Acesso, uso e reuso	<ul style="list-style-type: none"> Promover o acesso de acordo com os padrões e as necessidades do usuário Publicar dados de pesquisa 	Huvila (2016)
				<ul style="list-style-type: none"> Arquivar Auxiliar no compartilhamento dos dados de pesquisa 	Humphrey <i>et al.</i> (2000)
				<ul style="list-style-type: none"> Promover o Acesso Restrições de uso 	Dooley (2015)
				<ul style="list-style-type: none"> Arquivar dados e documentação referente aos dados de pesquisa Promover o Acesso 	Rutz e Collins (2005)

Fonte: elaborado pela autora.

6.4 Avaliação e destinação

Noonan e Chute (2014) consideram que as atividades de avaliação e seleção são atividades importantes que o arquivista pode oferecer aos pesquisadores. Os pesquisadores entendem claramente que seus dados têm um valor primordial para eles. Ao formar parcerias e trabalhar com pesquisadores, os arquivistas podem fornecer treinamento e orientação quanto à identificação dos potenciais valores de evidência e informativos (secundários) dos dados de pesquisa.

No contexto arquivístico, a avaliação é um processo de análise e seleção de documentos que visa estabelecer os prazos de guarda e a destinação final deles, definindo quais documentos serão preservados para fins administrativos ou de pesquisa e em que momento poderão ser eliminados ou recolhidos ao arquivo permanente, segundo o valor e o potencial de uso que possuem para a administração que os gerou e para a sociedade (INDOLFO, 2007).

As considerações de avaliação consciente resultam do exame deliberado do conteúdo, contexto e estrutura dos dados, ao mesmo tempo que atendem aos interesses expressos pelo pesquisador. Isso pode levar ao desenvolvimento de critérios prudentes para a classificação, armazenamento, preservação e recuperação de dados. Por isso, pode ser útil aos pesquisadores ter arquivistas orientando-os nas decisões de avaliação dos dados de pesquisa quanto à

preservação de suas pesquisas e dados. Ao considerar o valor das informações, incluindo suas circunstâncias de criação e potencial de reutilização, os pesquisadores devem ser capazes de chegar a uma decisão mais clara sobre os dados que precisam ser preservados e acessados a longo prazo (NOONAN; CHUTE, 2014).

Desse modo, nas etapas de arquivamento (DCC) e preservação (DataONE), ficam implícitos as ações de transferência e o recolhimento arquivístico, ou seja, a destinação de documentos. Os documentos arquivísticos digitais em fase corrente e intermediária devem, preferencialmente, ser gerenciados por meio de um SIGAD, a fim de se garantir o controle do ciclo de vida, o cumprimento da destinação prevista e a manutenção da autenticidade e da relação orgânica, características fundamentais desses documentos. Já nessas fases, os produtores precisam tomar cuidados especiais, previstos em um plano de preservação digital, com relação aos documentos digitais que serão mantidos por médio e longo prazos, de forma a garantir sua autenticidade e seu acesso. A partir da destinação para guarda permanente, ocorre uma alteração na cadeia de custódia, passando a responsabilidade pela preservação dos documentos dos produtores para a instância de guarda. Os documentos digitais em fase permanente são dependentes de um bom sistema informatizado que apoie o tratamento técnico adequado, incluindo arranjo, descrição e acesso, de forma a assegurar a manutenção da autenticidade e da relação orgânica desses documentos. A preservação dos documentos arquivísticos digitais, nas fases corrente, intermediária e permanente, deve estar associada a um repositório digital confiável. Em outras palavras, os arquivos devem dispor de repositórios digitais confiáveis para a gestão, a preservação e o acesso de documentos digitais (CONARQ, 2015).

Dessa maneira, segundo Humphrey *et al.* (2000), estabelecer práticas reais de arquivamento de dados dentro de um método de pesquisa convencional é outra tarefa educacional do arquivista. Criar uma consciência do valor do arquivamento de dados entre os membros de uma equipe de pesquisa não é suficiente, e a implementação de práticas que conduzam a produtos de arquivamento também é necessária. É função do arquivista garantir que os dados sejam preparados com vistas à preservação de longo prazo. Nesse sentido, os procedimentos de arquivamento devem ser incorporados como parte do método desde o início do projeto de pesquisa.

Na abordagem das etapas de preservação e migração (DCC), pressupõe-se a definição de estratégias de preservação dos documentos arquivísticos para mantê-los acessíveis e utilizáveis pelo tempo que for necessário, garantindo sua longevidade, funcionalidade e acesso contínuo. Devem ser asseguradas as características dos documentos, tais como autenticidade e

acessibilidade, pela adoção de estratégias institucionais e técnicas proativas de produção e preservação que garantam sua perenidade. Essas estratégias são estabelecidas por uma política de preservação (CONARQ, 2011). A preservação deve abranger técnicas utilizadas para evitar os riscos provenientes da obsolescência tecnológica, como a migração, que consiste no conjunto de procedimentos e técnicas para assegurar a capacidade de os objetos digitais serem acessados em face das mudanças tecnológicas, ou seja, a transferência de um objeto digital:

- a) de um suporte que está se tornando obsoleto, fisicamente deteriorado ou instável para um suporte mais novo;
- b) de um formato obsoleto para um formato mais atual ou padronizado;
- c) de uma plataforma computacional em vias de descontinuidade para uma outra mais moderna (CONARQ, 2011).

Dessa maneira, para Noonan e Chute (2014), os arquivistas devem estar presentes desde o início quando as instituições embarcarem em atividades de curadoria de dados, independentemente dos tipos de dados que estejam sendo criados. Eles podem oferecer uma boa visão sobre avaliação, acesso, aquisição e preservação digital e complementar as iniciativas digitais de bibliotecários e da Equipe de TI, que geralmente faz parte do processo.

Na percepção de Ramírez (2011), os arquivistas enfrentam desafios de preservação digital. Como a informação digital não está mais em um meio fixo, mas pode ser um conjunto dinâmico e inter-relacionado de arquivos, o profissional enfrenta desafios de proveniência, autenticidade e preservação digital para que o usuário possa ver os dados como o usuário original os visualizou. A teoria e a prática arquivística podem ajudar a resolver questões em torno da preservação digital.

Com relação às etapas de armazenamento (DCC), asseguuração (DataONE), processamento e arquivo (DDI), elas compõem a operação de arquivamento. Trata-se de uma tarefa arquivística de maior importância, uma vez que a função primordial dos arquivos é disponibilizar as informações contidas nos documentos para a tomada de decisão, o que só é possível se os documentos tiverem sido classificados, através de métodos adequados, de forma a ficarem protegidos e serem facilmente localizados. Assim, mais importante do que guardar (arquivar) é achar (recuperar as informações), rapidamente, e no momento desejado (CONARQ, 2011; PAES, 2004). A seguir o Quadro 7.

Ramírez (2011) argumenta que o conhecimento do arquivista em gerenciar, descrever e organizar proporciona o acesso ágil às informações armazenadas em repositórios. Essa abordagem pode ser aplicada aos dados digitais, especialmente quando se tem o reuso dos dados.

Quadro 7 – Avaliação e destinação.

Gestão Arquivística	Ciclo de vida dos dados de pesquisa		Ciclo da curadoria	Competências do Arquivista	Autores
	DDI	DataONE	DCC		
Avaliação e destinação (eliminação ou guarda permanente)	Arquivo	Assegurar Preservar	Armazenamento Preservação e migração Arquivamento Avaliação, seleção e eliminação	<ul style="list-style-type: none"> Promover o reuso dos dados de pesquisa 	Huvila (2016)
				<ul style="list-style-type: none"> Garantir a preservação de longo prazo Promover o reuso dos dados de pesquisa 	Humphrey <i>et al.</i> (2000)
				<ul style="list-style-type: none"> Selecionar e Avaliar (identificar valores de evidência e informativos — secundários — dos dados de pesquisa) Desenvolver critérios prudentes para a classificação, armazenamento, preservação e recuperação de dados de pesquisa Promover a Preservação digital de longo prazo 	Noonan Chute (2014)
				<ul style="list-style-type: none"> Garantir Preservação de longo prazo 	Dooley (2015)
				<ul style="list-style-type: none"> Fomentar a descoberta de dados de pesquisa Garantir o registro dos dados de pesquisa na base de dados 	Rutz e Collins (2005)
				<ul style="list-style-type: none"> Promover a preservação digital 	Ramírez (2011)

Fonte: elaborado pela autora.

Percebe-se que os conhecimentos do arquivista em organização, descrição, acesso, restrição de uso, utilização, preservação, estratégias de preservação, políticas, gestão, arquivamento, proveniência, autenticidade, integridade, avaliação e seleção, para citar alguns campos, são atividades inerentes do profissional na gestão dos documentos arquivísticos que podem ser aplicados na gestão dos dados de pesquisa.

