

Dados de pesquisa na Arquivologia: uma reflexão

Tânia Barbosa Salles Gava^I

^I Universidade Federal do Espírito Santo, Vitória, ES, Brasil;
tania.gava@ufes.br; <https://orcid.org/0000-0002-3920-5623>

Daniel Flores^{II}

^{II} Universidade Federal do Alagoas, Maceió, AL, Brasil;
dfloresbr@gmail.com; <https://orcid.org/0000-0001-8888-2834>

Diana Vilas Boas Souto Aleixo^{III}

^{III} Universidade Federal do Espírito Santo, Vitória, ES, Brasil;
diana.aleixo@ufes.br; <https://orcid.org/0000-0003-4156-354X>

Henrique Monteiro Cristovão^{IV}

^{IV} Universidade Federal do Espírito Santo, Vitória, ES, Brasil;
henrique.cristovao@ufes.br; <https://orcid.org/0000-0003-2011-7022>

Luciana Itida Ferrari^V

^V Universidade Federal do Espírito Santo, Vitória, ES, Brasil;
luciana.ferrari@ufes.br; <https://orcid.org/0000-0002-8223-039X>

Margarete Farias de Moraes^{VI}

^{VI} Universidade Federal do Espírito Santo, Vitória, ES, Brasil;
margarete.moraes@ufes.br; <https://orcid.org/0000-0002-4110-4610>

Resumo: Este artigo tem como objetivo fazer uma reflexão sobre os dados de pesquisa no contexto da Arquivologia e a importância da produção, gestão, transmissão e preservação dos dados em ambientes adequados, mantendo uma cadeia de custódia digital segura e ininterrupta para que os dados se mantenham autênticos e confiáveis ao longo do tempo, com fins de reuso. Trata-se de uma pesquisa de natureza qualitativa, do tipo exploratória, que fez uso da pesquisa bibliográfica e documental. O artigo apresenta semelhanças e diferenças nas definições dos conceitos relacionados ao contexto dos Repositórios Arquivísticos Digitais Confiáveis, para acervos documentais, e dos Repositórios de Dados Científicos, para conjuntos de dados, e sobre a importância da interoperabilidade entre eles, uma vez que cada tipo de repositório possui suas especificidades. O artigo reflete a importância da gestão, preservação, acesso e reuso dos documentos arquivísticos digitais em formato de dados, tanto para reuso quanto para apoio aos gestores em suas tomadas de decisão, destacando a importância dos Planos de Gestão de Dados. Concluímos que embora semelhantes, muitos conceitos diferem quando aplicados no contexto dos Repositórios Arquivísticos Digitais Confiáveis e dos Repositórios de Dados Científicos; que a Arquivologia

precisa se atentar para os documentos arquivísticos digitais em formato de dados; e que a interoperabilidade entre os Repositórios Arquivísticos Digitais Confiáveis e Repositórios de Dados Científicos trará muitos benefícios aos consumidores dessas informações.

Palavras-chave: dados de pesquisa; documento arquivístico digital; repositórios arquivísticos digitais confiáveis; repositórios de dados científicos; preservação digital

1 Introdução

Com a transformação digital sofrida principalmente a partir da década de 1990 os documentos produzidos em ambiente analógico começaram a ser migrados ou produzidos cada vez mais em ambientes digitais. Desta forma, sob o risco de perda desse patrimônio documental, surgem iniciativas voltadas à preservação digital de documentos. Na perspectiva desta pesquisa, considera-se patrimônio documental aquele “[...] que se encontra em bibliotecas e arquivos e que constitui uma parte essencial dessa memória [do mundo] e que reflete a diversidade de povos, idiomas e culturas” (Abdelaziz, 1998, p. 5, tradução nossa).

Uma resposta da ciência, em especial da Ciência da Informação (CI), foi o desenvolvimento de normas, modelos e padrões para a implementação de Repositórios Digitais Confiáveis (RDC) a fim de preservar objetos digitais e seus componentes. Dentre os modelos conceituais destaca-se o Modelo OAIS – ISO 14721:2012 (ISO, 2012a), que visa identificar os componentes de um sistema de informação dedicado à preservação digital, descreve suas interfaces internas e externas, e os objetos de informação manipulados em seu interior.

A transformação digital propiciou um volume cada vez maior de dados produzidos em ambientes digitais, tanto pelos cidadãos quanto pelas instituições, nos mais diversos segmentos da sociedade; e demanda de uma organização das informações com uma granularidade cada vez menor, de forma orientada aos dados (*data driven*), estimulando possibilidades na oferta de serviços diferenciados e inovadores aos usuários. Aliado a essa transformação, surge o fenômeno *dataism*, ou dataísmo, que conduz a observação dos contextos informacionais com foco nos dados, a fim de obter mais precisão e revelar padrões de comportamento que ainda não foram notados (Brooks, 2013).

Segundo Semeler e Pinto (2019, p. 115) “A geração de dados está atrelada a todas as coisas usadas no dia a dia. Os dados são colecionados sobre qualquer coisa, a qualquer momento e em qualquer lugar.” Como consequência, há uma preocupação em analisar os diferentes tipos de dados gerados pelo uso, cada vez mais constante das Tecnologias de Informação e Comunicação (TIC), e verificar quais precisam ser preservados, e como preservá-los, sob o risco de perda de informações importantes para as gerações futuras.

Nesse contexto, surge a curadoria digital que pode ser “[...] compreendida como um conjunto de atividades de gestão e preservação de dados, cujo objetivo é disponibilizá-los para acesso de forma rápida e a qualquer momento.” (Silva *et al.*, 2021, p. 568). Os estudos relacionados à curadoria digital tiveram início com a preocupação voltada à gestão de dados de pesquisa no âmbito da *eScience* ou *e-science*. Segundo Costa e Cunha (2014) existem outras denominações para o termo, dentre as quais destacam-se: ciência orientada a dados, computação fortemente orientada a dados, ciberinfraestrutura ou quarto paradigma etc.; e refere-se ao tratamento de um grande volume de dados no âmbito científico, gerando a necessidade de repositórios digitais (plataformas de preservação) para a gestão desse tipo específico de dado, ou seja, dados gerados no decorrer do desenvolvimento de atividades de pesquisa e com finalidade de reuso:

Dados e informações digitais gerados pelas atividades de pesquisa necessitam de cuidados específicos, tornando-se necessário a criação de novos modelos de custódia e de gestão de conteúdos científicos digitais que incluam ações de arquivamento seguro, preservação, formas de acrescentar valor a esses conteúdos e de otimização da sua capacidade de reuso. No intuito de por em prática soluções para o problema, observa-se, no âmbito de várias disciplinas, um esforço em torno do desenvolvimento de repositórios digitais orientados especialmente para uma gestão ativa de dados de pesquisa (Sayão; Sales, 2012, p. 180).

Uma vez que as instituições estão produzindo cada vez mais dados, coletados ou gerados, no decorrer de atividades institucionais de pesquisa, é importante que a Arquivologia comece a enxergá-los como um tipo ou componente de documento arquivístico, sob o risco de perda de conteúdos relevantes ao patrimônio científico e social. Ademais, a ausência de ações no campo arquivístico de políticas de gestão arquivística, preservação e acesso destes

dados faz com que outras áreas do conhecimento assumam o papel de responsável, o que pode ocasionar em perdas na constituição do patrimônio documental arquivístico de uma nação.

Sendo assim, ao analisar os conceitos de Dados de Pesquisa e de Documento Arquivístico Digital (DAD), este artigo tem como objetivo principal refletir sobre os dados de pesquisa no contexto da Arquivologia e a importância da produção, gestão, transmissão e preservação dos dados em ambientes digitais, mantendo uma cadeia de custódia digital segura e ininterrupta, para que eles permaneçam autênticos e confiáveis ao longo do tempo, e para fins de reuso, trazendo à reflexão o conceito de três tipos de documentos: o DAD; o DAD com dados, e o DAD em dados.

Embora a Arquivologia se preocupe com qualquer tipo de dado gerado no decorrer das atividades de uma instituição, o artigo se limita ao estudo dos dados gerados em ambiente digital, e no decorrer do desenvolvimento de atividades de pesquisa científica, tendo como princípio o reuso em dados abertos.

