

CARACTERIZAÇÃO DE REPOSITÓRIOS DIGITAIS DATAVERSE conforme o Modelo OAIS

CHARACTERIZATION OF DATAVERSE DIGITAL REPOSITORIES according to the OAIS Model



Resumo

Introdução: Na temática de padrões de modelagem de sistemas de arquivamento, especificamente OAIS (*Open Archival Information System*), parece correto afirmar que, por se tratar de um modelo abstrato e que não contempla nenhum tipo de implementação, vários desafios surgem quando se busca entender e visualizar sua aplicação prática. Assim, faz-se necessário situar no contexto do modelo OAIS uma solução tecnológica de repositórios digitais considerando aspectos referentes às suas políticas de gestão, acesso e uso além de possíveis processos de curadoria e preparação dos dados. **Objetivo:** Este estudo busca detalhar, com base nas entidades do modelo OAIS, os componentes essenciais de um repositório digital que utiliza o software Dataverse, responsáveis pela gestão do conteúdo digital armazenado. **Metodologia:** Análise qualitativa, pesquisa bibliográfica e documental. **Resultados:** Relação entre elementos que compõem o modelo OAIS, considerando o fluxo básico de ingestão, depósito, arquivamento e acesso aos dados, e os recursos do software Dataverse, especificações de funcionamento do repositório e ações de curadoria, acesso e uso. **Conclusão:** Entende-se que são muitos os desafios ao se pensar no ecossistema da gestão de dados científicos. No entanto, considerar somente tecnologia não é suficiente ou garantia de que as recomendações dos padrões e modelos internacionais estão sendo seguidos. Fica evidente no modelo OAIS a importância da equipe gestora do repositório para além das funcionalidades oferecidas pelo software. Trazer o modelo OAIS para uma realidade prática parece favorecer seu entendimento, especialmente para os profissionais que estão na fase inicial de implementação de uma solução de gestão de dados científicos.

Palavras-chave: Modelo OAIS. Repositórios digitais. Dataverse. Preservação digital.

Abstract

Introduction: On the subject of archival systems modeling standards, specifically OAIS (Open Archival Information System) in the digital repositories scope, it seems correct to affirm that, as it is an abstract model and does not include any type of implementation, several challenges may arise one seeks to understand and visualize the practical application of the model. Thus, it is necessary to try to place a technological solution for digital repositories in the context of the OAIS model, considering aspects related to their management, access and use policies, as well as possible processes of curation and data preparation by the depositor. **Objective:** This study seeks to detail, based on the entities of the OAIS model, the essential components of a digital repository that uses the Dataverse software, responsible for the effective management of stored digital data. **Methodology:** This is a qualitative analysis, based on bibliographic and documentary research. **Results:** Establish a relationship between elements that make up the OAIS model, considering the basic flow of ingestion, deposit, archiving and access to data, and the Dataverse software resources, repository operation specifications and curation actions, access and use by the users. **Conclusion:** It is understood that there are many challenges when thinking about the scientific data management ecosystem. However, thinking only about technology is not enough or a guarantee that the recommendations of international standards and models are being followed. It is evident in the OAIS model the importance of the repository management team beyond the functionalities offered by the software. Bringing the OAIS model to a practical reality seems to favor its understanding, especially for professionals who are in the initial phase of implementing a scientific data management solution.

Keywords: OAIS Model. Digital repositories. Dataverse. Digital preservation.



Laura V. Rodrigues Rezende

Universidade Federal de Goiás

E-mail: lauravil.rr@gmail.com

Goiânia – GO / Brasil

RBP

Revista Brasileira de
Preservação Digital

RBP

Brazilian Journal of
Digital PreservationCREDIT¹

• Conceituação	REZENDE, L. V. R.
• Curadoria de dados	REZENDE, L. V. R.
• Análise formal	REZENDE, L. V. R.
• Investigação	REZENDE, L. V. R.
• Metodologia	REZENDE, L. V. R.
• Administração de projetos	REZENDE, L. V. R.
• Recursos	REZENDE, L. V. R.
• Supervisão	REZENDE, L. V. R.
• Validação	REZENDE, L. V. R.
• Visualização	REZENDE, L. V. R.
• Redação – rascunho original	REZENDE, L. V. R.
• Redação – revisão e edição	REZENDE, L. V. R.



LICENÇA DE USO

Os autores cedem à [Revista Brasileira de Preservação Digital](#) os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a Licença Creative Commons Attribution (CC BY) 4.0 International. Esta licença permite que terceiros remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os autores têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

PUBLISHERS

Universidade Estadual de Campinas – Sistema de Bibliotecas / Instituto Brasileiro de Informação em Ciência e Tecnologia – Rede Brasileira de Serviços de Preservação Digital – Cariniana. As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES

Gildenir Carolino Santos, Miguel Angel Márdero Arellano.