A coleta, organização e preservação de recursos a longo prazo são a razão de ser dos arquivos e arquivistas. A comunidade arquivística, no entanto, deixou de lado os dados científicos, presumindo que eles estivessem fora dos limites de suas preocupações profissionais. Os cientistas, por outro lado, reconhecem cada vez mais que não têm as habilidades e os conhecimentos necessários para atender às demandas que lhes são impostas no tocante à curadoria de dados e estão buscando a ajuda de “arquivistas de dados”. Isso representa uma oportunidade significativa para os arquivistas (AKMON *et al.*, 2011).

Ramírez (2011) cita outras habilidades e conhecimentos que os arquivistas podem aplicar na gestão de dados de pesquisa: proveniência de dados de pesquisa e autenticidade:

A ciência arquivística se concentra em avaliar, selecionar, descrever, gerenciar, cuidar da fonte (proveniência), autenticar e preservar. Proveniência dos dados — o rastreamento de todos os contextos e transformações pelos quais os dados passaram — é a chave para verificar a autenticidade e confiabilidade dos arquivos de dados. (RAMIREZ, 2011, p. 22, tradução nossa).

Na percepção de Almeida (2015), a proveniência de dados vem ganhando destaque no contexto científico, tanto para garantir a origem dos dados como para avaliar a sua acurácia, além de permitir a manutenção e recuperação destes, mantendo seu significado. Buneman, Khanna e Chiew (2001) afirmam que a proveniência de dados, igualmente chamada de linhagem, genealogia ou *pedigree*, compreende os metadados que descrevem o histórico de como aquele dado foi produzido ou derivado, que contém a descrição de “como”, “quando”, “onde”, “por que” e “quem” criou ou obteve um determinado dado.

Segundo Bose e Frew (2005), a pesquisa científica depende tanto da disseminação e troca de conjuntos de dados quanto da publicação de conclusões. O rastreamento preciso da linhagem (origem e histórico de processamento subsequente) dos conjuntos de dados científicos é, portanto, fundamental para a documentação completa do trabalho científico.

Na arquivística, a proveniência se dá por meio da descrição, que determina como foi criado o documento arquivístico, identificando-se o produtor, data, tipo do documento, suporte e contexto de criação.

Duchein (1986) esclarece, no tocante à proveniência, que “[...] é essencial para a apreciação de um documento qualquer saber com exatidão quem o produziu, em que circunstâncias, no desenrolar de que ação, com que objetivo, a quem se destinava, como e quando foi recebido por seu destinatário e por que meios chegou até nós” (DUCHEIN, 1986, p. 17).

Com relação à autenticidade e confiabilidade na arquivística citadas por Ramírez (2011), Santos (2015,) refere-se a ela como o conceito de controle sobre a criação, manutenção e custódia do documento arquivístico, o que em análise inicial evitaria sua alteração não autorizada, e simultaneamente permitiria a identificação da autoria de qualquer modificação nele realizada. Tal conceito envolve preponderantemente sua relação com o produtor e sua capacidade de ser utilizado para os fins de sua criação.

O conceito de confiabilidade definido por Rondinelli (2017) tem a ver com a veracidade de seu conteúdo, sendo esta de total responsabilidade de seu produtor, representado pela pessoa

responsável por elaborá-lo — isto é, seu autor —, bem como a pessoa responsável pela sua gestão.

A aplicabilidade da proveniência arquivística, da autenticidade e da confiabilidade no contexto dos dados de pesquisa pode garantir a integridade desses dados, dessa maneira possibilitando o processo que permite a identificação do dado de pesquisa de forma a refletir todo o seu método de criação e contexto em que foi gerado, como também a disponibilização das informações de modo que se tenha uma visão total dos acontecimentos e fatos que marcaram sua produção, reconhecendo-se o valor das informações que façam sentido ao produtor e a quem interessar o acesso ao dado de pesquisa, e tornando-as acessíveis e compreensíveis para o reuso.

A atuação do arquivista na gestão de dados de pesquisa justifica-se na medida em que a contribuição do profissional se materializa no auxílio à elaboração do plano de gestão de dados, na política de preservação e ações de preservação que garantam a proveniência, autenticidade, integridade e confiabilidade dos dados de pesquisa, na determinação e normalização de formatos de preservação a longo prazo, na contribuição com a produção, avaliação, destinação dos dados de pesquisa, na identificação do valor dos dados (primário e secundário), na seleção e planejamento de repositórios confiáveis visando ao arquivamento de longo prazo, na contribuição com a representação arquivística dos dados de pesquisa, no auxílio à adesão a questões éticas, legais e de *copyright*, na orientação sobre o armazenamento seguro e *backup*, na promoção da descoberta, do acesso e do reuso dos dados.

6.5 Perfil do Arquivista no contexto da gestão de dados de pesquisa

Em harmonia com o discurso desta pesquisa, torna-se necessário estabelecer o perfil do arquivista de dados de pesquisa. Mueller (1989) define perfil profissional como o conjunto de conhecimentos, qualidades e competências próprias dos integrantes de uma profissão, o que se relaciona à ideia de função profissional — o perfil é delineado pelas habilidades, competências e atitudes necessárias para o desempenho da função profissional.

Durand (1998, 2000) fundamenta a noção de competência em três dimensões: *Knowledge* (conhecimento), *Know-How* (habilidade) e *Attitudes* (atitude), o que consiste em unir questões técnicas e cognitivas necessárias para a realização de um determinado trabalho.

Fleury e Fleury (2001), nessa perspectiva, definem competência como um conjunto de conhecimentos, habilidades e atitudes (isto é, um conjunto de capacidade humanas). A junção das três iniciais (CHA), segundo Urbanavicius Júnior *et al.* (2006), é tudo que uma função/cargo

de uma empresa exige para que o serviço/produto seja bem administrado e de boa qualidade. No entanto, essas atribuições precisam estar bem definidas e atualizadas. Rabaglio (2001), citado por Urbanavicius Júnior *et al.* (2006), aponta significados para essas letras, assim como segue:

- **C**: saber (conhecimentos adquiridos no decorrer da vida, nas escolas, universidades, cursos etc. Ex.: conhecimento da concorrência e de técnicas de negociação);
- **H**: saber fazer (capacidade de realizar determinada tarefa, física ou mental. Ex.: análise da concorrência e negociação);
- **A**: querer fazer (comportamentos que temos diante de situações do nosso cotidiano e das tarefas que desempenhamos no nosso dia a dia. Ex.: participar da concorrência e fazer negociações).

A é uma organização internacional de profissionais que trabalham com tecnologia da informação e serviços de dados para apoiar a pesquisa e o ensino. Tais profissionais são das áreas de tecnologia da informação, bibliotecas, serviços de dados, pesquisa e ensino superior, de setores governamentais, sem fins lucrativos e privados de pesquisa, e atuam para descobrir dados, curá-los, disponibilizá-los e mantê-los disponíveis. Eles trabalham juntos para defender o uso e o gerenciamento responsáveis de dados, para construir uma comunidade mais ampla em torno dos dados de pesquisa e incentivar o desenvolvimento de profissionais de dados. A associação conta com um repositório de emprego, o *The IASSIST Jobs Repository* (IASSIST, 2021b).

Cabe ressaltar a dificuldade em localizar vagas relacionadas ao arquivista de dados de pesquisa, a começar pela nomenclatura do cargo, que não tem relação com dados de pesquisa em alguns casos. A seleção dos anúncios em certas situações ocorreu mediante a leitura da descrição do trabalho a ser realizado.

O LinkedIn foi selecionado para a pesquisa por ser uma rede profissional mundialmente conhecida. As vagas identificadas não tinham relação com a atuação do arquivista na gestão de dados de pesquisa, mas com o trabalho cotidiano do profissional. Por esse motivo, não foram integradas à pesquisa.

As vagas identificadas no *The IASSIST Jobs Repository*, da IASSIST, são mostradas a seguir:

Quadro 8 – Cargo de Arquivista de dados de pesquisa.