O artigo é resultado de uma investigação no escopo de um projeto de pesquisa, sendo de natureza qualitativa, do tipo exploratória, e visa aprofundar o conhecimento sobre a temática pesquisada, e faz uma reflexão sobre os dados de pesquisa no contexto da Arquivologia. Em relação à natureza das fontes utilizadas, trata-se de uma pesquisa bibliográfica e documental, cujas principais fontes foram textos e artigos científicos provenientes de organizações, pesquisadores nacionais e internacionais, e especialistas na área.

As fontes bibliográficas utilizadas foram obtidas por meio de consultas diretas e indiretas. De forma direta, optou-se pela Base de Dados em Ciência da Informação (BRAPCI) (BRAPCI, 2010-2024) por ser uma plataforma digital voltada à coleta, preservação e ao acesso à literatura científica na área da CI. As consultas indiretas ocorreram junto às principais referências dos materiais obtidos na consulta direta. Como delimitação da busca na Brapci usou-se a palavra-chave “dados de pesquisa”, com uma restrição temporal dos últimos cinco anos. A busca foi realizada em dez de novembro de 2021 e retornou 290 trabalhos. Após rigorosa leitura junto às duas primeiras páginas com os resultados recuperados, foram

selecionados cinco artigos (Quadro 1) como fontes principais.

Quadro 1 - Artigos selecionados na Brapci

Título	Autor(es)/ano
Repositório de dados de pesquisa para grupo de pesquisa: um projeto piloto	Vidotti <i>et al.</i> (2017)
A gestão de dados de pesquisa no contexto da <i>e-science</i> : benefícios, desafios e oportunidades para organizações de P&D	Bertin, Vosili e Drucker (2017)
Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados	Semeler e Pinto (2019)
Afinal, o que é dado de pesquisa?	Sayão e Sales (2020)
Características dos Repositórios de Dados científicos no Brasil	Paganine e Amaro (2020)

Fonte: Dados da pesquisa.

As demais fontes utilizadas na pesquisa foram obtidas a partir das referências dos artigos indicados no Quadro 1.

Foram utilizadas como fontes documentais publicações do Conselho Nacional de Arquivos (CONARQ) (2015a, 2015b, 2020, 2022, 2023a, 2023b), Arquivo Nacional (AN) (2016), *The International Research on Permanent Authentic Records in Electronic Systems* (InterPARES Project) (2010), *International Council on Archives* (ICA) (c2024), *Digital Curation Centre* (DCC) (c2023), *Organisation for Economic Co-operation and Development* (OECD) (2007) e *National Research Council* (NRC) (1999) dos Estados Unidos (EUA) .

2 O conceito de dados de pesquisa

Como acontece com muitos conceitos dentro de uma vasta área do conhecimento, existem diversas definições para o conceito de “dados de pesquisa” na CI. Neste sentido, esta seção tem como objetivo apresentar algumas das definições de dados de pesquisa, uma vez que “[...] de disciplina a disciplina, de pesquisador a pesquisador, a expressão dados de pesquisa significa coisas diferentes e depende do contexto científico em que será utilizada.” (Semeler; Pinto, 2019, p. 120):

De fato, o termo “dado de pesquisa” tem uma amplitude de significados que vão se transformando de acordo com domínios científicos específicos, objetos de pesquisas, metodologias de geração e coleta de dados e muitas outras variáveis. Pode ser o resultado de um experimento realizado num ambiente controlado de

laboratório, um estudo empírico na área de ciências sociais ou a observação de um fenômeno cultural ou da erupção de um vulcão num determinado momento e lugar. Dados digitais de pesquisa ocorrem na forma de diferentes tipos de dados, como números, figuras, vídeos, softwares; com diferentes níveis de agregação e de processamento, como dados crus ou primários, dados intermediários e dados processados e integrados; e em diferentes formatos de arquivos e mídias (Sayão; Sales, 2020, p. 33).

Inicialmente é importante definir o termo “dado”. Segundo o Dicionário de Terminologia Multilíngue do *Internacional Council on Archives* (ICA), (ICA, c2024), dado é definido como uma “Representação de todo e qualquer elemento de conteúdo cognitivo, passível de ser comunicada, processada e interpretada de forma manual ou automática.” Os dados também podem ser definidos como “[...] fatos, números, letras e símbolos que descrevem um objeto, ideia, condição, situação, ou outros fatores.” (National Research Council, 1999, p. 15, tradução nossa). Ou seja:

[...] uma unidade de conteúdo necessariamente relacionada a determinado contexto e composta pela tríade entidade, atributo e valor, de tal forma que, mesmo que não esteja explícito o detalhamento sobre o contexto do conteúdo, ele deverá estar disponível de modo implícito no utilizador, permitindo, portanto, sua plena interpretação (Santos; Sant’Ana, 2013, p. 205).

Já os dados de pesquisa, também chamados de dados científicos, possuem várias definições. Algumas delas serão apresentadas a seguir, mas é importante destacar que dados de pesquisa podem ser objetos digitais ou não, mas que nesse estudo serão considerados apenas os dados de pesquisa gerados em ambiente digital.

O termo “dados de pesquisa” foi formalmente definido, em âmbito internacional, pela Organization Economic Co-operation and Development (OECD), em 2007, por meio da declaração de acesso aos dados científicos com financiamento, e inclui tanto os dados analisados quanto os metadados que os descrevem (Semeler; Pinto, 2019). A Organização define o termo como registros de fatos usados como fontes primárias para a pesquisa científica e que são comumente aceitos na comunidade como necessários para validar resultados da pesquisa.

Semeler e Pinto (2019, p. 115) apresentam os dados de pesquisa científica como o “[...] resultado de qualquer investigação sistemática que envolva processos de observação, experimentação ou simulação de procedimentos de pesquisa científica.” Além disso, são consideradas “[...] evidências e/ou insumos de pesquisa que são coletados, observados, registrados e/ou criados para fins de análise e que podem produzir resultados de pesquisa para um estudo científico” (Semeler; Pinto, 2019, p. 119). Segundo os autores supracitados, dados de pesquisa podem ser:

Documentos, questionários, avaliações, registros de casos, protocolos de estudo, planilhas, notas de laboratório, notas de campo, diários, filmes, imagens, arquivos digitais de áudio e vídeo, sequências genéticas, coordenadas geográficas, banco de dados, algoritmos, metodologias, protocolos, entre outros tipos de manifestação de pesquisa (Semeler; Pinto, 2019, p. 116).

Para Pampel *et al.* (2013, p. 1, tradução nossa), dados de pesquisa são definidos como:

[...] dados digitais sendo uma parte (descritiva) ou o resultado de um processo de pesquisa. Esse processo abrange todas as etapas da pesquisa, desde a geração dos dados da pesquisa, que pode ser em um experimento nas ciências, um estudo empírico nas ciências sociais ou observações de fenômeno cultural, até a publicação dos resultados da pesquisa. Os dados de pesquisa digital ocorrem em diferentes tipos de dados, níveis de agregação e formatos de dados, informados pelas disciplinas de pesquisa e seus métodos.

Sayão e Sales (2020, p. 32) abordam que dados de pesquisa podem ser:

[...] dados brutos coletados diretamente por um instrumento ou um sensor e agregados a partir de múltiplas fontes; ou podem ser produtos de um modelo teórico, simulação ou visualização; ou de experimentos conduzidos na bancada de um laboratório [...].

Além disso, Cris Grant (2017, tradução nossa), destaca que:

[...] embora os dados de pesquisa sejam muitas vezes pensados em termos de pesquisas científica, eles podem também representar outros contextos, por exemplo, dados de censos, histórias orais e pesquisas longitudinais de Ciências Sociais e resultados de pesquisas em artes e humanidades.

A partir das definições apresentadas pode-se compreender que os dados de pesquisa estão presentes em diferentes contextos e áreas do conhecimento,

sendo muito importantes no cenário científico.

3 Dados de pesquisa no contexto da Arquivologia

Segundo o *Glossário de documentos arquivísticos digitais* desenvolvido pela Câmara Técnica de Documentos Eletrônicos (CTDE) um documento é uma “Unidade de registro de informações, qualquer que seja o formato ou o suporte” (CONARQ, 2020, p. 23). O glossário define também documento arquivístico como um “Documento produzido (elaborado ou recebido), no curso de uma atividade prática, como instrumento ou resultado de tal atividade, e retido para ação ou referência” (CONARQ, 2020, p. 24). Por sua vez, documento arquivístico digital é um “Documento digital reconhecido e tratado como um documento arquivístico” (CONARQ, 2020, p. 25); e um documento digital é uma “Informação registrada, codificada em dígitos binários, acessível e interpretável por meio de sistema computacional” (CONARQ, 2020, p. 25).