Submetido em: 30/06/2022 – Aceito em: 10/06/2022 – Publicado em: 12/07/2022

¹ Sobre o CRediT, consulte o site e conheça outros papéis: <https://casrai.org/credit/>

1 Introdução

Os repositórios de dados permitem trazer um “novo” olhar para a relevância dos dados científicos, visto que permitem descortinar, para os pesquisadores e para a própria Ciência, possibilidades outras de (re)elaboração e (re)confirmação das pesquisas, além de servir como base para outras investigações (GALVINO; ROSA; OLIVEIRA, 2020) favorecendo a garantia de integridade, aumentando a eficiência da pesquisa e possibilitando a comparação de novos dados gerados a partir de outras investigações dotadas de dados coletados anteriormente. Para essas disposições, dentre as tecnologias disponíveis, o software Dataverse se apresenta como um dos mais completos e em conformidade com princípios e padrões internacionais, dentre eles os princípios FAIR endossados e projetados por representantes da academia, indústria, agências de fomento, editores acadêmicos, dentre outros. Podem ser considerados como diretrizes para aqueles que desejam aprimorar a reutilização de seus acervos de dados (WILKINSON, 2016). O conjunto de princípios pontua que os dados devem ser localizáveis (os metadados e os dados devem ser fáceis de encontrar tanto para humanos quanto para computadores), acessíveis (os usuários precisam saber como podem acessar os dados, possivelmente incluindo autenticação e autorização), interoperáveis (os dados precisam ser compatíveis e interoperar com outros aplicativos) e reutilizável (metadados e dados devem ser bem descritos para que possam ser replicados e/ou combinados em diferentes configurações) (GOFAIR, 2020). Um outro modelo conceitual que representou um relevante avanço na concepção de sistemas digitais de arquivamento foi o OAIS, ou *Open Archival Information System* ou Sistema Aberto para Arquivamento de Informação (SAAI). Trata-se de um modelo abstrato dotado de flexibilidade considerável aos desenvolvedores de sistemas de arquivamento, provendo uma visão abstrata teórica de como a informação arquivística pode fluir de uma entidade para outra sem a perda de nenhum dos elementos que a compõem.

Na temática de padrões de modelagem de sistemas de arquivamento, especificamente o OAIS, parece correto afirmar que, por se tratar de um modelo abstrato e que não contempla nenhum tipo de implementação, vários desafios podem surgir quando se busca entender e visualizar a aplicação prática do modelo. Assim, faz-se necessário tentar situar no contexto do modelo OAIS uma solução tecnológica de repositórios digitais e, para este estudo optou-se pelo software Dataverse.

Esta escolha ocorreu com base no relatório de 2018 - Acesso aberto a dados de pesquisa no Brasil: soluções tecnológicas (ROCHA, 2018) que aponta o Dataverse como o software de armazenamento e compartilhamento de dados de pesquisa mais utilizado. Ele possui recursos facilmente configuráveis para definir vários tipos de ambientes institucionais e diferentes características para repositórios, incluindo diferentes hierarquias organizacionais e gerenciamento de políticas para unidades ou grupos geradores de dados, vários metadados e esquemas de licença.

Um repositório Dataverse é a instalação de software, que hospeda vários arquivos virtuais chamados dataverses. Cada dataverse contém os *datasets*, que são os conjuntos de dados (e também pode conter outros dataverses), e cada *dataset* contém metadados descritivos e arquivos de dados (incluindo documentação e código que acompanham os dados) (THE DATAVERSE PROJECT, 2022).

O software Dataverse está agora na versão 5.0 e continua a melhorar seu suporte aos princípios FAIR, particularmente no suporte a: identificadores persistentes (no *dataset* e nível de arquivo), com URL e metadados registrados, metadados

personalizáveis (incluindo suporte para vários padrões) exportáveis em vários formatos, controle de versão para *datasets* e arquivos, suporte a dados vinculados, termos de acesso e uso de dados e conversão de arquivos para formatos reutilizáveis. É importante observar que, embora o software Dataverse ajude a tornar os dados em conformidade com os princípios FAIR, os produtores de dados e os gerentes de coleção devem contribuir para esse objetivo usando metadados de comunidade e padrões de vocabulários controlados apropriados.

Diante do exposto, tem-se a seguinte indagação: Como os repositórios digitais desenvolvidos no software Dataverse apresentam conformidade com o modelo OAIS?

