HIPÁTIA
MODELO DE PRESERVAÇÃO

Metadados e funcionalidades de *softwares* para apoiar cumprimento da LGPD e de restrições de acesso a dados no contexto do modelo de preservação digital Hipátia para implementação de Repositórios Arquivísticos Digitais Confiáveis (RDC-Arq)

Relatório Técnico

Preservação digital e gestão documental apoiada na implantação do modelo RDC Arq – TRT da 4ª Região

Brasília
2023

©2023 INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA

Este Relatório é licenciado sob uma licença Creative Commons Attribution 4.0, sendo permitida a reprodução parcial ou total desde que mencionada a fonte.

EQUIPE TÉCNICA

Diretor do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)

Tiago Emmanuel Nunes Braga

Coordenador-Geral Substituto de Tecnologias de Informação e Informática (CGTI)

Alexandre Faria de Oliveira

Coordenador do Projeto

Alexandre Faria de Oliveira

Autor

Cássio Felipe de Oliveira Pires

Este Relatório Técnico é um produto do projeto Preservação digital e gestão documental apoiada na implantação do modelo RDC Arq – TRT4

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto Brasileiro de Informação em Ciência e Tecnologia ou do Ministério da Ciência, Tecnologia e Inovações.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

Sumário

1. Introdução 04

2. Objetivos 08

2.1 Objetivo Geral 09

2.2 Objetivos Específicos 09

3. Contextualização e Resultados 10

3.1 Metadados de Direitos: PREMIS e a entidade "Rights" 11

3.1.1 Unidade semântica "rightsBasis" de PREMIS Rights 15

3.1.2 Documentando informação sobre LGPD com PREMIS Rights 19

3.2 Funcionalidades dos *softwares* Archivematica e AtoM 22

3.2.1 Identificação das funcionalidades do software Archivematica 22

3.2.2 Análise das funcionalidades do software Archivematica 24

3.2.3 Identificação das funcionalidades do software AtoM 26

3.2.4 Análise das funcionalidades do software AtoM identificadas 28

3.3 Testes de funcionalidades 29

3.3.1 Testes com Archivematica: estratégias de disseminação 30

3.3.2 Testes com AtoM: restrições de exibição de objetos digitais 41

3.4 Recomendações 44

4. Considerações finais 47

5. Referências 49

1.
Introdução

1. Introdução

O modelo de preservação digital Hipátia (BRAGA; MÁRDERO ARELLANO, 2022) consolidou-se, nos últimos anos, como a proposta do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) para a estruturação de Repositórios Arquivísticos Digitais Confiáveis, conhecidos pela sigla RDC-Arq (BRASIL, 2015). O Ibict tem desenvolvido pesquisas aplicadas do modelo Hipátia com parceiros de órgãos do Poder Executivo e, em especial, do Poder Judiciário, a exemplo do Tribunal de Justiça de Rondônia (TJRO) e do Tribunal Regional do Trabalho da 4ª Região (TRT4), entre outros. Destaca-se que, no Poder Judiciário, a adoção de RDC-Arq é obrigatória conforme as Resoluções CNJ nº 324/2020 e 408/2021, e deve estar baseado em *softwares* livres, gratuitos e de código aberto (CONSELHO NACIONAL DE JUSTIÇA, 2023).

No modelo Hipátia, além do BarraPres, tecnologia de *software* livre nacional desenvolvida pelo Ibict, os *softwares* AtoM (Access to Memory) e Archivematica têm desempenhado papel crucial na implementação de funcionalidades previstas no modelo de referência *Open Archival Information System* - OAIS (CCSDS, 2012), norma ISO 14721:2012, e no cumprimento de requisitos tecnológicos para repositórios previstos na *Audit and Certification of Trustworthy Digital Repositories* - ACTDR (CCSDS, 2011), norma ISO 16363:2012. Ambas, OAIS e ACTDR, fundamentam requisitos exigidos para a constituição de um RDC-Arq pelo Conselho Nacional de Arquivos (CONARQ) (BRASIL, 2015). Os recursos providos pelos *softwares* AtoM e Archivematica buscaram observar o modelo OAIS e requisitos tecnológicos de ACTDR, o que possibilitou que ambos se tornassem alternativas gratuitas viáveis na implementação da infraestrutura tecnológica de repositórios arquivísticos digitais, quando devidamente configurados, integrados e inseridos em uma infraestrutura organizacional adequada.

A implementação de um repositório nos termos das normas ISO constitui um desafio que perpassa não apenas o cumprimento de complexos requisitos tecnológicos, mas também administrativos e organizacionais (ROCHA; PIRES, 2020). Análise realizada por Braga *et al.* (2022) evidencia que, na proposta de revisão dos requisitos de RDC-Arq pelo CONARQ (BRASIL, 2022) há elementos que não são aplicáveis pelo modelo Hipátia, pois dependem de iniciativas administrativas

por parte da gestão da organização que implementa o repositório. Há principalmente aspectos da infraestrutura organizacional de um repositório digital que não são objetos diretos do modelo Hipátia. Em função de sua capacidade singular de inovação na dimensão tecnológica, o Hipátia privilegiou requisitos de RDC-Arq ligados, principalmente, ao gerenciamento do documento digital e à aplicação prática de princípios de preservação digital.

Nesse contexto de desafios à implementação de repositórios, há uma constante tensão entre o acesso à informação e a necessidade de cumprimento de leis e de normas que exigem a proteção de dados e o sigilo de informações. O "Manual de digitalização de documentos do Poder Judiciário" (CONSELHO NACIONAL DE JUSTIÇA, 2023, p. 15) destaca a coexistência de acesso e sigilo ao mencionar a Constituição Federal:

> Entre os direitos fundamentais que possuem relação direta com a gestão da documentação do Poder Judiciário, sobressaem a garantia de acesso à informação de interesse particular, coletivo ou geral, ressalvadas as hipóteses em que o sigilo seja imprescindível à segurança da sociedade e do Estado (art. 5º, XIV e XXXIII), a proteção à intimidade (art. 5º, X) e de dados pessoais, inclusive nos meios digitais (art. 5º, LXXIX).

Além de observar esses direitos fundamentais previstos na Carta Magna, um RDC-Arq também precisa cumprir a Lei nº 12.527, de 18 de novembro de 2011 (Lei de Acesso à Informação, LAI); a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet); e, logicamente, a Lei nº 13.709 de 2018 (Lei Geral de Proteção de Dados, LGPD). Outrossim, um RDC-Arq do Poder Judiciário também deve estar alinhado às normativas do Poder Judiciário, tais como: Resolução CNJ nº 215/2015, que dispõe sobre o acesso à informação e a aplicação da Lei nº 12.527/2011; Resolução CNJ nº 363/2021, que estabelece medidas para o processo de adequação à Lei Geral de Proteção de Dados a serem adotadas pelos tribunais; Resolução CNJ nº 324/2020, que institui diretrizes e normas de Gestão de Memória e de Gestão Documental e dispõe sobre o Programa Nacional de Gestão Documental e Memória do Poder Judiciário – Proname; Resolução CNJ nº 408/2021, que dispõe sobre o recebimento, o armazenamento e o acesso a documentos digitais relativos a autos de processos administrativos e judiciais; e, também, Resolução CNJ nº 469/2022, que estabelece diretrizes e normas sobre a digitalização de

documentos judiciais e administrativos e de gestão de documentos digitalizados do Poder Judiciário.

