

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO
DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO
CURSO BACHARELADO EM BIBLIOTECONOMIA**

MARIA FERNANDA DE NOVAIS PETRY

**PROVENIÊNCIA NA SUBMISSÃO DE INFORMAÇÃO EM DATAVERSE E
DSpace: ANÁLISE DAS ATIVIDADES, AGENTES E METADADOS**

PORTO ALEGRE-RS

2024

MARIA FERNANDA DE NOVAIS PETRY

**PROVENIÊNCIA NA SUBMISSÃO DE INFORMAÇÃO EM DATAVERSE E
DSpace: ANÁLISE DAS ATIVIDADES, AGENTES E METADADOS**

Trabalho de Conclusão de Curso, apresentado como requisito parcial para obtenção do título de Bacharel em Biblioteconomia, pela Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul – UFRGS.

Orientador: Prof. Dr. Rafael Port da Rocha

PORTO ALEGRE-RS

2024

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Dr. Carlos André Bulhões Mendes

Vice-Reitora: Profa. Dra. Patricia Helena Lucas Pranke

FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO

Diretora: Profa. Dra. Ana Maria Mielniczuk de Moura

Vice-Diretora: Profa. Dra. Vera Regina Schmitz

DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO

Chefe: Prof. Dr. Rene Faustino Gabriel Júnior

Chefe substituta: Profa. Dra. Caterina Marta Groposo Pavão

COMISSÃO DE GRADUAÇÃO DO CURSO DE BIBLIOTECONOMIA

Coordenadora: Profa. Dra. Maria Lúcia Dias

Coordenadora substituta: Profa. Dra. Helen Rose Flores de Flores

CIP - Catalogação na Publicação

de Novais Petry, Maria Fernanda
Proveniência na submissão de informação em
Dataverse e DSpace: análise das atividades, agentes e
metadados / Maria Fernanda de Novais Petry. -- 2024.
78 f.
Orientador: Rafael Port da Rocha.

Trabalho de conclusão de curso (Graduação) --
Universidade Federal do Rio Grande do Sul, Faculdade
de Biblioteconomia e Comunicação, Curso de
Biblioteconomia, Porto Alegre, BR-RS, 2024.

1. Biblioteconomia. 2. Repositório digital. I. da
Rocha, Rafael Port, orient. II. Título.

MARIA FERNANDA DE NOVAIS PETRY

**PROVENIÊNCIA NA SUBMISSÃO DE INFORMAÇÃO EM DATAVERSE E
DSpace: ANÁLISE DAS ATIVIDADES, AGENTES E METADADOS**

Trabalho de Conclusão de Curso, apresentado como requisito parcial para obtenção do título de Bacharel em Biblioteconomia, pela Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul – UFRGS.

Aprovado em: 16 de julho de 2024.

Banca examinadora:

Prof. Dr. Rafael Port da Rocha
Universidade Federal do Rio Grande do Sul (UFRGS)
Orientador

Prof. Dr. Rene Faustino Gabriel Junior
Universidade Federal do Rio Grande do Sul (UFRGS)
Examinador

Lucieli Francini Barni
PPGCIN - Universidade Federal do Rio Grande do Sul (UFRGS)
Examinadora

RESUMO

Pela facilidade na criação e disseminação de informações digitais, se faz presente uma grande necessidade de conhecimento sobre a origem e a integridade das informações publicadas. Este estudo analisa como as plataformas DSpace e Dataverse abordam a proveniência das informações durante o processo de submissão e sua guarda. Assim, por meio de um estudo de caso, esta pesquisa busca comparar a submissão de informações em relação ao controle e registro das informações de proveniência nos processos de submissão dos repositórios Dataverse e DSpace. Os resultados desta pesquisa são apresentados de forma descritiva a fim de promover uma melhor compreensão sobre as plataformas, não buscando ranqueá-las.

Palavras-Chave: Repositório Digital; Autenticidade; Segurança Da Informação; Biblioteconomia.

ABSTRACT

Due to the ease in the creation and dissemination of digital information, there is a great need for knowledge about the origin and integrity of published information. This study analyzes how the DSpace and Dataverse platforms address the provenance of information during the submission process and its safekeeping. Thus, through a case study, this research compares the submission of information, in relation to the control and registration of provenance information in the submission processes of the Dataverse and DSpace repositories. The results of this research are presented in a descriptive way in order to promote a better understanding of the platforms, not seeking to rank them.

Keywords: Digital repository; Authenticity; Information Security; Library Studies.

LISTA DE ILUSTRAÇÕES

Figuras

Figura 1: Modelo OAIS funcional.....	27
Figura 2: Modelo de proveniência W3C.....	29
Figura 3: Organização de um repositório DSpace.....	35
Figura 4: Passos possíveis na construção do processo de submissão do DSpace..	37
Figura 5: Instância de fluxo de submissão no DSpace.....	39
Figura 6: Organização do Dataverse.....	39
Figura 7: Ações possíveis em uma submissão em Dataverse.....	41
Figura 8: Instância de submissão em Dataverse.....	43
Figura 9: Situação de submissões.....	49
Figura 10: Metadados de proveniência DSpace.....	49
Figura 11: Metadados de um Dataset.....	51
Figura 12: Contribuidores no Dataverse.....	51
Figura 13: Notificações de Usuário.....	52
Figura 14: Diferença entre o campo e o metadado Contribuidor.....	57
Figura 15: Ação de publicação do item em DSpace.....	59

Quadros

Quadro 1: Elementos Dublin Core.....	18
Quadro 2: Elementos Data Cite obrigatórios.....	19
Quadro 3: Metodologia aplicada para cada objetivo.....	32
Quadro 4: Papéis do Dataverse.....	42
Quadro 5: Recorte do formulário sobre autenticação e identificação de agentes.....	44
Quadro 6: Recorte do formulário sobre as atividades e agentes no depósito.....	47
Quadro 7: Recorte do formulário sobre representação e registro de objetos.....	51
Quadro 8: Pacote SIP/AIP sobre representação e registro de objetos.....	52
Quadro 9: Metadados de proveniência de atividades em DSpace e Dataverse.....	53
Quadro 10: Recorte do formulário sobre alterações do AIP.....	55

LISTA DE ABREVIATURAS E SIGLAS

AIP	Archival Information Package (Pacote para Arquivamento de Informação ou PAI)
BDJur	Biblioteca Digital Jurídica
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CCSDS	Consultative Committee for Space Data Systems (Comitê Consultor para Sistemas de Dados Espaciais)
CONARQ	Conselho Nacional de Arquivos
DC	Dublin Core
DDI	Data Documentation Initiative (Iniciativa de Documentação de Dados)
DIP	Dissemination Information Package (Pacote de Disseminação de Informação ou PDI)
DOI	Digital Object Identifier (Identificador de Objeto Digital)
DSA	Data Seal of Approval (Selo de aprovação de dados)
DUnAs	Repositório de Dados de Investigação da Universidade de Aveiro
EDUFBA	Editores da Universidade Federal da Bahia
ENAP	Escola Nacional de Administração Pública
HP	Hewlett-Packard
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
IQSS	Institute for Quantitative Social Science (Instituto de Ciências Sociais Quantitativas)
ISO	International Organization for Standardization (Organização Internacional de Standards)

LDAP	Lightweight Directory Access Protocol (Protocolo Leve de Acesso a Diretórios)
MIT	Massachusetts Institute of Technology (Instituto de Tecnologia do Massachusetts)
NISO	National Information Standards Organization (Organização Nacional de Standards de Informação)
OAIS	Open Archival Information System (Sistema Aberto para Arquivamento de Informação ou SAAI)
OCLC	Online Computer Library Center (Centro de biblioteca informática online)
ORCID	Open Researcher and Contributor ID (Sistema Aberto de Identificação de Pesquisador)
PDI	Preservation Description Information (Informação Descritiva de Preservação ou IDP)
PREMIS	Preservation Metadata: Implementation Strategies (Estratégias de Implementação de Metadados de Preservação)
RD	Repositório digital
RLG	Research Libraries Group, Inc. (Grupo de Pesquisa de Bibliotecas)
SIP	Submission Information Package (Pacotes de Submissão de Informação ou PSI)
STJ	Tribunal Superior de Justiça
URI	Identificador Único de Recurso
W3C	World Wide Web Consortium
WDS	World Data System of the International Science Council (Sistema Mundial de Dados do Conselho de Ciência Internacional)

SUMÁRIO

1 INTRODUÇÃO.....	12
1.1 Problema.....	13
1.2 Objetivo.....	13
1.2.1 Objetivos específicos.....	13
1.3 Justificativa.....	14
2 REFERENCIAL TEÓRICO.....	15
2.1 Preservação Digital.....	15
2.2 Metadados.....	16
2.2.1 Esquema de metadados.....	17
2.2.2 Tipos de metadados.....	20
2.3 Repositório Digital.....	21
2.4 Modelo de Referência para Repositório Digital OAIS.....	24
2.5 Proveniência e Autenticidade.....	27
3 METODOLOGIA.....	30
3.1 Método de estudo de caso.....	30
3.2 Coleta de dados.....	31
4 SUBMISSÃO EM DSPACE E DATAVERSE.....	35
4.1 DSpace.....	35
4.1.1 Fluxo de submissão.....	36
4.1.2 Grupos e Autorizações no Fluxo de submissão.....	37
4.1.3 Submissão de um Item.....	38
4.2 Dataverse.....	39
4.2.1 Submissão em Dataverse.....	39
4.2.2 Grupos e Papéis no submissão.....	41
4.2.3 Submissão de um Dataset.....	42
5 A PROVENIÊNCIA NA SUBMISSÃO EM DSPACE E DATAVERSE.....	44
5.1 Autenticação e Identificação de Usuários da Submissão.....	44

5.1.1 Autenticação e Identificação de Usuário no DSpace.....	45
5.1.2 Autenticação e Identificação no Dataverse.....	46
5.1.3 Análise da Autenticação e Identificação de Usuários no Depósito.....	46
5.2 Atividades e Agentes no Depósito.....	47
5.2.1 Atividades e agentes no depósito do DSpace.....	48
5.2.2 Atividades e agentes no depósito do Dataverse.....	50
5.2.3 Análise das atividades e agentes no depósito.....	52
5.3 Representação dos objetos depositados, metadados e informações de proveniência.....	53
5.3.1 Pacotes no DSpace e Dataverse.....	54
5.3.2 Metadados de proveniência em DSpace e Dataverse.....	56
5.3.3 Análise da representação dos objetos digitais.....	58
5.4 Gerenciamento e registro de alterações dos objetos digitais.....	59
5.4.1 Gerenciamento e registro de alterações no DSpace.....	59
5.4.2 Gerenciamento e registro de alterações no Dataverse.....	61
5.4.3 Análise do gerenciamento e registro de alterações.....	61
6 CONCLUSÃO.....	63
REFERÊNCIAS.....	66
APÊNDICE A.....	72

1 INTRODUÇÃO

Nos meios digitais se faz cada vez mais presente a necessidade de segurança e qualidade de informação, por consequência da quantidade de informação que aumenta consistentemente com a facilidade do acesso à internet.

Repositórios digitais são, hoje em dia, uma grande fonte de acesso à produção acadêmica, bem como são grandes difusores da mesma. Ferreira (2006), afirma que, a fim da adoção deste tipo de ferramenta, é fundamental que se tenha simplicidade na criação e disseminação de materiais, bem como é essencial que se tenha qualidade nos resultados de buscas.

O uso de repositórios digitais pode facilitar muito o acesso à informação científica e uma de suas melhores qualidades é a capacidade de permitir que diversos indivíduos contribuam com suas publicações. Como consequência dessa característica, se faz necessário uma avaliação sobre o controle que se tem sobre a inclusão e modificação de documentos em repositórios digitais.

A criação de repositórios digitais requer uma plataforma ou software para estruturar e garantir o acesso ao repositório. Para um repositório digital, uma plataforma que aceita informações de diversos contribuidores, é essencial que se saiba a proveniência das informações depositadas.

As informações de proveniência de um documento podem registrar sua origem, quem possui seus direitos, quem o criou, e as mudanças que foram realizadas no documento (Harvey; Oliver, 2016, tradução nossa). A fim de manter a qualidade e coesão das informações depositadas em um repositório, a proveniência se torna essencial, pois ainda torna necessária a autenticação e a identificação do depositante, o que permite um maior controle sobre a qualidade e confiabilidade das informações depositadas.

Os softwares de repositórios digitais auxiliam no registro e acesso de informações de proveniência de diversas formas, algumas são: a promoção ao acesso às informações de proveniência por meio de sua publicação nas plataformas de repositórios; por meio do controle de acesso a estas publicações e por meio do registro, automático ou manual, de alterações aos dados ou documentos na própria plataforma.

Esta pesquisa é focada em duas plataformas, o DSpace e o Dataverse, que são softwares que possibilitam a criação de repositórios digitais por meio de suas aplicações de acesso aberto.

Embora o Dataverse e o DSpace sejam aplicações similares, possuem propósitos diferentes. Outras diferenças, aquelas presentes no processo de submissão e na guarda das informações de proveniência entre os dois sistemas serão observadas a fim de auxiliar na escolha do repositório que possa se adaptar melhor às suas necessidades e prioridades.

1.1 Problema

De que forma os repositórios digitais DSpace e Dataverse registram a proveniência das informações no momento em que estas são depositadas em um repositório? De que forma Dataverse e DSpace permitem a configuração de processos de submissão controlados, nos quais somente pessoas autorizadas participam? De que forma Dataverse e DSpace registram as informações de proveniência referentes à submissão?

1.2 Objetivo

Compreender a submissão de informações em relação ao controle e registro das informações de proveniência nos fluxos de submissão dos repositórios Dataverse e DSpace.

1.2.1 Objetivos específicos

- A) Analisar a autenticação e as informações de proveniência dos agentes que atuam na submissão como usuários em Dataverse e DSpace;
- B) Identificar e analisar as atividades e os agentes envolvidos na submissão em DSpace e Dataverse;
- C) Analisar a representação dos objetos, metadados e informações de proveniência em Dataverse e Dspace;
- D) Analisar o gerenciamento e o registro de mudanças nos objetos digitais no DSpace e no Dataverse.

1.3 Justificativa

O uso de repositórios digitais atualmente está em constante crescimento, pensando nisso, esta pesquisa busca apontar os pontos fortes e fracos que podem aparecer nessas plataformas, com foco na autenticação de usuários e segurança e qualidade de informações de proveniência.

É necessária uma gama mais ampla de pesquisas sobre temas como o abordado nesta pesquisa, tanto pela constante inovação em termos de tecnologia, quanto para descobrir um método cada vez mais preciso para lidar com a alarmante obsolescência de suportes que prejudicam o acesso a longo prazo aos documentos, com o objetivo de garantir a integridade e qualidade dos documentos depositados nesses repositórios. A ciência da informação está cada dia mais e mais ligada a tecnologias digitais e, com isso, traz a necessidade de profissionais da informação se inteirarem em recursos digitais e eletrônicos que possam melhorar a atuação do bibliotecário, facilitar o acesso aos serviços das unidades de informação e, conseqüentemente, oferecer uma experiência melhor a cada usuário.

Durante a realização desta pesquisa notamos que muitas das pesquisas, publicações científicas e técnicas sobre repositórios digitais são feitas sobre plataformas individuais. Considerando apenas as características de um único sistema, sem gerar comparações com outros repositórios de outros softwares, as necessidades de um usuário podem não ser atendidas. Portanto, a fim de auxiliar na seleção de um repositório que melhor atenda às necessidades de um usuário em potencial, se faz presente a necessidade de tais comparações entre diferentes repositórios.

2 REFERENCIAL TEÓRICO

Neste capítulo serão apresentados os principais aspectos que fornecerão um fundamento teórico para a realização da análise e um breve contexto para explicitar a importância na pesquisa.

2.1 Preservação Digital

Para melhor introduzir o conceito de preservação digital, pode-se começar com o entendimento sobre o que é um documento e o que é um documento digital. Um documento é uma “unidade de registro de informações, qualquer que seja o formato ou o suporte” (Arquivo Nacional, 2005). Ou seja, todo e qualquer registro, seja em livros, sites, pedras, tecidos, etc., pode ser considerado um documento.

Os documentos digitais são mais específicos, são um tipo de documento eletrônico, pois só podem ser acessados por meio eletrônico. Em outras palavras, um documento digital é um “documento codificado em dígitos binários, acessível por meio de sistema computacional” (Arquivo Nacional, 2005). Esses dígitos binários se manifestam como um arquivo que pode ser reproduzido de forma compreensível para um indivíduo.

A preservação digital é um conjunto de ações que leva a implementação de estratégias de preservação, a fim de manter a estabilidade e acessibilidade do documento digital. Ferreira (2006) ainda atribui, mais especificamente, que a preservação digital também deve manter “[...] qualidades de autenticidade suficientes para que possa ser interpretada no futuro recorrendo a uma plataforma tecnológica diferente da utilizada no momento da sua criação.”

A preservação digital envolve diversas técnicas, sendo muitas destas relacionadas aos suportes físicos da informação.

Hoje em dia, a preservação digital enfrenta inúmeros desafios para sua prática, pois

[...] a dinamicidade e a efemeridade dos ambientes digitais impõem perdas rápidas e definitivas de registros importantes disponíveis online, retratando um dos desafios globais do século XXI de garantir a preservação e o acesso contínuo à uma memória pessoal, corporativa e cultural digital. (Formenton; Gracioso, 2020)

Um dos maiores desafios que enfrenta a preservação digital é o desconhecimento sobre as técnicas de preservação (Arellano, 2004). Sant'Anna (2001, *apud* Arellano, 2004) considera as instituições de arquivos responsáveis pela preservação e restauração dos suportes físicos, a fim de manter suas informações disponíveis, podendo ser traduzido para o ambiente de repositórios digitais, de forma que as instituições responsáveis por tais repositórios as responsáveis por manter as informações depositadas acessíveis. Arellano ainda considera que

A preservação digital compreende os mecanismos que permitem o armazenamento em repositórios de dados digitais que garantiriam a perenidade dos seus conteúdos.

As condições básicas à preservação digital seriam, então, a adoção desses métodos e tecnologias que integrariam a preservação física, lógica e intelectual dos objetos digitais. (Arellano, 2004)

Cada técnica de preservação implica um custo diferente e pode demandar mais tempo para sua aplicação, dependendo do tipo de objeto digital e o quão desatualizado é seu suporte, pode ser necessária a aplicação de uma técnica específica para recuperação do objeto, seja apenas a recuperação do acesso ao mesmo até a reconstrução de sua estrutura. Uma das formas mais utilizadas para garantir o acesso à documentos após a obsolescência de seus formatos é a migração de formato (a técnica de migrar arquivos para formatos ou suportes mais novos).