Busca-se detalhar, com base nas entidades do modelo OAIS, os componentes essenciais de um repositório digital de software Dataverse responsáveis pela efetiva gestão do conteúdo digital armazenado.

Trata-se de uma análise qualitativa, com base em pesquisa bibliográfica e documental, que estabelecerá relação entre elementos e categorizará conforme o modelo OAIS, considerando o fluxo básico de ingestão, depósito, arquivamento e disponibilização de dados, os recursos do software Dataverse, especificações de funcionamento do repositório e ações de curadoria, acesso e uso por parte dos usuários.

2 Revisão da literatura

Conforme observado em inúmeros trabalhos sobre a guarda da informação em formato digital, nota-se que apenas promover a guarda destes documentos em soluções tecnológicas e aplicações de software não tem sido suficiente para assegurar a longevidade de acesso e preservação destes documentos resguardados. Tal fato se deve à constatação certa de que tanto os novos recursos de software quanto as novas tecnologias representadas pelos hardwares se tornam obsoletas em um tempo relativamente curto, estando ainda suscetíveis à danos e acidentes não previstos e que podem encurtar ainda mais sua perspectiva de funcionamento em relação ao tempo de uso estimado.

Desta forma, uma vez identificadas as particularidades dos objetos digitais em sua relação de dependência com os recursos de software e hardware, assim como as necessidades de acesso e preservação da informação neles contidas, observa-se que fundamentadas em diversos estudos sobre a temática da preservação digital, inúmeras foram as propostas de padronização de esquemas de metadados, iniciativas de criação de modelos de referência, sendo que o *Open Archival Information System* (OAIS) é considerado um dos mais difundidos mundialmente. O modelo de referência OAIS apresentado na figura 3 trata de:

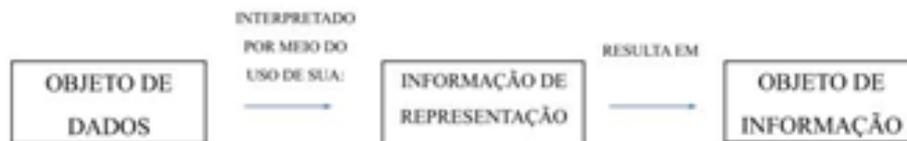
[...] uma iniciativa ISO (*International Organization for Standardization*), desde junho 2003, que define em alto nível um sistema aberto de arquivamento de informação, sendo um arquivo compreendido como uma organização de pessoas e sistemas, que aceitou a responsabilidade de preservar informação e torná-la disponível a uma comunidade alvo. Um arquivo distingue-se de outros usos do termo arquivo por cumprir um conjunto de responsabilidades, como as definidas na norma. (ABNT, 2007, p. vi).

O termo “aberto” do modelo não está relacionado ao acesso irrestrito ao arquivo e sim ao desenvolvimento de recomendações e padrões relacionados por meio de fóruns abertos à toda a comunidade de interesse ao nível mundial.

O modelo OAIS aborda um conjunto completo de funções arquivísticas para a preservação da informação, incluindo admissão, arquivamento, gerenciamento de dados, acesso e disseminação. O modelo também contempla a migração da informação digital para novas mídias e formatos, os modelos de dados usados para representar a informação, o papel do software na preservação da informação e o intercâmbio da informação digital entre arquivos. Identifica as interfaces, tanto internas quanto externas, para as funções arquivísticas e descreve alguns serviços de alto nível nessas interfaces. (ABNT, 2007, p. vi).

No contexto do modelo OAIS, a informação significa qualquer tipo de conhecimento intercambiável, sendo sempre expressa por algum tipo de dado. O esquema a seguir ilustra a obtenção da informação digital (objeto de informação) a partir da identificação dos bits (objeto de dados) e a informação de representação aplicada a esses bits.

Figura 1. OAIS - Obtenção de informação a partir de dados

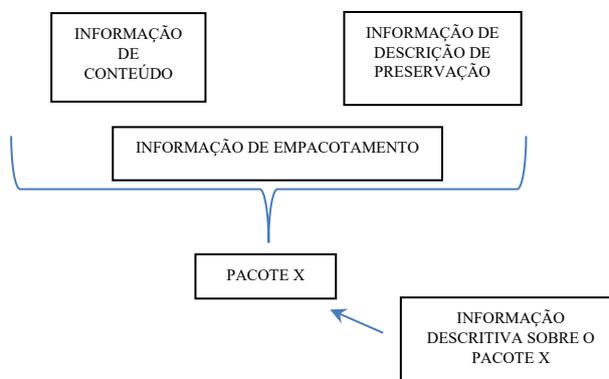


Fonte: adaptado de ABNT, 2007, p. 10

Um outro conceito importante para o modelo é o de pacote de informação,

Este pacote significa um recipiente conceitual de dois tipos de informação: informação de conteúdo e informação de descrição de preservação. Estas duas informações são encapsuladas e identificadas pela informação de empacotamento. Este pacote resultante pode ser encontrado em virtude da informação descritiva. (ABNT, 2007, p. 11)