Observa-se que, embora exista amplo arcabouço legal e normativo sobre acesso e sigilo em repositórios, há uma carência de estudos em relação às funcionalidades que os *softwares* livres, gratuitos e de código aberto utilizados nos repositórios dispõem, nativamente, para apoiar as organizações no acesso à informação respeitando hipóteses de sigilo.

A partir da breve discussão apresentada até este ponto, pondera-se: as tecnologias utilizadas pelo modelo de preservação digital Hipátia para implementação de RDC-Arq provêm recursos e funcionalidades nativas capazes de garantir o acesso à informação respeitando hipóteses de sigilo e de proteção de dados previstas na legislação e em normativas infralegais do Poder Judiciário que tratam do tema?

Esta questão orienta a investigação realizada, que se dedica a dois tipos de tecnologias do modelo: o padrão de metadados de preservação digital *Preservation Metadata: Implementation Strategies* (PREMIS) e os *softwares* AtoM e Archivematica, sob a perspectiva de instâncias integradas. Na próxima seção, apresenta-se o objetivo geral e os objetivos específicos deste estudo.

2.
Objetivos

2. Objetivos

Na presente seção, apresenta-se o objetivo geral e os objetivos específicos da investigação realizada.

2.1 Objetivo Geral

Elaborar recomendações para a implementação do modelo Hipátia a partir de análises que verifiquem o potencial de funcionalidades e recursos de *softwares* e de padrões de metadados para prover acesso à informação respeitando hipóteses de necessidade de sigilo e de proteção de dados.

2.2 Objetivos Específicos

1. Analisar unidades semânticas da tecnologia de metadados de preservação digital PREMIS subordinadas à entidade " *Rights* ".

2. Elaborar modelo exemplificativo de metadados PREMIS de Direitos para registrar restrição de acesso com base na LGPD em uma transferência para Archivematica.

3. Identificar e listar recursos e funcionalidades das tecnologias de *softwares* AtoM (versão 2.7.1) e Archivematica (versão 1.13.2) que possam auxiliar um RDC-Arq do Poder Judiciário a prover acesso à informação, protegendo dados pessoais e cumprindo exceções de sigilo de dados contidos em documentos digitais preservados.

4. Analisar as funcionalidades e os recursos identificados no objetivo anterior, discutindo hipóteses nas quais possam auxiliar em restrições de acesso e na proteção a dados pessoais.

5. Realizar testes nos *softwares* AtoM e Archivematica, incluindo, nesses testes, um modelo para registro de metadados PREMIS de Direitos, com a finalidade de identificar as capacidades de suprir, parcial ou plenamente, o imperativo de prover acesso à informação respeitando hipóteses de sigilo e de proteção de dados pessoais.

3.

Contextualização e Resultados

3. Contextualização e Resultados

Devido à abrangência do problema ora investigado, dividiu-se esta seção em quatro subseções. Na subseção 3.1, intitulada " Metadados de Direitos: PREMIS e a entidade '*Rights*'", a análise contextualiza e expõe os resultados relativos aos objetivos específicos 1 e 2, que se referem à tecnologia de metadados padrão PREMIS, acrônimo de *Preservation Metadata: Implementation Strategies* (LIBRARY OF CONGRESS, 2022). As subseções 3.2 e 3.3 tratam de tecnologias de *software*. Na subseção 3.2, designada "Funcionalidades dos *softwares* Archivematica e AtoM ", discute-se os resultados que dizem respeito aos objetivos específicos 3 e 4. E, na subseção 3.3, chamada de "Testes de funcionalidades", aborda-se os resultados de testes conforme o objetivo específico 5.

Por fim, na subseção 3.4, recomenda-se hipóteses de ações para planejamento e implementação de RDC-Arq considerando os resultados expostos no decorrer da seção 3.

3.1 Metadados de Direitos: PREMIS e a entidade "Rights"

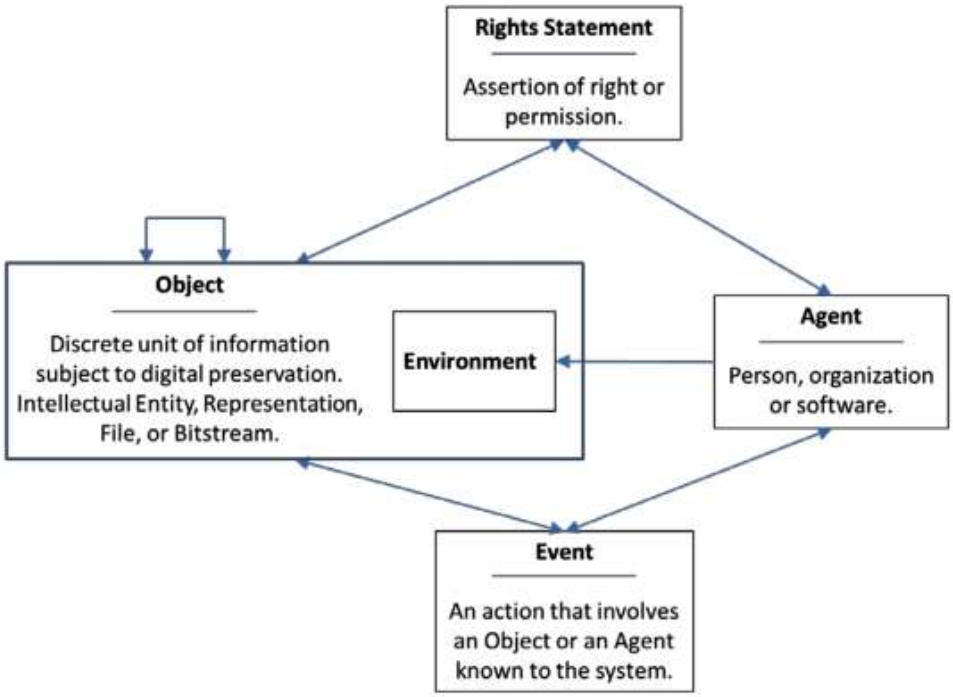
No Archivematica e no AtoM, *softwares* utilizados na implementação do modelo Hipátia, existem funcionalidades para caracterizar e descrever Direitos [*Rights*] de documentos digitais com o uso de metadados. Para essa finalidade, esses *softwares* implementam "unidades semânticas"[1] do padrão de metadados PREMIS.

O modelo de metadados PREMIS é o padrão *de facto* de metadados de preservação digital (DAPPERT, 2013), pois consolidou-se como um sinônimo de "metadado de preservação". PREMIS está na versão 3.0 e consiste em um dicionário de dados, um *schema* XML e documentação de apoio. Inicialmente desenvolvido pelo

1 "Unidades semânticas" são as propriedades das entidades de PREMIS (Objetos, Eventos, Direitos e Agentes). O termo "unidade semântica" é utilizado em detrimento do termo "elemento de metadado" (PREMIS EDITORIAL COMMITTEE, 2015). Por exemplo, rightsBasis, que serve para registrar um fundamento de direito, é um elemento de metadado, isto é, uma unidade semântica, da entidade Rights.

The Preservation Metadata: Implementation Strategies Working Group, as atualizações e revisões de PREMIS são atualmente coordenadas pelo *PREMIS Editorial Committee.* O padrão PREMIS é baseado em um modelo de dados que estabelece quatro entidades de topo, isto é, que não estão subordinadas a quaisquer outras entidades: *Objects, Events, Rights* e *Agents*. Em tradução literal, Objetos, Eventos, Direitos e Agentes. Cada unidade semântica do dicionário de dados é uma propriedade de uma das quatro entidades. A Figura 1 representa o modelo de dados de PREMIS com as quatro entidades.