Todas as técnicas de preservação requerem o uso dos metadados de preservação, para registrar, por exemplo, as ações de preservação.

2.2 Metadados

Metadados podem ser descritos como uma “informação descritiva sobre o contexto, qualidade, condições ou características de um recurso, dado ou objeto que tem a finalidade de facilitar sua recuperação, autenticação, avaliação, preservação ou interoperabilidade” (Senso; Piñero, 2003, tradução nossa).

Senso e Piñero (2003) ainda apontam que o processo de catalogação pode ser considerado um processo de geração de metadados. No entanto, os autores também consideram que Caplan (1995 *apud* Senso; Piñero, 2003) utiliza o termo

metadados a fim de não implicar definições e regras de catalogação, que são conhecimentos muito específicos dos profissionais da informação.

Arellano (2004), explica que os metadados também servem para localização do objeto digital. Mais especificamente, como Gilliland-Swetland (1998 *apud* Senso; Piñero, 2003) destaca, o uso dos esquemas de metadados aumenta a chance de um documento ser acessado, ou seja, é mais fácil recuperá-lo. Além disso, metadados possibilitam que uma busca seja feita em múltiplas bases de dados simultaneamente.

Como consequência desta facilitação ao acesso de publicações por meio de metadados, Senso e Piñero (2003) apontam, como característica importante, que o uso dos metadados promove uma maior disseminação da informação. No entanto, também notam que, por meio desses metadados, é possível estabelecer os direitos legais e restrições ao acesso e uso da informação. Além disso, metadados podem ser usados para controlar versões diferentes de objetos de informação.

Outra forma importante em que os metadados podem ser utilizados, é para fins de preservação digital, sendo que podem ser considerados a “chave para garantir que os recursos [digitais] irão sobreviver e continuar acessíveis no futuro” (NISO, 2004, tradução nossa).

2.2.1 Esquema de metadados

Os esquemas de metadados são também chamados de padrões, sistemas de metadados ou vocabulários são utilizados como forma de criar *standards* para os metadados, pois

Os metadados só são úteis se podem ser compreendidos pelos softwares e pelas pessoas que os utilizam. Para ajudar nesse entendimento, as organizações frequentemente predefinem conjuntos de metadados para atender a determinadas necessidades e publicam essas definições para consulta dos profissionais que projetam os sistemas (e, às vezes, dos usuários finais) (NISO, 2004, tradução nossa).

Segundo Lorist e Van der Meer (2001, *apud* Pöttker; Ferneda; Moreiro-González, 2018), o uso de esquemas de metadados aumenta a interoperabilidade e a interdisciplinaridade de uma plataforma, além de serem compreensíveis tanto para usuários quanto para computadores.

O esquema de metadados de uso mais comum na área em Repositórios Digitais de Publicações é o Dublin Core. Este esquema é adaptado para arquivos baseados na Web e foi desenvolvido de forma simples e concisa, tendo como objetivo permitir o seu uso por não catalogadores especializados. Originalmente possuindo 13 campos, que foram, posteriormente, ampliados para 15 (NISO, 2004). Este esquema é considerado simples e de fácil utilização, e a compreensão de seus campos não requer conhecimento especializado de uma área específica, conforme podemos observar os campos descritos na tabela 1. O Dublin Core é o esquema de metadados que faz parte da instalação padrão de DSpace. Dataverse também permite a exportação de metadados em Dublin Core.

Quadro 1: Elementos Dublin Core

Campo DC (Dublin Core)	Descrição
<i>dc.title</i>	Título
<i>dc.subject</i>	Palavras-chave/assunto
<i>dc.description</i>	Descrição geral
<i>dc.type</i>	Tipo de documento
<i>dc.source</i>	Fonte do documento
<i>dc.relation</i>	Documentos relacionados
<i>dc.coverage</i>	Cobertura do documento
<i>dc.creator</i>	Criador do arquivo
<i>dc.publisher</i>	Publicador
<i>dc.contributor</i>	Contribuidor
<i>dc.rights</i>	Direitos autorais
<i>dc.date</i>	Data
<i>dc.format</i>	Formato do arquivo
<i>dc.identifier</i>	Identificador
<i>dc.language</i>	Linguagem do arquivo

Fonte: Autor, 2024

Outro esquema de metadados é o Data Cite, que é relevante no contexto dos repositórios de dados de pesquisa. “O esquema de metadados DataCite é uma lista das principais propriedades de metadados escolhidas para identificação precisa e consistente de um recurso para fins de citação e recuperação” (Data Cite, 2024). Como pode ser visto no quadro 2, que ilustra os metadados obrigatórios do esquema, o termo propriedade é utilizado para descrever os campos de metadados. Este é o padrão disponibilizado pela plataforma Dataverse.

Quadro 2: Elementos Data Cite obrigatórios

ID (identificador)	Propriedade
1	<i>Identifier</i> (Identificador)
2	<i>Creator</i> (Criador)
3	<i>Title</i> (Título)
4	<i>Publisher</i> (Publicador)
5	<i>PublicationYear</i> (Ano de Publicação)
10	<i>ResourceType</i> (Tipo de recurso)

Fonte: Autor, 2024.

O esquema Data Cite, apesar de apresentar poucos elementos obrigatórios, oferece, no total, 20 campos de propriedade, sendo 6 deles obrigatórios (tabela 2) e outros 14 campos opcionais. Algumas propriedades são diferentes dos elementos do Dublin Core, como as propriedades de versão, localização geográfica e financiamento (Data Cite, 2024).

Dublin Core e DataCite são esquema de metadados descritivos, pois visam o acesso e a recuperação. Esquemas de metadados também são usados para auxiliar a preservação dos objetos digitais, como o esquema Preservation Metadata: Implementation Strategies (Estratégias de Implementação de Metadados de Preservação, PREMIS).

O esquema de metadados PREMIS foi criado pelas companhias Online Computer Library Center (Centro de biblioteca informática online, OCLC) e Research Libraries Group, Inc. (Grupo de Pesquisa de Bibliotecas, RLG) entre 2003 e 2005. Este trabalho conjunto das duas organizações resultou no Dicionário de Dados PREMIS, que “define um conjunto principal de elementos de metadados [...] que os

repositórios devem considerar para desempenhar suas funções de preservação.” (Caplan, 2009).

2.2.2 Tipos de metadados

Os metadados podem ser caracterizados como metadados descritivos, que são aqueles que possuem informações básicas necessárias para recuperação e identificação, como autoria, ano de publicação, local de publicação, editora, entre outros dados. Metadados também podem ser classificados como metadados estruturais, que indicam como o documento deve ser organizado; ou como metadados de administração, que possuem informações sobre direitos autorais, tipo de documento, criação do documento. Inserido na classificação de metadado administrativo, se encontram também os metadados de preservação e técnicos (NISO, 2004).

Os metadados descritivos são, segundo Barbedo, Corujo e Sant’Ana (2011), aqueles que têm, como objetivo, a pesquisa, a recuperação e a identificação. Estes metadados se referem a título, resumo, palavras-chave, autores, assuntos, etc. De forma simples, são metadados que descrevem informações importantes para a identificação do objeto digital.

Os metadados estruturais, de acordo com Sayão (2010, *apud* Formenton; Gracioso, 2022), são os que vinculam partes do objeto de forma hierárquica, a fim de apresentar o documento de forma coesa, para que seja apresentado ao leitor. Barbedo, Corujo e Sant’Ana (2011) identificam que alguns dos metadados estruturais podem representar a ordem e a estrutura dos textos, imagens e áudios, presentes no objeto. Formenton e Gracioso (2022) ainda definem estrutura como a “disposição dos atributos e relações entre elementos”, o que pode também definir de forma simples o que são os metadados estruturais.

Sayão (2010) descreve que metadados administrativos “fornecem informações que apoiam os processos de gestão do ciclo de vida dos recursos informacionais”. Esses metadados ainda podem ser classificados em dois tipos: os metadados técnicos e os metadados de preservação. Segundo Alves (2010), tendo como base as definições de Gilliland-Swetland (1999); Rosetto (2003) e Senso e Piñero (2003), o metadado administrativo “[...] fornece informações como: data de

criação dos recursos, tipos de arquivos, formas de acesso, controle de direitos e reproduções, informação sobre registros legais, informação sobre localização etc.”

Metadados técnicos descrevem as características físicas dos documentos, ao invés de aspectos intelectuais do objeto digital (PREMIS, 2015). Esses metadados podem representar “[...] formato do arquivo, resolução, tamanho e outros.” (Rodrigues, 2008).

Tendo com base as definições de metadados de preservação das instituições OCLC e RLG, Arellano (2004) descreve que:

Os metadados de preservação são uma forma especializada de administrar metadados que podem ser usados como um meio de estocar a informação técnica que suporta a preservação dos objetos digitais. Os metadados para preservação visam a apoiar e facilitar a retenção a longo prazo da informação digital (Arellano, 2004).

A preservação digital envolve, como definido anteriormente, uma série de ações que prolongam a estabilidade e o acesso de um objeto digital. Os registros das ações de preservação (como mudanças de formato de arquivo), edições, correções,

Todas as ações realizadas em um objeto digital, como mudança de formato, edições, versionamento e correções, devem ser registradas por meio de metadados de preservação. Ao registrar ações que representam o histórico do objeto, esses metadados descrevem aspectos da proveniência deste objeto. Isso também auxilia como evidência da garantia da autenticidade do mesmo.

Esses metadados são necessários para a tomada de ações que visam manter a integridade e autenticidade dos documentos. Ou seja, “os metadados de preservação podem ser definidos, de uma forma simples e direta, como a informação que apoia e documenta a preservação de longo prazo de materiais digitais” (Sayão, 2010).

2.3 Repositório Digital

O Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT, 2016 *apud* Sobral; Santos, 2017) caracteriza: “Os repositórios digitais (RDs) são bases de dados online que reúnem de maneira organizada a produção científica de uma

instituição ou área temática. Os RDs armazenam arquivos de diversos formatos.”
Silva e Costa (2021) consideram que

Pela sua capacidade de agregação e organização e como instrumento de socialização de comunidades e pesquisadores, os RD têm sua relevância como recurso informacional e se tornam um dispositivo de troca de experiências e compartilhamento de dados, se configurando parte essencial das pesquisas em escala global, portanto, podemos afirmar que este tipo de repositório contribui demasiadamente para colaboração e comunicação entre pesquisadores da comunidade científica. (Silva; Costa, 2021)

Ainda é possível caracterizar os repositórios digitais como

coleções digitais que armazenam, preservam e tornam disponível a produção intelectual de uma ou mais universidades, sem qualquer custo para o produtor e consumidor da informação (Sarmiento et al, 2005, *apud* Camargo e Vidotti, 2008)

Dentro da definição dos repositórios digitais, existem a categoria dos repositórios digitais confiáveis. Por caracterização da OCLC e RLG, em seu documento “*Trusted Digital Repositories: Attributes and Responsibilities*” (Repositórios Digitais Confiáveis: Atributos e Responsabilidades) de 2002, Souza e Aganette (2020) descrevem que “[...] o conceito de repositórios digitais confiáveis inclui em sua definição a preservação digital, a confiabilidade e a autenticidade das informações”. É possível ressaltar que o conceito de confiança

está vinculado à garantia de que as informações armazenadas são íntegras e autênticas, refletem o conteúdo criado pelos autores que depositaram a confiança de manter seus dados em uma base de dados para a recuperação por vários usuários (Souza; Aganette, 2020)

Pelas dificuldades em manter a autenticidade dos objetos digitais, por sua facilidade de distribuição, modificação e duplicação, alcançar o “status” de repositório digital confiável se tornou um desafio (Conselho Nacional de Arquivos, CONARQ, 2012, *apud* Souza; Aganette, 2020). Arellano e Oliveira (2016), descrevem que

Um repositório digital confiável deve seguir: princípios, políticas, fluxos e prevenção. O resultado da observância desses padrões produz não apenas um plano de preservação digital, mas, uma documentação substancial com dados para curadoria, que sustentem um planejamento de preservação digital adequado para a organização (Arellano; Oliveira, 2016)

Tendo como base as definições da OCLC/RLG (2002), a descrição de Arellano e Oliveira (2020) de repositórios digitais confiáveis e a conceituação de confiabilidade de Souza e Aganette (2020), observa-se a necessidade de um modelo referencial para repositórios digitais.

Assim, a necessidade de reconhecer e manter organizações que garantam confiança de documentos digitais, explorada desde final dos anos de 1990, refletiu na constituição do Open Archival Information System(OAIS), modelo de referência do Sistema de Informações de Arquivamento Aberto. (Souza; Aganette, 2020)

Relevante à confiabilidade de repositórios de dados de pesquisa, a Core Trust Seal é uma organização criada em 2017 que definiu requisitos para repositórios confiáveis, a instituição foi uma ação conjunta do Data Seal of Approval (DSA), World Data System of the International Science Council (WDS) (Core Trust Seal, [s.d], tradução nossa). Esses requisitos, se cumpridos, podem render ao repositório um certificado que o identifica como um repositório confiável.

Duas plataformas relevantes tanto para este trabalho quanto para o tópico de repositórios digitais confiáveis são o DSpace e o Dataverse, “[...] por serem as soluções mais utilizadas e adotadas por repositórios que obtiveram certificação de repositório confiável” (Rocha, et al, 2021), segundo o diretório de Repositórios de Dados Re3Data de 2018.

O software DSpace foi uma criação conjunta da empresa Hewlett-Packard (HP) e o Instituto de Tecnologia de Massachusetts (MIT) que foi lançado em 2002 e, segundo o IBICT (Dspace, [s.d]) além de apresentar um modo de dar visibilidade e habilidade de gerenciar e armazenar a produção intelectual, oferece o gerenciamento de diversos formatos digitais, que possibilita a preservação a longo prazo, customização e interoperabilidade com outros softwares. Blattmann e Weber (2008) ainda especificam que

O sistema DSpace possibilita gerenciar os diferentes e complexos recursos no formato digital [...] para coletar, preservar, indexar e distribuir os itens digitais de ambientes acadêmicos (departamentos, laboratórios, centros, escolas, programas). Já em novembro de 2002 mais de 1500 instituições realizaram o download do DSpace. (Blattmann; Weber, 2008).

O Projeto Dataverse é desenvolvido pelo Institute for Quantitative Social Science (IQSS) em colaboração com a biblioteca e a organização de Tecnologia da

Informação da Universidade de Harvard, desde 2006. A plataforma Dataverse é um software aberto que busca automatizar partes do trabalho de um arquivista e prover serviços e creditar os autores dos documentos. O projeto Dataverse foi criado com a intenção de compartilhar, preservar e analisar **dados de pesquisa** (About Dataverse, [s.d.]). Segundo o Repositório de Dados de Investigação da Universidade de Aveiro (DUnAs, 2023, *apud* Silva; Marques, 2023), o sistema ainda possui características como a atribuição automática de URLs persistentes, esquemas de metadados gerais e específicos, a atribuição de diferentes níveis de acesso à arquivos depositados (fechado, restrito, embargado e aberto), o versionamento dos arquivos, contendo todas as mudanças em um histórico acessível e métricas de uso. Silva e Marques (2023) ainda consideram que

Sua utilização é garantia de segurança digital, preservação dos dados científicos, otimização de tempo de futuras pesquisas que necessitem de dados utilizados anteriormente, redução dos custos, produção de metadados precisos, facilidade para a citação dos dados nele depositados etc. (Silva; Marques, 2023)

2.4 Modelo de Referência para Repositório Digital OAIS

O OAIS (Open. Archival Information System)¹ “[...] é um esquema conceitual que disciplina e orienta um sistema de arquivo dedicado à preservação e manutenção do acesso a informações digitais por longo prazo” (CONARQ, [s.d.]). Como mencionado, também “serve como modelo conceitual para interpretação e desenvolvimento de um repositório no contexto da preservação digital” (Pires; Rocha, 2020) que apresenta um ambiente de repositório e simplifica as relações e processos.

O CONARQ também descreve que o objetivo do OAIS é auxiliar na compreensão e conhecimento sobre definições importantes para a guarda de objetos digitais. Santos e Flores (2019) indicam que o desenvolvimento dos conceitos definidos neste modelo referencial ocorreu em fóruns abertos, sendo que o Consultative Committee for Space Data System (Comitê Consultor para Sistemas de Dados Espaciais ou CCSDS) foi responsável por estes fóruns. Arellano (2004) descreve, baseado nos documentos do CCSDS (1998), que um repositório

¹ No Brasil, OAIS é atualmente especificado pela norma NBR ISO 14721:2021, sendo chamado de Sistema Aberto de Arquivamento de Informação (SAAI).

nos termos do OAIS significa considerar os arquivos como organizações de pessoas e sistemas, que aceitaram a responsabilidade de preservar a informação e torná-la disponível e melhorar a comunicação e produtividade entre diferentes comunidades (Arellano, 2004)

O modelo OAIS se consolidou na norma ISO 14721. No Brasil, foi introduzido por meio da norma NBR 15472 de 2007² e substituído pela norma NBR ISO 15471 de 2021, que descreve a organização de um repositório digital que busca a preservação de objetos digitais a longo prazo. As normas definem os processos de preservação que devem estar presentes para que um repositório seja considerado confiável, bem como define os requisitos, diretrizes, estruturas e padrões. Santos e Flores (2015) ressaltam que “[...] o OAIS não é um repositório digital e sim, um modelo conceitual que especifica como deverá funcionar um repositório digital confiável. E por esta razão, o modelo conceitual proposto poderá ser implementado em diferentes contextos tecnológicos.”

O Modelo OAIS especifica o ambiente do repositório, e suas funções e modelo de informação.

O CCSDS (2012) define como ambiente OAIS, o modelo que identifica as funções exercidas externamente a um repositório, sendo estas:

- Produtor: O indivíduo ou entidade que fornece a informação a ser preservada (CCSDS, 2012, tradução nossa);
- Gestão ou Administração: Possui a função de estabelecer políticas institucionais no OAIS, não são a gestão do repositório, mas da instituição a qual pertence (CCSDS, 2012, tradução nossa);
- Consumidor: O indivíduo ou entidade que utiliza a informação preservada (CCSDS, 2012, tradução nossa).