Figura 2. Pacote de informação e seus relacionamentos



Fonte: adaptado de ABNT, 2007, p. 11

A informação de conteúdo é a informação-alvo original da preservação. Segundo o modelo OAIS, a informação de descrição de preservação é aplicada à

informação de conteúdo e de extrema importância para sua preservação garantindo sua identificação e entendimento do ambiente onde foi produzida. A informação de descrição de preservação é dividida em quatro tipos, quais sejam:

- Proveniência: descreve a fonte da informação de conteúdo, especificamente quem teve a sua custódia desde sua origem e seu histórico (incluindo histórico de processamento);
- Contexto: descreve como a informação de conteúdo relaciona-se com outras informações fora do pacote de informação;
- Referência: fornece um ou mais identificadores, ou sistemas de identificação, que permitem que a informação de conteúdo seja identificada de forma única;
- Fixidez: fornece proteção à informação de conteúdo contra alterações não documentadas.

Figura 3. O modelo de referência *Open Archival Information System (OAIS)* ou Sistema Aberto para Arquivamento de Informação (SAAI): Entidades Funcionais



Fonte: ABNT, 2007, p. 19

A seguir uma breve descrição dos tipos de pacotes de informações e das entidades funcionais do modelo OAIS. Vale ressaltar que além das entidades a serem descritas, diversos serviços comuns também fazem parte do modelo e são considerados uma entidade funcional à parte, não menos importante, porém, devido à sua extensão, não serão descritos neste estudo:

- Pacote de Submissão de Informação (PSI): Pacote enviado ao SAAI pelo produtor da informação. Geralmente ele possui informação de conteúdo e informação de descrição de preservação (IDP).
- Pacote de Arquivamento de Informação (PAI): O PSI é transformado em um ou mais pacotes de arquivamento de informação para preservação. O PAI possui o conjunto completo de informações de descrição de preservação para a informação de conteúdo associada;

- Pacote de Disseminação de Informação (PDI): O PDI é o pacote que será enviado ao consumidor de informação atendendo à sua solicitação, ou seja, o PAI se transforma em PDI. O PDI pode incluir séries de PAI e pode possuir ou não IDP completa. A informação de empacotamento necessariamente está presente para permitir que o consumidor distinga claramente a informação que solicitou.
- Recepção, Ingestão ou Submissão: Serviços e processos necessários para aceitar PSI dos produtores (ou de componentes internos sob controle da entidade administração do sistema) e preparar os conteúdos para arquivamento e gerenciamento.
- Armazenamento ou arquivamento: Serviços e processos necessários para arquivar, manter e recuperar PAI.
- Gerenciamento de dados: Serviços e processos necessários para incluir, manter e acessar tanto a informação descritiva, que identifica e documenta os conteúdos, quanto os dados administrativos usados para gerenciar o SAAI.
- Administração do sistema: Serviços e processos necessários para operar o sistema como um todo. É responsável ainda por estabelecer e manter os padrões e políticas do SAAI, por fornecer apoio ao cliente e por ativar solicitações armazenadas.
- Planejamento da Preservação: Serviços e processos necessários para monitorar o ambiente e fornecer recomendações para assegurar que a informação armazenada permaneça acessível por longo prazo à comunidade alvo, mesmo que o ambiente computacional original se torne obsoleto.
- Acesso: Serviços e processos necessários para apoiar os consumidores na determinação da existência, descrição, localização e liberação da informação armazenada e permitir que os consumidores solicitem e recebam produtos de informação.

O modelo de referência é utilizado no contexto de estratégias voltadas para a padronização dos processos que conduzem a conferência da preservação da informação digital suportada por meios de acesso e difusão como os repositórios institucionais, fornecendo uma ampla visão das etapas necessárias ao tratamento dos recursos para acesso, considerando desta forma os atores envolvidos neste processo e que consistem basicamente em quatro entidades básicas: “produtores, consumidores, administração e o arquivo propriamente dito” (MÁRDERO ARELLANO, 2004, p. 19). Por constituir-se um modelo referencial, o esquema não conduz à implementação de fato de sistemas de informação, mas permite que se tenha controle das condições a serem consideradas em um projeto de preservação utilizando os metadados como suporte à preservação dos documentos em contexto, por exemplo, do fluxo de inclusão em repositórios.