No entanto, para entender o conceito de documento arquivístico no contexto digital, a equipe do Projeto InterPARES (2010) buscou subsídios na análise diplomática para identificar outras características necessárias a um documento arquivístico. Sendo elas:

1) **forma fixa**, significando que o conteúdo binário do documento arquivístico deve ser armazenado de maneira que sua mensagem possa ser apresentada com a mesma forma documental que tinha quando retida pela primeira vez; 2) **conteúdo estável**; 3) **relações explícitas com outros documentos arquivísticos dentro ou fora do sistema digital**, por meio de um código de classificação ou outro identificador único; 4) **contexto administrativo identificável**; 5) **autor, destinatário, e escritor**; e 6) **ação**, na qual o documento participa ou que o documento apóia em termos de procedimentos ou como parte do processo de tomada de decisão. (Duranti, 2005, p. 7, grifo nosso).

O Projeto InterPARES 2 (2010) teve uma preocupação especial com a apresentação documental (forma documental) de materiais digitais, associando-a à manutenção e avaliação da confiabilidade e acurácia dos documentos arquivísticos digitais, uma vez que conversões ou migrações de dados, inevitáveis nos processos de preservação digital, podem resultar em mudanças na forma documental. Assim, as diretrizes do produtor, material resultante do Projeto

InterPARES 2, recomendam:

[...] estabelecer primeiramente a forma dos documentos associados com cada atividade ou procedimento, e depois identificar as características essenciais (isto é, os elementos extrínsecos e intrínsecos) de cada apresentação ou forma documental (InterPares Project 2, 2010).

O projeto considera como elementos extrínsecos aqueles:

[...] que constituem sua aparência externa, inclusive as características de apresentação, como fonte, gráficos, imagens, sons, *layouts*, *hyperlinks*, resoluções de imagens etc., assim como selos, assinaturas digitais, carimbos de tempo e sinais especiais (marcas d'água digitais, logotipos, timbres etc.) (InterPares Project 2, 2010, p. 3).

E como elementos intrínsecos entende-se os:

[...] Elementos de um documento arquivístico que expressam a ação da qual ele participa e seu contexto imediato, inclusive os nomes das pessoas envolvidas na sua produção, o nome e descrição da ação ou assunto ao qual ele pertence, a(s) data(s) de produção e transmissão etc (InterPares Project 2, 2010, p. 3).

O projeto também relata que “[...] para assegurar que a autenticidade possa ser presumida e mantida ao longo do tempo, deve-se definir e conservar a identidade dos materiais e proteger sua integridade.” (InterPares Project 2, 2010, p. 3). A identidade seria o “Conjunto de características de um documento ou de um documento arquivístico que o identifica de forma única e o distingue dos demais [...]” (InterPares Project 2, 2010, p. 8). E a integridade dos materiais digitais é a “Qualidade de ser completo e inalterado em todos os aspectos essenciais; junto com a identidade, é um componente da autenticidade.” (InterPares Project 2, 2010, p. 10).

Um documento arquivístico digital também pode ser composto por um ou mais componentes digitais. Um componente digital é um “Objeto digital que é parte de um ou mais documentos digitais, incluindo os metadados necessários para ordenar, estruturar ou manifestar seu conteúdo e forma, que requer determinadas ações de preservação [...]” (CONARQ, 2020, p. 18). O Glossário de Documentos Arquivísticos Digitais (CONARQ, 2020) apresenta o exemplo de uma fotografia digital como tendo apenas um componente digital, que é o próprio

arquivo com a imagem; outro exemplo é o caso de um documento multimídia que possui diversos componentes digitais, os quais são os arquivos com o código executável, os textos, as imagens e os registros sonoros.

Uma vez que a Arquivologia define um documento de arquivo ou documento arquivístico como um documento vinculado a uma atividade ou ação, e no caso deste estudo o foco são os documentos produzidos em ambiente digital, e que os dados de pesquisa são produzidos no decorrer de atividades de pesquisa, este artigo também objetiva refletir sobre três tipos de Documentos: (1) o DAD (Documento Arquivístico Digital); (2) o DAD com dados, que possui um conjunto de dados como um de seus componentes digitais; e (3) o DAD em dados, cujo formato é exclusivamente na forma de dados. No entanto, a presente pesquisa se concentra em um escopo bem definido, que se refere aos DAD produzidos em ambiente digital no desenvolvimento de atividades de pesquisa, visto sua importância no progresso da ciência, sob o risco de perda de informações importantes que poderiam ser reutilizadas em pesquisas futuras.

Considera-se o DAD como o documento arquivístico digital em sua forma manifestada, que é o tipo de documento comumente tratado na Arquivologia. Um documento manifestado é a “Visualização ou apresentação do documento arquivístico de uma forma compreensível para uma pessoa ou outro sistema.” (Duranti; Thibodeau, 2006¹, p. 402-459 *apud* CONARQ, 2020, p. 25). Comenta-se que a Arquivologia tem seu foco na produção, migração e arquivamento de documentos arquivísticos digitais manifestados, principalmente no formato de texto.

Um DAD com dados é um documento arquivístico digital que possui como componente digital um conjunto de dados de pesquisa, em semelhança ao exemplo citado de um documento multimídia.

Por outro lado, o DAD em dados, a exemplo da fotografia, possui apenas um componente digital, que é o próprio conjunto de dados que representa a fotografia.

Nesse contexto, há de se fazer algumas considerações, o primeiro ponto refere-se ao fato que cada vez mais tem se produzido dados em ambiente digital

no curso das atividades institucionais e que tais dados, aqui considerados como um tipo de documento arquivístico, têm especificidades de produção, manutenção, armazenamento, preservação e acesso, as quais precisam ser consideradas para o seu devido reuso no futuro. O segundo ponto remete ao fato que nem a prática arquivística, tampouco os modelos teóricos de Repositórios Digitais Confiáveis, que no caso do Brasil estão definidos pela Resolução n. 43 do CONARQ (CONARQ, 2023b), apresentam de forma explícita a preocupação com esse tipo de documento. E ressalta-se que embora as definições de dados de pesquisa deem margem para outras interpretações, aqui considera-se os dados de pesquisa como dados abertos, em formatos amplamente utilizados e homologados pela área, produzidos no curso de atividades de pesquisa e com princípio de reuso.

Para entender a importância da Arquivologia se atentar para documentos arquivísticos em formato de dados, sua gestão e preservação, cita-se o exemplo de uma fotografia. Se a fotografia for preservada no formato .jpeg, .png, TIFF etc., ela não poderá ser reusada no futuro. Para que ela possa ser reusada, deve ser preservada em um formato que permita sua edição, por exemplo, em formato .NEF ou .RAW. Deste modo, mesmo que a política de preservação institucional requeira que o documento seja preservado em sua forma manifestada, em formatos estáticos, deve-se pensar na preservação de componentes digitais, que virão encapsulados aos documentos manifestados, em formatos de dados (de elaboração) para reuso.

Outro exemplo seria a preservação de um Plano de Classificação de Documentos (PCD). Se o PCD estiver publicado em formato .PDF/A, como um documento manifestado na forma de texto, não será possível importá-lo automaticamente em um sistema de informação para reuso. No entanto, se o documento possuir um componente digital no formato de dados tabulares, como .XLS ou .CSV, ou um arquivo de exportação do AtoM² ou no padrão de metadados *Encoded Archival Description* (EAD)³, haverá a possibilidade de reuso no futuro. O PCD terá a forma documental diplomática manifestada, terá dados para reuso, dados abertos, e a possibilidade de importação por um software de descrição arquivística, como é o caso do AtoM, ou de um Sistema

Informatizado de Gestão Arquivística de Documentos (SIGAD). No entanto, se o PCD não for manifestado diplomaticamente, ou seja, se ele estiver somente em formato de dados (tais como os formatos .CSV, .XLS, EAD, .XML, dentre outros formatos e padrões), ele também poderá ser usado como dados abertos para reuso, importação, manipulação, análise e cruzamento de dados entre sistemas de informação.

Documentos no padrão PDF/A também aceitam a incorporação de outros documentos digitais, classificados como sendo de padrão A-2, A-3 e A-4, e contribuem fortemente para a prática da publicação ampliada mais estável, uma vez que os dados de pesquisa ficam assegurados dentro do próprio documento PDF original (Cristovão; Batista; Rocha, 2023). A exemplo cita-se uma planta de engenharia manifestada em PDF/A, que pode ter um componente digital encapsulado no formato .DWG para uso no AutoCad, ou em outro formato, a depender do sistema usado na instituição custodiadora.