O CCSDS (2012) descreve, através do modelo de informação, que a informação a ser preservada, ao ser submetida por um produtor e ao ser distribuída a um consumidor, passa por “transmissões”, identificando que, durante essas transmissões, a informação do objeto digital está encapsulada em pacotes de informação. Os pacotes de informação são compostos pela informação do conteúdo (informação a ser preservada) e a informação de descrição de preservação (**PDI**³ ou

² A NBR 15472:2007 foi cancelada e substituída pela NBR ISO 14721 em 2021 (ABNT, 2021).

³ Informação Descritiva de Preservação (IDP), traduzindo para o português.

Preservation Description Information), sendo possível o acesso pela informação descritiva.

No modelo funcional OAIS, que identifica as funções a serem realizadas no sistema, existem três tipos de pacotes de informação. O pacote de submissão de informação (**SIP**)⁴ é a informação acordada a ser enviada pelo Produtor ao OAIS, normalmente sendo uma mistura de informações de conteúdo e alguma PDI. O pacote de arquivamento de informação (**AIP**)⁵ é resultado de um ou mais SIP, pode conter outros AIPs e é gerado para fins de preservação, por isso, também possui um PDI completo. Por fim, o pacote de disseminação de informação (**DIP**)⁶ é um AIP completo, parte de um ou múltiplos AIPs que são concedidos para acesso do Consumidor, podendo ou não apresentar seu PDI completo (CCSDS, 2012, tradução nossa).

Estes pacotes de informação são a informação encapsulada que é “transmitida” e gerada nos processos presentes no modelo OAIS. Esses processos, apresentados no modelo funcional do OAIS (figura 1), são chamados de entidades funcionais e são:

A Entidade Funcional **Ingestão** “[...] engloba os serviços necessários para preparar os conteúdos de armazenamento e gerenciamento de dados dentro do Arquivo OAIS, através do Pacote de Submissão de Informação (SIP).” (Flores; Pradebon; Cé, 2017). Algumas funções de ingestão incluem receber SIPs, realizar garantia de qualidade em SIPs e gerar o AIP. (CCSDS, 2012, tradução nossa).

A Entidade Funcional **Armazenamento** realiza o armazenamento, manutenção e recuperação de AIPs. Algumas funções de armazenamento são a realização de verificações de erros, fornecimento de AIPs para acesso e a recuperação e preservação de AIPs (CCSDS, 2012, tradução nossa).

A Entidade Funcional **Gestão de Dados** (ou gestão de metadados) realiza o serviço de preencher, manter e acessar a informação descritiva que identifica e documenta o acervo do repositório e os dados administrativos usados para gerenciar o OAIS (CCSDS, 2012, tradução nossa).

A **Administração** engloba os serviços gerais de operações do sistema, esta entidade, entre suas funções, inclui a criação e negociação de acordos de

⁴ Pacote de Submissão de Informação (PSI), traduzindo para o português.

⁵ Pacote de Arquivamento de Informação (PAI), traduzindo para o português.

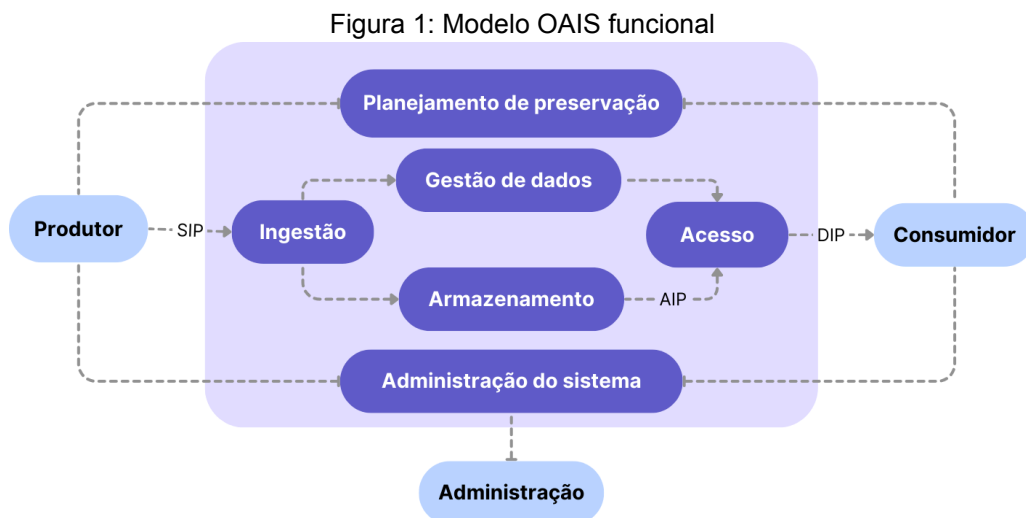
⁶ Pacote de Disseminação de Informação (PDI), traduzindo para o português.

submissão com os Produtores, migração e atualização dos conteúdos do repositório, atender aos pedidos de acesso à AIPs e atendimento ao Consumidor, entre outras (CCSDS, 2012, tradução nossa).

O **Planejamento de Preservação** monitora o ambiente OAIS, criando planos para preservação e dando recomendações a fim de preservar os objetos digitais. Algumas de suas funções são a recomendação de migração ou atualização dos documentos, avaliação periódica do conteúdo do repositório, realização de relatórios periódicos de análise de risco e monitorando mudanças na tecnologia ambiente, entre outras (CCSDS, 2012, tradução nossa).

A Entidade Funcional **Acesso** ajuda os Consumidores a descobrir, requerer e receber as informações depositadas no OAIS. Além de auxiliarem os Consumidores para acessarem a informação, também controlam o acesso à informação protegida.

Na figura 1 é possível identificar, de forma simplificada, os agentes, processos e funções apresentados no modelo funcional do OAIS.



Fonte: Autor, 2024.

2.5 Proveniência e Autenticidade

A autenticidade de um documento pode ser caracterizada como a

credibilidade de um documento enquanto documento, isto é, qualidade de um documento ser o que diz ser e que está livre de adulteração ou qualquer outro tipo de corrupção. A autenticidade é composta de identidade e integridade. (CONARQ, 2020)

A CONARQ (2012) ainda identifica que a autenticidade de um documento é composta por sua identidade e sua integridade. Sendo a identidade “[...] o conjunto dos atributos de um documento arquivístico que o caracterizam como único e o diferenciam de outros documentos arquivístico” (CONARQ, 2012) e a integridade depende da completude do documento e seu estado não corrompido ou alterado sem autorização e documentação (CONARQ, 2020).

A integridade de um documento pode ser verificada por meio do Checksum (soma de verificação), que

é uma sequência de números e letras usada para verificar a integridade dos dados, ou seja, se um arquivo é exatamente o mesmo depois de uma transferência, verificar se não foi alterado por terceiros ou se não está corrompido. (SciELO, 2023)

O processo de submeter um documento para um repositório digital pode ser simples ou complexo, pois envolve acordos entre o produtor, o repositório e a instituição que o detém. A importância da autenticidade neste processo se dá por diversos fatores, como direitos autorais e políticas institucionais para submissão de publicações. Na busca de sanar preocupações com estes tópicos, é necessário que os documentos submetidos pelo depositante sejam autênticos.

A autenticidade de um documento pode ser entendida como o quão confiável são os dados depositados em um repositório (Core Trust Seal, 2019, tradução nossa). Toda e qualquer alteração nos dados, no formato e ações realizadas por qualquer fim no objeto digital, devem ser registrados e apresentados no histórico do documento como informação de proveniência, a fim consolidar a autenticidade do documento.

No modelo OAIS, as informações de proveniência são descritas como o “histórico” do documento, envolvendo a origem da informação e todas as alterações realizadas. O repositório de ambiente OAIS assume a responsabilidade de registrar e guardar as informações de proveniência desde o ponto em que o produtor submete o SIP no momento da Ingestão (CCSDS, 2012).

As informações de proveniência possibilitam a checagem de qualidade e confiabilidade do documento por meio das informações dos indivíduos, instituições ou entidades que estão envolvidas na produção dos dados a serem depositados (World Wide Web Consortium, 2013, tradução nossa).

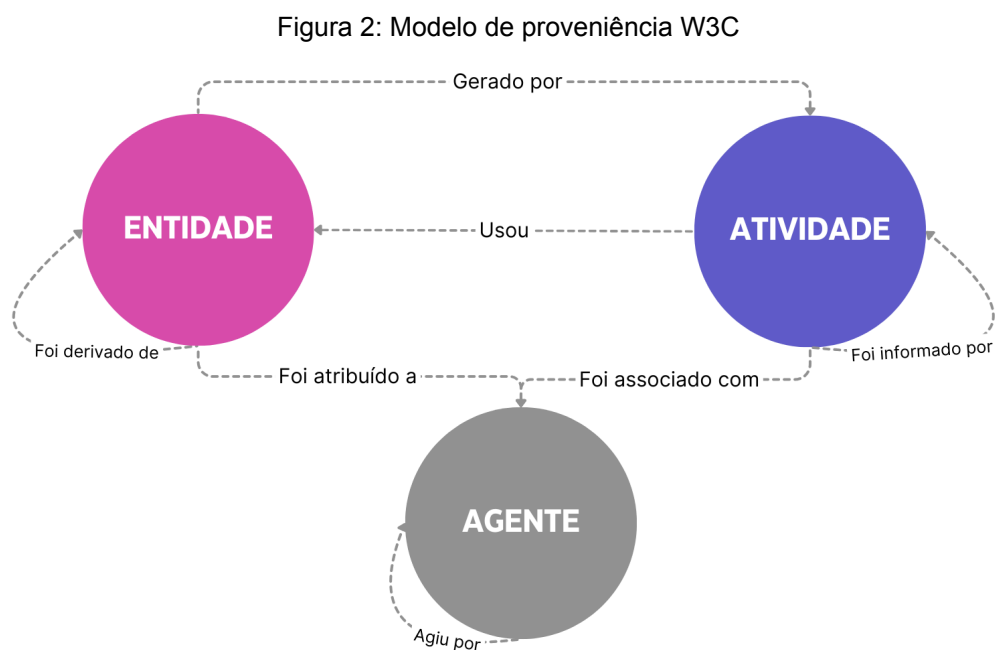
World Wide Web Consortium (W3C) é uma instituição que promoveu relevantes contribuições para definição de requisitos de autenticidade e proveniência. Desenvolveu protocolos e guias para auxiliar a evolução da rede de internet com foco em acessibilidade, segurança, internalização e privacidade. Para atingir esses objetivos, a W3C promoveu um modelo recomendado para dados de proveniência, chamado “PROV-DM: The PROV Data Model” em 2013. (W3C, [s.d], tradução nossa). Neste documento, é definido que “Na sua essência, a proveniência descreve o uso e a produção de entidades por atividades, que podem ser influenciadas de várias maneiras pelos agentes” (W3C, 2013)

O PROV-DM (W3C, 2013) define entidade como a informação ou coisa que deve ser preservada, que se deve obter informações de proveniência sobre, quem sofre as atividades ou é influenciada pelos agentes.

As atividades são descritas como ações que ocorrem a ou em conjunto com a entidade. Algumas atividades são: a geração da entidade; a derivação da entidade; a associação de uma atividade; a atribuição de uma entidade a um agente, etc.

O agente é o responsável, seja pela entidade, por uma atividade realizada sobre a entidade ou sobre outro agente. O agente ainda pode ser uma entidade ou atividade particular. (W3C, 2013, tradução nossa).

As funções descritas e as atividades definidas no documento são ilustradas na figura 2.



Fonte: Autor, 2024.

3 METODOLOGIA

Neste capítulo serão descritos os instrumentos e métodos utilizados a fim de alcançar o objetivo geral com plenitude. No quadro abaixo serão especificadas as etapas realizadas a fim de cumprir cada objetivo específico. Serão utilizadas como fonte para este estudo a documentação técnica dos sistemas e experimentação prática em DSpace e Dataverse.

3.1 Método de estudo de caso

A pesquisa descritiva tem como objetivo descrever fatos e fenômenos de uma realidade e incluem métodos como: “estudos de caso, análise documental, pesquisa ex-post-facto” (Gerhardt; Silveira, 2009).

Robert Edward Stake (2007, *apud* Andrade et al, 2017) identifica que o estudo de caso é uma metodologia que possui foco em aspectos importantes para o problema investigado e que este tipo de estudo é realizado em um tempo limitado, além de também identificar que este método ainda busca apresentar, por meio de uma descrição detalhada, uma visão mais clara de eventos. Andrade et al (2017), com base nas publicações de Robert K. Yin (2010) e Coimbra e Martins (2014), caracteriza que o estudo de caso é “[...] um método com capacidade de produzir evidências com base em distintas técnicas quantitativas e/ou qualitativas de coleta e análise de dados”.

Santa Anna (2018) define que “o estudo de caso constitui um tipo de investigação realizada em um ambiente específico, por meio de técnicas variadas, a fim de coletar dados sobre um fenômeno, processo, objeto ou atividade, considerando o contexto situacional.”

Fazendo uso deste método, um repositório teste será criado nas duas plataformas e o estudo será feito sobre este caso artificial, no qual serão realizados os processos que devem render respostas que objetivam alcançar os objetivos deste estudo. A fim de utilizar este método de forma a tornar mais claro os eventos estudados, foi criado um formulário para coleta de dados com questões relevantes ao cumprimento dos objetivos que foram respondidas de forma qualitativa.

O estudo seguiu o roteiro apresentado no quadro 3, onde é possível observar claramente o objetivo específico a ser cumprido pelo segmento de questões descritas de forma simples.

Quadro 3: Metodologia aplicada para cada objetivo

OBJETIVO	METODOLOGIA
Analisar a autenticação e as informações de proveniência dos agentes da submissão que atuam como usuários em Dataverse e DSpace;	Estudo de Caso: <ul style="list-style-type: none"> Quais são as informações de cadastro requeridas a um usuário e como ele é autenticado e autorizado? De que forma os usuários são autenticados e as ações de cadastro e autenticação são registradas
Identificar e analisar as atividades e agentes envolvidos nas atividades de submissão do DSpace e do Dataverse	Estudo de caso <ul style="list-style-type: none"> Como é realizada e registrada a criação, revisão e edição do SIP e do AIP na submissão de informações?
Analisar a representação dos objetos, metadados e informações de proveniência em Dataverse e Dspace;	Estudo de caso: <ul style="list-style-type: none"> Quais são as informações de proveniência registradas e de que forma estas são representadas em metadados?
Analisar o gerenciamento e registro de mudanças nos objetos digitais no DSpace e no Dataverse.	Estudo de caso: <ul style="list-style-type: none"> Como são gerenciadas e registradas as alterações ao AIP?

Fonte: Autor, 2024.

3.2 Coleta de dados

Para responder ao problema identificado neste trabalho, os dados foram coletados em dois exemplos de repositórios digitais, um em DSpace e outro em Dataverse, os dois utilizando a configuração padrão de instalação. Essa coleta foi orientada por um formulário de coleta. Esse formulário permite a identificação e caracterização das atividades, dos agentes e do gerenciamento da submissão, assim como das representações dos objetos e dos metadados depositados em Dataverse e DSpace, com enfoque nas informações de proveniência. O estudo envolveu as plataformas DSpace e Dataverse em suas configurações padrões, disponibilizadas em suas instalações básicas.

O uso deste formulário para coleta dos dados foi escolhido pois, como Gonçalves e Santos (2023) reiteram, Babbie (1999) define que podem ser

classificados de acordo com a sua finalidade em pesquisas: i) exploratórias, que buscam obter uma perspectiva inicial sobre determinado tema; ii) descritivas, que pretendem descrever o público-alvo da investigação; e iii) explicativas, voltadas a explicar as razões de mudanças em um determinado fenômeno (Gonçalves; Santos, 2023)

Thomas, Oenning e Goulart (2018), apesar de caracterizarem o uso de formulários em estudos da saúde, podem ter sua definição de que a “aplicação de instrumentos, tais como questionários e formulários, é uma maneira eficaz para levantamento de dados populacionais com relativa praticidade logística” traduzida para outras áreas sem grandes diferenças semânticas.

O formulário foi dividido em quatro quadros, cada um destes corresponde à coleta de dados necessária para alcançar cada objetivo específico da pesquisa.

O quadro 1-A, intitulado “Autenticação e informações de proveniência dos agentes da submissão”, busca alcançar o objetivo específico A (Analisar a autenticação e as informações de proveniência dos agentes da submissão que atuam como usuários em Dataverse e DSpace) e desenvolve questões como:

- Quais são as informações de cadastro fornecidas pelos usuários em seus cadastros?
- Como a identidade do usuário é verificada?
- A realização da ação de verificação da identidade do usuário é registrada?
- Mudanças nos dados de cadastro do usuário são verificadas/validadas?
- Mudanças nos dados de cadastro do usuário são registradas?
- De que forma um usuário é autorizado a realizar uma atividade sobre um objeto?

O quadro 2-A, intitulado “Identificação e descrição das atividades e agentes na submissão”, busca alcançar o objetivo B, identificar e analisar as atividades e agentes envolvidos na submissões em DSpace e Dataverse, por meio de questões como:

- Criação do Pacote de Submissão SIP:
 - Como a criação do SIP é realizada?
 - Qual o agente? Somente agentes autorizados podem atuar?
 - Sua realização é registrada?
- Revisão e Edição do Pacote de Submissão (SIP):
 - Como é realizada?
 - Sua realização é registrada?
- Geração do Pacote de Armazenamento de Informação (AIP):
 - Como é realizada?
 - Sua realização é registrada?

O quadro 3-A, intitulado “Descrição da representação dos objetos, metadados e informações de proveniência em Dataverse e Dspace”, possui como finalidade auxiliar no cumprimento do objetivo C, analisar a representação dos objetos, metadados e informações de proveniência em Dataverse e Dspace. Apresenta, de forma menos simplificada, questões como:

- Pacote SIP/AIP:
 - Representa arquivos e pastas?
 - Representa metadados?
 - Representa metadados descritivos?
 - Representa metadados estruturais?
 - Representa metadados administrativos?
 - Descreve nome de arquivo?
 - Descreve tamanho de arquivo?
 - Descreva a data de criação do arquivo?
 - Representa Checksum de arquivo?
 - Descreve o formato do arquivo?
 - Atribui identificador persistente de objeto ao arquivo?
- Metadados de Proveniência:
 - Quais metadados descrevem agentes?
 - Quais metadados que descrevem atividades?
 - Quais metadados que descrevem o artefato/objeto?