Segundo acrescenta Sayão (2010, p. 13), o modelo de referência OAIS “é uma infraestrutura conceitual que descreve o ambiente, as interfaces externas, os componentes funcionais e os objetos de informação, associados com um sistema responsável pela preservação de longo prazo de materiais digitais”. Por conseguinte, resta às unidades administradoras destas coleções de cunho científico, acadêmico, arquivístico, administrativo, cultural e histórico, se atentarem para a importância de preservar os documentos digitais, adotando assim soluções de metadados específicas para estas finalidades, assim como associá-los a outras práticas de preservação que

culminam na garantia da segurança, preservação e acesso futuro ao capital intelectual e social de que dispõem.

A essência da preservação digital inserida no contexto do modelo OAIS, considerando suas entidades funcionais e seus processos, sugere um fluxo informacional com papéis bem definidos para produtor, administração e consumidor de informação basicamente, interagindo com o SAAI (arquivo) favorecendo a implementação de um sistema de arquivamento para a preservação a longo prazo, incluindo os repositórios de dados científicos que utilizam o software Dataverse. Esta relação será apresentada no presente estudo visando favorecer o entendimento do modelo OAIS por meio de um esquema de implementação envolvendo infraestrutura conceitual, tecnológica, de curadoria e utilização.

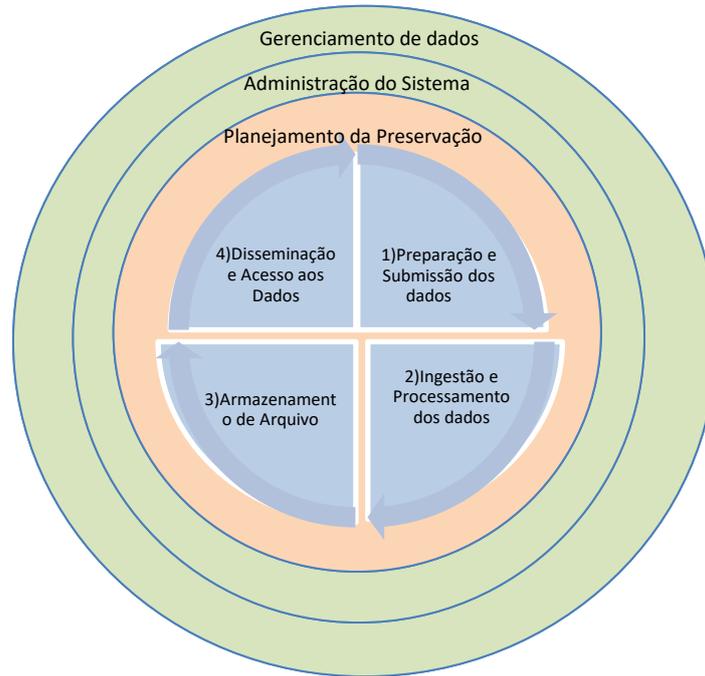
5 Discursão e Resultados

Ao se propor a situar o modelo OAIS no contexto dos repositórios digitais implementados utilizando-se o software Dataverse, faz-se necessário inicialmente especificar as funcionalidades do modelo conforme o fluxo de trajetória dos dados considerando o ecossistema de gestão de dados científicos utilizando-se o software Dataverse. Os processos de cada funcionalidade foram agrupados em categorias assim denominadas:

- **Software:** referente às funcionalidades oferecidas pelo software Dataverse;
- **Infraestrutura do Repositório:** referente às regras de funcionamento do repositório estabelecidas por seus gestores;
- **Curadoria:** envolve os produtores e usuários que acessam e depositam os dados.

A figura a seguir ilustra as funcionalidades de um ecossistema de gestão de dados científicos que utiliza o software Dataverse trazendo também as entidades funcionais consideradas pelo modelo OAIS.

Figura 4. Ecossistema de Gestão de Dados Científicos utilizando o software Dataverse e entidades funcionais do modelo OAIS



Fonte: Elaborado pela autora

A seguir serão descritas as funcionalidades apresentadas na figura 4 conforme as categorias previamente definidas, os recursos oferecidos pelo software Dataverse, a infraestrutura do repositório e a curadoria.

5.1 Funcionalidade: Gerenciamento de dados

A funcionalidade “Gerenciamento de Dados” do ponto de vista do ciclo de vida de uma pesquisa científica é responsável por todo o sistema de arquivamento de maneira geral. Especificando tecnicamente conforme o modelo OAIS, ela é responsável pela manutenção e funcionamento das bases de dados, processos necessários para a inclusão, manutenção e acesso aos dados e metadados além da geração de dados estatísticos e relatórios.