Além disso, quando os documentos estão no formato de dados, isto é, DAD em dados, eles podem ser matéria-prima para ferramentas de *data mining* e *data warehouse*, utilizados pelos gestores como apoio à tomada de decisão em sintonia com o fenômeno do dataísmo.

4 Repositórios arquivísticos digitais confiáveis x repositório digitais confiáveis de dados

Para que um DAD possa servir de fonte de prova, evidências, testemunho, memória, patrimônio, garantia de direitos e exercício pleno da cidadania, seja nativo digital ou representante digital, ele deve ser mantido autêntico e confiável desde sua produção, durante sua transmissão no tempo e espaço, até sua preservação, mantendo uma cadeia de custódia digital arquivística segura e ininterrupta (Gava; Flores, 2021). Quando um destes elementos é negligenciado, a autenticidade estará em risco. Para isso, é importante que se desenvolvam procedimentos que estabeleçam controles sobre a produção, transmissão, manutenção e preservação dos documentos arquivísticos digitais; tais procedimentos devem ser implementados nos ambientes em que os dados são

produzidos, preservados e acessados.

A exemplo do que acontece com os SIGAD, definido pelo Modelo de requisitos para sistemas informatizados de gestão arquivística de documentos (e-ARQ Brasil) (CONARQ, 2022) para o contexto dos documentos arquivísticos produzidos nas esferas do Poder Executivo e Legislativo e do GestãoDoc (Sistema Informatizado de Gestão de Processos e Documentos), definido pelo MoReq-Jus (Brasil, 2009) para os documentos do Poder Judiciário, é importante que os dados de pesquisa também sejam geridos em ambientes adequados, com base em requisitos definidos e homologados pela comunidade de interesse, que irão manter a presunção de autenticidade e confiabilidade dos dados ao longo do tempo.

Sistemas de gestão arquivística, sejam eles para documentos ou dados, devem estar associados a Repositórios Digitais Confiáveis, cujo objetivo é a preservação dos objetos digitais. Na Ciência da Informação existem normas, modelos e padrões que devem ser seguidos para a devida preservação dos objetos digitais e seus componentes. Uma dessas normas é a ISO 16363:2012 (ISO, 2012b) que define requisitos para a avaliação da confiabilidade de Repositórios Digitais de organizações públicas ou privadas. Os requisitos estão organizados em três conjuntos: (1) Infraestrutura organizacional; (2) Gerenciamento do documento digital; e (3) Tecnologia, infraestrutura técnica e segurança. Além disso, é necessário que o conjunto de requisitos que tratam do gerenciamento do documento digital adote as diretrizes do Modelo OAIS (ISO 14721:2012) (ISO, 2012a). Destaca-se que uma organização só pode ser considerada um RDC se for certificada pelas normas ISO citadas, e que no Brasil, até o presente momento, ainda não temos nenhuma entidade de auditoria e certificação de RDC.

Para ser uma entidade certificadora, a instituição deve atender à ISO 16919:2014 (ISO, 2014), que é uma recomendação técnica criada pelo *Consultative Committee for Space Data Systems* (CCSDS) (CCSDS, 2021), que estabelece requisitos para as entidades de auditoria e certificação de um RDC. Neste sentido, o CONARQ criou, em 2021, uma câmara técnica consultiva com o objetivo de elaborar requisitos de certificação e regras de auditoria para os Repositórios Arquivísticos Digitais Confiáveis (RDC-Arq), dentre seus objetivos

estão: elaborar uma lista de critérios e requisitos a serem cumpridos por um RDC-Arq, para que ele seja considerado aderente às resoluções do CONARQ; definir a metodologia para o diagnóstico, a auditoria e a autocertificação de RDC-Arq, baseada nas normas ISO 16363:2012 (ISO,2012b) e 16.919:2014 (ISO, 2014) e nas resoluções do CONARQ; definir a metodologia para aferição da maturidade em preservação digital; e definir a metodologia para monitoramento de RDC-Arq.

Em relação aos RDC, no Brasil surge em 2015 a Resolução n. 43 (CONARQ, 2023b), como uma resposta do CONARQ para a comunidade arquivística, no sentido de orientar a implementação de RDC-Arq para o arquivamento e manutenção de documentos arquivísticos digitais em suas fases corrente, intermediária e permanente, nos órgãos e entidades integrantes do Sistema Nacional de Arquivos (SINAR), a fim de manter a autenticidade e confiabilidade dos documentos ao longo do tempo. A resolução teve como base várias normas, padrões e modelos, dentre eles a ISO 16363:2012 (ISO,2012b) e a ISO 14721:2012 (Modelo OAIS) (ISO,2012a). O objetivo da resolução é apoiar a preservação dos documentos arquivísticos digitais em todo o seu ciclo de vida, e não somente na fase permanente, como preconizava a Resolução n. 39 do CONARQ (2023a):

O termo RDC-Arq surgiu na Resolução n. 43 do Conarq, de 4 de setembro de 2015, que deu nova redação à Resolução n.º 39, de 29 de abril de 2014. A resolução n. 39 os denominava de Repositórios Digitais Confiáveis de Documentos Arquivísticos. A mudança de nomenclatura da Resolução n. 39 para a Resolução n.º 43 traz em si um significado muito importante, o de se ter um Repositório Digital Confiável (RDC) adjetivado, ou seja, um RDC Arquivístico, com características específicas. Ou seja, um Repositório Arquivístico deve incorporar em seu funcionamento princípios e normas arquivísticas, e não apenas ser um RDC para documentos arquivísticos, no sentido de gerenciar e preservar apenas um tipo específico de material digital, dentre tantos outros diferentes tipos de materiais digitais. (Gava; Flores, 2020, p. 78).

Como explicado por Gava e Flores (2020), um RDC-Arq é um tipo específico de Repositório Digital Confiável, especializado em documentos arquivísticos digitais. No entanto, a resolução não faz nenhuma menção a tipos específicos de documentos e como eles devem ser tratados. Contudo, com o avanço das pesquisas em Ciência da Informação, muitos pesquisadores já têm

atentado para a necessidade de repositórios digitais específicos para dados, em especial para os dados de pesquisa. Segundo Sayão e Sales (2016, p. 96), os repositórios de dados têm como objetivo “[...] garantir o acesso contínuo e aberto - agora e no futuro - aos resultados de pesquisa que se manifestam na forma de dados, e que são considerados parte importante do patrimônio digital da humanidade [...]”.

4.1 Comparando os conceitos no contexto de RDC-Arq e repositórios de dados científicos

Embora ambos devam ser Repositórios Digitais Confiáveis (RDC), um RDC-Arq é um RDC para documentos arquivísticos, ao passo que o RDC de dados científicos é específico para dados de pesquisa. Além disso, nota-se que um mesmo conceito pode ter definições diferentes em cada um desses contextos. O Quadro 2 apresenta alguns conceitos importantes, e sua definição, no cenário de documentos arquivísticos digitais e dados de pesquisa (dados científicos).