Figura 1. Modelo de dados de PREMIS



Fonte: PREMIS EDITORIAL COMMITTEE, 2015, p. 6.

Observa-se, por meio da Figura 1, que uma Declaração de Direitos está relacionada a Agentes e a Objetos. As setas bidirecionais significam que, nas Declarações de Direitos, há unidades semânticas relacionando-as a Objetos e a Agentes do repositório, e **vice-versa**. No presente relatório, trata-se especificamente da entidade *Rights,* que tem duas unidades semânticas de nível superior: *rightsStatement* e *rightsExtension.* Isto é, "se o repositório registra informação de Direitos [*Rights*], deve estar presente *rightsStatement* ou *rightsExtension* " (PREMIS EDITORIAL COMMITTEE, 2015, p. 178).[2]

2 Todas as traduções de citações deste relatório foram realizadas pelo autor.

Ambas as unidades, *rightsStatement* e *rightsExtension* são contêineres, designação dada a unidades semânticas que agrupam outras unidades semânticas que podem ou não serem também contêineres. Além disso, as duas são opcionais porque o repositório tem três alternativas frente à documentação de Direitos: poderá utilizar (1) *rightsStatement* ou (2) *rightsExtension* ou, ainda, (3) nenhuma delas. No primeiro caso, a unidade *rightsStatement* compreende unidades semânticas definidas pelo próprio padrão PREMIS com o intuito de caracterizar a declaração sobre Direitos que se pretende fazer. No segundo caso, a unidade *rightsExtension* serve como contêiner para unidades semânticas definidas externamente ao padrão PREMIS. Se esta unidade for utilizada, a informação de Direitos que se pretende caracterizar dependerá de um ou mais padrões de metadados diferentes de PREMIS.

No último caso, se nenhum dos dois contêineres, *rightsStatement* ou *rightsExtension,* for utilizado, isso significa que o repositório não registra informação sobre Direitos com metadados PREMIS. Nessa hipótese, a entidade *Rights* não deve ser incluída nos metadados de preservação. O que não é recomendado, pois o padrão PREMIS encoraja que os repositórios registrem informação de Direitos. Entretanto, em alguns casos, isso pode ser impossível. Por exemplo, quando a informação é desconhecida.

É de interesse prioritário neste estudo, portanto, a unidade *rightsStatement* cuja definição é "Documentação de Direitos ou de restrições de fato do repositório para performar um ou mais atos" (PREMIS EDITORIAL COMMITTEE, 2015, p. 181). Os componentes semânticos de *rightsStatement* são: *rightsStatementIdentifier, rightsBasis, copyrightInformation, licenseInformation, statuteInformation, otherRightsInformation, rightsGranted, linkingObjectIdentifier, linkingAgentIdentifier.* Todos esses componentes são contêineres, isto é, compreendem outras unidades semânticas, à exceção de um deles: *rightsBasis,* que possibilita o registro do fundamento legal ou normativo da declaração de Direitos, objeto da próxima seção do relatório.

A Figura 2 reúne todas as unidades semânticas de *Rights* organizada em forma de classificação. As letras "O" ou "M", entre parênteses, significam que a unidade é opcional ou mandatória. As letras "R" ou "NR", indicam se é repetível ou não repetível.

Figura 2. Unidades semânticas da entidade Rights de PREMIS

4.1 rightsStatement (O, R)
- 4.1.1 rightsStatementIdentifier (M, NR)
  - 4.1.1.1 rightsStatementIdentifierType (M, NR)
  - 4.1.1.2 rightsStatementIdentifierValue (M, NR)
- 4.1.2 rightsBasis (M, NR)
- 4.1.3 copyrightInformation (O, NR)
  - 4.1.3.1 copyrightStatus (M, NR)
  - 4.1.3.2 copyrightJurisdiction (M, NR)
  - 4.1.3.3 copyrightStatusDeterminationDate (O, NR)
  - 4.1.3.4 copyrightNote (O, R)
  - 4.1.3.5 copyrightDocumentationIdentifier (O, R)
    - 4.1.3.5.1 copyrightDocumentationIdentifierType (M, NR)
    - 4.1.3.5.2 copyrightDocumentationIdentifierValue (M, NR)
    - 4.1.3.5.3 copyrightDocumentationRole (O, NR)
  - 4.1.3.6 copyrightApplicableDates (O, NR)
    - 4.1.3.6.1 startDate (O, NR)
    - 4.1.3.6.2 endDate (O, NR)
- 4.1.4 licenseInformation (O, NR)
  - 4.1.4.1 licenseDocumentationIdentifier (O, R)
    - 4.1.4.1.1 licenseDocumentationIdentifierType (M, NR)
    - 4.1.4.1.2 licenseDocumentationIdentifierValue (M, NR)
    - 4.1.4.1.3 licenseDocumentationRole (O, NR)
  - 4.1.4.2 licenseTerms (O, NR)
  - 4.1.4.3 licenseNote (O, R)
  - 4.1.4.4 licenseApplicableDates (O, NR)
    - 4.1.4.4.1 startDate (O, NR)
    - 4.1.4.4.2 endDate (O, NR)
- 4.1.5 statuteInformation (O, R)
  - 4.1.5.1 statuteJurisdiction (M, NR)
  - 4.1.5.2 statuteCitation (M, NR)
  - 4.1.5.3 statuteInformationDeterminationDate (O, NR)
  - 4.1.5.4 statuteNote (O, R)
  - 4.1.5.5 statuteDocumentationIdentifier (O, R)
    - 4.1.5.5.1 statuteDocumentationIdentifierType (M, NR)
    - 4.1.5.5.2 statuteDocumentationIdentifierValue (M, NR)
    - 4.1.5.5.3 statuteDocumentationRole (O, NR)
  - 4.1.5.6 statuteApplicableDates (O, NR)
    - 4.1.5.6.1 startDate (O, NR)
    - 4.1.5.6.2 endDate (O, NR)
- 4.1.6 otherRightsInformation (O, NR)
  - 4.1.6.1 otherRightsDocumentationIdentifier (O, R)
    - 4.1.6.1.1 otherRightsDocumentationIdentifierType (M, NR)
    - 4.1.6.1.2 otherRightsDocumentationIdentifierValue (M, NR)
    - 4.1.6.1.3 otherRightsDocumentationRole (O, NR)
  - 4.1.6.2 otherRightsBasis (M, NR)
  - 4.1.6.3 otherRightsApplicableDates (O, NR)
    - 4.1.6.3.1 startDate (O, NR)
    - 4.1.6.3.2 endDate (O, NR)
  - 4.1.6.4 otherRightsNote (O, R)
- 4.1.7 rightsGranted (O, R)
  - 4.1.7.1 act (M, NR)
  - 4.1.7.2 restriction (O, R)
  - 4.1.7.3 termOfGrant (O, NR)
    - 4.1.7.3.1 startDate (M, NR)
    - 4.1.7.3.2 endDate (O, NR)
  - 4.1.7.4 termOfRestriction (O, NR)
    - 4.1.7.4.1 startDate (M, NR)
    - 4.1.7.4.2 endDate (O, NR)
  - 4.1.7.5 rightsGrantedNote (O, R)
- 4.1.8 linkingObjectIdentifier (O, R)
  - 4.1.8.1 linkingObjectIdentifierType (M, NR)
  - 4.1.8.2 linkingObjectIdentifierValue (M, NR)
  - 4.1.8.3 linkingObjectRole (O, R)
- 4.1.9 linkingAgentIdentifier (O, R)
  - 4.1.9.1 linkingAgentIdentifierType (M, NR)
  - 4.1.9.2 linkingAgentIdentifierValue (M, NR)
  - 4.1.9.3 linkingAgentRole (O, R)

4.2 rightsExtension (O, R)

Fonte: adaptado de PREMIS EDITORIAL COMMITTEE, 2015, p. 178-180.