Cargo: Arquivista de dados de pesquisa		
Empregador: Instituto Odum da Universidade da Carolina do Norte Publicado no site da IASSIST Site: https://iassistdata.org/jobs-repository/		
Resumo do trabalho: O objetivo principal desta posição é: 1) executar gerenciamento de dados baseados em padrões e fluxos de trabalho de curadoria; 2) estabelecer programas de treinamento que apoiem a adoção pelos pesquisadores das melhores práticas de gerenciamento de dados como parte da prática normativa de pesquisa; 3) aumentar a conscientização sobre as ferramentas e serviços do <i>Odum Institute Data Archive</i> ; 4) expandir as coleções de arquivos de dados de pesquisa por meio de solicitações a produtores locais e externos de dados de ciências sociais que atendam aos critérios da política de desenvolvimento de coleções; e 5) fazer contribuições significativas de pesquisa e desenvolvimento para a profissão de arquivista de dados para ajudar a melhorar a prática da curadoria de dados.		
Conhecimento	Habilidade	Atitude
<ul style="list-style-type: none"> • Gerenciamento de dados e curadoria no contexto do ciclo de vida da pesquisa, de acordo com os padrões estabelecidos e as melhores práticas • A publicação científica da <i>Open Science</i> • Práticas de gerenciamento e curadoria de dados durante a concepção do projeto de pesquisa • Padrões de arquivamento de dados vigentes • Pacotes de software estatístico, incluindo Stata, R e SPSS 	<ul style="list-style-type: none"> • Desenvolvimento e fornecimento de treinamento em gerenciamento de dados para públicos de diversos conhecimentos • Coleta de dados, análise de dados e compartilhamento de dados 	<ul style="list-style-type: none"> • Capacidade de comunicar conceitos de gerenciamento de dados e curadoria verbalmente e por escrito para vários públicos • Aptidão para obter o domínio de outras linguagens de programação complexas e códigos

Fonte: dados coletados na IASSIST (2021b, *online*, tradução nossa).

Quadro 9 – Cargo de Arquivista Assistente.

Cargo: Arquivista assistente		
Empregador: Universidade de Michigan, ICPSR Publicado no site da IASSIST Site: https://iassistdata.org/jobs-repository/		
Resumo do trabalho: Arquivista Assistente para aprimorar os metadados e auxiliar no planejamento da preservação digital. Auxiliar no repositório, planejar políticas e procedimentos para acesso confiável e de longo prazo a recursos digitais.		
Conhecimento	Habilidade	Atitude
<ul style="list-style-type: none"> • Catalogação descritiva de coleções de dados de ciências sociais • Criação, edição e transformação de metadados • Prática no uso de MARC, Dublin Core, Schema.org, DDI e outros padrões e esquemas de metadados • Curadoria de coleções de dados de ciências sociais 	<ul style="list-style-type: none"> • Gestão e coordenação • Organizacional e de gerenciamento de projetos • Revisão e limpeza dos metadados • Aderência aos padrões de metadados • Orientação sobre padrões de metadados • Auxílio no desenvolvimento e implementação de procedimentos e fluxos de trabalho • Participação na implementação da DDI • Planejamento de preservação • Criação, manutenção e apresentação de políticas, procedimentos, fluxos de trabalho e estratégias de preservação digital 	<ul style="list-style-type: none"> • Comunicação verbal, escrita e interpessoal • Capacidade para trabalhar em equipe

Fonte: dados coletados na IASSIST (2021b, *online*, tradução nossa).

Quadro 10 – Cargo de Arquivista Associado.

Cargo: Arquivista associado		
Empregador: Universidade de Michigan, ICPSR Publicado no site da IASSIST Site: https://iassistdata.org/jobs-repository/		
Resumo do trabalho: Arquivista Associado para desenvolver e aprimorar os metadados e liderar o planejamento da preservação digital, revisar e implementar políticas e procedimentos para acesso confiável e de longo prazo aos recursos digitais.		
Conhecimento	Habilidade	Atitude
<ul style="list-style-type: none"> • Criação, edição e transformação de metadados em um ambiente de repositório digital • ISO 14721 (Modelo de Referência OAIS) e ISO 16363 (Auditoria e Certificação de Repositórios Digitais Confiáveis) • Curadoria de coleções de dados de ciências sociais • Catalogação descritiva de coleções de dados de ciências sociais • Linguagens de <i>script</i> para manipulação de dados (por exemplo, Python, XSLT) • DDI, MARC, Dublin Core, Schema.org, DCAT e outros padrões e esquemas de metadados 	<ul style="list-style-type: none"> • Organizacional e de gerenciamento de projetos • Desenvolvimento de padrões de metadados • Pesquisa e implementação de novos padrões de metadados e políticas para novos e crescentes tipos de dados • Planejamento de Preservação • Liderança na criação e implementação de políticas, procedimentos, fluxos de trabalho e estratégias de preservação digital, especialmente para apoiar a conformidade com a ISO 14721 (Modelo de Referência OAIS) e a ISO 16363 (Auditoria e Certificação de Repositórios Digitais Confiáveis) • Revisão e implementação de novos padrões e políticas de metadados 	<ul style="list-style-type: none"> • Comunicação verbal, escrita e interpessoal • Capacidade para trabalhar em equipe

Fonte: dados coletados na IASSIST (2021b, *online*, tradução nossa).

Quadro 11 – Cargo de Arquivista de dados Sênior

Cargo: Arquivista de dados Sênior		
Empregador: Cornell University Publicado no site da IASSIST Site: https://iassistdata.org/jobs-repository/		
Resumo do trabalho: Este profissional garante que os dados aceitos no Arquivo sejam mantidos usando as melhores práticas em métodos de arquivamento e sejam oferecidos nos formatos atuais de modo a permitir tanto a preservação quanto a disseminação a longo prazo. O Arquivista: desenvolve e adota padrões de arquivamento para as coleções do Arquivo; gerencia e coordena vários fluxos de trabalho dentro das coleções de arquivos de acordo com a expansão da curadoria de dados e requisitos de acessibilidade on-line; oferece suporte aos usuários na utilização de arquivos de dados para análise secundária; coordena as operações diárias que garantem a qualidade, disponibilidade e abrangência do banco de dados; planeja, implementa e dirige fluxos de trabalho de produção, além de trabalhar em conjunto com doadores e provedores de dados para desenvolver coleções e atender aos usuários no tocante a localização, acesso e uso dos arquivos de dados para análise secundária. O arquivista desempenha um papel importante no suporte ao gerenciamento de longo prazo de dados de pesquisa em todo o ciclo de vida dos dados.		
Conhecimento	Habilidade	Atitude
<ul style="list-style-type: none"> • Gerenciamento de arquivos e fluxos de trabalho • Teoria e prática de gerenciamento de arquivos e registros • Teoria e prática da preservação de dados • Estratégias e tecnologias utilizadas pela comunidade arquivística para o gerenciamento de dados • Padrões relevantes para a descrição de dados de arquivo, incluindo DDI, e familiaridade com outros padrões de metadados, como METS e PREMIS • Métodos e habilidades de pesquisa 	<ul style="list-style-type: none"> • Organizacional e capacidade de planejar, coordenar e implementar projetos complexos • Análises científicas • Implementação de políticas, padrões e procedimentos para gerenciamento de dados em um arquivo 	<ul style="list-style-type: none"> • Comunicação interpessoal, oral e escrita • Capacidade de trabalhar de forma independente e colaborativa com uma variedade de funcionários em um ambiente que muda rapidamente

Fonte: dados coletados na IASSIST (2021b, *online*, tradução nossa).

Quadro 12 – Cargo de Arquivista Chefe.

Cargo: Arquivista chefe		
Empregador: <i>The Social and Economic Survey Research Institute</i> , Universidade do Qatar Publicado no site da IASSIST Site: https://iassistdata.org/jobs-repository/		
Resumo do trabalho: O Arquivista trabalhará em estreita colaboração com os pesquisadores e a equipe de pesquisa no processamento, limpeza, documentação, disseminação, preservação e suporte de conjuntos de dados coletados, monitorará o uso de dados, avaliará a aderência à disseminação e às políticas de dados e suporte, além de promover o uso de conjuntos de dados.		
Conhecimento	Habilidade	Atitude
<ul style="list-style-type: none"> • Pacotes estatísticos como SAS, SPSS, STATA e R • Processamento e documentação de conjuntos de dados • Processo de pesquisa 	<ul style="list-style-type: none"> • Liderar a Unidade de Arquivo do Instituto, além de supervisionar a equipe de processamento de dados e documentação • Conhecer os projetos e programas de pesquisa • Avaliar e recomendar políticas e procedimentos para processamento, documentação, arquivamento, disseminação e formatos de armazenamento e documentação • Receber conjuntos de dados e documentação em vários formatos de armazenamento e converter esses formatos de armazenamento nos usados pelo SESRI para processamento, disseminação e arquivamento • Processar questionários finais para divulgação ao público no site • Processar e limpar conjuntos de dados • Verificar erros de padrão, registros duplicados, conduta imprópria do entrevistador, problemas de coleta de dados e questões metodológicas • Desenvolver livros de código em estreita coordenação com o pesquisador, incluindo metadados, descrições de estudo, informações de desenho de estudo, formatos de citação, relatórios metodológicos, notas de pesquisa, instruções de uso de dados, resumo, notas de processamento, relatórios de erros e referências bibliográficas • Especificar recursos e trabalhar com membros da equipe de Tecnologia da Informação no desenvolvimento de ferramentas de disseminação personalizadas ou na avaliação e implementação de pacotes de software e ferramentas existentes, conforme apropriado • Garantir a adesão às políticas de divulgação, citação e uso de dados • Identificar ferramentas e procedimentos que serão benéficos para o processamento, uso e preservação dos dados e documentação • Fornecer relatórios regulares ao Chefe de Pesquisa sobre estatísticas de uso e <i>download</i>, o tamanho e a atividade da comunidade de usuários, novas entradas bibliográficas e tendências no uso de conjuntos de dados no arquivo 	<ul style="list-style-type: none"> • Organizado e com atenção aos detalhes • Trabalho em conjunto com os pesquisadores e membros da equipe de pesquisa para sugerir e implementar estratégias • Atendimento ao usuário • Orientação da equipe

Fonte: dados coletados na IASSIST (2021b, *online*, tradução nossa).