Quadro 2 - Comparação entre os conceitos: documento arquivísticos e dados de pesquisa

	Documento Arquivístico	Dados de Pesquisa Científica
Definição	“Documento produzido (elaborado ou recebido), no curso de uma atividade prática, como instrumento ou resultado de tal atividade, e retido para ação ou referência.” (CONARQ, 2020, p. 24).	“São o resultado de qualquer investigação sistemática que envolva processos de observação, experimentação ou simulação de procedimentos de pesquisa científica.” (Semeler; Pinto, 2019, p. 116).
Formato	Podem estar ou não em formato digital	Podem estar ou não em formato digital
Tipo	Texto estruturado com formatação, imagem matricial, imagem vetorial, áudio, audiovisual, mensagem de correio eletrônico, apresentação (slides), planilha e base de dados relacional, dentre outros tipos mais complexos de documentos em formato digital, como multimídia e páginas web (Arquivo Nacional, 2016, p. 11).	“[...] dados brutos coletados diretamente por um instrumento ou um sensor e agregados a partir de múltiplas fontes; ou podem ser produtos de um modelo teórico, simulação ou visualização; ou de experimentos conduzidos na bancada de um laboratório [...]” (Sayão; Sales, 2020, p. 32)
Ciclo de vida	“Sucessivas fases por que passam os documentos arquivísticos, da sua produção à guarda permanente ou eliminação” (CONARQ, 2020, p. 16). Essas fases são a corrente, intermediária e permanente, e referem-se à chamada Teoria das três idades, que é uma estratégia	Existem vários modelos de ciclo de vida, que descrevem as etapas necessárias para a gestão dos dados de pesquisa. Um desses modelos é do <i>Digital Curation Centre (DCC)</i> (c2023), que define as seguintes etapas: conceituação; criação; acesso e uso; avaliação e seleção; descarte;

	organizacional que divide o ciclo de vida dos documentos arquivísticos em três fases ou idades.	ingestão; ações de preservação; reavaliação; armazenamento; acesso e reuso; e transformação.
Preservação	Sua preservação deve perpassar todo o ciclo de vida dos documentos, desde sua produção até sua guarda final, por meio do uso de um RDC-Arq associado a um SIGAD, nas fases corrente e intermediária, e de um RDC-Arq na fase permanente. (CONARQ, 2015b).	A preservação dos dados de pesquisa é feito pela curadoria digital, que é o conjunto de atividades de gestão e preservação de dados, em todo o ciclo de vida dos dados, cujo objetivo é disponibilizá-los para acesso de forma rápida e a qualquer momento e envolve a implantação de Repositórios de Dados Científicos ou Dados de Pesquisa (Silva <i>et al.</i> ; 2021, p. 568; Vidotti <i>et al.</i> , 2017).
Política	Os documentos devem ser respaldados por uma política de gestão arquivística que contemple uma política de preservação digital e planos de preservação digital.	Dados de pesquisa devem estar respaldados por um Plano de Gestão de Dados, que descreve o ciclo de vida de gestão dos dados que serão coletados, processados ou gerados por um projeto de pesquisa (Sayão; Sales, 2015).

Fonte: Elaborado pelos autores.

Observando o Quadro 2, vê-se que um documento arquivístico é um documento produzido no decorrer de uma atividade prática, retido para ação ou referência, e que no caso dos dados de pesquisa, eles podem ser considerados um tipo particular de documento, pois são produzidos no decorrer de uma pesquisa científica e devem ser preservados para apoio a pesquisas futuras. Documentos arquivísticos e dados de pesquisa podem ser produzidos ou não em ambientes digitais, mas nesse estudo foram considerados apenas documentos e dados produzidos em ambiente digital.

Existem também diferentes tipos de documentos arquivísticos e dados de pesquisa, mas diferente do que acontece com os tipos de documentos, os exemplos apresentados na literatura para tipos de dados de pesquisa podem diferir muito de autor para autor.

Em relação ao ciclo de vida, o termo é usado tanto para documentos quanto para dados de pesquisa. No entanto, o ciclo de vida dos dados de pesquisa tem a ver com a curadoria digital, que é uma área recente, que diz respeito ao gerenciamento e preservação de dados ao longo do tempo (Digital Curation Centre, c2023), ao passo que o ciclo de vida dos documentos arquivísticos refere-se às sucessivas fases (corrente, intermediária e permanente) por que passam os

documentos arquivísticos, que compreende desde a produção até sua guarda permanente ou eliminação.

Em relação à preservação dos documentos arquivísticos digitais, o Brasil possui uma resolução (CONARQ, 2015a) que prevê a implementação de Repositórios Arquivísticos Digitais Confiáveis (RDC-Arq) para a preservação dos documentos. Além disso, há uma orientação técnica (CONARQ, 2015b) que sugere o uso de um RDC-Arq associado a um SIGAD nas fases corrente e intermediária, e seu uso obrigatório na fase permanente, para fins de preservação. No que se refere aos dados de pesquisa, a preservação dos dados é feita pela curadoria digital, que é o conjunto de estratégias, abordagens tecnológicas e atividades que:

[...] envolve a gestão atuante e a preservação de recursos digitais durante todo o ciclo de vida de interesse do mundo acadêmico e científico, tendo como perspectiva o desafio temporal de atender a gerações atuais e futuras de usuários” (Sayão; Sales, 2012, p. 184).

Associados às metodologias utilizadas pela curadoria digital estão os processos de arquivamento digital e de preservação digital, que envolvem a implantação de RDC para dados científicos ou dados de pesquisa.

No tocante às políticas, um RDC-Arq deve estar sempre associado a uma política e aos planos de preservação digital, tendo como base uma política de gestão arquivística. No Brasil, em 2010, foi publicada a primeira versão da Política de Preservação Digital do AN Digital, que é o Programa Permanente de Preservação e Acesso a Documentos Arquivísticos Digitais do Arquivo Nacional. Uma segunda versão foi publicada em 2016 (Arquivo Nacional, 2016).

Em relação aos dados de pesquisa, o planejamento da preservação dos dados deve ser feito por meio dos Planos de Gestão de Dados (PGD) e da administração de todas as etapas do ciclo de vida dos dados (Silva *et al.*; 2021, p. 573). Um PGD tem como objetivo fazer um planejamento de como gerar e coletar dados de alta qualidade e sustentáveis ao longo do tempo, explicitando como os dados serão tratados durante o desenvolvimento do projeto de pesquisa e depois de sua finalização (Sales; Sayão, 2019). Ou seja, um PGD deve descrever quais dados serão coletados ou gerados, quais metodologias e padrões devem ser

utilizados nestes processos, sob quais condições os dados devem ser compartilhados e outorgados acessíveis ou abertos para a comunidade de pesquisa e como eles serão curados e preservados.

No Brasil, cresce cada vez mais a exigência, pelas agências de fomento e instituições de pesquisa, de um PGD como um documento obrigatório na solicitação de apoio às pesquisas. A exemplo cita-se a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), que apresenta o “Plano de Gestão de Dados (PGD)” como um dos anexos obrigatórios nas submissões de propostas. Embora sem um modelo próprio, a FAPESP disponibiliza algumas diretrizes para o desenvolvimento de um PGD.

A adoção de PGD é algo que precisa de mais discussão no campo da Arquivologia, a qual deve se atentar cada vez mais ao tema, associando-os às políticas de gestão, preservação e acesso a documentos arquivísticos digitais, e enxergando os documentos arquivísticos digitais em formato de dados como sua responsabilidade, e não somente da Biblioteconomia, sob risco de perda de um patrimônio documental tão importante para reuso no futuro e para apoio à tomada de decisão pelos gestores.

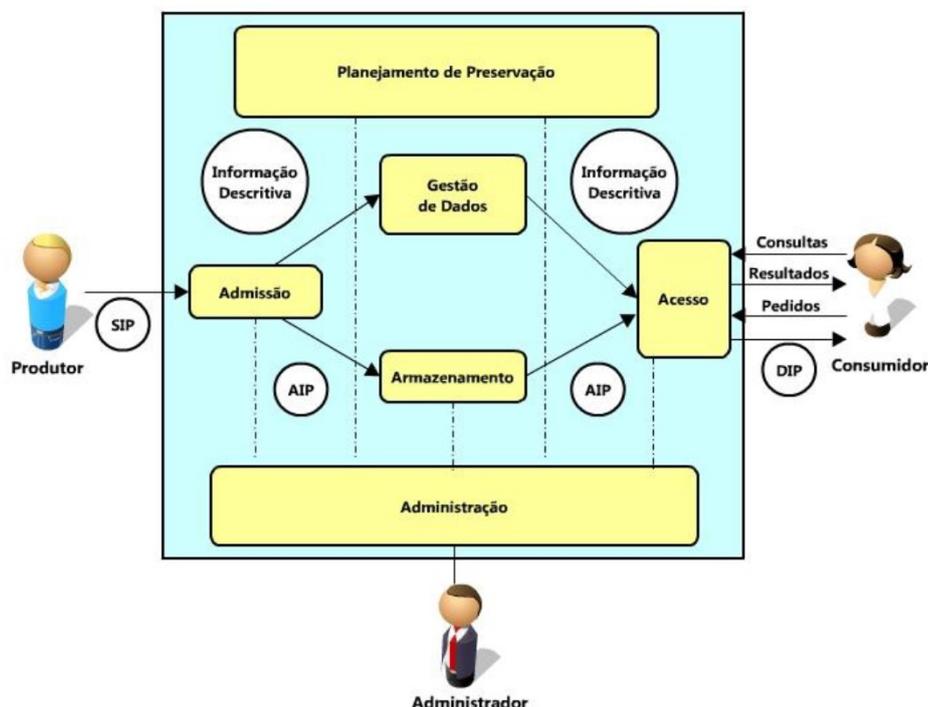
4.2 Interoperabilidade entre um RDC-Arq e um repositórios de dados científicos

Os RDC-Arq são Repositórios Arquivísticos Digitais Confiáveis para o arquivamento e manutenção de documentos arquivísticos digitais em todo o seu ciclo de vida, com o objetivo de manter a autenticidade e confiabilidade dos documentos ao longo do tempo. Em relação aos dados de pesquisa, Paganine e Amaro (2020) apresentam que, para que os dados e coleções de dados de pesquisa transmitam conhecimento no tempo e no espaço e possam ser reusados ao longo do tempo, é necessária a implantação de uma infraestrutura tecnológica e gerencial, permanente e sustentável, que permita que eles sejam cuidados ao longo de todo o seu ciclo de vida. E que no centro dessa infraestrutura estão os repositórios digitais de dados de pesquisa, que são sistemas voltados para apoiar

a seleção, catalogação, arquivamento, acesso e compartilhamento de dados de pesquisa.