Em síntese, o padrão PREMIS possibilita que o repositório documente informação sobre Direitos. Para fazê-lo com os próprios recursos semânticos de PREMIS, o contêiner *rightsStatement* precisa ser utilizado. Isso é feito em detrimento de *rightsExtension,* pois *rightsStatement* é a unidade semântica de PREMIS para registrar uma declaração de direitos utilizando metadados disponibilizados pelo próprio padrão. Do mesmo modo, *rightsBasis* é a unidade semântica usada para registrar o fundamento da declaração de Direitos.

Em decorrência das declarações de Direitos registradas para um documento digital, um repositório pode, em tese, estabelecer permissões e restrições sobre esse documento (MCLELLAN, 2016, p. 152):

> uma vez que um fundamento de Direitos [rightsBasis] é determinado, permissões e restrições podem ser registradas e associadas com aquele fundamento. PREMIS faz isso permitindo que um ou mais atos possam ser associados com um fundamento de Direitos [rightsBasis] e que cada ato possa ser refinado por restrições, se existir alguma. Portanto, uma simples declaração de Direitos em uma implementação de PREMIS pode consistir em um rightsBasis (como *Copyright*); um ato (como replicar) e informação sobre quaisquer restrições em relação a este ato (por exemplo, replicação permitida apenas para a finalidade de produzir cópias de preservação).

PREMIS possibilita que um mesmo objeto de preservação esteja sujeito a diferentes declarações de Direitos, com distintos fundamentos legais e, consequentemente, diversas permissões e restrições. Contudo, observa-se que o registro de permissões e de restrições, para atingir o seu pleno potencial em um repositório, exige que os *softwares* que implementam PREMIS tenham certas funcionalidades. Por exemplo, sejam capazes de definir quais metadados de Direitos serão efetivamente utilizados pelo *software* para subsidiar processos de automação e, além disso, quais serão preenchidos com conteúdo textual que não terá esta finalidade (MCLELLAN, 2016).

No decorrer das investigações em documentação e dos testes nos *softwares* realizados no âmbito deste estudo, percebeu-se que o código de Archivematica e do AtoM prevê esta distinção entre quais valores de PREMIS devem ser textuais e quais devem ser vocabulários ou valores controlados. Não obstante, os processos de automação de ações desses *softwares* com base em metadados PREMIS de Direitos não estão plenamente desenvolvidos, como evidenciado pelos testes descritos nas seções subsequentes. Esta informação é relevante porque o desenvolvimento dos metadados PREMIS de Direitos está vinculado ao objetivo de automação das ações do repositório (COYLE, 2006), para além da documentação visando à informação humana.

3.1.1 Unidade semântica "rightsBasis" de PREMIS Rights

A unidade *rightsBasis* é, de certo modo, o núcleo do *rightsStatement* porque *rightsBasis* permite que o repositório registre qual é o fundamento legal ou normativo de uma *rightsStatement.* Entre as notas de uso de *rightsStatement*,

destaca-se que este metadado deve ser repetido em duas circunstâncias: quando o(s) ato(s) (parte de *rightsGranted*) que se pretende descrever tiverem mais de um fundamento (*rightsBasis*) ou quando diferentes atos têm diferentes fundamentos.

O padrão PREMIS prevê quatro diferentes tipos de fundamentos (*rightsBasis*) para uma declaração de Direitos (*rightsStatement*): *copyright*, *license, statute* e *other*. **Esta informação é de suma importância para um RDC-Arq que contém processos judiciais e outros documentos públicos por dois motivos centrais.** Primeiro, ao fazer uma declaração de direitos sobre um documento, o repositório precisa definir qual é a categoria na qual se enquadra o fundamento legal ou normativo de sua declaração de direitos. Segundo, o valor atribuído à unidade *RightsBasis* definirá se a declaração de Direitos terá informação sobre direitos autorais (*copyrightInformation*), sobre licenças (*licenseInformation*), sobre legislação (*statuteInformation*) ou outro tipo de Direitos (*otherRightsInformation*). Por exemplo, se o valor atribuído a *rightsBasis* for *copyright,* o contêiner *copyrightInformation* será usado por AtoM ou por Archivematica para caracterizar a informação de *copyright,* e assim respectivamente para cada uma das quatro hipóteses.

Estabelecer qual dos quatro tipos de fundamento (*rightsBasis*) que se aplica em uma declaração de Direitos sobre um objeto de preservação requer conhecimento a respeito das normas ou regras que se pretende registrar com os metadados. Essa tarefa é ainda mais complexa quando o ordenamento jurídico nacional é diferente daquele de origem do padrão PREMIS. Os termos *copyright, statute* e *license* têm sentidos específicos no ordenamento jurídico dos Estados Unidos da América, país de origem das organizações da maioria dos membros que desenvolveram inicialmente o padrão de metadados, reunidos no PREMIS *Working Group*

Copyright é, sucintamente, "o direito legal exclusivo para reproduzir, publicar, vender, ou distribuir a matéria e forma de alguma coisa" (*COPYRIGHT*, 2022). Esse direito é garantido por meio do *U.S. Copyright Act,* uma legislação federal que resguarda "todos os trabalhos autorais fixados em um meio de expressão tangível e que estejam dentro da matéria de *copyright* [...]" (*COPYRIGHT*, 2022).

Statute é uma

> lei promulgada por uma legislatura. Os statutes são também chamados atos, tais como o Civil Rights Act de 1964 ou o Sarbanes-Oxley Act. Leis federais devem ser passadas pelas duas casas do Congresso, a Casa dos Representantes e o Senado, e elas geralmente precisam da aprovação do Presidente antes de entrarem em vigor (STATUTE, 2020).

Por fim, *License,* refere-se geralmente a

> [...] uma permissão concedida por uma autoridade qualificada permitindo ao titular da licença [licensee] fazer alguma coisa que poderia de outro modo ser proibido. [...] Na legislação de propriedade intelectual ("IP"), licenças permitem a indivíduos e entidades usarem patentes, trademarks e obras criativas de propriedade de outros sem infringir os direitos de propriedade de seus detentores (LICENSE, 2020).

De acordo com o dicionário de dados de PREMIS (PREMIS EDITORIAL COMMITTEE, 2015), *copyrightInformation* compreende informação sobre o *status* de *copyright* de um ou mais objetos de preservação do repositório, cuja tradução para o português é, usualmente, "direitos autorais". *LicenseInformation* refere-se a uma licença ou a um acordo que garantiu determinadas permissões relacionadas a um objeto de preservação, por exemplo, um documento. *statuteInformation* possibilita o registro de informação sobre a legislação que permite o uso do objeto de preservação. Como o próprio termo indica, *otherRightsInformation* permite caracterizar informação sobre Direitos que se aplicam ao objeto de preservação e que não se enquadram em nenhuma das três categorias anteriores.