Por fim, o Quadro 4-A, com o título de “Gerenciamento e registro de mudanças nos objetos digitais”, busca cumprir o objetivo D, analisar o gerenciamento e registro de mudanças nos objetos digitais no DSpace e no Dataverse. Apresenta questões como:

- Mudanças no AIP
 - Como é realizada?
 - Mudanças são autorizadas?
 - Mudanças são autorizadas em metadados gerados automaticamente?
 - O sistema registra as mudanças?
 - O sistema registra quem realizou as mudanças?
 - O sistema controla versões?

As perguntas presentes nestes quadros possuem respostas qualitativas descritivas ou objetivas e foram preenchidas de acordo com as informações coletadas do estudo de caso criado, informações da documentação técnica das plataformas DSpace e Dataverse e informações técnicas dos documentos extraídos destes repositórios. O formulário preenchido se encontra no Apêndice A (pg. 72).

4 SUBMISSÃO EM DSPACE E DATAVERSE

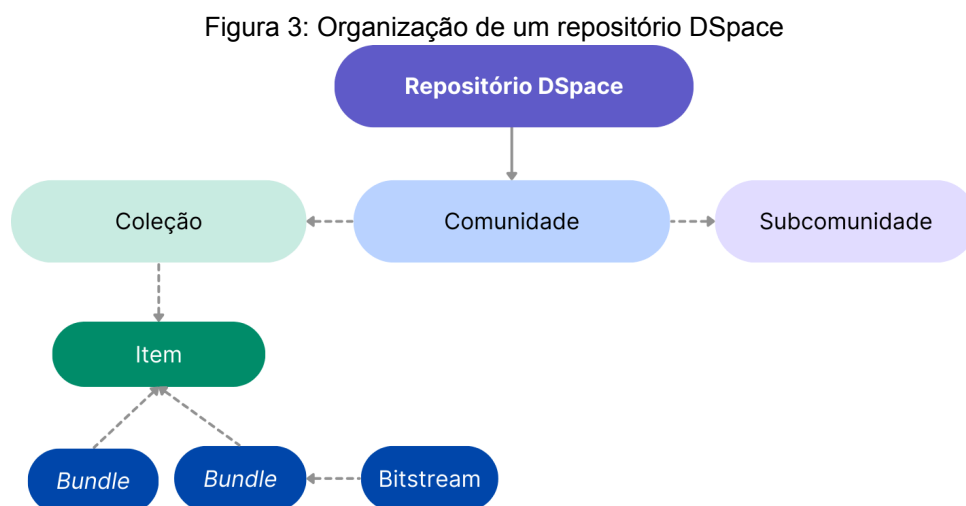
Neste capítulo serão apresentados o ambiente, as atividades e os agentes envolvidos no processo de submissão nos repositórios Dataverse e DSpace.

4.1 DSpace

A Plataforma DSpace possui uma estrutura organizacional que permite a representação e a organização hierárquica de várias coleções. Shintaku e Meirelles descrevem que

A estrutura informacional do DSpace, pelo qual o acervo do repositório é disponibilizado, é hierárquico, composto por Comunidades, Coleções e Itens. Essa estrutura não apenas permite a organização de acervo, mas também, facilita a recuperação dos objetos digitais depositados. Assim, fornecendo uma estrutura que, apesar de rígida, é muito própria para manter os objetos digitais de forma fácil de construir e manter. (Shintaku; Meirelles, 2010)

Em DSpace, os itens depositados são reunidos em coleções. Estas coleções são organizadas em comunidades, e uma comunidade pode ser composta por coleções ou subcomunidades. A figura 3 apresenta de que forma comunidades, subcomunidades e coleções podem ser organizadas e o que compõe um item.



Fonte: Autor, 2024.

DSpace permite a definição de Esquemas de Metadados, sendo que o Esquema Dublin Core é pré-instalado e usado como base. Todos os itens de uma coleção são

descritos por um mesmo conjunto de metadados, passam por um mesmo fluxo de submissão e estão sujeitos às mesmas políticas de submissão, gestão e acesso (Rocha et al., 2018).

4.1.1 Fluxo de submissão

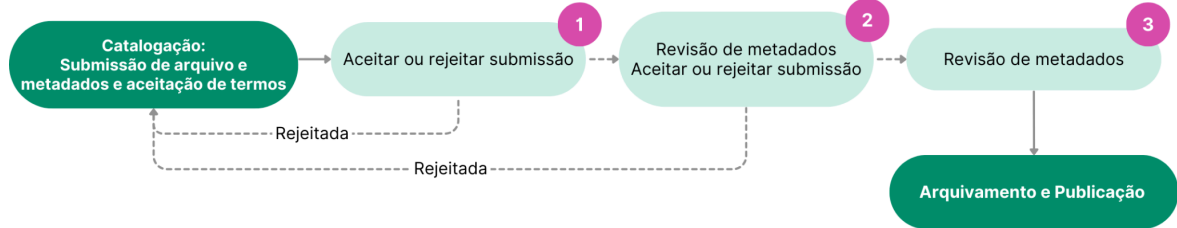
A Plataforma DSpace permite a configuração de um fluxo de submissão para cada coleção. Segundo Shintaku e Meirelles (2010), o fluxo de submissão do DSpace é composto por três atividades:

- **Catologação:** corresponde a realização da submissão pelo usuário, com a carga dos arquivos, preenchimento dos metadados e aceitação do termo de submissão;
- **Avaliação:** nesta atividade do fluxo, são avaliadas a pertinência e a qualidade do documento para o repositório. Nesta etapa, a submissão realizada pelo usuário pode ser rejeitada;
- **Revisão de Metadados:** nesta atividade, os metadados preenchidos pelo usuário são revisados. Essa atividade pode ser configurada para ocorrer de duas formas, no fluxo: com a rejeição, na qual a submissão do usuário pode ser rejeitada por falta de qualidade nos metadados; ou sem a rejeição, na qual a submissão não pode mais ser rejeitada por critérios relacionados aos metadados.

Em caso de rejeição em alguma atividade do fluxo, o processo de submissão é encerrado. Sua continuidade somente será possível através de uma nova submissão.

A figura 4 exemplifica as etapas possíveis para a construção de um fluxo: a Catologação, Avaliação (1), Revisão de Metadados com Possibilidade de Rejeição (2) e Edição de Metadados sem Possibilidade de Rejeição (3). Durante o processo de submissão, o item submetido pode ser rejeitado nas etapas 1 e 2.

Figura 4: Passos possíveis na construção do processo de submissão do DSpace



Fonte: Autor, 2024.

Em um fluxo de submissão, a atividade de catalogação é sempre obrigatória (figura 4). Em um fluxo em que há revisão de metadados com possibilidade de rejeição, a atividade Revisão de Metadados com Rejeição (2) deve ser usada. Já em um fluxo em que há revisão de metadados sem possibilidade de rejeição, a atividade Revisão de Metadados sem Rejeição deve ser usada (3). Apesar de ser possível, a Revisão de Metadados com Rejeição (2) e Revisão de Metadados sem Rejeição (3) não devem ocorrer juntas em um mesmo fluxo.

É recomendado pelo manual do DSpace da Editora da Universidade Federal da Bahia (EDUFBA), a especificação de um fluxo com as etapas de avaliação e revisão de metadados, para repositórios em que haja grande necessidade de controle de conteúdo (Shintaku; Meirelles, 2010).

Após a submissão passar pelas atividades do fluxo de submissão construídas para uma coleção do repositório, a submissão é arquivada e publicada na coleção no repositório.

Sob a ótica do Modelo de Referência OAIS, a atividade do fluxo de Catalogação compreende na criação e na submissão do Pacote de Submissão (SIP) pelo usuário submetedor. As atividades do fluxo de Avaliação e Edição de Metadados fazem parte da Entidade Funcional de Ingestão de OAIS. Sua aprovação compreende na criação e arquivamento do AIP.

4.1.2 Grupos e Autorizações no Fluxo de submissão

Com o objetivo de controlar os agentes que realizam as atividades de Catalogação, Avaliação e Revisão de Metadados do fluxo de submissão, DSpace permite a criação de grupos de usuários e a atribuição de autorizações a esses grupos. Dessa forma somente usuários de grupos autorizados podem realizar atividades

especificadas no fluxo. Ou seja, a execução das atividades de cada etapa do fluxo são dependentes da participação do usuário no grupo que possui a Autorização para realizá-las. É importante mencionar que um usuário pode fazer parte de mais de um grupo. Nesta situação, suas permissões são agregadas.

DSpace também permite que a administração de coleções e seus fluxos sejam realizadas somente por grupos de usuários autorizados. Membros desses grupos podem alterar os metadados da coleção, restringir o acesso à coleção, modificar o fluxo, inclusive autorizações. Entretanto, grupos que são autorizados a administrar coleções não estão autorizados a criar grupos ou gerenciar usuários. Isso é feito por usuários que são membros do grupo de administradores do repositório.

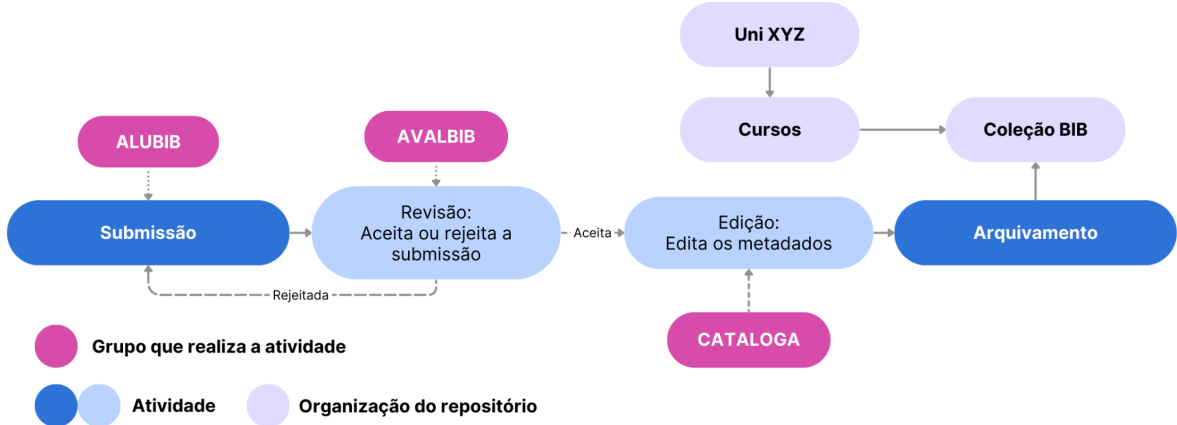
4.1.3 Submissão de um Item

Nessa seção será exemplificada uma instância de submissão de um item a uma coleção de DSpace, objetivando ilustrar as interações entre os agentes envolvidos nas atividades do fluxo.

Para este estudo, foi utilizado um exemplo hipotético da coleção BIB (figura 5). A coleção BIB faz parte da subcomunidade Cursos, que, por sua vez, é parte comunidade da universidade “Uni XYZ”. Esta coleção é composta por trabalhos de graduação de alunos de um curso de biblioteconomia. Nela, os alunos do Grupo do Curso de Biblioteconomia (ALUBIB) podem submeter trabalhos que julgam relevantes. Estas submissões são avaliadas por uma Comissão de Avaliadores representada pelo grupo AVALBIB. Se aceitas, os metadados são revisados pela equipe de catalogadores representada pelo grupo CATALOGA. Nesse fluxo, os catalogadores não possuem a possibilidade de rejeitar a submissão.

Resumidamente, os usuários pertencentes ao grupo ALUBIB estão autorizados a submeter itens à coleção BIB; estes itens serão avaliados pelos usuários do grupo AVALBIB, e a revisão final dos metadados será feita por usuários do grupo CATALOGA. Após passar por estas etapas, o item é arquivado e publicado, como é ilustrado na figura 5.

Figura 5: Instância de fluxo de submissão no DSpace



Fonte: Autor, 2024

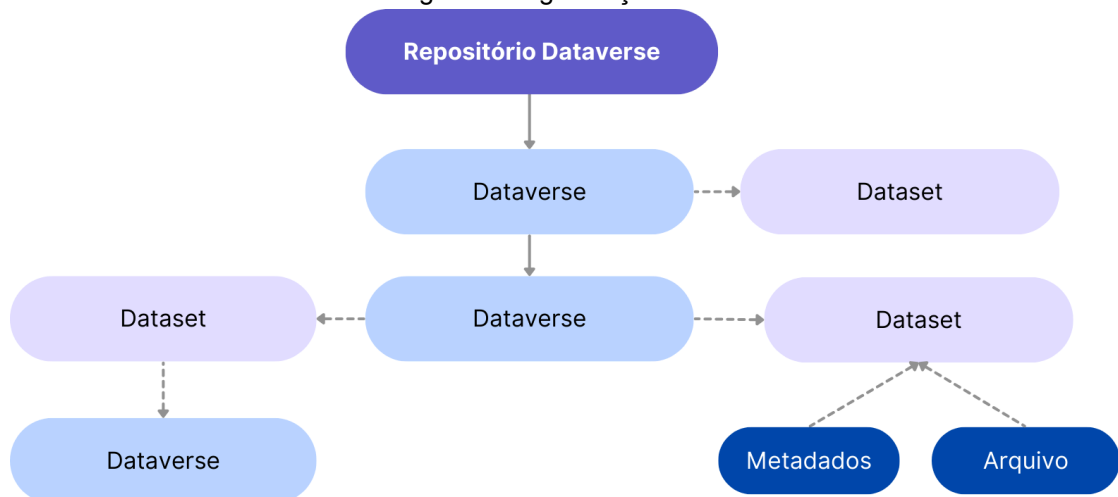
4.2 Dataverse

Rezende (2012) concebe que a plataforma Dataverse é um repositório 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). (Rezende, 2012)

Esta organização composta por dataverses e datasets pode ser observada de forma simplificada na figura 6, bem como a composição de um dataset.

Figura 6: Organização do Dataverse



Fonte: Autor, 2024

4.2.1 Submissão em Dataverse

O Dataverse tem um fluxo de submissão que ocorre por meio de papéis (descritos na seção 4.2.2), onde o papel do usuário que submete um dataset define se poderá publicar seu próprio dataset ou se seguirá o fluxo (Dataverse Project, [s.d.]).

Na plataforma Dataverse, para gerenciar e operar o repositório, é possível realizar sete ações: Adicionar (dataverse ou dataset), Editar (dataverse ou dataset), Ver (dataverse ou dataset), Publicar (dataverse ou dataset), Deletar (dataverse ou dataset), Gerenciar Autorizações (dataverse e dataset) e Baixar (arquivo)

Em relação ao fluxo de submissão de um dataset, é possível identificar quatro, das sete ações: Adicionar, Editar, Publicar e Deletar:

- Adicionar: é a ação que permite criar um dataset, incluindo seus metadados e arquivos;
- Editar: é a atividade que permite editar um dataset;
- Publicar: é a atividade que torna um dataset acessível ao público.
- Deletar: é a ação que permite que um datasets não publicado seja excluído.

Além destas quatro ações, a plataforma Dataverse também apresenta as seguintes ações relacionadas ao dataset:

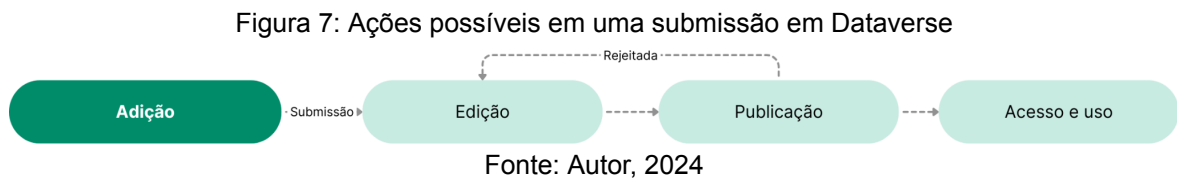
- Ver não publicados: permite datasets não publicados sejam acessíveis;
- Gerenciar grupos, papéis e permissões: é a ação que atribui grupos, papéis e autorizações para a realização das oito ações descritas.

A única ação que é atribuída especificamente a um arquivo presente no dataset é a atividade de *Download*, que permite que este arquivo seja baixado.

O primeiro passo de um processo de submissão sempre ocorre pela realização da ação de Adição do Dataset. Na Adição, o dataset adicionado não será publicado automaticamente. Sua publicação somente poderá ocorrer após a sua edição, que é realizada por meio da ação de Edição do Dataset.

No final da ação de Edição de Dataset, o dataset é enviado para publicação. Na ação de Publicação de Dataset, o dataset é aprovado e publicado. Entretanto, caso não

seja aprovado, o dataset é enviado novamente para edição. A figura 7 apresenta o encadeamento das ações da Plataforma Dataverse para a submissão.



Sob a ótica do OAIS, a ação de Adição de Dataset se compara à submissão do pacote de submissão (SIP) por parte do produtor. A ação de Edição de Dataset compreende na revisão do SIP. A avaliação do SIP, sua aprovação e a geração do pacote de armazenamento (AIP) compreendem na ação de Publicação.

4.2.2 Grupos e Papéis no submissão

Na plataforma Dataverse, ações são agrupadas em papéis. Existem oito papéis previamente definidos: o Criador de Dataverse, o Criador de Dataset, o Criador de Dataverse e Dataset, o Colaborador, o Curador, o Administrador, o Membro e o Downloader (quadro 4):

- Criador de Dataverse: este papel permite realizar a ação de Adição de Dataverse.
- Criador de Dataset: este papel permite realizar ação de Adição de Dataset;
- Criador de Dataverse e Dataset: este papel permite realizar tanto ação de Adição de Dataverses quanto a de Adição de Datasets;
- Colaborador: este papel permite a realização das ações de Ver Dataverses e Datasets não Publicados, Editar Dataset e Fazer o Download de Arquivos;
- Curador: este papel permite realização das ações de ver Dataverses não Publicados e Adicionar Dataverses; pode também realizar ações de controlar permissões, adicionar, editar, publicar, deletar e ver datasets não publicados, além de poder realizar o download de arquivos;
- Administrador: este papel permite a realização de todas as ações possíveis, tanto nos datasets, quanto nos dataverses, além da realização de downloads de arquivos;

- Membro: este papel permite realização das ações de Ver Dataverses e Ver Datasets não Publicados e permite a realização de downloads;
- Downloader: este papel permite a realização da ação de download de arquivos.