Por serem Repositórios Digitais Confiáveis, tanto o RDC-Arq quanto os Repositórios de Dados Científicos devem atender ao Modelo OAIS (ISO 14721:2012) (ISO,2012a). O Modelo OAIS define três entidades externas (produtor, administrador e consumidor) e três tipos de pacotes de informação: SIP, AIP e DIP (Figura 1).

Figura 1 - O Modelo de Referência OAIS



Fonte: Conselho Nacional de Arquivos (2015a, p. 20).

O pacote *Submission Information Package* (SIP) é o pacote de submissão, e contém os objetos em seu formato original. O pacote *Archival Information Package* (AIP) é o pacote de arquivamento, e contém os objetos em seu formato de preservação. O pacote *Dissemination Information Package* (DIP) é o pacote de disseminação, e contém os objetos em formatos mais adequados para sua difusão e acesso.

Embora possuam o mesmo objetivo, que é a preservação adequada dos

objetos digitais e seus metadados, os RDC-Arq e Repositórios de Dados Científicos possuem suas especificidades. Os RDC-Arq foram projetados para lidar com qualquer tipo de documento arquivístico digital, e por serem tão gerais, a Resolução n. 43 do CONARQ (2023b) não apresenta detalhes sobre como tratar os diferentes tipos de documentos em seus diferentes formatos. Ao passo que os Repositórios de Dados Científicos foram projetados para lidar especificamente com dados de pesquisa. Ao discutir o escopo dos documentos arquivísticos digitais produzidos no decorrer do desenvolvimento de pesquisas científicas, eles podem ser produzidos em formato de dados como, por exemplo, no formato de dados tabulares ou geoespaciais.

É importante ressaltar que embora as definições de dados de pesquisa deem margem para outras interpretações, consideram-se os dados de pesquisa como dados abertos, em formatos amplamente utilizados e homologados pela área, produzidos no curso de atividades de pesquisa e com princípio de reuso. Nesse contexto, é importante pensar na interoperabilidade desses dois tipos de repositórios, onde cada um poderia se concentrar no tipo de objeto digital de sua especialidade. Ressalta-se que, embora possam ser produzidos em diferentes formatos, os repositórios lidam com os objetos em seus formatos de preservação, e que para serem reusados no futuro, esses objetos devem estar em formatos que permitam seu reuso no futuro.

Nesse sentido, o presente artigo apresenta duas plataformas de preservação como alternativa para a interoperabilidade entre os RDC-Arq e os Repositórios de Dados Científicos: o Archivematica e o Dataverse, deixando a cargo do Archivematica os acervos documentais e para o Dataverse os conjuntos de dados.

O Archivematica é um sistema de preservação e acesso a longo prazo de materiais digitais, de código aberto e baseado na Web, que está em conformidade com o Modelo OAIS, foi desenvolvido pela Artefactual Systems Inc. e adota padrões bem estabelecidos tais como: *Metadata Encoding & Transmission Standard* (METS), *Preservation Metadata: Implementation Strategies*

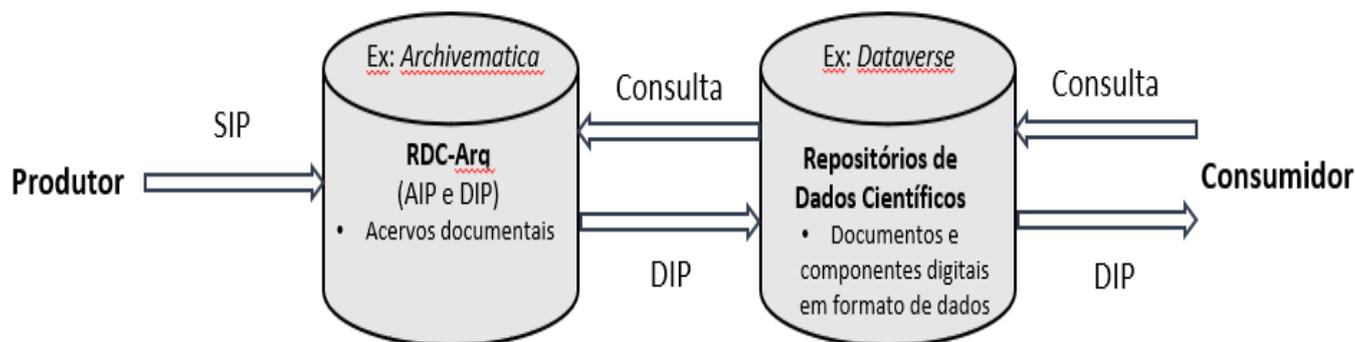
(PREMIS), *Dublin Core*, a especificação *BagIt* da Library of Congress e outras normas reconhecidas internacionalmente (Gava; Flores, 2021).

Em relação aos pacotes do Modelo OAIS, o Archivematica tem seu foco em gerar Pacotes de Informação de Arquivamento (AIP) confiáveis, autênticos, seguros e independentes do sistema de armazenamento utilizado. O Archivematica é usado como uma plataforma de preservação digital para documentos arquivísticos digitais, sendo atualmente a plataforma adotada pelo Arquivo Nacional do Brasil.

O Dataverse é um software de repositório de dados de código aberto, desenvolvido por uma comunidade de desenvolvedores no âmbito do Projeto Dataverse, para compartilhamento, arquivamento e publicação de dados. Embora o Dataverse trabalhe com diferentes formatos, a fim de atender as demandas de diferentes consumidores, o ambiente possui recursos especiais para tratar arquivos em formatos tabulares e geoespaciais, gerando representações em formato uniforme e aberto (.tab), como também confere a integridade e armazena dados geoespaciais em formatos vetoriais, permitindo a integração dos dados com ferramentas externas (Rocha *et al.*, 2021). Os locais de origem de transferência do Dataverse também podem ser configurados para exibir todos os conjuntos de dados disponíveis ou um subconjunto deles:

O Dataverse permite que metadados e arquivos que compõem *datasets* sejam extraídos do ambiente via interface de programação (API). Isso possibilita que ações de preservação digital sejam realizadas por aplicativos externos ao Dataverse. Esta API pode ser configurada para integração com o ambiente Archivematica, que assume a responsabilidade de preservar os *datasets* a longo prazo, enquanto o Dataverse assume a função de dar acesso aos dados (Rocha *et al.*, 2021, p. 17).

A Figura 2 apresenta uma ilustração da interoperabilidade entre o Archivematica e o Dataverse.

Figura 2 - Interoperabilidade entre RDC-Arq e Repositório de Dados Científicos

Fonte: Elaborada pelos autores.

Sendo assim, uma vez que o Archivematica tem sido muito usado como uma plataforma de preservação digital para documentos arquivísticos digitais, o Archivematica pode ser integrado ao Dataverse para preservar os Documentos Arquivísticos Digitais (pacotes AIP) que não estiverem em formato de dados, e interoperar com o Dataverse para que ele apresente aos usuários (consumidor), por meio dos pacotes DIP, os documentos e componentes digitais que estiverem exclusivamente no formato de dados.