Se a declaração de Direitos (*rightsStatement*) tiver como fundamento (*rightsBasis*) o *copyright,* o contêiner implementado (*copyrightInformation*) possibilitará o registro de uma série de informações, a começar pelo *status* do objeto no momento de registro da informação de *copyright,* que poderá ser protegido por *copyright,* desconhecido ou em domínio público. O manual *Understanding* PREMIS (CAPLAN, 2021b), em suas versões em espanhol (CAPLAN, 2009b), francês (CAPLAN, 2009a) e português (CAPLAN, 2021a) não traduziram o termo *copyright.*

Por outro lado, se a declaração de Direitos (*rightsStatement*) tiver como fundamento (*rightsBasis*) uma *license,* o contêiner implementado *licenseInformation* possibilitará o registro de informações sobre a licença. Por exemplo, permitirá a inserção de texto sobre o conteúdo da licença que atribui determinadas permissões em relação ao objeto de preservação. Esse texto pode ser, inclusive, a reprodução literal do texto do documento da licença. O manual *Understanding* PREMIS (CAPLAN, 2021b), em suas versões em espanhol (CAPLAN, 2009b), francês (CAPLAN, 2009a) e português (CAPLAN, 2021a) traduziram o termo *license,* respectivamente, para *licencia, licence* e licença.

Em uma terceira hipótese, se a declaração de Direitos (*rightsStatement*) tiver como fundamento (*rightsBasis*) um *statute,* então, PREMIS registrará informação sobre o " *statute* " (*statuteInformation*), por exemplo, a jurisdição e o nome do " *statute* ". Nesta hipótese, há uma questão de tradução relevante para o problema investigado. Em português, traduziu-se literalmente o termo *statute* para "estatuto" no *Understanding* PREMIS em português (CAPLAN, 2021a) e também na tradução para o português de versões dos *softwares* AtoM (2.7.1) e Archivematica (1.13.2). Contudo, as versões de *Understanding* PREMIS em espanhol (CAPLAN, 2009b) e em francês (CAPLAN, 2009a) apresentam o termo *statute,* respectivamente, como *legislación* e *texte législatif,* mais fiéis ao sentido do termo original que define *statute* como uma "lei promulgada por uma legislatura". Dessa forma, *statute* é melhor compreendido em português como "legislação".

Para fins de documentar a LGPD e outras leis relevantes, como a Lei de Acesso à Informação, o termo *statute* parece ser semanticamente o mais adequado em uma declaração de Direitos com metadados PREMIS. Por outro lado, normas infralegais, tais como decretos, resoluções, portarias e instruções normativas, parecem ter como fundamento (*rightsBasis*) categoria não especificada no vocabulário de PREMIS (" *other* ").

Ressalta-se que AtoM e Archivematica utilizam além das quatro categorias de fundamentos de Direitos (*rightsBasis*) discutidas nesta seção, outras duas: *policy* e *donor,* não previstas originalmente no padrão PREMIS. Não obstante, a escolha de uma delas para fazer uma declaração de Direitos utilizando o Archivematica resulta, na prática, no registro do fundamento de Direitos (*rightsBasis*) com o valor " *other* ", previsto em PREMIS.

Entende-se que a administração do repositório deve definir se utilizará uma alternativa provida pelo *software* ou não. Na primeira hipótese, poderá registrar o fundamento de declarações de Direitos baseadas em normas infralegais como " *policy* ", na segunda, registrará como " *other* " de modo fiel à terminologia de PREMIS. Em qualquer um dos casos, não haverá repercussões imediatas na automação de ações do repositório com funcionalidades nativas de Archivematica e AtoM, no entanto, recomenda-se que o repositório tenha um padrão convencionado.

3.1.2 Documentando informação sobre LGPD com PREMIS Rights

Conforme veiculado no *site* oficial mantido pela Biblioteca do Congresso dos EUA: "[...] PREMIS é implementado em projetos de preservação digital ao redor do mundo, e apoio para PREMIS é incorporado em um grande número de ferramentas e sistemas de preservação comerciais e código aberto" (LIBRARY OF CONGRESS, 2022), a exemplo de AtoM e Archivematica, utilizados pelo modelo de repositório arquivístico digital Hipátia.

Archivematica e AtoM podem apoiar, até certo ponto, o registro e o controle de informação sobre LGPD e outras leis e normas de interesse por meio da entidade *Rights* do padrão de metadados PREMIS. Para isso, utiliza-se modelo de arquivo com metadados PREMIS *Rights* e seus respectivos valores seguindo as especificações da documentação do Archivematica (ARTEFACTUAL SYSTEMS, *s. d.*). Este arquivo deve ser uma planilha com até 18 colunas, uma para cada elemento de descrição de Direitos previsto no Archivematica.

No Quadro 1, representa-se o termo de topo de cada uma das 18 colunas, o tipo de valores esperado nessa coluna e o valor efetivamente utilizado nos testes de estratégias de disseminação realizados no presente estudo (seção 3.3).

Quadro 1. Estrutura do arquivo *rights.csv*

Título da coluna em *rights.csv*	Tipo de valor esperado	Valor utilizado nos testes de estratégias de disseminação de processo judicial
file	Nome do arquivo	objects/processo_trabalhista.pdf
basis	O fundamento/base do Direito [*rightsBasis*]: *copyright, statute, license, policy, donor, other.*	statute
status	*copyrighted, public domain, unknown.*	[não se aplica ao fundamento " *statute* ", deixar em branco]
determination_date	Data em que se determinou que a legislação autorizava a permissão em questão, no formato ISO 8061: AAAA-MM-DD.	2023-06-19
jurisdiction	País ou corpo político que promulgou a lei conforme vocabulário controlado ISO 3166.	BR
start_date	Data de início da aplicação da legislação ao conteúdo, no formato ISO 8061: AAAA-MM-DD.	2022-09-18
end_date	Data em que a legislação deixou de se aplicar ao conteúdo, no formato ISO 8061: AAAA-MM-DD.	[não se aplica, pois LGPD permanece em vigor]
terms	Descrição de texto de licença ou de acordo que atribuiu uma determinada permissão.	[não se aplica ao fundamento " statute ", deixar em branco]
citation	Designação da legislação.	Lei nº 13.709, de 14 de agosto de 2018 - Lei Geral de Proteção de Dados Pessoais (LGPD)
note	Informação adicional sobre a legislação.	Registro de aplicação da LGPD obrigatória no repositório, conforme determinação do Comitê Gestor de Proteção de Dados Pessoais (CGPD) do Tribunal, criado pela Resolução CNJ nº 363, de 12 de janeiro de 2021.
grant_act	*delete, disseminate, migrate, modify, modify, replicate, use*	*disseminate*

Título da coluna em *rights.csv*	Tipo de valor esperado	Valor utilizado nos testes de estratégias de disseminação de processo judicial
grant_restriction	*Allow, Disallow, Conditional*	*Disallow*
grant_start_date	Data de início da permissão ou restrição, no formato ISO 8061: AAAA-MM-DD.	2023-06-19
grant_end_date	Data de fim da permissão ou da restrição, no formato ISO 8061: AAAA-MM-DD.	[não se aplica, pois o prazo é indefinido]
grant_note	Informação adicional sobre a permissão/restrição.	[não utilizado]
doc_id_type	Compõem o container *statuteDocumentationIdentifi er* que compreende "uma designação usada para identificar de forma unívoca a documentação que apoio os Direitos especificados atribuídos pela legislação dentro do sistema do repositório" (PREMIS EDITORIAL COMMITTEE, 2015, p. 213). É opcional.	[não utilizados no teste porque são opcionais]
doc_id_value		
doc_id_role		

Fonte: elaborado pelo autor com base em testes realizados em instância do *software* Archivematica 1.13.2 e na sua documentação oficial.