Na busca por mais evidências da atuação do arquivista de dados de pesquisa, foi realizada uma pesquisa no site institucional dos empregadores citados na descrição das vagas de emprego: Instituto Odum da Universidade da Carolina do Norte, *Inter-university Consortium for Political and Social Research* (ICPSR) da Universidade de Michigan, Cornell University e *The Social and Economic Survey Research Institute* da Universidade do Qatar.

Fundado em 1924, o *Odum Institute for Research in Social Science*¹⁰ apoia a missão de ensino e pesquisa em ciências sociais na Universidade da Carolina do Norte. Ele opera um arquivo de dados de renome mundial, sendo líder em ferramentas de arquivamento e desenvolvimento de suporte. O instituto auxilia pesquisadores no desenvolvimento de propostas de financiamento e fornece suporte completo ao ciclo de vida da pesquisa (ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE, 2022).

Na equipe do instituto há duas arquivistas, as quais fornecem consultoria e suporte nas seguintes áreas: Doações para as coleções de arquivos do Instituto Odum, Suporte ao UNC Dataverse, Gerenciamento e curadoria de dados, Planos de gerenciamento de dados, Arquivamento e compartilhamento de dados, Desenvolvimento e implementação de políticas de dados, Práticas recomendadas de pesquisa reproduzíveis, Educação e treinamento em dados, Localização e acesso a dados, Repositório de Dados (ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE, 2022).

O ICPSR¹¹ é uma unidade do Instituto de Pesquisa Social da Universidade de Michigan. Trata-se de um consórcio internacional de mais de 750 instituições acadêmicas e organizações de pesquisa que oferece liderança e treinamento em acesso a dados, curadoria e métodos de análise à comunidade de pesquisa em ciências sociais. Ele mantém um banco de dados de mais de 250.000 arquivos de pesquisas nas ciências sociais e comportamentais, hospedando 21 coleções especializadas de dados sobre educação, envelhecimento, justiça criminal, abuso de substâncias, terrorismo e outros campos. Além disso, colabora com vários financiadores, incluindo agências e fundações estatísticas dos EUA, para criar coleções de dados temáticos e projetos de pesquisa e administração de dados. A pesquisa patrocinada pelo ICPSR se concentra nos desafios emergentes da curadoria digital e da ciência de dados (ICPSR, 2022).

No site institucional, na parte que se refere às informações institucionais, mais precisamente onde estão listadas as pessoas que fazem parte da equipe do ICPSR, dois arquivistas são mencionados como membros da equipe. Na descrição do perfil profissional, não

¹⁰ <https://odum.unc.edu/>.

¹¹ <https://www.icpsr.umich.edu/web/pages/>.

constam as atividades desempenhadas por eles, mas há uma breve biografia que destaca os projetos realizados pelos arquivistas e uma relação de publicações (ICPSR, 2022).

Um dos arquivistas participou de vários projetos colaborativos, tendo atuado como Diretor de Processamento de Dados da Pesquisa Nacional de Crescimento Familiar e como Coinvestigador Principal da Série Integrada de Pesquisa em Fertilidade. O outro ocupou o cargo de Arquivista e dirigiu uma unidade do ICPSR, tendo sido responsável pela supervisão de atividades nas áreas de Metadados, Publicações, Desenvolvimento de Website, suporte ao Usuário e Desenvolvimento. Ele atuou na diretoria da DDI, que visava estabelecer um padrão de metadados para as ciências sociais e comportamentais. Além disso, participa de projetos internacionais relacionados a gestão de dados, incluindo o *Data Seal of Approval*, a *Research Data Alliance* e várias iniciativas para promover a citação de dados (ICPSR, 2022).

Cornell¹² é uma universidade de pesquisa privada e parceira da Universidade do Estado de Nova York. Como instituição federal de concessão de terras no estado de Nova York, ela tem a responsabilidade — única dentro da Ivy League — de contribuir em todas as áreas do conhecimento de forma a priorizar o engajamento público para ajudar a melhorar a qualidade de vida nesse estado, na nação e no mundo inteiro. No site institucional, nenhuma informação referente à equipe que trabalha na instituição foi encontrada (CORNELL UNIVERSITY, 2022).

O *Social and Economic Survey Research Institute* (SESRI)¹³, uma iniciativa de pesquisa científica social da Universidade do Qatar, foi fundado em outubro de 2008 com o apoio entusiástico da liderança da Universidade do Qatar. A missão do SESRI é fornecer dados sólidos e confiáveis para orientar a formulação de políticas, estabelecimento de prioridades e planejamento baseado em evidências nos setores sociais e econômicos. No site institucional, não constam informações sobre o profissional arquivista que lida com dados de pesquisa (SESRI, 2022).

Observa-se que, nas vagas ofertadas, nem todos os conhecimentos solicitados são desconhecidos do arquivista, podendo-se citar o PREMIS, OAIS e questões relacionadas a repositórios. Além das habilidades e atitudes mencionadas por Bellotto (2005), o arquivista do século XXI não deve se esquecer das qualidades esperadas do profissional. Em qualquer situação, tempo e lugar, com ou sem tecnologia, ele deve possuir a capacidade de análise e síntese, a aptidão para esclarecer situações complexas, a habilidade de se expressar claramente, tanto na forma escrita quanto na verbal, a competência para tomar decisões e a capacidade de adaptação à realidade das condições de seu tempo e lugar.

¹² <https://www.cornell.edu/>.

¹³ <http://sesri.qu.edu.qa/research/sesri>.

Outras qualidades ainda são elencadas por Bellotto (2005): adaptabilidade pragmatismo, curiosidade intelectual, rigor, método, continuidade, capacidade de compreensão e de escuta relativamente ao produtor, pesquisador, e ao cidadão. Em outros casos, conhecimento, habilidades e atitudes são solicitados dependendo da área de conhecimento em que a vaga de emprego é ofertada, o que mostra que o arquivista de dados de pesquisa deve adquirir conhecimento pertinente à área do conhecimento na qual ele irá atuar.

Percebe-se, nas vagas selecionadas no item conhecimento, que os padrões PREMIS e OAIS são requeridos. No item habilidades, as atividades relacionadas são gestão, coordenação, estratégias de preservação digital, avaliação e recomendação de políticas, implementação de políticas, padrões e procedimentos para gerenciamento de dados em um arquivo. No item atitude, trabalho em equipe e orientação de equipe são alguns dos requisitos solicitados. Em alguns casos, outros conhecimentos, habilidades e atitudes são solicitados dependendo da área de conhecimento em que a vaga de emprego é ofertada.

A pesquisa de vagas no mercado de trabalho demonstrou a baixa visibilidade da atuação do arquivista de dados de pesquisa. Ao que parece, existe a necessidade de atuação de profissionais nesse contexto, mas ainda não há uma ideia clara do papel do arquivista na gestão de dados de pesquisa.

7 CONCLUSÃO

A pesquisa foi desafiadora porque a temática arquivista de dados de pesquisa no Brasil é pouco conhecida enquanto no âmbito internacional é um assunto mais palpável. Foram poucos artigos e informações relacionados ao tema. Outro ponto que chamou atenção foi a questão da nomenclatura. O termo arquivista de dados é um termo muito abrangente. Tanto pode estar relacionado ao arquivista de dados de pesquisa como arquivista de dados voltado para a área computacional, por exemplo. O termo “arquivista de dados de pesquisa” foi empregado na pesquisa para especificar a atuação do profissional na gestão de dados de pesquisa, tendo em vista que na literatura o termo usado “arquivista de dados” é mais amplo, o que dificultou a identificação de autores relacionados à temática.