5 Considerações finais

Com a transformação digital experimentada principalmente a partir da década de 1990, os cidadãos e instituições começaram a produzir cada vez mais documentos em ambiente digital. Sob o risco de perda desse acervo digital, começaram a surgir iniciativas para a preservação dos documentos digitais e seus componentes. Uma resposta da ciência foi o desenvolvimento de normas, modelos e padrões para a implementação de Repositórios Digitais Confiáveis (RDC) para a preservação de objetos digitais, componentes digitais e metadados.

As instituições também começaram a produzir cada vez mais dados, no decorrer de suas atividades, onde destacam-se as atividades de pesquisa, uma vez que os dados são produzidos no decorrer de atividades institucionais de pesquisa, muitas vezes com financiamento público, é importante que a Arquivologia olhe

para tais conjuntos de dados como um tipo de documento arquivístico, que precisa ser inserido no contexto das políticas de gestão, preservação e acesso.

Neste contexto, este artigo teve como objetivo principal fazer uma reflexão sobre os dados de pesquisa no contexto da Arquivologia e sobre a importância da produção, gestão, transmissão e preservação dos dados em ambientes adequados, mantendo uma cadeia de custódia digital segura e ininterrupta, para que eles se mantenham autênticos e confiáveis ao longo do tempo, com princípio de reuso.

O artigo também teve como objetivo refletir sobre três tipos de Documentos: (1) o DAD (Documento Arquivístico Digital); (2) o DAD com dados, que possui um conjunto de dados como um de seus componentes digitais; e (3) o DAD em dados, cujo formato é exclusivamente na forma de dados. Outra questão abordada refere-se às convergências e semelhanças dos conceitos relacionados ao contexto dos RDC-Arq, para acervos documentais, e dos Repositórios de Dados Científicos, para conjuntos de dados, além da importância da interoperabilidade entre eles, uma vez que cada um desses tipos de repositórios possuem suas especificidades.

O debate também enfatizou a importância da gestão, preservação, acesso e reuso dos documentos arquivísticos digitais que estão em formato de dados, e seus metadados, tanto para reuso quanto para serem usados como matéria-prima em ferramentas de apoio aos gestores em suas tomadas de decisão.

Destacou-se também a importância da inclusão dos PGD, tanto em debates quanto em aplicações, na área arquivística, como parte de políticas de gestão, preservação e acesso, descrevendo quais dados precisam ser coletados ou gerados, quais metodologias e padrões devem ser utilizados nos processos, as condições para que estes dados sejam compartilhados e outorgados acessíveis ou abertos para a comunidade de pesquisa, como eles devem ser curados e preservados, dentre outras questões.

Conclui-se que embora semelhantes, muitos conceitos diferem em suas definições quando aplicados no contexto dos RDC-Arq e dos Repositórios de Dados Científicos. Infere-se que a Arquivologia precisa estar atenta aos

documentos arquivísticos digitais em formato de dados, sob perda desse importante patrimônio documental, à adoção de Planos de Gestão de Dados junto às suas políticas arquivísticas; e sobretudo, que a interoperabilidade entre esses dois tipos de Repositórios trará muitos benefícios, principalmente aos consumidores dessas informações, devido as especificidades que cada um possui com relação à gestão, preservação e acesso aos objetos digitais sob sua responsabilidade.

Referências

ABDELAZIZ, Abid. **Memoria del mundo**: conservando nuestro patrimonio documental. México: Comisión Nacional de los Estados Unidos Mexicanos para la UNESCO, 1998.

ARQUIVO NACIONAL (Brasil). **AN Digital**: política de preservação digital. Rio de Janeiro, 2016.

ARTEFACTUAL SYSTEMS. AtoM. Canada, c2022. Disponível em: <https://www.accesstomemory.org/pt-br/docs/2.3/user-manual/overview/intro/>. Acesso em: 20 fev. 2024.

BASE DE DADOS EM CIÊNCIA DA INFORMAÇÃO (BRAPCI). Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010-2024.

BERTIN, Patrícia Rocha Bello; VISOLI, Marcos Cezar; DRUCKER, Debora Pignatari. A gestão de dados de pesquisa no contexto da e-science: benefícios, desafios e oportunidades para organizações de P&D. **PontodeAcesso**, 2017, 11.2: 34-48. Disponível em: <https://periodicos.ufba.br/index.php/revistaici/article/view/21449>. Acesso em: 20 fev. 2024.

BRASIL. Conselho Nacional de Justiça. **Modelo de requisitos para sistemas informatizados de gestão de processos e documentos do Poder Judiciário**. Brasília: Conselho Nacional de Justiça, 2009.

BROOKS, David. The Philosophy of data. **The New York Times**, New York, 4 Feb. 2013. Disponível em: <https://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html>. Acesso em: 20 fev. 2024.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ) (Brasil). **Diretrizes para a implementação de repositórios arquivísticos digitais confiáveis - RDC-Arq**. Rio de Janeiro: Arquivo Nacional, 2015a.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ) (Brasil). **Orientação Técnica n. 3:** cenários de uso de RDC-Arq em conjunto com o SIGAD. Rio de Janeiro: Arquivo Nacional, 2015b.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ) (Brasil). **Glossário:** documentos arquivísticos digitais. Versão 8.0. Rio de Janeiro: Arquivo Nacional, 2020.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ) (Brasil). **e-ARQ Brasil:** modelo de requisitos para sistemas informatizados de gestão arquivística de documentos. Rio de Janeiro: Arquivo Nacional, 2022.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ) (Brasil). **Resolução n. 39, de 29 de abril de 2014.** [Revogada pela Resolução n. 51, de 25 de agosto de 2023]. Rio de Janeiro, 2023a.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ) (Brasil). **Resolução n. 43, de 4 de setembro de 2015.** [Revogada pela Resolução n. 51, de 25 de agosto de 2023]. Rio de Janeiro, 2023b.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEM (CCSDS). Reference Model for an Open Archival Information System (OAIS). Magenta Book. Washington, Jun. 2012. Disponível em: <https://public.ccsds.org/pubs/650x0m2.pdf>. Acesso em: 20 fev. 2014.

COSTA, Maira Murrieta; CUNHA, Murilo Bastos. O bibliotecário no tratamento de dados oriundos da e-science: considerações iniciais. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 19, n. 3, p. 189-206, 2014. Disponível em: <https://doi.org/10.1590/1981-5344/1900>. Acesso em: 20 fev. 2024.

CRISTOVÃO, Henrique; BATISTA, Willian Alves; ROCHA, Bruna Morêto Sibaldo. Documentos digitais em formato PDF autocontidos, autorreferenciados e autodocumentados como suporte à publicação ampliada. **ÁGORA: Arquivologia em debate**, Florianópolis, v. 33, n. 67, p. 1-25, 2023. Disponível em: <https://agora.emnuvens.com.br/ra/article/view/1151>. Acesso em: 15 fev. 2024.

DIGITAL CURATION CENTRE (DCC). **What is digital curation?** Edimburgo, c2023. Disponível em: <https://www.dcc.ac.uk/about/digital-curation>. Acesso em: 20 fev. 2024.

DURANTI, Luciana. Rumo a uma teoria arquivística de preservação digital: as descobertas conceituais do projeto InterPARES. **Arquivo & Administração**, Rio de Janeiro, v. 4, n. 1, p. 5-18, 2005.

GAVA, Tânia Barbosa Salles; FLORES, Daniel. Cadeia de Custódia Digital

Arquivística - CCDA. *In*: SIMPÓSIO INTERNACIONAL DE ARQUIVOS, 8., 2020, São Paulo. **Anais** [...]. São Paulo: ARQ-SP, 2020. Disponível em: <https://even3.blob.core.windows.net/anais/336974.pdf>. Acesso em: 20 fev. 2024.

GAVA, Tânia Barbosa Salles; FLORES; Daniel. O papel do Archivematica no RDC-Arq e possíveis cenários de uso. **ÁGORA: Arquivologia em debate**, Florianópolis, v. 31, n. 63, p. 1-21, 2021. Disponível em: <https://agora.emnuvens.com.br/ra/article/view/1018>. Acesso em: 20 fev. 2024.

GRANT, Rebeca. Recordkeeping and research data management: a review of perspectives. **Records Management Journal**, Dublin, v. 27, n. 2, p. 159-174, 2017. Disponível em: <http://hdl.handle.net/10197/8769>. Acesso em: 20 fev. 2024.