O Quadro 1 representa apenas uma proposta exemplificativa de como elaborar um arquivo *rights.csv*. Os valores efetivos atribuídos aos metadados precisam estar de acordo com as determinações da administração do repositório.

Nas seções subsequentes, identifica-se e discute-se funcionalidades de AtoM e de Archivematica, inclusive, aquelas que possibilitam o registro de metadados PREMIS de Direitos com o modelo exemplificado no Quadro 1. A finalidade de identificar e analisar essas funcionalidades (objetivos específicos 3 e 4) é subsidiar os testes realizados nos *softwares* (objetivo específico 5). Utiliza-se o modelo do Quadro 1, apresentado nesta seção, integrado aos testes das funcionalidades, cujos resultados compõem a parte final deste relatório.

3.2 Funcionalidades dos *softwares* Archivematica e AtoM

Na presente subseção, identifica-se e procede-se à análise de funcionalidades do *software* Archivematica (subseções 3.2.1 e 3.2.2) e AtoM (subseções 3.2.3 e 3.2.4) que podem auxiliar no acesso à informação respeitando hipóteses de sigilo e de proteção de dados com diferentes finalidades.

3.2.1 Identificação das funcionalidades do *software* Archivematica

O Quadro 2 contém uma lista de funcionalidades identificadas no *software* Archivematica com potencial para auxiliar na problemática ora em análise.

Quadro 2. Funcionalidades identificadas no Archivematica, associadas à respectiva aba do *software* e ao link da documentação

Funcionalidades	Aba(s) da *dashboard*	Links da documentação associada
1. Importar metadados de Direitos em arquivo CSV	Transferir [*Transfer*]	https://www.archivematica.org/en/docs/archivematica-1.13/user-m anual/transfer/transfer/#transfers-with-metadata https://www.archivematica.org/en/docs/archivematica-1.13/user-m anual/transfer/import-metadata/#rights-csv https://www.archivematica.org/pt-br/docs/archivematica-1.13/user manual/metadata/premis/#premis-rights
2. Adicionar manualmente metadados de direitos (Rights)	Admissão [*Ingest*]	https://www.archivematica.org/pt-br/docs/archivematica-1.13/user manual/ingest/ingest/#add-rights https://www.archivematica.org/pt-br/docs/archivematica-1.13/user manual/metadata/premis/#premis-rights
3. Transferir material com derivadas de acesso	Transferir [*Transfer*]	https://www.archivematica.org/en/docs/archivematica-1.13/user-m anual/transfer/transfer/#transfer-derivatives https://www.archivematica.org/en/docs/archivematica-1.13/user-m anual/ingest/manual-normalization/#man-norm-before

Funcionalidades	Aba(s) da *dashboard*	Links da documentação associada
4. Transferir material com matrizes secundárias [*mezzanine files / service files*]	Transferir [*Transfer*]; Admissão [*Ingest*]	https://www.archivematica.org/pt-br/docs/archivematica-1.13/user manual/transfer/transfer/#transfer-service-files https://www.archivematica.org/en/docs/archivematica-1.13/user-m anual/ingest/manual-normalization/#man-norm-before
5. Examinar conteúdo (*examine contents*) no processamento de transferência	Transferir [*Transfer*]; Administração [*Administration*]	https://www.archivematica.org/pt-br/docs/archivematica-1.13/user manual/administer/dashboard-admin/#dashboard-processing
6. Transcrever arquivos (*Transcribe files OCR*) no processamento de transferência	Transferir [*Transfer*]; Administração [*Administration*]	https://www.archivematica.org/pt-br/docs/archivematica-1.13/user manual/ingest/inges-t/#transcribe-contents
7. Arquivo de configuração de transferência	Transferir [*Transfer*]; Admissão [*Ingest*]	https://www.archivematica.org/pt-br/docs/archivematica-1.13/user manual/administer/dashboard-admin/#dashboard-processing
8. Fazer o *upload* somente de metadados para o AtoM [*Metadata-only upload to AtoM*]	Armazenamento de arquivo [*Archival storage*]; Acesso [*Access*]	https://www.archivematica.org/pt-br/docs/archivematica-1.13/user manual/access/access/#upload-metadata-AtoM
9. Criptografar AIP Armazenamento	de arquivo [*Archival storage*]; Administração [*Administration*];	https://www.archivematica.org/pt-br/docs/archivematica-1.13/user manual/archival-storage/archival-storage/#aip-encryption https://www.archivematica.org/en/docs/storage-service-0.19/admini strators/#gpg

Fonte: elaborado pelo autor com base em testes realizados em instância do *software* Archivematica 1.13.2 e na sua documentação oficial.

As primeiras duas funcionalidades do Quadro 2 referem-se aos metadados de Direitos [*Rights*], tratados na seção 1 do relatório. No Archivematica, é possível fazer a importação de metadados de Direitos contidos em uma planilha eletrônica de formato CSV, transferida junto com os objetos de preservação. A planilha deve seguir o modelo padrão disponível na documentação do

software, conforme os termos da coluna da esquerda do Quadro 1. Os valores contidos no arquivo *rights.csv* serão incorporados aos metadados do objeto no repositório. Alternativamente, é possível inserir manualmente valores para metadados de Direitos (entre outros tipos) durante o processo de admissão [*ingest*] no repositório.

A terceira e a quarta funcionalidades do Quadro 2 referem-se à normalização manual de objetos de preservação antes da admissão no repositório. Em uma das hipóteses, as matrizes secundárias, também chamadas de *mezzanine* ou *service files* pela documentação do Archivematica, servirão como representantes da matriz primária (ou "arquivo original") ao serem reconhecidas pelo *software*. Nesse caso, quando os objetos forem submetidos ao microsserviço de "normalização", na aba de "Admissão", o Archivematica utilizará as matrizes secundárias fornecidas pelo usuário, e não as primárias, para produzir derivadas de acesso. De modo análogo, o Archivematica pode reconhecer e utilizar derivadas de acesso também fornecidas pelo usuário ao invés de produzi-las a partir da matriz primária, como destaca a quarta funcionalidade do Quadro 2.

A funcionalidade cinco trata do microsserviço de "exame de conteúdo", executado durante a transferência, que se refere à possibilidade de autorizar o uso da ferramenta " *bulk extractor* " (GARFINKEL, 2023), responsável por realizar o reconhecimento de padrões nos dados. De modo contrário, a ferramenta de OCR que é mencionada na funcionalidade seis do Quadro 2 pode, eventualmente, prejudicar uma tentativa de proteger dados pessoais ao invés de auxiliar.

A sétima funcionalidade do Quadro 2 permite incluir um arquivo *processingMCP.xml* na transferência com a finalidade de automatizar as decisões tomadas no processo. Uma oitava funcionalidade permite que o usuário envie para o AtoM apenas metadados, ao invés de enviar derivadas de acesso.

Por fim, a nona funcionalidade identificada no Quadro 2 possibilita a criptografia do pacote de arquivamento de informação do repositório, o AIP, com o intuito de reforçar a segurança da informação armazenada.

3.2.2 Análise das funcionalidades do *software* Archivematica

Nesta seção do relatório desenvolve-se uma análise das funcionalidades identificadas por meio da descrição de hipóteses que podem auxiliar na proteção de

dados pessoais. Os objetivos das funcionalidades 1 e 2 do Quadro 2, que são registrar metadados de Direitos, já foram objeto da seção 3.1, que trata de PREMIS *Rights*.