Quadro 1. Funcionalidade: Gerenciamento de dados - Categorias e processos

Gestão de Dados Científicos - Categorias	Gestão de Dados Científicos – Processos
Software	<ul style="list-style-type: none"> Referente à base de dados gerenciada pelo Dataverse Dataverse fornece dados estatísticos e relatórios;
Infraestrutura do Repositório	<ul style="list-style-type: none"> Acompanhamento e avaliação do funcionamento do repositório;
Curadoria	<ul style="list-style-type: none"> Engloba todo o processo de curadoria dos dados;

Fonte: Elaborado pela autora

5.2 Funcionalidade Administração do Sistema

A funcionalidade “Administração do Sistema”, do ponto de vista do ciclo de vida de uma pesquisa científica é responsável pelo acompanhamento diário do funcionamento do repositório que engloba monitoramento de performance e atualizações do sistema. Coordena os processos que fazem parte de todas as outras funcionalidades.

Quadro 2. Funcionalidade: Administração do Sistema - Categorias e processos

Gestão de Dados Científicos - Categorias	Gestão de Dados Científicos – Processos
Software	<ul style="list-style-type: none"> • Desenvolvimento contínuo do software com atualizações;
Infraestrutura do Repositório	<ul style="list-style-type: none"> • Acompanhamento e avaliação diária de funcionamento do repositório que engloba monitoramento de performance e atualizações do sistema; • Contribuem com o desenvolvimento do software (Comunidades de práticas); • Estabelecem a interface entre produtores e consumidores de dados.
Curadoria	<ul style="list-style-type: none"> • Acompanham seus <i>datasets</i>; • Contribuem com o desenvolvimento do software.

Fonte: Elaborado pela autora

5.3 Funcionalidade Planejamento da Preservação

A funcionalidade “Planejamento da Preservação”, do ponto de vista do ciclo de vida de uma pesquisa científica é responsável pelo acompanhamento do ambiente externo verificando a existência de ameaças e buscando melhorias na segurança do sistema.

Quadro 3. Funcionalidade: Planejamento da Preservação - Categorias e processos

Gestão de Dados Científicos - Categorias	Gestão de Dados Científicos – Processos
Software	<ul style="list-style-type: none"> • Verificações automáticas de integridade (<i>checksum</i>); • Favorece uso de formatos de arquivos preserváveis;
Infraestrutura do Repositório	<ul style="list-style-type: none"> • Criação e desenvolvimento de um plano de preservação; • Contribuem com Comunidades de práticas visando estabelecer melhores práticas para a preservação; • Definem e atribuem adequadamente os papéis de uso do software de acordo com a realidade da instituição;
Curadoria	Produtores: <ul style="list-style-type: none"> • Uso de formatos de arquivos preserváveis; • Designam os atores que poderão manipular os <i>datasets</i>;

Fonte: Elaborado pela autora

5.4. Funcionalidade: Preparação e Submissão dos dados

Do ponto de vista do ciclo de vida de uma pesquisa científica, este processo de preparação e submissão dos dados se inicia na etapa de planejamento da pesquisa. O ideal seria que o pesquisador, ao planejar a realização de sua investigação considere também pensar no que irá fazer com seus dados. No contexto do modelo OAIS, considera-se a entidade funcional “Recepção” (Ingestão) como equivalente às ações realizadas nesta funcionalidade.

Quadro 4. Preparação e Submissão dos dados - Categorias e processos

Gestão de Dados Científicos - Categorias	Gestão de Dados Científicos – Processos
Software	<ul style="list-style-type: none"> • Dataverse exige que os depositantes façam cadastro no sistema; • Dataverse identifica campos de metadados obrigatórios e exige preenchimento;
Infraestrutura do Repositório	<ul style="list-style-type: none"> • Determina e verifica conformidade com tipos de dados aceitos para depósito; <ul style="list-style-type: none"> ➢ Tipos de dados do ponto de vista da natureza (imagem, áudio, vídeo, texto), origem (observacionais, computacionais, experimentais) e formatos (preferencialmente preserváveis); • Estabelece licenças de acesso e uso dos dados; • Determina e verifica se os requisitos de submissão foram atendidos;
Curadoria	<p>Produtor de dados:</p> <ul style="list-style-type: none"> • Coleta ou seleciona os dados com suas análises, arquivos de documentação e códigos; • Realiza cadastro no Dataverse para depósito; • Estabelece termos de acesso e uso conforme política do repositório; • Descreve os dados por meio do preenchimento dos metadados; • Cria datasets; • submete os arquivos;