Os únicos papéis que permitem a publicação de um dataset pelo próprio criador são os de administrador e de curador. As características de cada papel são sintetizadas no quadro 4. Para este quadro se utilizou a sigla DV, para Dataverse, e DS, para dataset.

Quadro 4: Papéis do Dataverse

AÇÃO EM DATAVERSE (DV)	Administrador	Curador	Colaborador	Criador DV	Criador DS	Criador DV/DS	Membro	Downloader
Permissões	✓							
Adicionar	✓	✓		✓		✓		
Editar	✓							
Publicar	✓							
Ver não publicados	✓	✓	✓				✓	
Deletar não publicados	✓							
AÇÕES EM DATASET (DS)								
Permissões	✓	✓						
Adicionar	✓	✓			✓	✓		
Editar	✓	✓	✓					
Publicar	✓	✓						
Ver não publicados	✓	✓	✓				✓	
Deletar não publicados	✓	✓						
ARQUIVO								
Download	✓	✓	✓	✓			✓	✓

Fonte: Autor, 2024.

Dataverse permite a criação de grupos de usuários e a atribuição de papéis a esses grupos. Os usuários participantes de um grupo possuem a habilidade de realizar as ações permitidas ao papel atribuído ao grupo.

4.2.3 Submissão de um Dataset

Nessa seção será descrita uma instância de submissão de um dataset, objetivando ilustrar as interações entre os agentes envolvidos.

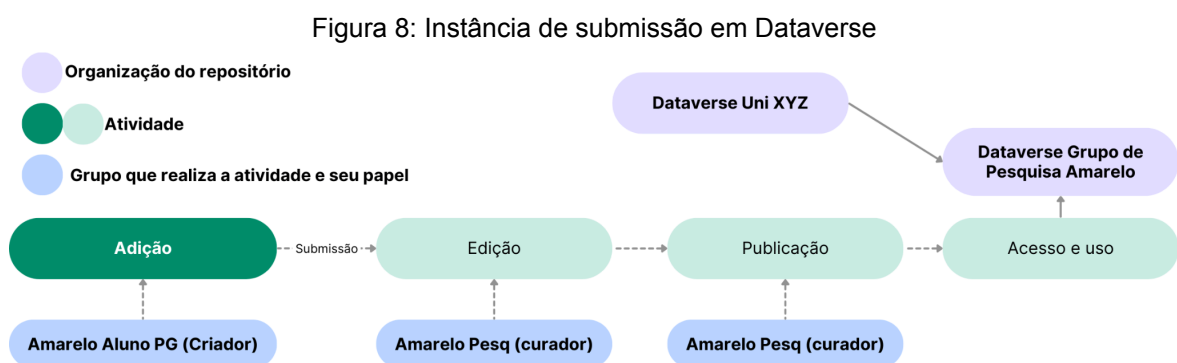
Para este estudo, foi utilizado um exemplo hipotético de repositório de dados: O Grupo de Pesquisa Amarelo. Como o Grupo de Pesquisa Amarelo é parte da Universidade XYZ, o Dataverse do Grupo de Pesquisa Amarelo se encontra dentro do Dataverse “Uni XYZ”. O Dataverse Grupo de Pesquisa Amarelo contém conjuntos de dados de pesquisas realizadas pelo referido grupo de pesquisa.

Os membros do Dataverse Grupo de Pesquisa Amarelo estão organizados em três grupos de usuários: Amarelo Administrador, Amarelo Pesquisador e Amarelo Aluno PG. Os usuários do grupo Amarelo Administrador possuem o papel de administrador do Dataverse. Os usuários grupo Amarelo Pesq possuem o papel de curador e os usuários do grupo Amarelo Aluno PG possuem o papel de colaborador e criador de dataset.

A figura 8 exemplifica uma submissão de um dataset ao Dataverse Grupo de Pesquisa Amarelo. Nesse processo, um usuário do grupo Amarelo Aluno PG adiciona o dataset, este que é composto por arquivos de dados e pelos metadados relacionados. Esta ação cria um rascunho que é submetido ao fluxo de submissão. Então, usuários do grupo Amarelo Pesq, que são curadores, revisam e editam os dados e os metadados do dataset. A seguir, um usuário do grupo Amarelo Pesq submete o Dataset para publicação. Então, usuários do grupo Amarelo Pesq, que são curadores, revisam e editam os dados e os metadados do dataset. A seguir, um usuário do grupo Amarelo Pesq submete o Dataset para publicação.

Membros do grupo Amarelo Pesq, por serem curadores, ainda possuem permissão para publicar dataset que foram submetidos para publicação. Então, um membro desse grupo publica o dataset submetido que, assim, estará livre para acesso e uso e não será mais um rascunho.

O grupo Administrador Amarelo pode realizar todas as etapas da submissão e está envolvido, nesta instância, na atribuição de papéis aos grupos de usuários.



Fonte: Autor, 2024.

5 A PROVENIÊNCIA NA SUBMISSÃO EM DSPACE E DATAVERSE

Neste capítulo serão analisados aspectos de proveniência na submissão no DSpace e no Dataverse, seguindo as questões apresentadas no formulário de coleta (Apêndice A). A análise envolve os seguintes aspectos, descritos na metodologia:

- Atividades e Agentes que atuam na submissão em DSpace e Dataverse;
- Autenticação e Identificação dos Usuários que atuam na submissão em DSpace e Dataverse;
- Metadados e Representação de Objetos depositados por meio da submissão em DSpace e Dataverse;
- Mudança em Objetos Depositados em DSpace e Dataverse.

5.1 Autenticação e Identificação de Usuários da Submissão

Esta seção identifica e analisa as informações, permissões e registros sobre os usuários que atuam nos processo de submissão de DSpace e Dataverse. Está relacionada ao Objetivo A deste trabalho.

A coleta dos dados foi realizada a partir de experimentos de submissões realizadas em Dataverse e DSpace, e utilizou o “Quadro 1-A: Autenticação e Informações de Proveniência dos Agentes da Submissão”, descrito no apêndice A. O quadro 5 apresenta de forma simplificada um excerto do instrumento de coleta.

Quadro 5: Recorte do formulário sobre autenticação e identificação de agentes da submissão

Agentes que são Usuários	
Quais são as informações fornecidas pelos usuários em seus cadastros?	A1
Como a identidade do usuário é verificada?	A2
A realização da ação de verificação da identidade do usuário é registrada?	A3

Mudanças nos dados de cadastro do usuário são verificadas/validadas?	A4
Mudanças nos dados de cadastro do usuário são registradas?	A5
De que forma um usuário é autorizado a realizar uma atividade sobre um objeto?	A6

Fonte: Autor, 2024.

5.1.1 Autenticação e Identificação de Usuário no DSpace

Respondendo aos questionamentos apresentados no quadro 5, identificou-se que, no sistema do DSpace, o cadastro de um usuário (Questão do Quadro 5 - A1) se dá com o fornecimento de informações básicas. DSpace requer do usuário, obrigatoriamente, seu nome, sobrenome e e-mail, como forma de contato. Opcionalmente, é possível realizar um cadastro de usuário com informações de identificação do Sistema Aberto de Identificação de Pesquisador ORCID (Open Researcher and Contributor ID). Também é possível a configuração de telefone para contato, como informação obrigatória ou opcional.

A verificação da identidade do usuário⁷ (Quadro 5 - A2) é feita após o cadastro. O e-mail do usuário é verificado por meio do envio de uma mensagem, na qual o mesmo deve confirmar o endereço. A realização da ação de verificação de identidade do usuário não é registrada no repositório (Quadro 5 - A3).

Quando se trata de modificações de dados pessoais pelos próprios usuários, em DSpace é permitido modificar nome, sobrenome, senha do *login* e o telefone de contato. Não é possível alterar o e-mail cadastrado. Nenhuma modificação possível requer verificação de identidade do usuário, além da confirmação por senha (Quadro 5 - A4), e essa ação de modificação não é registrada (Quadro 5 - A5).

Para que o usuário possa realizar operações no repositório, como editar, revisar ou administrar uma coleção, subcomunidade ou comunidade, o usuário necessita ser incluído em um grupo com autorização para tal atividade. Essa inclusão somente pode ser feita por um usuário do grupo de administradores (Quadro 5 - A6).

⁷ O estudo limita-se à configuração padrão de DSpace, como mencionado na Metodologia.

5.1.2 Autenticação e Identificação no Dataverse

Na plataforma Dataverse, o cadastro de um usuário se dá com informações básicas. Dataverse requer do usuário, obrigatoriamente: nome, sobrenome, nome de usuário e e-mail, como forma de contato. Opcionalmente, é possível realizar o cadastro com outras informações de identificação como vínculo institucional, cargo ou ORCID e cadastrar um telefone para contato (Quadro 5 - A1).

A respeito da verificação da identidade do usuário (Quadro 5 - A2) identificou-se que, após o cadastro, a confirmação do e-mail é realizada pelo envio de uma mensagem ao endereço cadastrado. A ação de verificação não é registrada no perfil do usuário (Quadro 5 - A3).

É possível, no entanto, utilizar aplicações externas de gerenciamento de identidade ou diretórios, como o Lightweight Directory Access Protocol (LDAP), Shibboleth e OAuth 2.0 (Autenticação Aberta) para autenticar usuários de forma mais acurada e rígida. Neste trabalho, apenas a instalação padrão, que não faz uso de aplicações externas, foi utilizada.

Referente a mudanças de dados pessoais no Dataverse (Quadro 5 - A4), é permitido modificar o nome, sobrenome, contato e e-mail. Se o e-mail é alterado, uma mensagem de verificação é enviada ao novo endereço. O único dado que não é possível modificar é o nome de usuário definido no cadastro. Essa ação de modificação não é registrada no perfil (Quadro 5 - A5).

A fim de permitir ao usuário que realize alguma ação dentro do repositório, que não seja apenas visualizar as publicações, o usuário deve ser incluído em um grupo que possua um papel. Essa inclusão pode ser feita por um administrador do repositório ou por um curador de um dataset (Quadro 5 - A6).

5.1.3 Análise da Autenticação e Identificação de Usuários no Depósito

Comparando o DSpace e o Dataverse no tópico de autenticação e identificação de usuários, em suas configurações básicas, é possível observar claramente que são processos simples, focados na verificação de seus e-mails.

A distinção que se faz sobre o cadastro é a maior diversidade de opções de identificação do Dataverse, sendo que o uso de um vínculo corporativo como identificação pode ser considerado uma autenticação mais confiável, por se tratar de

um vínculo externo oficial com a instituição responsável pelo repositório. A identificação pelo ID ORCID é outra opção mais confiável do que uma autenticação por e-mail, e que se encontra disponível tanto no Dataverse quanto no DSpace.

Outro ponto a ser considerado no Dataverse é que, apesar de requerer a verificação do e-mail, é possível acessar o repositório sem verificá-lo.

Se tratando das modificações de dados pessoais, os dois softwares as tratam de forma igual, sem registrar estas mudanças ou validá-las. No entanto, cada plataforma possui um campo de identificação não alterável (nome de usuário no Dataverse e e-mail no DSpace), o que pode oferecer alguma consistência no contato e proveniência das submissões dos usuários.

Seria uma opção interessante a validação das mudanças em informações de contato. Ou a necessidade de uma validação ou autorização de um administrador para mudar estes dados, bem como os dados de identificação, como nome e sobrenome.

No entanto, se houvesse apenas um registro permanente destas modificações, já renderiam mais credibilidade à proveniência das informações que são depositadas sob a responsabilidade do repositório.

5.2 Atividades e Agentes no Depósito

Esta seção identifica os recursos das plataformas DSpace e Dataverse para gerenciar as atividades de criação, revisão e edição do Pacote de Submissão de Informação (SIP), geração de Pacote de Armazenamento de Informação (AIP), bem como os agentes envolvidos nesses processos. O segmento de questões relacionado ao “Quadro 2-A: Identificação e descrição das atividades e agentes nas atividades” do formulário de coleta de dados do apêndice A busca alcançar o objetivo B deste projeto.

Um fragmento do instrumento de coleta foi reproduzido no quadro 6 a fim de melhor ilustrar as respostas que foram buscadas.

Quadro 6: Recorte do formulário sobre as atividades e agentes no depósito

Criação do Pacote de Submissão (SIP)	
Como a criação do SIP é realizada?	D1

Qual o agente? Somente agentes autorizados podem atuar?	D2
Sua realização é registrada?	D3
Revisão e Edição do Pacote de Submissão (SIP)	
Como é realizada?	D4
Sua realização é registrada?	D5
Geração do Pacote de Armazenamento (AIP)	
Como é realizada?	D6
Sua realização é registrada?	D7

Fonte: Autor, 2024

5.2.1 Atividades e agentes no depósito do DSpace

Na plataforma DSpace, a criação do SIP (Quadro 6 - D1), identificada como catalogação, se dá por meio da submissão de um item, que inclui o depósito de um arquivo e o registro de metadados por parte do depositante. O submetedor depositante ou, em termos do OAIS, produtor, deve ser um usuário cadastrado e que esteja incluído em um grupo que possua autorização para submeter publicações (Quadro 6 - D2), a autorização de um usuário ocorre por meio de sua inclusão em um grupo autorizado por um administrador, conforme foi descrito na seção 4.1.2 e 4.1.3 . Após a catalogação, o item é submetido às etapas seguintes do fluxo de submissão configurado, onde este SIP será revisado (Quadro 6 - D4), editado, publicado ou apenas publicado (Quadro 6 - D6).

DSpace permite que o depositante acompanhe sua submissão até o momento de publicação da mesma (Quadro 6 - D3, D5 e D7), conforme apresentado na figura 9. Nessa figura, a tela de DSpace relata as submissões de um usuário depositante, em que são apresentadas as submissões já publicadas (arquivadas) e uma submissão que está em fase de revisão. Entretanto, o depositante não tem acesso à informação sobre quem está realizando cada etapa. Após a conclusão de uma etapa de um fluxo de submissão, o registro das etapas anteriores do fluxo não está mais disponível ao depositante.

Figura 9: Situação de submissões

[Página inicial](#) → [Submissões](#)

Submissões e Tarefas do fluxo de submissões

Submissões

Você pode [Iniciar nova submissão](#)

O processo de submissão inclui a descrição de um item e realizar o upload do(s) arquivo(s) que o compõe. Cada comunidade ou coleção pode definir a sua própria política para submissão.

Submissões arquivadas

Estas são as submissões completas que foram aceitas para o DSpace.

Data de aceitação	Título	Coleção
2023-11-27	Alubib: Submission	BIB
2023-08-11	Prática em DSPACE Submetendo Item	BIB

Submissões para serem revisadas

Estas são as suas submissões concluídas que estão sendo analisadas pelos revisores da coleção.

Título	Coleção	Status
Submission Alubib	BIB	Submissão sendo revisada

Fonte: Print de repositório DSpace, 2024.

Após a submissão passar por todas as etapas do fluxo, o AIP é gerado e publicado. As etapas da submissão e os agentes envolvidos são registrados nos metadados de proveniência do item (metadado: *dc.description.provenance*), que são acessíveis por um administrador (Quadro 6 - D3, D5 e D7). Na figura 10 são apresentados os metadados que registram as ações de catalogação (*submission*) e revisão (*approved*) de um item. Esses metadados incluem informações como: o usuário, seu email, data, hora, os nomes dos arquivos e seus tamanho e checksums.

Figura 10: Metadados de proveniência DSpace

<input type="checkbox"/>	dc.description.provenance	Submitted by usu 01 (usu01@ufxz.br) on 2023-12-14T13:01:45Z No. of bitstreams: 1 DocumentoDigital - Digitalização - Digitalização de um Acervo.2023.pdf: 1327081 bytes, checksum: 637b0e6b25707be878b33d9f64b9ff7c (MD5)	en
<input type="checkbox"/>	dc.description.provenance	Approved for entry into archive by Ava 01 (av01@ufxz.br) on 2023-12-14T13:07:46Z (GMT) No. of bitstreams: 1 DocumentoDigital - Digitalização - Digitalização de um Acervo.2023.pdf: 1327081 bytes, checksum: 637b0e6b25707be878b33d9f64b9ff7c (MD5)	en

Fonte: Print DSpace, 2024.

5.2.2 Atividades e agentes no depósito do Dataverse

Para a criação do Pacote de Submissão (SIP) do Dataverse, o usuário cria um conjunto de dados, chamado de dataset, onde adiciona os arquivos a serem depositados, e registra os metadados relacionados. Após registrar os metadados e incluir os arquivos a serem depositados (Quadro 6 - D1), o dataset é editado (revisado) e submetido para publicação (Quadro 6 - D4 - D6).

No que se refere às autorizações (Quadro 6 - D2), somente agentes autorizados podem realizar nessas ações. Essa autorização ocorre por meio de grupos e papéis, conforme apresentado nas seções 4.2.2 e 4.2.3.

No que se refere ao registro da realização da ação da criação do SIP (Quadro 6 - D3), esta é registrada no próprio dataset, por meio dos metadados que identificam a Data de Criação. Em Dataverse, a Data de Criação é automaticamente registrada quando o dataset é criado, no entanto, é possível modificar esta data. Dataverse também preenche automaticamente o metadado Depositante (Quadro 6 - D3).

O metadado *Depositante* (Depositor) identifica quem depositou o conjunto de dados no repositório (Quadro 6 - D2). Na adição de um Dataset, o ambiente preenche automaticamente o metadado *Depositante* com os dados do cadastro do usuário que realizou a adição. Além do *Depositante*, o ambiente também preenche automaticamente os metadados de contato e autoria. Os valores desses metadados podem ser modificados. Na figura 11, os campos *Author*, *Contact* e *Depositor* foram preenchidos automaticamente com as informações do usuário que adicionou o Dataset: Paulo. O campo *Deposit Date* foi preenchido com a data de depósito. *Publication date* foi preenchido com a data de publicação.

Figura 11: Metadados de um Dataset

Citation Metadata	
Dataset Persistent ID	doi:10.80102/FK2/XSCSUM
Publication Date	2024-06-04
Title	Estudo 3 Amarelo
Author	R, Paulo (Uni Orange)
Contact	Use email button above to contact. R, Paulo (Uni Orange)
Description	Estudo 3 Amarelo..
Subject	Agricultural Sciences
Depositor	R, Paulo
Deposit Date	2024-06-04

Fonte: Print Dataverse, 2024

O campo *Contribuidor* (Contributors) registra os usuários envolvidos na criação, edição e publicação do dataset (Quadro 6 - D3, D5 e D7). O registro é feito automaticamente, mas não é possível modificá-lo. Essas informações não são registradas em metadados e não constam na exportação dos metadados para os padrões DDI e DataCite. O registro dos contribuidores é apresentado na aba *versões* (Versions). No exemplo da figura 12, observamos (no campo *Contribuidores*) as três pessoas envolvidas na criação, edição e publicação do Dataset V1: Paulo, Júlia e Maria. Entretanto, não é possível distinguir quem criou (Paulo), quem editou (Júlia) e quem publicou (Maria).