A normalização manual e prévia de objetos digitais, indicada nas funcionalidades 3 e 4 do Quadro 2, apresenta-se como uma alternativa para proteger dados inseridos no AIP, tornando-os indisponíveis no DIP. No caso de processos judiciais digitalizados e preparados previamente para admissão no repositório, a produção prévia de matrizes secundárias ou de derivadas de acesso permite que sejam realizados *workflows* de ocultação de dados pessoais sem alteração do "arquivo original". A matriz primária (ou "arquivo original") ficaria armazenada apenas no ambiente de preservação, encapsulada no AIP (de acesso restrito).

O exame de conteúdo (funcionalidade 5 do Quadro 2) pode, eventualmente, servir como uma ferramenta de apoio à localização de dados a serem protegidos, desde que a ferramenta *bulk extractor* seja antes devidamente configurada e testada para essa finalidade, o que requer um estudo específico. De acordo com a documentação do Archivematica, o *bulk extractor* é capaz de localizar padrões como números de cartão de crédito e de seguridade social. Contudo, dificilmente o exame de conteúdo proporcionado por essa ferramenta poderá garantir, por si só, que um processo possa ser amplamente divulgado por causa da quantidade virtualmente ilimitada de tipos de dados que podem estar documentados, por exemplo, em um processo judicial trabalhista. Ademais, é preciso submeter a ferramenta *bulk extractor* a testes específicos antes que se possa avançar nas conclusões sobre sua utilidade no contexto da problemática investigada. Nos testes do presente estudo, não se utilizou o *bulk extractor* porque ampliaria demasiadamente o foco da investigação.

A transcrição de conteúdo do SIP por meio de OCR (*Optical Character Recognition*), microsserviço 6 do Quadro 2, pode ser potencialmente prejudicial para a proteção de dados pessoais porque arquivos textuais contendo a transcrição dos conteúdos do objeto digital são adicionados ao AIP e ao DIP nesse processo. Em cenários nos quais o objetivo for manter certos dados restritos, pode ser necessário que esta funcionalidade seja desabilitada. Todavia o microsserviço afeta apenas imagens JPG e TIFF, pelo que foi possível localizar na documentação. Os AIPs de arquivos em PDF dos testes realizados no presente estudo corroboram esta afirmação porque não continham transcrições mesmo com esta opção ativada na admissão [*ingest*] no repositório.

Utilizar um arquivo de configuração no processo de transferência para o Archivematica, indicado na funcionalidade 7 do Quadro 2, é um meio de evitar a execução indesejada de microsserviços que possam expor indevidamente dados a serem protegidos. Isso é feito por meio da inclusão de um arquivo *processing-MCP.xml.* Ao mesmo tempo, também é uma forma de garantir que microsserviços necessários à proteção de dados sejam efetivamente acionados.

A funcionalidade 8 do Quadro 2 refere-se à possibilidade de *upload* para o AtoM somente de metadados. Desse modo, assim como nas funcionalidades 3 e 4, é uma alternativa para proteger dados inseridos no AIP, tornando-os ausentes no DIP. Na funcionalidade 4 do Quadro 2 isso é feito por meio de "arquivos de serviço" (matrizes secundárias). Na funcionalidade 3 do Quadro 2, realiza-se uma possível restrição de acesso a dados por meio de derivadas produzidas previamente e incluídas na transferência para o Archivematica. O *upload* somente de metadados é uma funcionalidade limitada porque não suporta o envio de metadados descritivos, tampouco de níveis de descrição arquivística, conforme a documentação do Archivematica 1.13. Portanto, trata-se tão somente da disponibilização, na interface do AtoM, de metadados técnicos do objeto digital. Um usuário do AtoM com as permissões necessárias precisará descrever e classificar corretamente esses itens após o *upload.*

Por fim, a funcionalidade 9 do Quadro 2, que se refere à criptografia do AIP, permite que dados protegidos tenham uma camada extra de segurança. Se a criptografia for empregada, o *download* do AIP descriptografado pode ainda ser feito por meio do *storage service* do Archivematica de acordo com a documentação.

3.2.3 Identificação das funcionalidades do *software* AtoM

O Quadro 3 contém uma lista de funcionalidades identificadas no *software* AtoM e que têm potencial para auxiliar na problemática investigada.

Quadro 3. Funcionalidades identificadas no AtoM associadas às respectivas páginas da aplicação e à documentação

Recurso ou funcionalidade	Página do AtoM	Links da documentação associada
1. Metadados de Direitos [*Rights*]: apagar, explorar, **exibir**, disseminar, migrar, modificar, replicar	Configurações de Permissões do Administrador; Criar novos direitos na página da descrição	https://www.accesstomemory.org/en/docs/2.7/user-manual/a dd-edit-content/rights/#
2. Notificação de privacidade	Páginas estáticas	https://www.accesstomemory.org/en/docs/2.7/user-manual/a dminister/static--pages/ https://www.accesstomemory.org/en/docs/2.7/user-manual/a dminister/settings/#privacy-notification
3. Cookies	Coleta de dados	https://www.accesstomemory.org/en/docs/2.7/user-manual/a dminister/settings/#privacy-notification
4. Google Analytics	Coleta de dados	https://www.accesstomemory.org/en/docs/2.7/user-manual/a dminister/settings/#privacy-notification
5. Dados de usuários registrados	Coleta de dados	https://www.accesstomemory.org/en/docs/2.7/user-manual/a dminister/settings/#privacy-notification

Fonte: elaborado pelo autor com base em testes realizados em instância do *software* AtoM 2.7.1 e na sua documentação oficial.

O AtoM permite a configuração de restrições de acesso à representação digital de um documento por meio da configuração de permissões de Direitos. Contudo, nem todos os elementos relativos a Direitos que são visíveis em páginas de configuração do AtoM estão funcionais. Isto é, administradores são capazes de visualizar opções de configurações de Direitos que ainda não foram totalmente desenvolvidas no *software* e, portanto, não resultam em ações efetivas do sistema.

Além das questões de Direitos, o *software* tem uma funcionalidade criada para que administradores possam cumprir leis e regulamentos que exigem que o *site* tenha uma postura ativa em informar para o usuário, já no início da sessão, a política de privacidade. Isso é importante porque o AtoM utiliza *cookies*, pode também utilizar o *Google Analytics* e, além disso, armazena alguns dados de usuários registrados, que têm *login* e senha (compõem o grupo de usuários "autenticados").

3.2.4 Análise das funcionalidades do *software* AtoM identificadas

No que diz respeito à notificação de privacidade, a mensagem possibilita que administradores criem seu próprio texto para informar de modo pró-ativo para os usuários sobre a política de privacidade ao acessarem o AtoM. Uma mensagem exemplificativa, em inglês, está disponível na instalação padrão.

Habilita-se a "notificação de privacidade" no menu de "Definições" [*Settings*] do Administrador. Se a mensagem estiver ativa, o usuário que acessar o *site* visualizará o texto correspondente em um *banner* já na página inicial. Por padrão, a notificação contém uma frase em inglês informando que o *site* utiliza *cookies* para melhorar a capacidade do navegador de carregar conteúdos. Esta mensagem está acompanhada de um *link* que leva o usuário para a página estática da "Política de Privacidade" que contém um texto explicativo editável. Do lado do administrador, a Política de Privacidade é acessível pelo menu "Administrador", em "Páginas estáticas", onde pode ser editada ou excluída. Destaca-se que textos editáveis, no AtoM, aceitam a inserção de *links* por meio do uso da linguagem de marcação *markdown,* o que favorece a customização.