INTERNATIONAL COUNCIL ON ARCHIVES (ICA). **Multilingual archival terminology**. Paris, c2024. Disponível em: <http://www.ciscra.org/mat/>. Acesso em: 20 fev. 2024.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO) ISO 14721:2012. Space data and information transfer systems: open archival information system – Reference model. Genebra, 2012a.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO 16363:2012. Space data and information transfer systems: audit and certification of trustworthy digital. Genebra, 2012b.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO 16.919:2014. Space data and information transfer systems. Requirements for bodies [...]. Genebra, 2014.

INTERNATIONAL RESEARCH ON PERMANENT AUTHENTIC RECORDS IN ELECTRONIC SYSTEMS (InterPARES 2 Project). **Diretrizes do Produtor**. A elaboração e a manutenção de materiais digitais: diretrizes para indivíduos. Vancouver: University of British Columbia, 2010. Disponível em: https://www.gov.br/arquivonacional/pt-br/canais_atendimento/imprensa/copy_of_noticias/serie-publicacoes-do-conarq-diretrizes-do-produtor-e-do-preservador. Acesso em: 20 fev. 2024.

NATIONAL RESEARCH COUNCIL (NRC). **A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases**. Washington, DC: The National Academies Press, 1999.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). **OECD Principles and guidelines for access to research data from public funding**. Paris, OECD, 2007.

PAGANINE, Lucas Nóbrega; AMARO, Bianca. Características dos repositórios de dados científicos no Brasil. **BIBLOS**, Rio Grande, v. 34, n. 1, p. 176-188,

2020. Disponível em: <https://doi.org/10.14295/biblos.v34i1.11132>. Acesso em: 20 fev. 2024.

PAMPEL, Heinz *et al.* Making research data repositories visible: the re3data.org registry. **PLoS One**, San Francisco, v. 8, n. 11, p. 1-10, 2013. Disponível em: <https://doi.org/10.1371/journal.pone.0078080>. Acesso em: 20 fev. 2024.

ROCHA, Rafael Port *et al.* Análise dos sistemas DSpace e Dataverse para repositórios de dados de pesquisa com acesso aberto. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 17, p. 1-25, 2021. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/1572/1261>. Acesso em: 20 fev. 2024.

SALES, Luana Farias; SAYÃO, Luís Fernando. **Dados de pesquisa: quem ama cuida**. Rio de Janeiro: Livro Aberto, 2019. Disponível em: <http://livroaberto.ibict.br/handle/123456789/1083>. Acesso em: 20 fev. 2024.

SANTOS, Plácida L. V. A. da Costa; SANT'ANA, Ricardo César Gonçalves. Dado e Granularidade na perspectiva da Informação e Tecnologia: uma interpretação pela Ciência da Informação. **Ciência da Informação**, Brasília, v. 42, n. 2, 2013. Disponível em: <https://doi.org/10.18225/ci.inf.v42i2.1382>. Acesso em: 20 fev. 2024.

SAYÃO, Luís Fernando; SALES, Luana Farias. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação e Sociedade**, João Pessoa, v. 22, n. 3, p. 179-191, 2012. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/12224/8586>. Acesso em: 20 fev. 2024.

SAYÃO, Luís Fernando; SALES, Luana Farias. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN. 2015.

SAYÃO, Luís Fernando; SALES, Luana Farias. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 21, n. 2, p. 90-115, 2016. Disponível em: <https://doi.org/10.5433/1981-8920.2016v21n2p90>. Acesso em: 20 fev. 2024.

SAYÃO, Luís Fernando; SALES, Luana Farias. Afinal, o que é dado de pesquisa? **BIBLOS**, Rio Grande, v. 34, n. 2, p. 1-20, 2020. Disponível em: <https://doi.org/10.14295/biblos.v34i2.11875>. Acesso em: 20 fev. 2024.

SEMELER, Alexandre Ribas; PINTO, Adilson Luiz. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia e dados. **Ciência da Informação**, Brasília, v. 48, n. 1, p. 113-129, 2019. Disponível em: <https://doi.org/10.18225/ci.inf.v48i1.4461>. Acesso em: 20 fev. 2024.

SILVA, Amanda Marissa Soares; *et al.* Curadoria digital e arquivologia: olhares sobre o documento arquivístico digital. **Revista Ibero-Americana de Ciência**

da Informação, Brasília, v. 14, p. 567-582, 2021. Disponível em:
<https://doi.org/10.26512/rici.v14.n2.2021.37558>. Acesso em: 20 fev. 2024.

SOCIETY OF AMERICAN ARCHIVISTS. **Encoded Archival Description** (EAD). Chicago, c2024. Disponível em:
<https://www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-standards-ts-eas/encoded-archival-description-ead>. Acesso em: 15 fev. 2024.

VIDOTTI, Silvana Aparecida Borsetti Gregório; *et al.* Repositório de dados de pesquisa para grupo de pesquisa: um projeto piloto. **Informação & Tecnologia**, São Paulo, v. 4, n. 2, p. 221-242, 2017. Disponível em:
<https://doi.org/10.22478/ufpb.2358-3908.2017v4n2.40198>. Acesso em: 20 fev. 2024.

Research data in Archival Science: a reflection

Abstract: This article aims to reflect on research data in the context of Archival Science and on the importance of production, management, transmission, and preservation of data in suitable environments, maintaining a safe and uninterrupted digital chain of custody so that data remain authentic and reliable over time and for reuse purposes. This is qualitative, exploratory research. Regarding the nature of the sources used, this is a bibliographic and documentary research. The article presents similarities and differences in the definitions of concepts related to the context of Reliable Digital Archival Repositories, for documentary collections, and Scientific Data Repositories, for datasets, and on the importance of interoperability between them, since each type of repository has its specificities. The article reflects on the importance of management, preservation, access, and reuse of digital archival documents in data format, both for reuse and to support managers in their decision-making, highlighting the importance of Data Management Plans. We conclude that although similar, many concepts differ when applied in the context of Trusted Digital Archival Repositories and Scientific Data Repositories. That Archival Science needs to pay attention to digital archival documents in data format, and interoperability between Reliable Digital Archival Repositories and Scientific Data Repositories will bring many benefits to consumers of this information.

Keywords: research data; digital archival document; reliable digital archival repositories; scientific data repositories; digital preservation

Recebido: 11/06/2023

Aceito: 21/02/2024

Declaração de autoria:

Concepção e elaboração do estudo: Tânia Barbosa Salles Gava; Daniel Flores; Diana Vilas Boas Souto Aleixo; Henrique Monteiro Cristovão; Luciana Itida Ferrari; Margarete Farias de Moraes.

Coleta de dados: Tânia Barbosa Salles Gava; Daniel Flores; Diana Vilas Boas Souto Aleixo; Henrique Monteiro Cristovão; Luciana Itida Ferrari; Margarete Farias de Moraes.

Análise e interpretação de dados: Tânia Barbosa Salles Gava; Daniel Flores; Diana Vilas Boas Souto Aleixo; Henrique Monteiro Cristovão; Luciana Itida Ferrari; Margarete Farias de Moraes.

Redação: Tânia Barbosa Salles Gava; Daniel Flores; Diana Vilas Boas Souto Aleixo; Henrique Monteiro Cristovão; Luciana Itida Ferrari; Margarete Farias de Moraes.

Revisão crítica do manuscrito: Tânia Barbosa Salles Gava; Daniel Flores; Diana Vilas Boas Souto Aleixo; Henrique Monteiro Cristovão; Luciana Itida Ferrari; Margarete Farias de Moraes.

Como citar

GAVA, Tânia Barbosa Salles; FLORES, Daniel; ALEIXO, Diana Vilas Boas Souto; CRISTOVÃO, Henrique Monteiro; FERRARI, Luciana Itida; MORAES, Margarete Farias. Dados de pesquisa na Arquivologia: uma reflexão. **Em Questão**, Porto Alegre, v. 30, e-135857, 2024. DOI: <https://doi.org/10.1590/1808-5245.30.135857>



¹ DURANTI, Luciana; THIBODEAU, Ken. The concept of record in interactive, experiential and dynamic environments: the view of InterPARES. **Archival Science**, Netherlands, v. 6, p.13-68, 2006. Disponível em: <https://doi.org/10.1007/s10502-006-9021-7>. *Apud* CONAR (2020, p. 25).

² O AtoM é um software livre e com o código-fonte aberto, desenvolvido para ser uma interface de acesso aos documentos arquivísticos (Artefactual Systems, c2022).

³ A EAD é um padrão de codificação usado na descrição arquivística de documentos na web que auxilia nos processos de busca e arquivamento (Society of American Archivists, c2024).