A análise dos ciclos de vida dos dados de pesquisa DDI e DataONE e da curadoria DCC permitiu constatar que a teoria e a prática arquivística inserem o arquivista na gestão de dados de pesquisa, possibilitando a esse profissional a aplicação dos processos arquivísticos de gestão à preservação a longo prazo no tratamento dos dados de pesquisa, com destaque para o conhecimento e prática em avaliação, seleção, descrição, gestão e preservação, o uso secundário dos dados de pesquisa; além de deixar claro que a atuação do arquivista deve começar nas primeiras discussões e no planejamento da gestão de dados de pesquisa, mostrando que no tocante à questão da pesquisa: “Qual o papel do arquivista na gestão de dados de pesquisa?”, esse papel começa antes da gestão, se estendendo à preservação e ao reuso dos dados de pesquisa. Orientar o pesquisador no desenvolvimento da pesquisa com segurança e organização.

A prática e o conhecimento do arquivista, como avaliação, seleção, e o perfil do profissional de gestão para a preservação são mencionados pelos pesquisadores nos artigos. Os pesquisadores não têm uma ideia clara da atuação do arquivista. Citam a atuação de forma intuitiva. Os artigos utilizados na pesquisa são de pesquisadores discorrendo sobre a importância e contribuição do arquivista na gestão de dados de pesquisa.

Conceber a atuação do arquivista para além das instituições arquivísticas, em conjunto com a revisitação das bases teóricas com um novo olhar, permite que o profissional vislumbre novas áreas de atuação, ou talvez não se trate de uma área nova. O perfil do arquivista de dados construído a partir da análise das vagas de emprego evidencia como são poucas as vagas ofertadas no mercado de trabalho. Nas bases do mercado de trabalho alguns conhecimentos arquivísticos são solicitados, tais como padrão do OAIS, metadados de preservação PREMIS e questões relacionadas a repositórios.

Esta pesquisa focou questões tais como a pouca ou nenhuma visibilidade do arquivista de dados de pesquisa na comunidade arquivística. A continuidade dos estudos deve ser desenvolvida para todas as questões passíveis de interlocução com a atuação do arquivista na gestão de dados de pesquisa, com o intuito de orientar melhor o profissional que irá atuar nesse contexto e fazer com que o conhecimento e a prática arquivista sejam aderentes ao universo dos dados de pesquisa, promovendo a atuação e a visibilidade do profissional no tratamento desses dados. Esse debate merece ser mais aprofundado em pesquisas futuras, tendo em vista a dificuldade de encontrar literatura relacionada à temática da pesquisa. Essa talvez seja uma perspectiva ainda pouco percebida, com potencial de investigação produtiva com articulações devidamente fundamentadas na arquivística.

REFERÊNCIAS

ACADEMIA BRASILEIRA DE CIÊNCIAS (ABC). **ABC lança documento sobre abertura e gestão de dados para a ciência**. Rio de Janeiro, 3 set. 2020. Disponível em: <https://www.abc.org.br/2020/09/03/abc-lanca-documento-sobre-abertura-e-gestao-de-dados-para-a-ciencia/>. Acesso em: 18 maio 2020.

AKMON, D. *et al.* The Application of Archival Concepts to a Data-Intensive Environment: Working with Scientists to Understand Data Management and Preservation Needs. **Archival Science**, Ann Arbor, v. 11, n. 3/4, p. 329-348, 2011. DOI: <https://doi.org/10.1007/s10502-011-9151-4>. Disponível em: <https://deepblue.lib.umich.edu/handle/2027.42/86738>. Acesso em: 7 fev. 2022.

ALMEIDA, R. P. **Gerenciamento de dados de proveniência de Workflow de bioinformática com banco de dados baseados em Grafo**. 2015. 62 f. Dissertação (Mestrado em Informática) – Instituto de Ciências Exatas, Universidade de Brasília, Brasília, DF, 2015. Disponível em: https://bdtd.ibict.br/vufind/Record/UNB_c8fc3992d93e16edb257033d774b65e8. Acesso em: 6 fev. 2022.

ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 132 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010. Disponível em: https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/alves_rcv_do_mar.pdf. Acesso em: 7 fev. 2022.

APPEL, A. L.; MACIEL, M. L.; ALBAGLI, S. A e-Science e as novas práticas de produção colaborativa do conhecimento científico. **Revista Internacional de Ciência y Sociedad**, [Madri], v. 3, n. 1, p. 41-52, 2016.

ARAÚJO, D. G. *et al.* Contribuições para a gestão de dados científicos: análise comparativa entre modelos de ciclo de vida dos dados. **Liinc Em Revista**, Rio de Janeiro, v. 15, n. 2, p. 32-51, 2019. DOI: <https://doi.org/10.18617/liinc.v15i2.4686>. Disponível em: <http://revista.ibict.br/liinc/article/view/4686>. Acesso em: 7 fev. 2022.

ARAUJO, R. O.; FINAMOR, M. S. Curadoria digital: papéis e responsabilidades do arquivista. **Informação@Profissões**, Londrina, v. 6, n. 1, p. 44-68, 2017. DOI: <http://dx.doi.org/10.5433/2317-4390.2017v6n1p44>. Disponível em: <https://www.uel.br/revistas/uel/index.php/infoprof/article/view/31381>. Acesso em: 7 fev. 2022.

AROVILIUS, R. *et al.* **Management and preservation of scientific records and data**. Paris: ICA/SUV, 2010. Disponível em: https://www.ica.org/sites/default/files/SUV_Management-and-preservation-scientific-records-Handbook_2010-12_EN.pdf. Acesso em: 15 jun. 2022.

ARQUIVO PORTUGUÊS DE INFORMAÇÃO SOCIAL (APIS). **Preparar os dados e a documentação**. Lisboa, [2021]. Disponível em: <http://www.apis.ics.ulisboa.pt/preparar-os-dados/>. Acesso em: 4 fev. 2022.

BALL, A. **A review of data management lifecycle models**. Bath: University of Bath, 2012. Disponível em: <https://researchportal.bath.ac.uk/en/publications/review-of-data-management-lifecycle-models>. Acesso em: 4 fev. 2022.

BELLOTTO, H. L. **Arquivos permanentes: tratamento documental**. Rio de Janeiro: Editora FGV, 2005.

BERNARDES, I. P.; DELATORRE, H. **Gestão documental aplicada**. São Paulo: Arquivo Público do Estado de São Paulo, 2008.

BOSE, R.; FREW, J. Lineage retrieval for scientific data processing: a survey. **ACM Computing Surveys**, New York, v. 37, n. 1, p. 1-28, 2005. DOI: <https://doi.org/10.1145/1057977.1057978>.

BRASIL. **Lei nº 8.159, de 8 de janeiro de 1991**. Dispõe sobre a política nacional de arquivos públicos e privados e dá outras providências. Brasília, DF: Presidência da República, 1991. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/18159.htm. Acesso em: 4 fev. 2022.

BUNEMAN, P.; KHANNA, S.; CHIEW, W. Why and Where: a Characterization of Data Provenance. *In*: INTERNATIONAL CONFERENCE ON DATABASE THEORY – ICDT, 8th., 2001, London. **Proceedings** [...]. London: LNCS, p. 316-330. DOI: https://doi.org/10.1007/3-540-44503-X_20. Disponível em: <https://pdfs.semanticscholar.org/b647/2ee11749ef70713bfb0e322a9ec27523ed88.pdf>. Acesso em: 7 fev. 2022.

CAPLAN, P. **Understanding PREMIS**. Washington, DC: Library of Congress, 2009. Disponível em: <http://www.loc.gov/standards/premis/understanding-premis.pdf>. Acesso em: 5 fev. 2022.

CESSDA. Bergen, [2022]. Disponível em: <https://www.cessda.eu/>. Acesso em: 7 fev. 2022.