Figura 12: Contribuidores no Dataverse

Conórcio Cores > Repositório Laranja > Grupo de Pesquisa Amarelo >

Estudo 3 Amarelo

Version 1.0

R, Paulo, 2024, "Estudo 3 Amarelo", <https://doi.org/10.80102/FK2/XSCSUM>, Conórcio Cores, V1, UNF:6:KayLKAZ1mOF2pYUrvF9bDA::= [fileUNF]

Cite Dataset - Learn about Data Citation Standards

Access Dataset -
Edit Dataset -
Link Dataset
Contact Owner Share

Dataset Metrics
0 Downloads

Description Estudo 3 Amarelo.
Subject Agricultural Sciences

Files Metadata Terms Versions

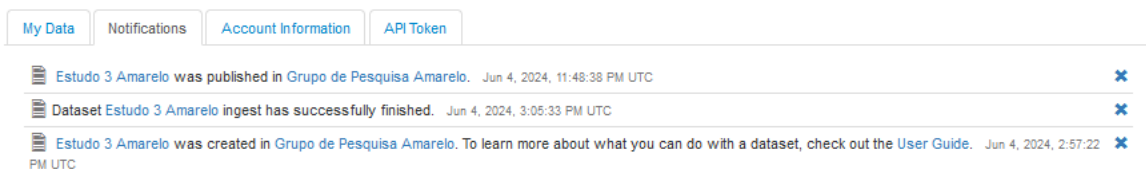
Dataset	Summary	Contributors	Published
1.0	This is the first published version.	Paulo R, Júlia M, Maria F	Jun 4, 2024

Fonte: Print Dataverse, 2024

Sobre a revisão e edição do SIP (Quadro 6 - D4), identifica-se que estas atividades devem ser realizadas por um usuário que possui o papel de colaborador (somente edição), curador ou administrador. Estas atividades não são registradas (Quadro 6 - D5) em metadados, no entanto, os usuários que realizaram a revisão e a edição são identificados em *Contribuidores* (na aba Versões do Dataset).

Em Dataverse, um usuário é notificado quando ações foram realizadas em Datasets aos quais está autorizado a atuar. Essas notificações são registros de ações, mas podem ser apagadas pelo usuário e não informam os agentes. A figura 13 apresenta notificações do usuário Pedro, que adicionou o dataset apresentado na figura 12 .

Figura 13: Notificações de Usuário



Fonte: Print Dataverse, 2024.

A geração do AIP (Quadro 6 - D6) ocorre em conjunto com a ação de publicação do dataset, sendo a atividade de geração do AIP registrada (Quadro 6 - D7) pelos metadados Data de Distribuição e Data da Versão em em Contributor.

5.2.3 Análise das atividades e agentes no depósito

Comparando os registros, atividades e agentes envolvidos no depósito das em Dataverse e DSpace, é possível observar semelhanças e diferenças no registro de dados envolvidos nas atividades e agentes no depósito. As duas plataformas possuem um controle sobre os usuários autorizados a agir sobre qualquer depósito. Em DSpace, somente usuários autenticados e autorizados podem catalogar, aprovar e editar metadados de um item. Similarmente, em Dataverse, somente usuários autenticados e autorizados podem adicionar, editar e publicar um dataset.

Sobre adição, edição e publicação do SIP, o registro permanente dos envolvidos é mais detalhado no DSpace, por meio do metadado

dc.description.provenance, como se pôde observar na figura 10. Esse registro somente pode ser acessado por administradores, mas pode ser alterado.

Em Dataverse, as pessoas envolvidas nas ações de criar, editar ou publicar um dataset são registradas como *Colaborador*, este campo pode ser visto na aba *Versões* de um dataset. No entanto, esse registro não é representado na forma de metadado e não inclui o tipo de ação realizada por cada *Colaborador*.

Outra diferença que se observa é com relação a alterações realizadas e versionamento. DSpace permite que usuários autorizados modifiquem documentos ou metadados de um item e não exige que uma alteração seja sujeita a um novo fluxo de aprovação e revisão.

Já em Dataverse, ações de edição em dataset exigem uma ação de publicação. Esta publicação realiza o versionamento automático do dataset. Em dataverse, qualquer alteração em um dataset só é efetivada através da publicação de uma nova versão, com o registro dos atores e das mudanças no dataset (arquivos, metadados), e com a preservação das versões anteriores.

Para fins de proveniência, seriam opções interessantes que o Dataverse disponibilizasse a identificação das ações realizadas por cada usuário que atuou na submissão de um dataset. Para o DSpace 5 seria útil que os dados de proveniência fossem acessíveis para consulta de outros usuários que não sejam administradores e que não fosse permitido editar os metadados de *dc.description.provenance*, que descrevem as ações realizadas.

5.3 Representação dos objetos depositados, metadados e informações de proveniência

Nesta seção será descrita a representação dos objetos depositados nos repositórios e seus metadados, na perspectiva da proveniência e autenticidade. Observa tanto a capacidade de representar informações em metadados (descritivos, estruturais e administrativos), quanto o registro de informações relacionadas aos arquivos, como nome, datas, checksum, formato e identificador.

O segmento de questões apresentadas no quadro 7 está relacionado ao “Quadro 3-A: Descrição da representação dos objetos, metadados e informações de proveniência em Dataverse e DSpace”, do apêndice A. Esse segmento de questões

busca concluir o objetivo C deste projeto, e o excerto de questões reproduzidas no quadro abaixo buscam contextualizar as respostas descritas a seguir.

Quadro 7: Recorte do formulário sobre representação e registro de objetos

Pacote SIP/AIP	
Representa arquivos e pastas?	R1
Representa metadados?	R2
Representa metadados descritivos?	R3
Representa metadados estruturais?	R4
Representa metadados administrativos?	R5
Descreve nome de arquivo?	R6
Descreve tamanho de arquivo?	R7
Descreva a data de criação do arquivo?	R8
Representa Checksum de arquivo?	R9
Descreve o formato do arquivo?	R10
Atribui identificador persistente de objeto ao arquivo?	R11
Metadados de Proveniência	
Quais metadados descrevem agentes?	R12
Quais metadados que descrevem atividades?	R13
Quais metadados que descrevem o artefato/objeto?	R14

Fonte: Autor, 2024.

5.3.1 Pacotes no DSpace e Dataverse

Conforme os questionamentos sobre os pacotes de submissão (SIP) e armazenamento (AIP) do quadro 7, podemos afirmar que tanto DSpace quanto Dataverse permitem que objetos armazenados sejam constituídos por arquivos, e que esses arquivos possam ser organizados em pastas. Isso viabiliza o armazenamento de objetos complexos, compostos por vários arquivos em pastas. (Quadro 7 - R1).

Ambas plataformas permitem também a representação de metadados descritivos (Quadro 7 - R2 e R3). Análises mais detalhadas sobre os metadados são apresentadas na seção 5.3.2.

Com relação a um arquivo armazenado, ambas plataformas descrevem o seu nome do arquivo (Quadro 7 - R6), seu tamanho (Quadro 7 - R7) e o seu formato (Quadro 7 - R10). As duas plataformas descrevem o Checksum do arquivo (Quadro 7 - R9) e a data de criação do mesmo (Quadro 7 - R8). O checksum permite conferir se o arquivo armazenado não foi adulterado. Um ponto a ser comentado sobre a data de criação é que o Dataverse apenas a registra, não a representa como um metadado.

Se tratando de metadados estruturais (Quadro 7 - R4), em Dataverse o padrão de metadados DDI (Data Documentation Initiative) permite a descrição da estrutura de um dataset, que é composta por arquivos. Em Dataverse, metadados exportados em DDI contém descrições dos arquivos que compõem um estudo. DSpace permite o armazenamento de arquivos em pastas e estes arquivos são descritos por metadados, conforme já observado. DSpace não exporta metadados que contém descrição da estrutura de um item, isto é, sua composição por meio de pastas e arquivos.

Sobre os metadados administrativos (Quadro 7 - R5), o Dataverse registra automaticamente os agentes que realizaram qualquer atividade no processo de submissão do dataset (no campo *Colaborador*), bem como as informações de datas de publicação e datas de versionamento, conforme já destacado na seção 5.2.2 ao abordar atividades e agentes da submissão. Esse registro automático ainda identifica quem criou o dataset (no metadado *Depositor*). No entanto, como já mencionado, no campo *Colaborador*, nenhuma distinção é feita entre para identificar se o usuário atuou como revisor, publicador ou editor.

O DSpace registra o publicador e a data de publicação e ainda especifica, no metadado de proveniência (*dc.provenance*), quando cada atividade relacionada a submissão foi realizada e quem realizou a avaliação e submissão do item (Quadro 7 - R5), conforme já destacado na seção 5.2.2, que aborda atividades e agentes da submissão

Em se tratando de identificadores persistentes (Quadro 7 - R11), o Dataverse atribui um identificador persistente para o dataset (como DOI) e também possibilita a configuração de identificador de objeto persistente para o arquivo (opcional). Já o DSpace apenas atribui identificador persistente para o item. O quadro 8 sintetiza Dataverse e DSpace quanto à representação de objetos (pacotes).

Quadro 8: Pacote SIP/AIP sobre representação e registro de objetos

Pacote SIP/AIP	DSpace	Dataverse
Representa arquivos e pastas?	Sim	Sim
Representa metadados?	Sim	Sim
Representa metadados descritivos?	Sim	Sim
Representa metadados estruturais?	Não	Sim
Representa metadados administrativos?	Parcialmente	Parcialmente
Descreve nome de arquivo?	Sim	Sim
Descreve tamanho de arquivo?	Sim	Sim
Descreva a data de Criação de arquivo?	Sim	Sim
Representa Checksum de arquivo?	Sim	Sim
Descreve o formato do arquivo?	Sim	Sim
Atribui identificador persistente de objeto ao arquivo?	Não	Sim

Fonte: Autor, 2024.

5.3.2 Metadados de proveniência em DSpace e Dataverse

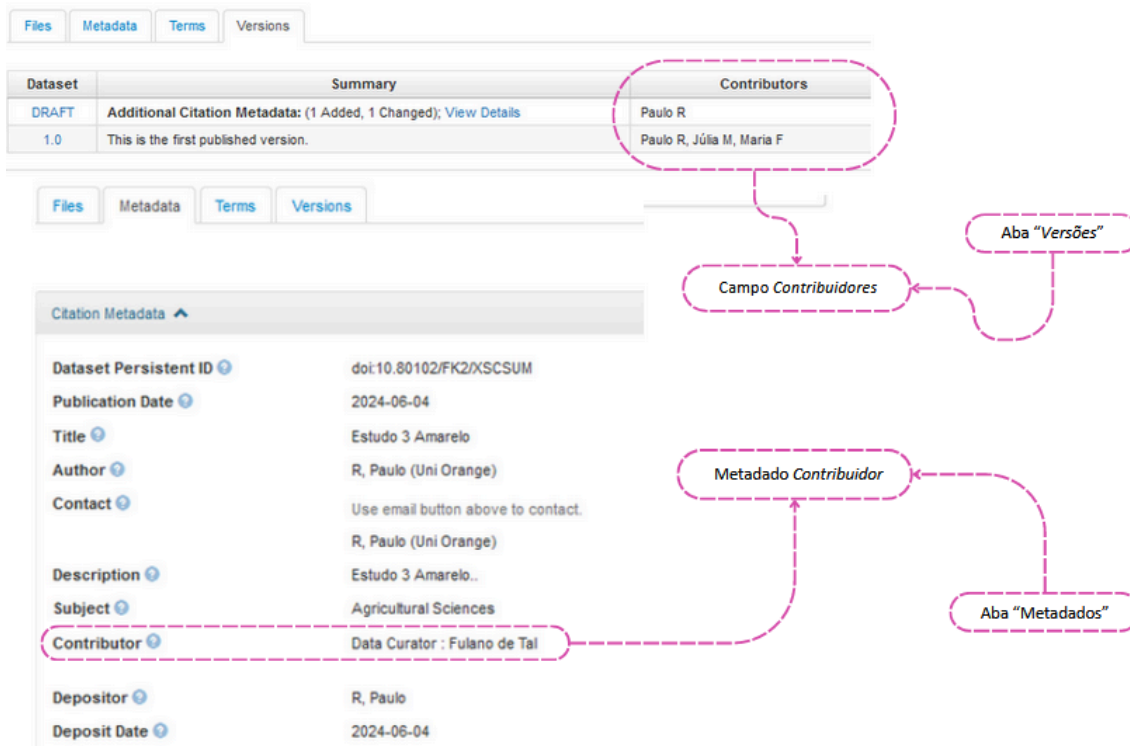
A seção 5.2 apresenta as atividades e os agentes envolvidos na submissão. Essa seção analisa as atividades e os agentes sob o aspecto dos metadados. Seguindo os questionamentos do quadro 7, com relação aos metadados de proveniência que descrevem agentes e atividades no DSpace, os metadados que descrevem os agentes envolvidos na submissão são registrados nos metadados de proveniência (*dc.provenance*).

O DSpace descreve quem revisou os metadados e quem publicou o item, conforme já apresentado na seção 5.2.1. Outros agentes podem ser inseridos manualmente nos campos de *contribuidor*, *publicador* e de *direitos* (*dc.contributor*, *dc.publisher*, *dc.rights*, respectivamente).

No Dataverse, o campo *Contribuidor* identifica automaticamente os agentes envolvidos na submissão (depósito, edição e publicação). Entretanto, as ações e os

agentes de edição e publicação não são representados como metadados e o metadado *Depositante* pode ser modificado. Por outro lado, Dataverse permite o registro ou inclusão manual dos metadados *Contribuidor* (*contributor*, diferente do campo *Contribuidores* na aba *versões*, como pode ser visto na figura 14) distribuidor (*distributor*), produtor e de publicador (*publisher*).

Figura 14: Diferença entre o campo e o metadado Contribuidor



Fonte: Prints Dataverse, Autor, 2024.

Os metadados que descrevem as atividades no Dataverse e DSpace (quadro 7) foram agrupados no quadro 9.

Quadro 9: Metadados de proveniência de atividades em DSpace e Dataverse

Metadado DSpace	Descrição
<i>dc.date.accessioned</i>	Data de posse do documento pelo repositório
<i>dc.date.available</i>	Data ou período em que o item se tornou público
<i>dc.date.created</i>	Data de criação do conteúdo do arquivo
<i>dc.date.issued</i>	data de publicação ou distribuição
<i>dc.date.submitted</i>	Data de submissão, recomendado para teses e dissertações

<i>dc.description.provenance</i>	O histórico de custódia do item desde a sua criação, incluindo quaisquer alterações realizadas.
Metadados Dataverse	Descrição
<i>Data de produção</i>	Data em que o conteúdo do arquivo foi criado
<i>Data de Depósito</i>	Data da submissão do dataset ao repositório
<i>Data (Período) de Coleta</i>	Data ou período em que os dados contidos no arquivo foram gerados/coletados
<i>Data de Distribuição</i>	A data em que o Dataset foi disponibilizado para o público
<i>Data da Versão</i>	Data de cada versão do dataset

Fonte: Autor, 2024.

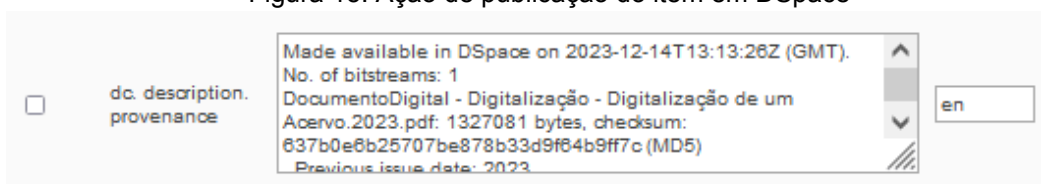
Se tratando dos metadados que descrevem os identificadores dos objetos digitais (quadro 7), as duas plataformas descrevem e representam o Identificador Uniforme de Recursos (URI), que pode ser seus identificadores persistentes (DOI ou Handle).

5.3.3 Análise da representação dos objetos digitais

Comparando a representação dos objetos digitais, dos metadados e informações de proveniência do DSpace e do Dataverse, é possível afirmar que existem poucas diferenças nos dados representados.

Em Dataverse, como já mencionado na seção 5.2.3 e 5.3.1, se percebe a falta da identificação das ações realizadas pelos agentes envolvidos na submissão. Já DSpace discrimina o submetedor e o avaliador, delegando a publicação como uma atividade passiva, como se pode ver na figura 15, onde descreve a ação como “Feito público em DSpace em” (tradução nossa).

Figura 15: Ação de publicação do item em DSpace



Fonte: Print DSpace, 2023.

Outro ponto que se deve considerar é que muitos dos metadados descritos neste capítulo são opcionais, como os metadados *colaborator* e *publisher* do Dataverse e os metadados *dc.contributor*, *dc.publisher*, *dc.rights* do DSpace, o que pode resultar na falta de informações essenciais para provar a integridade do objeto digital.

5.4 Gerenciamento e registro de alterações dos objetos digitais

Nesta seção será descrito o gerenciamento e o registro de alterações dos objetos digitais. O segmento de questões apresentadas no quadro 10 está relacionado ao quadro “Quadro 4-A: Gerenciamento e registro de mudanças nos objetos digitais”, do apêndice A. Esse segmento de questões busca concluir o objetivo D deste trabalho o excerto de questões reproduzidas no quadro abaixo buscam contextualizar as respostas descritas a seguir.

Quadro 10: Recorte do formulário sobre alterações do AIP

Mudanças no AIP	
Como é realizada?	A1
Mudanças são autorizadas?	A2
Mudanças são autorizadas em metadados gerados automaticamente?	A3
O sistema registra as mudanças?	A4
O sistema registra quem realizou as mudanças?	A5
O sistema controla versões?	A6

Fonte: Autor, 2024.

5.4.1 Gerenciamento e registro de alterações no DSpace

A alteração do AIP, como questionada no quadro 10, é realizada sobre um item já publicado, isto é, sobre um item que passou pelo fluxo de submissão (Quadro 10, A1).