Fonte: Elaborado pela autora

5.5 Funcionalidade: Ingestão e Processamento dos dados

A ingestão e processamento dos dados se inicia quando o produtor finaliza o processo de submissão. No contexto do Dataverse, após o processo de submissão inicial, os dados permanecem em uma “versão rascunho” (*draft*), aguardando alguma possível conferência de conformidade das informações submetidas por parte de outra entidade (gestores do repositório, equipe de curadoria). Em relação ao modelo OAIS, é possível estabelecer que a entidade funcional “Gerenciamento dos Dados” realiza processos necessários para incluir e manter os dados, informação descritiva e administrativa no sistema. Para esta funcionalidade é possível considerar também processos atribuídos à entidade funcional “Armazenamento”.

Quadro 5. Ingestão e Processamento dos Dados - Categorias e processos

Gestão de Dados Científicos – Categorias	Gestão de Dados Científicos – Processos
Software	<ul style="list-style-type: none"> • Dataverse realiza leitura dos arquivos de dados e podendo converter dados tabulares e outros em formato de arquivo não proprietário; • Dataverse extrai e permite pré-visualização de arquivos de tabelas; • Dataverse processa e armazena os dados depositados; • Dataverse atribui identificadores persistentes; • Dataverse verifica integridade e autenticidade (<i>checksum</i>);
Infraestrutura do Repositório	<ul style="list-style-type: none"> • Determinam rotina de verificação de depósitos (detecção de spams, revisão de depósitos realizados por pesquisadores); • Determinam quais são os campos de metadados de preenchimento obrigatórios; • Determinam rotina de armazenamento dos dados em hardware específico; • Quando não existe auto depósito: Publicam os dados submetidos pelos pesquisadores após conferência;
Curadoria	<ul style="list-style-type: none"> • Produtores de dados (auto depósito): Publicam definitivamente os dados submetidos;

Fonte: Elaborado pela autora

5.6 Funcionalidade: Armazenamento de Arquivo

Esta funcionalidade é caracterizada pelo processo efetivo de armazenamento dos arquivos de dados. Considera-se realizada após a publicação definitiva dos dados. No caso em que o repositório gerencia de maneira centralizada o processo de depósito, a equipe gestora encaminha esta publicação definitiva. Em relação ao modelo OAIS, além dos processos necessários para o arquivamento, são realizadas nesta entidade funcional os processos necessários para manter e recuperar os dados.

Quadro 6. Armazenamento de Arquivo - Categorias e processos

Gestão de Dados Científicos - Categorias	Gestão de Dados Científicos – Processos
Software	<ul style="list-style-type: none"> • Dataverse mantém conexão entre a base de dados e o sistema de armazenamento; • Dataverse cria checagens de integridade dos dados (<i>checksum</i>);
Infraestrutura do Repositório	<ul style="list-style-type: none"> • Arquivos de dados e metadados são armazenados e backup é realizado conforme política institucional de preservação digital (quando houver);
Curadoria	<ul style="list-style-type: none"> • Produtores de dados (auto depósito): Uma vez que os dados publicados não podem ser removidos, utilizam os recursos de Versionamento ou desabilitar acesso aos dados, quando for o caso (<i>deaccession</i>); Os metadados permanecem disponíveis.

Fonte: Elaborado pela autora

5.7 Funcionalidade: Disseminação e Acesso

A disseminação e acesso aos dados é caracterizada pelos processos que proporcionam a busca e navegação pelos dados científicos do repositório. Nesta funcionalidade também pode-se incluir as implementações de mecanismos de segurança e controle de acesso com permissões diferenciadas e configuração da interface de usuário. Em relação ao modelo OAIS, é possível estabelecer que a entidade funcional “Acesso” realiza processos necessários para apoiar os consumidores de informação na solicitação e recepção de dados selecionados.

Quadro 7. Disseminação e Acesso - Categorias e processos

Gestão de Dados Científicos - Categorias	Gestão de Dados Científicos – Processos
Software	<ul style="list-style-type: none"> • Dataverse disponibiliza recurso de busca multifacetada; • Disponibiliza recurso de cadastro prévio de usuários de dados (<i>guestbook</i>); • Dataverse disponibiliza pré-visualização de alguns formatos de dados (texto, tabelas, dentre outros); • Dataverse disponibiliza várias opções de download dos arquivos de dados e metadados;
Infraestrutura do Repositório	<ul style="list-style-type: none"> • Estabelece e informa as licenças de acesso e uso dos dados;
Curadoria	<p>Produtores de dados (auto depósito):</p> <ul style="list-style-type: none"> • Determinam qual a licença de acesso e uso dos dados conforme política do repositório; • Podem habilitar recurso de solicitar cadastro prévio dos usuários que desejam realizar download dos dados; <p>Usuários de dados:</p> <ul style="list-style-type: none"> • Podem pré-visualizar dados em formatos específicos; • Decidem opções de download disponibilizadas; • Realizam buscas utilizando os recursos disponíveis;