CHEIN, M. *et al.* **Big Data: related technologies, challenges and future prospects**. Vancouver: Springer International Publishing, 2014.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Câmara Técnica de Documentos Eletrônicos (CTDE). **Glossário**. Versão 5.0. Rio de Janeiro: CONARQ, ago. 2009. Disponível em: https://www.gov.br/conarq/pt-br/assuntos/camaras-tecnicas-setoriais-inativas/camara-tecnica-de-documentos-eletronicos-ctde/2009_glossario_v5.pdf. Acesso em: 7 fev. 2022.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Câmara Técnica de documentos eletrônicos. **e-ARQ Brasil: modelo de requisitos para sistemas informatizados de gestão arquivística de documentos**. Rio de Janeiro: Arquivo Nacional, 2011.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Carta para a preservação do patrimônio arquivístico digital**. Brasília, DF: Unesco, 2005. Disponível em: https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/conarq_carta_preservacao_patrimonio_arquivistico_digital.pdf. Acesso em: 4 fev. 2022.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Diretrizes para implementação de repositórios arquivísticos digitais confiáveis- RDC-Arq**: Resolução nº 43, de 04 de setembro de 2015. Rio de Janeiro: CONARQ, 2015. Disponível em: https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/conarq_diretrizes_rdc_arq_resolucao_43.pdf. Acesso em: 7 fev. 2022.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEM (CCSDS). **Reference model for an Open Archival Information System (OAIS)**: Recommended practice CCSDS 650.0-M-2. Washington, DC: Magenta Book, 2012. Disponível em: <https://public.ccsds.org/pubs/650x0m2.pdf>. Acesso em: 5 fev. 2022.

CORETRUSTSEAL. [Edinburgh, 2022]. Disponível em: <https://www.coretrustseal.org/>. Acesso em: 7 fev. 2022.

CORNELL UNIVERSITY. Ithaca, [2022]. Disponível em: <https://www.cornell.edu/>. Acesso em: 5 fev. 2022.

CORRÊA, F. C. **Gestión de datos de investigación**. Barcelona: Editorial UOC, 2016.

DATA ONE. **About DataONE**. Santa Barbara, [2021]. Disponível em: <https://www.dataone.org/about/>. Acesso em: 4 fev. 2022.

DATA'S shameful neglect. **Nature**, London, v. 461, n. 7261, p. 145, 2009. DOI: <https://doi.org/10.1038/461145a>. Disponível em: <https://www.nature.com/articles/461145a.pdf>. Acesso em: 7 fev. 2022.

DELMAS, B. **Arquivos para quê?** São Paulo: Instituto Fernando Henrique Cardoso, 2010.

DINIZ, B. C.; DIAS, G. A.; SILVA, A. K. A. Os princípios FAIR como diretrizes para maximizar o uso e (re)uso dos dados de pesquisa: traçando um paralelo com a gestão de documentos arquivísticos digitais. In: WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA UFPB, 2., 2018, João Pessoa. **Anais [...]**. Recife: UFPB, 2018. Disponível em: https://dadosabertos.info/enhanced_publications/idt/papers/39.pdf. Acesso em: 15 jun. 2022.

DOOLEY, J. **The Archival Advantage**: integrating archival expertise into management of born-digital library materials. Dublin: OCLC Research, 2015. Disponível em: <https://www.oclc.org/content/dam/research/publications/2015/oclcresearch-archival-advantage-2015-a4.pdf>. Acesso em: 5 fev. 2022.

DRUCKER, D. P. **A integração da informação sobre biodiversidade e ecossistemas para embasar políticas de conservação**: o projeto Biota Gradiente Funcional como estudo de caso. 2012. 165 f. Tese (Doutorado) – Instituto de Filosofia e Ciências Humanas, Universidade Estadual de Campinas, Campinas, 2012.

DUCHEIN, M. La profesión de archivero entre el pasado y el future. **Lligall: Revista Catalana d'Arxivística**, [Barcelona], n. 4, p. 13-24, 1991.

DUCHEIN, M. O respeito aos fundos em arquivística: princípios teóricos e problemas práticos. **Arquivo & Administração**, [Brasília, DF], v. 10/14, n. 2, 1986. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/49818>. Acesso em: 6 fev. 2022.

DURAND, T. Forms of Incompetence. *In*: INTERNATIONAL CONFERENCE ON COMPETENCE-BASED MANAGEMENT, 4th, 1998, Oslo. **Proceedings** [...]. Oslo: Norwegian School of Management, 1998.

DURAND, T. L'alchimie de la Compétence. **Revue Française de Gestion**, [Paris], v. 32, n. 160, 2000.

DURANTI, L. Registros documentais contemporâneos como provas de ação. **Estudos Históricos**, Rio de Janeiro, v. 7, n. 13, p. 49-64, 1994.

SANTOS, Paulo Roberto Elian dos. Arquivologia, laboratórios e ciência aberta: contribuições e desafios para a gestão de documentos e dados. **Acervo**, Rio de Janeiro, v. 34, n. 3, p. 1-22, set./dez. 2021. Disponível em: https://revista.an.gov.br/index.php/revistaace_rvo/article/view/1772/1681. Acesso em 5 mai. 2022.

EUROPEAN COMMISSION. **Data, Information and Knowledge Management at the European Commission**. Brussels, 2016. Disponível em: https://ec.europa.eu/info/about-european-commission/service-standards-and-principles/transparency/freedom-information/access-documents/information-and-document-management_en. Acesso em: 5 fev. 2022.

FLEURY, M. T. L.; FLEURY, A. Construindo o conceito de competência. **Revista de Administração Contemporânea**, Rio de Janeiro, v. 5, n. esp., 2001. DOI: <https://doi.org/10.1590/S1415-6552001000500010>. Disponível em: <https://www.scielo.br/j/rac/a/C5TyphygpYbyWmdqKJCTMkN/?lang=pt>. Acesso em: 7 fev. 2022.

FLORES, D.; ROCCO, B. C. B.; SANTOS, H. M. Cadeia de custódia para documentos arquivísticos digitais. **Acervo: Revista do Arquivo Nacional**, Rio de Janeiro, v. 29, n. 2, p. 117-132, 2016. Disponível em: <https://brapci.inf.br/index.php/res/v/40511>. Acesso em: 6 fev. 2022.

FORCE 11. **About FSCI@UCSD**. [San Diego, 2021]. Disponível em: <https://force11.org/fsci/post/about/>. Acesso em: 4 fev. 2022.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO (FAPESP). **Gestão de dados de pesquisa**. São Paulo, [2020]. Disponível em: <https://fapesp.br/gestaodados>. Acesso em: 18 maio 2020.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GONSALVES, E. P. **Iniciação à pesquisa científica**. 3. ed. Campinas: Alínea, 2003.

GRAY, J. Jim Gray on eScience: A transformed scientific method. *In*: HEY, T.; TANSLEY, S.; TOLLE, K. (ed.). **The fourth paradigm: Data-intensive scientific discovery**. Redmond: Microsoft Research, 2009. p. xvii-xxxi.

GREEN, A. Conceptualizing the digital life cycle. *In: IASSIST*. [S. l., 2006]. Disponível em: <https://iassistdata.org/blog/2006/06/01/conceptualizing-digital-life-cycle/#>. Acesso em: 4 fev. 2022.

HENNING, P. *et al.* Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, João Pessoa, v. 14, n. 3, p. 175-191, 2019 Disponível em: <https://periodicos.ufpb.br/index.php/pcbci/article/view/46969>. Acesso em: 6 fev. 2022.

HIGGINS, S. Draft DCC curation lifecycle model. **International Journal of Digital Curation**, Edinburgh, v. 2, n. 2, p. 82-87, 2008. DOI: <https://doi.org/10.2218/ijdc.v2i2.30>. Disponível em: <http://www.ijdc.net/article/view/46>. Acesso em: 7 fev. 2022.

HUMPHREY, C. *et al.* Archivist on board: contributions to the research team. **Forum: Qualitative Social Research**, Berlin, v. 1, n. 3, p. 1-8, 2000. DOI: <https://doi.org/10.17169/fqs-1.3.1022>. Disponível em: <https://www.qualitative-research.net/index.php/fqs/article/view/1022>. Acesso em: 7 fev. 2022.

HUVILA, I. If We Just Knew Who Should Do It or the The Social Organization of the Archiving of Archaeology in Sweden. **Information Research**, Borås, v. 21, n. 2, 2016. Disponível em: <https://files.eric.ed.gov/fulltext/EJ1104372.pdf>. Acesso em: 5 fev. 2022.

INDOLFO, A. C. Gestão de documentos: uma renovação epistemológica no universo da arquivologia. **Arquivística.net**, [s. l.], v. 3, n. 2, p. 28-60, 2007. Disponível em: <https://www.brapci.inf.br/index.php/res/v/50444>. Acesso em: 6 fev. 2022.

INDOLFO, A. C. **Dimensões político-arquivísticas da avaliação de documentos na administração pública federal (2004-2012)**. 2013. 312 f. Tese (Doutorado em Ciência da Informação) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2013.

INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICE AND TECHNOLOGY (IASSIST). **About IASSIST**. [S. l., 2021a]. Disponível em: <https://iassistdata.org/about/index.html>. Acesso em: 27 mar. 2021.

INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICE AND TECHNOLOGY (IASSIST). **Jobs repository**. [S. l., 2021b]. Disponível em: <https://iassistdata.org/jobs-repository/>. Acesso em: 27 mar. 2021.