Para realizar alguma modificação em um item, é necessário que o usuário tenha a autorização de Administrador da coleção que contém o item (Quadro 10, A2), pois uma alteração não passa novamente pelo fluxo de submissão.

O administrador da coleção em que o item está armazenado tem permissão para editar os metadados, para incluir ou excluir arquivos ou pastas, assim como para restringir o acesso ao item, e alterar permissões para ler, editar, remover ou adicionar elementos do item. Administradores de coleções também podem incluir novos itens, não necessitando que os mesmos passem pelo fluxo de submissão.

Cada alteração em um item é imediatamente refletida em seu armazenamento, isto é, não há um processo adicional de aprovação das alterações (Quadro 10, A2).

Em se tratando de metadados gerados automaticamente (Quadro 10, A3), como os de proveniência (datas de publicação e publicador), o DSpace permite que um usuário administrador altere todos estes campos. Isso significa que essas informações de proveniência que foram geradas pelo sistema podem ser alteradas.

Se os metadados do item forem alterados, esta ação de alteração não será registrada (Quadro 10, A4 e A5), por isso não será possível saber sobre os envolvidos.

Se o arquivo do item for substituído (removido, editado e novamente adicionado), DSpace apresenta apenas os metadados do novo arquivo (incluindo Checksum), não registrando a exclusão do arquivo anterior, assim como os metadados desse arquivo.

Referente às versões (Quadro 10, A6), o DSpace 5, em sua configuração padrão, não controla versões. Somente a última versão é registrada, portanto, o estado de um item anterior a uma modificação não será mantido, se este for alterado. No entanto, embora este trabalho se limite às configurações padrões de instalação, existe a possibilidade de habilitar a criação de versões durante a instalação. Esta configuração permite que os administradores criem versões de um item, segundo Lyrasis (2014).

Estas versões podem incluir mudanças no arquivo e nos metadados e, após sua criação, apenas a versão atual pode ser editada. Ou seja, o submetedor original não pode mais realizar nenhuma edição pois o item não é mais atribuído como sua publicação e sim, como uma publicação do administrador que criou a nova versão. Enquanto é possível criar uma nova versão para modificar apenas os metadados, toda a ação de versionamento é submetida ao fluxo de submissão. A ação “*Editar metadados do Item*”, que é permitida aos administradores e que não submete estas

edições ao fluxo, se torna um pouco redundante se a configuração de versionamento for habilitada.

5.4.2 Gerenciamento e registro de alterações no Dataverse

A alteração (Quadro 10, A2) de um dataset publicado pode ser feita por um usuário que possua o papel de administrador, curador ou colaborador. Quando um dataset é alterado, a versão anterior é mantida, e o dataset em alteração é chamado de dataset draft (rascunho). Para ser publicado, este dataset draft deve passar novamente pelas etapas de avaliação e edição (Quadro 10, A1). Quando este dataset draft for publicado, os envolvidos na avaliação e edição são registrados no campo *Contribuidor* (na aba *versões*), além das atividades de publicação e criação da versão serem registradas nos metadados de *Data de Distribuição* e da *Data da Versão*, estes que podem ser visualizados quando exportamos os metadados de um dataset em DDI (Quadro 10, A4 e A5).

O repositório Dataverse permite a alteração dos metadados gerados automaticamente, exceto no metadado de datas de depósito, incluindo datas de versionamento (Quadro 10, A3).

O Dataverse realiza automaticamente o versionamento de datasets alterados e mantém todas as versões do dataset, como questionado no quadro 10, A6. A plataforma ainda registra quem realizou as mudanças e quem avaliou, editou e publicou o dataset modificado no campo de contribuidores (Quadro 10, A4 e A5). Não é possível, pelo campo de *Contribuidor*, distinguir as atividades que cada um realizou, como já mencionado nas seções anteriores. No entanto, a forma de versionamento utilizada no Dataverse identifica e atribui as mudanças feitas no dataset ao usuário que as realizou.

5.4.3 Análise do gerenciamento e registro de alterações

Analisando o Dataverse e o DSpace pelo gerenciamento e registro de alterações de objetos submetidos aos repositórios, é possível observar algumas diferenças entre os dois. Enquanto os dois registram as ações e os agentes de forma parecida, o Dataverse, além de realizar o versionamento de datasets (em sua instalação padrão), ainda oferece uma descrição detalhada de quem atuou e o que

foi modificado. O Dataverse ainda permite a modificação de alguns metadados gerados automaticamente como o metadado de Depositante, que deveria ser imutável.

O processo para modificar um item no DSpace é mais simples e direto. No entanto, DSpace também oferece a possibilidade de modificar dados que deveriam ser imutáveis, principalmente os de proveniência e datas. Outro ponto relevante sobre o DSpace é a falta de versionamento nas configurações de instalação padrão, que pode resultar na perda de documentos dependendo das alterações.

É uma possibilidade interessante que os metadados gerados automaticamente não sejam modificáveis para fins de integridade, como dados sobre datas, publicadores e proveniência.

6 CONCLUSÃO

Esse trabalho investigou a questão da proveniência em duas importantes plataformas, o DSpace e o Dataverse, e comparou a submissão de informações em relação ao controle e registro das informações de proveniência nos fluxos de submissão dos repositórios.

O estudo enfatizou a representação e o controle das informações de proveniência e a autorização e autenticação de usuários nas plataformas. Assim, é possível identificar que os resultados se encaixam em uma das quatro divisões de análise: A autenticação e as informações de proveniência dos agentes que atuam na submissão como usuários em Dataverse e DSpace; As atividades e os agentes envolvidos na submissão em DSpace e Dataverse; A representação dos objetos, metadados e informações de proveniência em Dataverse e Dspace; O gerenciamento e o registro de mudanças nos objetos digitais no DSpace e no Dataverse.

No que diz respeito à autenticação e as informações de proveniência dos agentes que atuam na submissão como usuários em Dataverse e DSpace, conclui-se que possível observar que a autenticação de usuários das duas plataformas é feita por informações básicas e por meio de passos simples, como uso de informações como email para confirmar a identidade do usuário. Esse tipo de validação de informações pode não ser suficiente para as exigências de um repositório. Entretanto, tanto DSpace quanto Dataverse apresentam alternativas de autenticação. Estas alternativas envolvem integração com outros ambientes (como ORCID) e são soluções mais complexas de implantação. Observa-se também, que tanto DSpace quanto Dataverse não registram a realização das atividades envolvidas no cadastro.

No que diz respeito às atividades e os agentes envolvidos na submissão em DSpace e Dataverse, ambas possuem atividades de submissão, permitem a especificação de quem pode atuar em cada atividade e controla quem pode realizar cada atividade. Em Dspace, isso é realizado por um fluxo de submissão, onde cada etapa possui uma especificação de qual atividade deve ser realizada e quem está autorizado a realizá-la. Em Dataverse, isso é feito por meio de papéis, que concedem uma gama de ações que o usuário pode realizar sobre um dataset ou Dataverse.

Em relação ao registro das atividades de submissão, DSpace registra as atividades do fluxo em metadados de proveniência (*dc.provenance*). Já Dataverse apresenta, na informação *Colaborador* (na aba *versões* de um dataset), os envolvidos nas ações de submissão, mas não identifica as ações realizadas por cada colaborador (adição, edição, publicação de dataset) e essa informação não é registrada como um metadado. Além disso, também é possível observar que os metadados de proveniência do DSpace registram as ações, incluindo pessoas envolvidas,, mas esses registros podem ser modificados.

No que diz respeito à representação dos objetos digitais no DSpace e no Dataverse, foi possível observar que as duas plataformas representam os objetos de forma similar, em pastas e arquivos, com metadados administrativos e descritivos, com arquivos sendo descritos por checksum, nome, tamanho e tipo de arquivo e data de criação. Também foi possível observar que, enquanto os dois atribuem identificadores persistentes, apenas Dataverse disponibiliza a opção de usar identificador persistente para cada arquivo. Outro ponto a ser notado é que, DSpace e Dataverse representam informações dos agentes que atuaram sobre um item ou dataset, no entanto, Dataverse não identifica ações, resumindo-se a descrever os agentes envolvidos nas ações, e não os representa essa descrição como metadado. Já o DSpace não disponibiliza a visualização do metadado de proveniência (onde identifica estes agentes) ao público e não identifica o agente publicador de um item.

Sobre o gerenciamento e o registro de mudanças nos objetos digitais no DSpace e no Dataverse, foi possível observar algumas diferenças, como o versionamento muito detalhado do Dataverse e a como em DSpace, não é necessário submeter mudanças em metadados à um fluxo de submissão, o que faz seu processo mais simples e direto. As duas plataformas apresentaram semelhanças no controle de gerenciamento, se tratando de autorizações de usuários que podem modificar um item ou dataset, o que permite, ao objeto, um certo nível de integridade, no entanto, não é uma prova da mesma.

Como o objetivo deste trabalho era a identificação e análise de processos, registros e agentes do DSpace e Dataverse, uma teoria sobre os resultados não foi formulada e os objetivos foram cumpridos plenamente.

Outras pesquisas que podem ser derivadas e úteis para a esta área seria uma análise mais aprofundada que envolva outras versões do DSpace, outras

aplicações de repositórios digitais e outras configurações além das instalações padrão das plataformas.

REFERÊNCIAS

ABNT NBR 15472:2007. **ABNT Catálogo**. Disponível em: <https://www.abntcatalogo.com.br/pnm.aspx?Q=N0dMOW9vOFp3YVNTSC9jVW9rMmhydVJXRWljUU5FaTQ=#hide1>. Acesso em 17 jul. 2024.

ABOUT Dataverse. **Dataverse Project**. Disponível em: <https://dataverse.org/about>. Acesso em 4 set. 2023.

ALVES, Rachel Cristina Vesú. **Metadados como elementos do processo de catalogação**. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, SP, 2010.

ANDRADE, Selma Regina de; et al. O estudo de caso como método de pesquisa em enfermagem: uma revisão integrativa. **Texto & Contexto - Enfermagem**, v. 26, n. 4.

ARELLANO, Miguel Ángel Márdero. Preservação de documentos digitais. **Ci. Inf.**, Brasília, v. 33, n. 2, p. 15-27, maio/ago. 2004. Disponível em: <https://www.scielo.br/j/ci/a/FLfgJvpH3PZKf3HbpKYchZr/?lang=pt>. Acesso em 12 jan. 2024.

ARELLANO, Miguel Ángel Márdero; OLIVEIRA, Alexandre Faria de. Gestão de repositórios de preservação digital. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 14, n. 3, p. 465–483, 2016. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8646346>. Acesso em: 26 fev. 2024.

ARQUIVO NACIONAL. **Dicionário brasileiro de terminologia arquivística**. Rio de Janeiro: Arquivo Nacional, 2005. Disponível em: https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/dicionario_de_terminologia_arquivistica.pdf. Acesso em 28 ago. 2023.

BARBEDO, Francisco; CORUJO, Luís; SANT'ANA, Mário **Recomendações para a produção de planos de preservação digital**. Lisboa: Direção-Geral do Livro, dos Arquivos e das Bibliotecas (DGLAB), 2011. Disponível em: http://arquivos.dglab.gov.pt/wp-content/uploads/sites/16/2014/02/Recomend_producao_PPD_V2.1.pdf. Acesso em: 9 fev. 2024.

BLATTMANN, Ursula; WEBER, Claudiane. DSpace como repositório digital na organização. **Revista ACB**, [s. l.], v. 13, n. 2, p. 467–485, 2008. Disponível em: <https://revista.acbsc.org.br/racb/article/view/593>. Acesso em: 10 jan. 2024.

BRASIL. Conselho Nacional de Arquivos. **Diretrizes para a Presunção de Autenticidade de Documentos Arquivísticos Digitais**. Brasília, 2012. Disponível em: https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/conarq_presuncao_autenticidade_completa.pdf. Acesso em 22 mar. 2024.

CAMARGO, Liriane Soares de Araújo de; VIDOTTI, Silvana Ap. Borsetti Gregorio. **Uma estratégia de avaliação de repositórios digitais**. São Paulo: CRUESP, 2008. Disponível em: https://www.researchgate.net/profile/Silvana-Vidotti/publication/267715334_UMA_ESTRATEGIA_DE_AVALIACAO_EM_REPOSITORIOS_DIGITAIS/links/555b177f08ae6fd2d8287cb2/UMA-ESTRATEGIA-DE-AVALIACAO-EM-REPOSITORIOS-DIGITAIS.pdf. Acesso em 3 set. 2023.

CAPLAN, Priscilla. **Entendendo o PREMIS**. [s.l.]: Library of Congress Network Development and MARC Standards Office, 2009. Disponível em: https://loc.gov/standards/premis/understandingPREMIS_portuguese_2021.pdf. Acesso em 26 out. 2023.

CONSELHO NACIONAL DE ARQUIVOS; Câmara Técnica de Documentos Eletrônicos. **Glossário Documentos Arquivísticos Digitais**. 8.ed. [s.l.]: CONARQ, 2020. Disponível em: https://www.gov.br/conarq/pt-br/assuntos/camaras-tecnicas-setoriais-inativas/camara-tecnica-de-documentos-eletronicos-ctde/glosctde_2020_08_07.pdf. Acesso em 29 ago.2023.

CONSELHO NACIONAL DE ARQUIVOS. **OAIS**: Open Archival Information System. [s.d.]. Disponível em: <https://www.gov.br/conarq/pt-br/assuntos/camaras-tecnicas-setoriais-inativas/camara-tecnica-de-documentos-eletronicos-ctde/oais.pdf>. Acesso em 2 set. 2023.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. **Reference Model for an Open Archival Information System (OAIS)**: Magenta Book. Washington, 2012. Disponível em: <https://public.ccsds.org/pubs/650x0m2.pdf>. Acesso em: 13 dez. 2023.

CORETRUSTSEAL Trustworthy Data Repositories Requirements 2020–2022. **Zenodo**, 2019. Disponível em: <https://doi.org/10.5281/zenodo.3638211>. Acesso em 17 jan. 2024.

DATAcite Metadata Schema for the Publication and Citation of Research Data and Other Research Outputs: Version 4.5. **DataCite**, 2024. Disponível em: <https://doi.org/10.14454/g8e5-6293>. Acesso em 15 mar. 2024.

DATAVERSE. **Instituto Brasileiro de Informação em Ciência e Tecnologia**, [s.d.]. Disponível em: https://www.gov.br/ibict/pt-br/assuntos/tecnologias-para-a-informacao/copy_of_dataverse. Acesso em 01 dez. 2023.

DSPACE. **Instituto Brasileiro de Informação em Ciência e Tecnologia**, [s.d.]. Disponível em: <https://www.gov.br/ibict/pt-br/assuntos/tecnologias-para-a-informacao/dspace>. Acesso em 29 ago. 2023.

DSPACE Release 5.1 Status. **Wiki LYRISIS**, 2015. Disponível em: <https://wiki.lyrasis.org/display/DSPACE/Dspace+Release+5.1+Status>. Acesso em 15 jan, 2024.

FERREIRA, Miguel. **Introdução à preservação digital**: conceitos, estratégias e actuais consensos. Guimarães, Portugal: Escola de Engenharia da Universidade do Minho, 2006.

FLORES, Daniel.; PRADEBON, Daiane Segabinazzi. CÉ, Graziella. Análise do conhecimento teórico-metodológico da preservação digital sob a ótica da OAIS, SAAI, ISO 14721 e NBR 15472. **Brazilian Journal of Information Science**, v. 11, n., 2017.

FORMENTON, Daniel; GRACIOSO, Luciana de Souza. Padrões de metadados no arquivamento da web: recursos tecnológicos para a garantia da preservação digital de websites arquivados. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 20, 2022.

FORMENTON, Daniel; GRACIOSO, Luciana de Souza. Preservação digital: desafios, requisitos, estratégias e produção científica. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 18, 2020. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8659259>. Acesso em: 11 fev. 2024.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo (org.). **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009.

GONÇALVES, Ana Paula Vasconcelos; SANTOS, Rodrigo Salles Pereira dos. A pesquisa quantitativa em sociologia: recursos e dilemas da realização de surveys online com trabalhadores durante a pandemia de COVID-19. **Sociologia & Antropologia**, v. 13, n. 2, 2023.

ITEM Level Versioning. **Wiki LYRISIS**, 2014. Disponível em: <https://wiki.lyrasis.org/display/DSDOC5x/Item+Level+Versioning#ItemLevelVersioning-OnlyAdministratorsandCollection/CommunityAdministratorsCanaddnewversions>. Acesso em 01 jun. 2024.

MENDES, Karina Dal Sasso; SILVEIRA, Renata Cristina de Campos Pereira; GALVÃO, Cristina Maria.. Revisão integrativa: método de pesquisa para a incorporação de evidências na saúde e na enfermagem. **Texto & Contexto - Enfermagem**, v. 17, n. 4, p. 758–764, out. 2008.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Técnicas de Pesquisa**. 9. ed. São Paulo: Atlas, 2021.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **Understanding Metadata**. Bethesda: NISO, 2004. Disponível em: <http://hdl.handle.net/10150/105486>. Acesso em 20 dez. 2023.

OLIVER, Gillian; HARVEY, Ross. **Digital Curation**. Chicago: ALA Neal-Schuman, 2016. 2.ed. Disponível em: <https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1616606&lang=pt-br&site=ehost-live>. Acesso em: 29 jan. 2024.

PEREIRA, Laís de Toledo Krücken; GODOY, Dalva Maria Alves; TERÇARIOL, Denise. Estudo de caso como procedimento de pesquisa científica: reflexão a partir da clínica fonoaudiológica. *Psicologia: Reflexão e Crítica*, v. 22, n. 3, p. 422–429, 2009.

PIRES, Cássio de Oliveira; ROCHA, Rafael Port da. Finalidade e Atividades da Curadoria Digital na Perspectiva de sua Implantação em uma Instituição. **Brazilian Journal of Information Science: Research trends**, vol.14, n.4, set./dez. 2020

PÖTTKER, Luciana Maria Vieira; FERNEDA, Edberto; MOREIRO-GONZÁLEZ, José Antonio. Mapeamento relacional entre padrões de metadados educacionais. **Perspectivas em Ciência da Informação**, v. 23, n. 3, p. 25–38, jul. 2018.

PREMIS data dictionary for preservation metadata. **PREMIS version 3.0**. 2015. Premis Editorial Committee: [s.l.], 2015. Disponível em: <https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>. Acesso em 20 fev. 2024.