Fonte: Elaborado pela autora

6 Conclusão

Entende-se que são muitos os desafios ao se pensar no ecossistema da gestão de dados científicos, especialmente no contexto plural onde são produzidos e coletados. No entanto, parece correto afirmar com clareza que ao se implementar um repositório de dados científicos utilizando-se um software com considerável robustez e abrangência como o Dataverse, tem-se um caminho sendo trilhado em grandes chances de sucesso. O que não se pode deixar de considerar é que somente tecnologia não é suficiente ou garantia de que as recomendações dos padrões e modelos internacionalmente estabelecidos estão sendo seguidos.

Fica evidente neste estudo a fundamental importância das decisões a serem tomadas e seguidas pela equipe gestora do repositório para além das funcionalidades oferecidas pelo software. Tal orientação também se mostra explícita no modelo OAIS.

As entidades funcionais se transbordam para além das tecnologias da informação e comunicação (TIC) exigindo da equipe gestora decisões a serem implementadas, estabelecidas e divulgadas para produtores e consumidores de informação desde o momento de planejamento e preparação dos dados, publicação, armazenamento, acesso e uso.

Embora o modelo OAIS seja uma referência conceitual fora do escopo de orientar a implementação de uma solução tecnológica, o exercício de trazê-lo para uma realidade prática, de um software consideravelmente bem aceito pela comunidade científica, como o Dataverse, parece favorecer seu entendimento, especialmente para os profissionais que estão na fase inicial de implementação de uma solução de gestão de dados científicos.

Buscou-se deslocar de um modelo com alto nível de abstração que demarca a organização estrutural de um sistema de arquivamento para uma realidade de aplicação prática, considerando, além dos recursos tecnológicos, os atores envolvidos nas entidades funcionais.

Finalizando, os sistemas de arquivamento devem ser implementados considerando modelos e padrões que orientem esta tarefa visando atender requisitos mínimos de qualidade. Destaca-se que tais normativas não são excludentes e podem ser utilizadas conjuntamente. O que não se pode perder de vista é a flexibilidade, amplitude e abstração destes modelos que demonstram que as realidades e contextos de produção e geração de dados são distintas e precisam ser preservadas. Estabelecer a dosagem certa destes elementos é o grande desafio na implementação de uma solução tecnológica de arquivamento considerando aspectos de preservação digital.

Referências

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS (ABNT). **NBR15472**.

Sistemas espaciais de dados e informação – modelo de referência para um sistema aberto de arquivamento de informação (SAAI). Rio de Janeiro: ABNT, 2007. 95 p.

DATAVERSE PROJECT. **About Dataverse Project**. Disponível em:

<https://dataverse.org/about> . Acesso: 26 maio 2022.

GALVINO, C. C. T.; ROSA, M. N. B.; OLIVEIRA, B. M. J. F. de. O movimento de Acesso Aberto e a Ciência Aberta: uma proposta de repositório de dados e memória na Universidade Federal de Alagoas. **Ci. Inf. Rev.**, Maceió, v. 7, n. 1, p. 34-45, jan./abr. 2020.

GO-FAIR. **FAIR principles**. Disponível em: <https://www.go-fair.org/fair-principles/> .

Acesso em 23 maio 2022.

MÁRDERO ARELLANO, M. A. Preservação de documentos digitais. **Ciência da Informação**, Brasília, v. 33, n. 2, p. 15-27, 2004. Disponível em:

<https://www.scielo.br/j/ci/a/FLfgJvpH3PZKf3HbpKYchZr/?format=pdf&lang=pt>.

Acesso em 25 maio 2022.

ROCHA, R. P. da *et al.* **Relatório de acesso aberto a dados de pesquisa no**

Brasil: soluções tecnológicas. Porto Alegre, RS: UFRGS, 2018. 75 p. Disponível

em: <https://lume.ufrgs.br/bitstream/handle/10183/185126/001082284.pdf?sequence=1&isAllowed=y> . Acesso em: 26 maio 2022.

SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 15, n. 30, p. 1-31, 2010. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2010v15n30p1/19527>. Acesso em: 24 maio 2022.

WILKINSON, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific data**, v. 3, n. 1, p. 1-9, 2016.