INTERPARES 2 PROJECT. **Diretrizes do produtor**: a elaboração e a manutenção de materiais digitais: diretrizes para indivíduos. Vancouver: InterPARES, 2007. Disponível em: http://www.interpares.org/ip2/display_file.cfm?doc=ip2_creator_guidelines_booklet--portuguese.pdf. Acesso em: 5 fev. 2022.

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH (ICPSR). Ann Arbor, [2022]. Disponível em: <https://www.icpsr.umich.edu/web/pages/>. Acesso em: 5 fev. 2022.

JARDIM, J. M. O conceito e a prática da gestão de documentos. **Acervo: Revista do Arquivo Nacional**, Rio de Janeiro, v. 2, n. 2, p. 35-42, 1987.

LANDIM, L. A.; JORENTE, M. J. V. Curadoria digital no contexto arquivístico. *In*: JORENTE, M. J. V.; VIDOTTI, S. A. B. G. (org.). **Anais do 3º Encontro Internacional Dados, Tecnologia e Informação**. Marília: UNESP, 2016. v. 2.

LINKEDIN. **Sobre o LinkedIn**. [S. l., 2021]. Disponível em: <https://about.linkedin.com/pt-br>. Acesso em: 27 mar. 2021.

LYON, L. *et al.* Applying translational principles to data science curriculum development. *In*: IPRES, 2015, Chapel Hill. **Proceedings** [...]. Chapel Hill, 2015. Disponível em: http://d-scholarship-dev.library.pitt.edu/27159/1/Applying_Translational_Principles_to_Dat.pdf. Acesso em: 4 fev. 2022.

MADEIRO, A. F.; DIAS, G. A. Arquivista de dados: análise do perfil profissional em sites de bancos internacionais de trabalho. **ÁGORA: Arquivologia em Debate**, Florianópolis, v. 30, n. 61, p. 649-664, 2020. Disponível em: <https://agora.emnuvens.com.br/ra/article/view/924>. Acesso em: 4 fev. 2022.

MÁRDERO ARELLANO, M. A. **Critérios para a preservação digital da informação científica**. 2008. 354 f. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Brasília, DF, 2008. Disponível em: https://repositorio.unb.br/bitstream/10482/1518/1/2008_MiguelAngelMarderoArellano.pdf. Acesso em: 6 fev. 2022.

MENEZES, A. F. H. N. *et al.* **Metodologia científica**: teoria e aplicação na educação a distância. Petrolina: Universidade Federal do Vale do São Francisco, 2019. Disponível em: <https://portais.univasf.edu.br/dacc/noticias/livro-univasf/metodologia-cientifica-teoria-e-aplicacao-na-educacao-a-distancia.pdf>. Acesso em: 6 fev. 2022.

MINAYO, M. C. S. (org.). **Pesquisa social**: teoria, método e criatividade. 18. ed. Petrópolis: Vozes, 2001.

MINAYO, M. C. S. O desafio da pesquisa social. *In*: MINAYO, M. C. S. (org.). **Pesquisa social**: teoria, método e criatividade. 14. ed. Petrópolis: Vozes, 1999. p. 51-66.

MÖLLER, K. Lifecycle models of data-centric systems and domains: the abstract datda lifecyle model. **Semantic Web**, Amsterdam, v. 4, n. 1, p. 67-88, 2013. DOI: <http://dx.doi.org/10.3233/SW2012-0060>.

MORO-CABERO, M. Datos y archivos: retos para una profesión agitada por la transformación digital. **Acervo: Revista do Arquivo Nacional**, Rio de Janeiro, v. 34, n. 3, p. 1-23, 2021. Disponível em: <https://revista.an.gov.br/index.php/revistaacervo/article/view/1735>. Acesso em: 7 fev. 2022.

MUELLER, S. P. M. Perfil do bibliotecário, serviços e responsabilidades na área de informação e formação profissional. **Revista de Biblioteconomia de Brasília**, Brasília, DF, v. 17, n. 1, p. 63-70, 1989. Disponível em: <https://www.brapci.inf.br/index.php/article/download/17684>. Acesso em: 7 fev. 2022.

NATIONAL ARCHIVES AND RECORDS ADMINISTRATION (NARA). **Appraisal policy of the National Archives**. College Park, Oct. 2007. Disponível em: <https://www.archives.gov/records-mgmt/scheduling/appraisal>. Acesso em: 5 fev. 2022.

NATIONAL ARCHIVES OF AUSTRALIAN (NAA). **Retaining, managing and disposing of data and datasets**. Parkes, [2021]. Disponível em: <https://www.naa.gov.au/information-management/disposing-information/retaining-managing-and-disposing-data-and-datasets>. Acesso em: 5 fev. 2022.

NATIONAL RESEARCH COUNCIL UK. **A question of balance: private rights and the public interest inscientific and technical databases**. Washington, DF: National Academy Press, 1999.

NATIONAL SCIENCE BOARD. **Long-lived digital data collections: enabling research and education in the 21st century**. Alexandria: National Science Foundation, 2005. Disponível em: <https://www.nsf.gov/geo/geo-data-policies/nsb-0540-1.pdf>. Acesso em: 4 fev. 2022.

NATIONAL SCIENCE FOUNDATION (NSF). Alexandria, [2022]. Disponível em: <https://www.nsf.gov/>. Acesso em: 7 fev. 2022.

NOONAN, D.; CHUTE, T. Data curation and the university archives. **The American Archivist**, Chicago, v. 77, n. 1, p. 201-240, 2014. DOI: <https://doi.org/10.17723/aarc.77.1.m49r46526847g587>. Disponível em: <https://meridian.allenpress.com/american-archivist/article/77/1/201/24298/Data-Curation-and-the-University-Archives>. Acesso em: 7 fev. 2022.

ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE. Chapel Hill, [2022]. Disponível em: <https://odum.unc.edu/>. Acesso em: 5 fev. 2022.

OLIVEIRA, A. C. S. *et al.* Quarto paradigma científico: adoção do data life cycle model para o gerenciamento de dados científicos abertos do experimento e-quilt prototype. *In*: CONFERÊNCIA LUSO-BRASILEIRA CONFOA, 6., 2015, Salvador. **Anais [...]** Salvador: UFB, 2015.

OLIVEIRA, A. C. S.; SILVA, E. M. Ciência aberta: dimensões para um novo fazer científico. **Informação & Informação**, Londrina, v. 21, n. 2, p. 5-39, 2016. DOI: <http://dx.doi.org/10.5433/1981-8920.2016v21n2p5>. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27666>. Acesso em: 7 fev. 2022.

ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OCDE). **OCDE principles and guidelines for access to research data from public funding**. Paris: OCDE, 2007. Disponível em: <https://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 4 fev. 2022.

PAES, M. L. **Arquivo: teoria e prática**. 3. ed. Rio de Janeiro: FGV, 2004.

PAMPEL, H. *et al.* Making research data repositories visible: the re3data.org Registry. **PLoS One**, San Francisco, v. 8, n. 11, p. e78080, 2013. DOI: <https://doi.org/10.1371/journal.pone.0078080>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0078080>. Acesso em: 7 fev. 2022.

PIWOWAR, H. A.; DAY, R. S.; FRIDSMA, D. B. Sharing detailed research data is associated with increased citation rate. **PLoS One**, San Francisco, v. 2, n. 3, p. e308, 2007. DOI: <http://dx.doi.org/10.1371/journal.pone.0000308>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000308>. Acesso em: 7 fev. 2022.

RAMÍREZ, M. L. Whose role is it anyway? A library practitioner's appraisal of the digital data deluge. **Bulletin of the American Society for Information Science and Technology**, v. 37, n. 5, p. 21-23, 2011. DOI: <https://doi.org/10.1002/bult.2011.1720370508>. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/10.1002/bult.2011.1720370508>. Acesso em: 7 fev. 2022.

RE3DATA.ORG: Registry of Research Data Repositories. West Lafayette, [2021]. Disponível em: <http://www.re3data.org/>. Acesso em: 4 out 2021.

RESEARCH COUNCILS UK. **Developing the UK's e-infrastructure for science and innovation**: report of the OSI e-Infrastructure Working Group. Swindon: RCUK, 2010. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.593.6788&rep=rep1&type=pdf>. Acesso em: 7 fev. 2022.

RESEARCH LIBRARIES GROUP (RLG). **Trusted digital repositories**: attributes and responsibilities: an RLG-OCLC report. Mountain View: RLG, 2002. Disponível em: <https://prd-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/atoms/files/repositories-1.pdf>. Acesso em: 7 fev. 2022.