PROV-DM: The PROV Data Model. 2013. **W3C Recommendation**: [s.l.], 2013. Disponível em: <https://www.w3.org/TR/2013/REC-prov-dm-20130430/>. Acesso em 10 mar. 2024.

REZENDE, Laura Vilela Rodrigues. Caracterização de repositórios digitais Dataverse conforme o Modelo OAIS. **Revista Brasileira de Preservação Digital**, Campinas, SP, v. 3, n. 00, 2012. Disponível em: <https://econtents.bc.unicamp.br/inpec/index.php/rebpred/article/view/16581>. Acesso em: 4 abr. 2024.

ROCHA, Rafael. Port ... et al. **Acesso aberto a dados de pesquisa no Brasil**: soluções tecnológicas: relatório 2018. Porto Alegre, RS: UFRGS, 2018. Disponível em: <http://hdl.handle.net/10183/185126>. Acesso em: 30 dez 2023.

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**, [S. l.], v. 17, p. 1–25, 2021. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/1572>. Acesso em: 23 mar. 2024.

RODRIGUES, Nelson de Almeida. Introdução ao METS: Preservação e Intercâmbio de Objetos Digitais. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, [s. l.], v. 13, n. 26, p. 172–187, 2008. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2008v13n26p172>. Acesso em: 15 mar. 2024.

SANTA ANNA, Jorge. Relações paradigmáticas entre Arquivologia e Ciência da Informação: a vivência prática de um arquivo. **Biblios**, Pittsburgh, n. 72, p. 51-66, jul. 2018. Disponível em http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1562-4730201800030004&lng=es&nrm=iso. Acesso em 20 mar. 2024.

SANTOS, Henrique Machado; FLORES, Daniel. Introdução aos conceitos básicos do modelo Open Archival Information System no contexto da arquivística. **Acervo**, [s. l.], v. 32, n. 1, p. 8-26, 2019. Disponível em: <https://revista.arquivonacional.gov.br/index.php/revistaacervo/article/view/1029>. Acesso em: 10 jan. 2024.

SANTOS, Henrique Machado dos; FLORES, Daniel. Repositórios digitais confiáveis para documentos arquivísticos: ponderações sobre a preservação em longo prazo. **Perspectivas em Ciência da Informação**, v. 20, n. 2, p. 198-218, abr. 2015.

SARMENTO, F, et al. Algumas considerações sobre as principais declarações que suportam o movimento Acesso Livre. In: In **World Congress on Health Information and Libraries**, 9, Salvador, Brasil, 2005. Disponível em: <https://hdl.handle.net/1822/4282>. Acesso em: 4 set. 2023.

SAYÃO, Luís Fernando. 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.

SciELO. **Guia de curadoria de dados de pesquisa para equipes editoriais**. SciELO, 2023. Disponível em: https://wp.scielo.org/wp-content/uploads/Guia_curadoria_pt.pdf. Acesso em 22 mar. 2024.

SENSO, José A; PIÑERO, Antonio de la Rosa. El concepto de metadato: algo más que descripción de recursos electrónicos. **Ciência da Informação**, v. 32, n. 2, p. 95-106, maio 2003.

SHINTAKU, Milton; MEIRELLES, Rodrigo. **Manual do DSpace**: administração de repositórios. Salvador: EDUFBA, 2010. Disponível em: <https://repositorio.ufba.br/handle/ri/769>. Acesso em 18 nov. 2023.

SILVA, Pollianna Marys de Souza e; COSTA, Levi Cadmiel Amaral da. O Dataverse e suas contribuições para a comunicação científica. **ÁGORA: Arquivologia em debate**, [s. l.], v. 31, n. 63, p. 1-14, 2021. Disponível em: <https://agora.emnuvens.com.br/ra/article/view/995>. Acesso em: 18 jan. 2024.

SILVA, Vanessa Batista; MARQUES, Tathiane Amaral. O uso do Dataverse como repositório de dados: uma análise prática. In: **Seminário Nacional de Bibliotecas Universitárias**, 22, 2023, Florianópolis. Anais... Disponível em: <https://portal.febab.org.br/snbu2023/article/view/2920>. Acesso em 15 jan. 2024.

SOBRAL, Renato Machado; SANTOS, Cibele Araújo Camargo Marques dos. Repositórios institucionais digitais de informação científica: implementação com o software Dspace como solução técnica. **Prisma**, v.35, p. 152-184, 2017. Disponível em: <https://www.eca.usp.br/acervo/producao-academica/002916776.pdf>. Acesso em 3 set. 2023.

SOBRE Repositórios Digitais. **Instituto Brasileiro de Informação em Ciência e Tecnologia**, [s.d.]. Disponível em: <http://www.ibict.br/informacao-para-ciencia-tecnologia-einovacao%20/repositorios-digitais>. Acesso em 17 out. 2023.

SOUZA, Luciana Gonçalves Silva; AGANETTE, Elisângela Cristina. Repositórios digitais confiáveis: uma revisão da literatura nacional e internacional publicada em periódicos científicos. **Informação & Sociedade**, [s. l.], v. 30, n. 1, 2020. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/45426>. Acesso em: 15 mar. 2024.

THOMAS, Débora Butka; OENNING, Nágila Soares Xavier; GOULART, Bárbara Niegia Garcia de. Aspectos essenciais na construção de instrumentos de coleta de dados em pesquisas primárias de saúde. **Revista CEFAC**, v. 20, n. 5, p. 657–664, set. 2018.

TRUJILLO FERRARI, Alfonso. **Metodologia da pesquisa científica**. Rio de Janeiro: McGraw-Hill do Brasil, 1982.

USER guide. **DATAVERSE**, [s.d.], 2023. Disponível em: <https://guides.dataverse.org/en/5.13/user/index.html>. Acesso em: 18 jan. 2024.

APÊNDICE A

Quadro 1-A: Autenticação e a informações de proveniência dos agentes da submissão

Agentes que são Usuários	Dataverse	Dspace
<p>Quais são as informações fornecidas pelos usuários em seus cadastros?</p> <p>X - Obrigatório O - Disponibilizado como opção</p>	<p><input checked="" type="checkbox"/> Nome, Sobrenome <input checked="" type="checkbox"/> Nome de usuário <input type="checkbox"/> Identidade, CPF, Passaporte <input checked="" type="checkbox"/> Email, Telefone <input type="checkbox"/> Local e Data de Nascimento <input type="checkbox"/> Vínculo/empresa <input type="checkbox"/> Cargo <input type="checkbox"/> Foto <input type="checkbox"/> Identificador em outro ambiente (como ORCID)</p>	<p><input checked="" type="checkbox"/> Nome, Sobrenome <input type="checkbox"/> Nome de usuário <input type="checkbox"/> Identidade, CPF, Passaporte <input checked="" type="checkbox"/> Email, Telefone <input type="checkbox"/> Local e Data de Nascimento <input type="checkbox"/> Vínculo/empresa <input type="checkbox"/> Cargo <input type="checkbox"/> Foto <input type="checkbox"/> Identificador em outro ambiente (como ORCID)</p>
<p>Como a identidade do usuário é verificada?</p> <p>X - Obrigatório O - Disponibilizado como opção</p>	<p><input checked="" type="checkbox"/> Por email <input type="checkbox"/> Por SMS <input type="checkbox"/> Pessoalmente <input type="checkbox"/> Por foto <input type="checkbox"/> Por confiança (cadastramento em outro ambiente, como ORCID) <input type="checkbox"/> É pré-cadastrado pela instituição</p>	<p><input checked="" type="checkbox"/> Por email <input type="checkbox"/> Por SMS <input type="checkbox"/> Pessoalmente <input type="checkbox"/> Por foto <input type="checkbox"/> Por confiança (cadastramento em outro ambiente, como ORCID) <input type="checkbox"/> É pré-cadastrado pela instituição</p>
<p>A realização da ação de verificação da identidade do usuário é registrada?</p>	<p>Não é registrada de nenhuma forma acessível aos usuários e administradores do repositório</p>	<p>Não é registrada de nenhuma forma acessível aos usuários e administradores do repositório</p>

Mudanças nos dados de cadastro do usuário são verificadas/validadas?	Mudanças de nome: Não verifica ou valida Mudanças e e-mail: envia verificação Não é possível mudar o nome de usuário	Mudanças de nome: Não verifica ou valida Não é possível modificar o e-mail Não verifica informações de contato
Mudanças nos dados de cadastro do usuário são registradas?	São notificadas na plataforma mas não registradas permanentemente	São notificadas na plataforma mas não registradas permanentemente
De que forma um usuário é autorizado a realizar uma atividade sobre um objeto?	Seu usuário é atribuído um papel ou incluído em um grupo que realiza este papel por um administrador/curador..	Seu usuário é incluído em um grupo que é autorizado a realizar uma atividade por um administrador.

Quadro 2-A: Identificação e descrição das atividades e agentes nas atividades

Atividade	Dataverse	Dspace
Criação do Pacote de Submissão (SIP)		
Como a criação do SIP é realizada?	Por meio da submissão de um dataset.	Pela atividade “Submissão” do fluxo de submissão.
Qual o agente? Somente agentes autorizados podem atuar?	Somente Usuários com papel de criador/curador/administrador podem adicionar o dataset.	Somente realizada por usuários de grupos autorizados a submeter itens.
Sua realização é registrada?	Sim, nos metadados: <i>Data de Criação e Publicador</i> e no campo <i>Contribuidor</i>	Sim, nos metadados: <i>Data de submissão (dc.provenance)</i> ; <i>Submetedor</i>
Revisão e Edição do Pacote de Submissão (SIP)		
Como é realizada?	Usuário de grupo autorizado a realizar o papel de Colaborador realiza a edição. Curadores/Administradores podem editar, revisar e avaliar o dataset.	Pela atividade “Avaliação” do fluxo de submissão, realizada por usuário de grupo autorizado Pela atividade “Revisão dos Metadados” do fluxo de submissão, realizada por usuário de grupo autorizado

Sua realização é registrada?	Sim, Usuários que avaliam, editam e publicam o dataset são registrados como <i>Contribuidores</i> .	Sim, na <i>Data de submissão</i> (provenance, para atividade de avaliação); <i>Avaliador</i> ; <i>Data de Publicação</i>
Geração do Pacote de Armazenamento (AIP)		
Como é realizada?	Usuário de grupo autorizado a realizar o papel de Curador ou Administrador publica o dataset submetido, o que gera o AIP.	O item é aprovado na etapa de Avaliação Então, o item é revisado na etapa de Revisão de Metadados Então, o item é publicado na etapa de revisão de metadados, o que gera o AIP.
Sua realização é registrada?	Sim, nos metadados: <i>Data de Distribuição e Data da Versão</i> e no campo <i>Contribuidor</i>	Nos dados de proveniência apenas

Quadro 3-A: Descrição da representação dos objetos, metadados e informações de proveniência em Dataverse e Dspace

Pacote SIP/AIP	Dataverse	Dspace
Representa arquivos e pastas?	Sim	Sim
Representa metadados?	Sim	Sim
Representa metadados descritivos?	Sim, nos metadados exportados e o DataCite permite a criação de novos metadados	Sim, nos metadados em Dublin Core e permite a criação de novos campos de metadados
Representa metadados estruturais?	Sim, descreve os arquivos que compõem o dataset e nos metadados exportados (elemento <i>fileDescr</i>)	Não
Representa metadados administrativos?	Parcialmente, Registra automaticamente alguns agentes que contribuíram no processo de submissão do conjunto de dados (publicador, depositante), datas e versões.	Parcialmente, o campo <i>dc.provenance</i> é preenchido automaticamente pelo ambiente para registrar: -Quem submeteu (e quando) -Quem aprovou (e quando) -Quando o item foi armazenado (AIP)

	-Quem criou dataset (metadado <i>publicador</i>) , quem editou os metadados (campo <i>contribuidor</i>), quem publicou (campo <i>contribuidor</i>)	
Descreve nome de arquivo?	Sim, descreve, e representa em metadados (Em DDI: <i>otherMat</i>)	Sim, em metadados de arquivo
Descreve tamanho de arquivo?	Sim, descreve, mas não representa em metadados	Sim, em metadados de arquivo
Descreva a data de criação do arquivo?	Sim, descreve, mas não representa em metadados	Sim <i>dc.date.accessioned</i> 2024-03-14T15:14:19Z <i>dc.date.available</i> 2024-03-14T15:14:19Z
Representa Checksum de arquivo?	Sim, descreve, mas não representa em metadados	Sim
Descreve o formato do arquivo?	Sim, descreve, e representa em metadados (<i>otherMat</i>)	Sim
Atribui identificador persistente de objeto ao arquivo?	Sim, quando configurado	Não, somente para o item
Metadados de Proveniência		
Quais metadados descrevem agentes?	<i>Produtor</i> : A entidade, pessoa ou organização, gerenciando as finanças ou outros processos administrativos envolvidos na criação do Dataset (Em DDI: <i>stdyDscr > citation > prodStmt > producer</i>) <i>Contribuidor</i> : A entidade, pessoa ou organização, responsável por coletar, gerenciar ou contribuir de outra forma para o desenvolvimento do Dataset (<i>stdyDscr > citation > rspStmt > othId (role)</i>)	Junto aos metadados que descrevem atividades <i>Submetedor</i> : “ <i>dc.description.provenance Submitted by DSpace ADM (dspace) on 2024-03-08T19:56:59Z No. of bitstreams: 1 DSPACE sob Olhar OAIS.2020.pdf: 2256564 bytes, checksum: 7f465351f1b46dd38c0d5b2dd78ffb06 (MD5) en</i> ” <i>Avaliador</i> :

	<p><i>Distribuidor:</i> A entidade, pessoa ou organização, designada para gerar cópias do Dataset, incluindo edições ou revisões (<i>stdyDscr > citation > distStmt > distrbtr</i>)</p> <p><i>Depositante:</i> A entidade, pessoa ou organização, que depositou o Dataset no repositório (<i>stdyDscr > citation > distStmt > deposit</i>)</p> <p><i>Publicador:</i> Dado gerado automaticamente pelo sistema em relação ao dataset.</p>	<p><i>“dc.description.provenance Approved for entry into archive by DSpace ADM (dspace) on 2024-03-14T15:12:16Z (GMT) No. of bitstreams: 1 DSPACE sob Olhar OAIS.2020.pdf: 2256564 bytes, checksum: 7f465351f1b46dd38c0d5b2dd78ffb06 (MD5) en”</i></p> <p>Outros agentes podem ser inseridos manualmente em:</p> <p><i>dc.contributor, dc.publisher, dc.rights</i></p>
Quais metadados que descrevem atividades?	<p><i>Data de produção:</i></p> <p><i>Data de Depósito</i> (<i>Em DDI: stdyDscr > citation > distStmt > depDate</i>)</p> <p><i>Data (Período) de Coleta:</i> (<i>stdyDscr > stdyInfo > sumDscr > collDate</i>)</p> <p><i>Data de Distribuição:</i> A data em que o Dataset foi disponibilizado para distribuição/apresentação (<i>stdyDscr > citation > distStmt > distDate</i>)</p> <p><i>Data da Versão:</i> (<i>docDscr > citation > verStmt > version (date)</i>)</p>	<p><i>dc.date.accessioned:</i> ex. <i>dc.date.accessioned</i> 2024-03-14T15:14:19Z</p> <p><i>dc.date.available:</i> ex. <i>dc.date.available</i> 2024-03-14T15:14:19Z</p> <p><i>dc.date.created</i></p> <p><i>dc.date.issued</i></p> <p><i>dc.date.submitted</i></p> <p><i>dc.description.provenance</i></p> <p><i>Submissão e Avaliação: dc.description.provenance</i></p> <p><i>Publicação:</i> <i>“dc.description.provenance Made available in DSpace on 2024-03-14T15:14:19Z (GMT). No. of</i></p>

		<p><i>bitstreams: 1 DSPACE sob Olhar OAIS.2020.pdf: 2256564 bytes, checksum: 7f465351f1b46dd38c0d5b2dd78ffb06 (MD5)</i></p> <p><i>Previous issue date: 2023"</i></p>
Quais metadados que descrevem o artefato/objeto?	<p><i>Identificador:</i> Identificador que seja único do dataset. (<i>Em DDI: tdyDscr > citation > titlStmnt > IDNo</i>)</p> <p><i>Autor:</i> Criador do dataset (<i>stdyDscr > citation > rspStmnt > AuthEnty</i>)</p> <p><i>Número da Versão:</i> Versão do dataset. <i>docDscr > citation > verStmnt > version</i></p> <p><i><fileDscr ID="f646"</i> <i>URI="https://dataverse.fgv.br/api/access/datafile/646"</i> <i>></i> <i></dataDscr></i></p> <p><i><otherMat ID="f581"</i> <i>URI="https://hdl.handle.net/10438.3/FK2/TXAZ56/MQPZUG" level="datafile"></i></p>	<p><i>dc.format.extent:</i> Tamanho ou duração do arquivo;</p> <p><i>dc.contributor.author</i></p> <p><i>dc.identifier:</i> <i>Catch-all</i> para identificadores inequívocos não definidos por formulário qualificado; Use <i>identifier.other</i> para um identificador conhecido comum a uma coleção local em vez de formulário não qualificado.</p> <p><i>dc.identifier.uri</i></p>

Quadro 4-A: Gerenciamento e registro de mudanças nos objetos digitais.

Mudanças no AIP	Dataverse	Dspace
Como é realizada?	Alterações no Dataset são realizadas (gerando um rascunho) que são publicadas após serem tratadas como submissão nova.	Usuários de grupos autorizados, como administradores, podem alterar os dados e metadados O fluxo de submissão não envolve alterações

Mudanças são autorizadas?	Sim, por pessoas autorizadas a publicar	Sim, por pessoas autorizadas a publicar
Mudanças são autorizadas em metadados gerados automaticamente?	Sim, por pessoas autorizadas Não permite alteração em: publicador e datas de depósito	Permite mudar, inclusive metadados de proveniência (<i>dc.description.provenance</i>)
O sistema registra as mudanças?	Sim, registra a versão e nos metadados: <i>Data de Distribuição</i> ; <i>Data da Versão</i> e no campo <i>Contributor</i>	Não registra
O sistema registra quem realizou as mudanças?	Parcialmente. registra todo os atores que atuaram em todas versões, Não identificando a atuação de cada ator em cada versão	Registra em <i>dc.provenance</i> quem submeteu e avaliou. Registra a data de publicação
O sistema controla versões?	Sim	Não (DSpace 5 padrão)