RILEY, J. **Understanding metadata**: what is metadata, and what is it for?. Baltimore: NISO, 2017. Disponível em https://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf. Acesso em: 5 fev. 2022.

ROCHA, C. L. Repositórios para a preservação de documentos arquivísticos digitais. **Acervo: Revista do Arquivo Nacional**, Rio de Janeiro, v. 28, n. 2, p. 180-191, 2015. Disponível em: <https://www.brapci.inf.br/index.php/res/v/40764>. Acesso em: 7 fev. 2022.

RODRIGUES, E.; SARAIVA, R. **Os repositórios de dados científicos: estado da arte**. [Porto]: RCAAP, 2010. (Projecto RCAAP: Relatório D24). Disponível em: http://projeto.rcaap.pt/index.php?option=com_repository&Itemid=2&func=startdown&id=271&lang=pt. Acesso em: 7 fev. 2022.

RODRIGUES, G. M. A formação do arquivista contemporâneo numa perspectiva histórica: impasses e desafios atuais. **Arquivo & Administração**, [Brasília, DF], v. 5, n. 2, p. 17-24, 2006.

RONDINELLI, R. C. **Gerenciamento arquivístico de documentos eletrônicos: uma abordagem teórica da diplomática arquivística contemporânea**. Rio de Janeiro: FGV, 2002.

RONDINELLI, R. C. **Gerenciamento arquivístico de documentos eletrônicos**. São Paulo: FGV, 2005.

RONDINELLI, R. C. **O documento arquivístico ante a realidade digital: uma revisão conceitual necessária**. Rio de Janeiro: Editora FGV, 2017.

ROUSSEAU, J. Y.; COUTURE, C. **Os fundamentos da disciplina arquivística**. Lisboa: Publicações Dom Quixote, 1998.

RUTZ, S. B.; COLLINS, D. W. An ocean data archivist's perspective during an ocean exploration cruise. *In*: OCEANS 2005 MTS/IEEE, 2005, Washington, DC. **Proceedings** [...]. Manhattan: IEEE, 2005. p. 2815-2819 DOI: <https://doi.org/10.1109/OCEANS.2005.1640201>.

SALES, L. F. **Integração semântica de publicações científicas e dados de pesquisa: proposta de modelo de publicação ampliada para a área de ciências nucleares**. 2014. 268 f. Tese (Doutorado em Ciência da Informação) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014. Disponível em: <https://ridi.ibict.br/handle/123456789/874>. Acesso em: 4 out 2021.

SANT'ANA, R. C. G. Ciclo de vida dos dados e o papel da ciência da informação. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais** [...]. Florianópolis: ENANCIB, 2013.

SANTOS, H. M.; FLORES, D. Políticas de preservação digital para documentos arquivísticos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 20, n. 4, p. 197-217, 2016. DOI: <https://doi.org/10.1590/1981-5344/2542>. Disponível em: <https://www.scielo.br/j/pci/a/s5JqMw9Mkk3VVgKpCK7GRZC/?lang=pt>. Acesso em: 7 fev. 2022.

SANTOS, P. R. E. **A arquivística no laboratório: história, teoria e métodos de uma disciplina**. 2008. 259 f. Tese (Doutorado em História) – Programa de Pós-graduação em História Social, Departamento de História, Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2008. Disponível em: https://www.teses.usp.br/teses/disponiveis/8/8138/tde-12022009-124733/publico/TESE_PAULO_ROBERTO_ELIAN_SANTOS.pdf. Acesso em: 6 fev. 2022.

SANTOS, P. R. E. Arquivologia, laboratórios e ciência aberta: contribuições e desafios para a gestão de documentos e dados. **Acervo**, Rio de Janeiro, v. 34, n. 3, p. 1-22, 2021. Disponível em: <https://revista.an.gov.br/index.php/revistaacervo/article/view/1772/1681>. Acesso em: 15 jun. 2022.

SANTOS, V. B. **A arquivística como disciplina científica: princípios, objetivos e objetos**. Salvador: 9Bravos, 2015.

SAYÃO, L. F.; SALES, L. F. Afinal, o que é dado de pesquisa?. **BIBLOS: Revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 34, n. 2, p. 32-51, 2020. Disponível em: <https://periodicos.furg.br/biblos/article/view/11875>. Acesso em: 7 fev. 2022.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 21, n. 2, p. 90-115, 2016. DOI: <http://dx.doi.org/10.5433/1981-8920.2016v21n2p90>. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939>. Acesso em: 7 fev. 2022.

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**, João Pessoa, v. 22, n. 3, p. 179-191, 2012. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/12224>. Acesso em: 6 fev. 2022.

SCHELLENBERG, T. R. **Arquivos modernos: princípios e técnicas**. 5. ed. Rio de Janeiro: FGV, 2006.

SCHELLENBERG, T. R. Problemas arquivísticos do governo brasileiro. **Acervo: Revista do Arquivo Nacional**, Rio de Janeiro, v. 28, n. 2, p. 287-300, 2015. Disponível em: <https://revista.an.gov.br/index.php/revistaacervo/article/view/630/677>. Acesso em: 4 fev. 2022.

SILVA, A. M. *et al.* **Arquivística: teoria de uma ciência da informação**. 2. ed. Porto: Edições Aforamento, 1999.

SILVA, W. A.; MARINHO, S.; SANTOS, P. K. Uma abordagem sistêmica aplicada à arquivística. **Arquivística.net**, Rio de Janeiro, v. 3, n. 1, p. 54-71, 2007. Disponível em: <http://www.brapci.inf.br/index.php/article/download/10724>. Acesso em: 4 fev. 2022.

SOCIAL AND ECONOMIC SURVEY RESEARCH INSTITUTE (SESRI). Doha, [2022]. Disponível em: <http://sesri.qu.edu.qa/research/sesri>. Acesso em: 5 fev. 2022.

SOUZA, K. I. M. **Arquivista, visibilidade profissional: formação, associativismo e mercado de trabalho**. Brasília, DF: Starprint, 2011.

UNIVERSIDADE DE SÃO PAULO (USP). Pró-reitora de Pesquisa. **Gestão de dados científicos**. São Paulo, [2019]. Disponível em: <http://prp.usp.br/gestao-de-dados-cientificos/?codmnu=9979>. Acesso em: 4 fev. 2022.

URBANAVICIUS JÚNIOR, V. *et al.* Identificação do conhecimento, habilidade e atitude (CHA) dos coordenadores de curso de uma Instituição de Ensino Superior. In: SIMPÓSIO DE EXCELÊNCIA EM GESTÃO E TECNOLOGIA – SEGET, 4., 2006, Rio de Janeiro. **Anais [...]**. Rio de Janeiro: UNIFEI, 2006.

UZWYSHYN, R. Online Research Data Repositories: The What, When, Why and How. **Computers in Libraries**, Westport, v. 36, n. 3, p. 18-21, 2016. Disponível em: <https://digital.library.txstate.edu/handle/10877/7597>. Acesso em: 7 fev. 2022.

VARDIGAN, M.; HEUS, P.; THOMAS, W. Data Documentation Initiative: Toward a Standard for the Social Sciences. **The International Journal of Digital Curation**, Edinburgh, v. 3, n. 1, p. 107-113, 2008. Disponível em: <http://www.ijdc.net/article/view/66/45>. Acesso em: 4 fev. 2022.

VERGARA, S. C. **Projetos e relatórios de pesquisa em administração**. 7. ed. São Paulo: Atlas, 2006.

WELLCOME TRUST. **Data, software and materials management and sharing policy**. London, 10 July 2017. Disponível em: <https://wellcome.org/grant-funding/guidance/data-software-materials-management-and-sharing-policy>. Acesso em: 7 fev. 2022.

WILKINSON, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, London, v. 3, id. 160018, 2016. DOI: <https://doi.org/10.1038/sdata.2016.18>. Disponível em: <https://www.nature.com/articles/sdata201618.pdf>. Acesso em: 7 fev. 2022.

YAMAOKA, E. J. Ontologia para mapeamento da dependência tecnológica de objetos digitais no contexto da curadoria e preservação digital. **AtoZ: Novas Práticas em Informação e Conhecimento**, Curitiba, v. 1, n. 2, p. 65-78, 2012. DOI: <http://dx.doi.org/10.5380/atoz.v1i2.41313>. Disponível em: <https://revistas.ufpr.br/atoz/article/view/41313>. Acesso em: 7 fev. 2022.

ZENK-MÖLTGEN, W. Metadaten und die Data Documentation Initiative (DDI). *In*: ALTENHÖNER, R.; OELLERS, C. (org.). **Langzeitarchivierung von Forschungsdaten: standards und disziplinspezifische Lösungen**. Berlin: Scivero, 2012. p. 111-126.