Nas mensagens nativas da instalação padrão do AtoM e na documentação do *software* há informações importantes a respeito de privacidade. No que diz respeito ao item 3 do Quadro 3, os *cookies,* não existe funcionalidade padrão no AtoM que permita aos usuários visitantes desabilitarem a coleta de *cookies* por meio de configurações da própria aplicação, como ocorre em outros *sites* que possibilitam ao usuário definir quais recursos deseja desligar. A documentação do AtoM 2.7 recomenda que usuários que não desejam *cookies* configurem seu navegador de Internet para não aceitá-los. Todavia, certas funcionalidades do AtoM não funcionarão nessa hipótese, por exemplo, a "Área de Transferência" [*Clipboard*], que serve para que o usuário salve e exporte conteúdos de interesse localizados durante suas pesquisas.

Em relação ao *Google Analytics,* o AtoM permite que seja configurada integração com esse recurso para fins de produção de estatísticas sobre páginas acessadas, uso do *site,* etc. Esses dados são armazenados pelo Google, e não pelo AtoM, logo, estão sujeitos aos termos de uso daquela empresa. O que a documentação do AtoM da versão 2.7 afirma é que os dados coletados em função de *cookies* ou do *Google Analytics* não são utilizados para nenhum outro fim, exceto os fins informados.

Ressalta-se que, conforme resolvido pelo CNJ, os tribunais devem estabelecer medidas para a disponibilização de "informação adequada sobre tratamento de dados pessoais nos termos do art. 9º da LGPD, por meio de: a) avisos de cookies no portal institucional de cada tribunal; b) política de privacidade para navegação na página da instituição" (CONSELHO NACIONAL DE JUSTIÇA, 2021). Entende-se que o sistema de acesso e disseminação baseado no AtoM, que é uma aplicação para a *web,* pode ser compreendido como parte do portal institucional ou página da instituição.

Sobre o item 5 do Quadro 3 (dados de usuários registrados), a documentação também informa que esses dados não são usados para nenhum outro fim senão possibilitar o *login* por parte dos usuários. Logicamente, para que isso seja possível, ao menos o e-mail, o nome de usuário e a senha devem estar armazenadas pelo AtoM. A documentação afirma que as senhas são criptografadas, contudo, nada diz sobre os nomes de usuário e os e-mails também armazenados. Por meio da interface do AtoM, um administrador autenticado não pode visualizar senhas, mas pode alterá-las, além de visualizar e editar as demais informações de cada usuário registrado.

3.3 Testes de funcionalidades

Nesta subseção, discute-se os resultados de testes realizados com os *softwares* AtoM e Archivematica. No caso do Archivematica, elaborou-se duas "estratégias de disseminação" que utilizam, cada uma delas, uma seleção de funcionalidades do Quadro 2. A primeira estratégia de disseminação transfere uma derivada de acesso previamente editada para o AtoM. Nessa hipótese, o documento digital com dados que não podem ser amplamente divulgados fica restrito ao AIP. A segunda estratégia de disseminação faz o *upload* somente de metadados para o AtoM, que não recebe nenhuma representação digital de documento que tenha dados que precisam ser protegidos.

Os testes com o *software* AtoM focaram em restrições de acesso por meio de permissões e configurações de Direitos realizadas diretamente na plataforma de acesso.

3.3.1 Testes com Archivematica: estratégias de disseminação

Na presente seção, analisa-se e discute-se duas estratégias de disseminação de informação, contendo dados com supostas restrições de acesso, usando funcionalidades do *software* Archivematica 1.13.2. Em ambos os casos, o objeto da estratégia é uma simulação de um documento "processo judicial" em formato PDF. Na primeira estratégia, designada "disseminação com transferência de derivada", incluiu-se na transferência uma "derivada de acesso" com dados pessoais do arquivo original do "processo trabalhista" previamente suprimidos.

Os objetivos dessa estratégia são: registrar de modo automatizado informação sobre Direitos no repositório por meio da inclusão de arquivo com metadados e respectivos valores no padrão PREMIS (modelo da seção 3.1 do relatório); fornecer, na transferência, uma derivada de acesso do documento original com dados previamente ocultados para fins de disseminação.

O Quadro 4 identifica as funcionalidades do Quadro 2 acionadas na "estratégia de disseminação com transferência de derivada" e justifica o porquê de acioná-las ou não.

Quadro 4. Teste da "estratégia de disseminação com transferência de derivada"

Recurso ou funcionalidade	Uso no teste da estratégia?	Justificativa
1. Importar metadados de Direitos em arquivo CSV	Sim	Documentar de modo automatizado informação sobre Direitos.
2. Adicionar manualmente metadados de direitos (*Rights*)	Não	Esta estratégia não tem como um de seus objetivos a inserção manual de informação sobre Direitos. Realiza-se este tipo de registro por meio de arquivo CSV incluso na transferência para o repositório.
3. Transferir material com derivadas de acesso	Sim	Esta estratégia não tem como um de seus objetivos transferir derivadas produzidas antes da transferência. Inclui-se na transferência uma matriz secundária que servirá como fonte para a produção de derivadas de acesso na admissão no repositório.

4. Transferir material com matrizes secundárias [*mezzanine files* / *service files*]	Não	O teste com matriz secundária (arquivo de "serviço") falhou, pois o Archivematica não realizou o upload do arquivo de serviço. São necessários mais estudos parainvestigar o motivo da falha. No entanto, os objetivos eram, na prática, os mesmos do teste com-transferência de derivada de acesso, que foi bem-sucedido.
5. Examinar conteúdo (*examine contents*) no processamento de transferência	Não	Esta estratégia não tem como um de seus objetivos utilizar a ferramenta bulk extractor para analisar dados pessoais porque não parte do princípio de que a inclusão desta ação em um pipeline pode prescindir de testes específicos que demonstrem a efetiva proteção de dados pessoais.
6. Transcrever arquivos (*Transcribe files OCR*) no processamento de transferência	Sim	Verificar se há transcrição de PDF.
7. Arquivo de configuração de transferência	Não	Esta estratégia não tem como objetivo regular a automatização de processos de transferência e admissão no repositório, embora, em tese, possa ser utilizada com alto grau de automatização.
8. Fazer o *upload* somente de metadados para o AtoM [*Metadata-only upload to AtoM*]	Não	Esta estratégia não tem como objetivo enviar apenas metadados para a plataforma de disseminação (AtoM), pois pretende também enviar arquivos no DIP.
9. Criptografar AIP	Não	Esta estratégia não tem como objetivo testar a segurança de dados pessoais armazenados no ambiente de preservação.

Fonte: elaborado pelo autor com base em testes realizados em instância do *software* Archivematica 1.13.2 e na sua documentação oficial.

No teste da "estratégia de disseminação com transferência de derivada" utilizou-se um diretório "processo" contendo: dois subdiretórios chamados, respectivamente, de "access" e de "metadata", além de um arquivo "processo_trabalhista2.pdf". O subdiretório "access" incluiu uma derivada de acesso do "processo_trabalhista2.pdf", com o mesmo nome de arquivo e extensão. O subdiretório "metadata" conteve um arquivo de planilha eletrônica "rights.csv", com metadados PREMIS de Direitos e seus respectivos valores, conforme as especificações do Quadro 1.

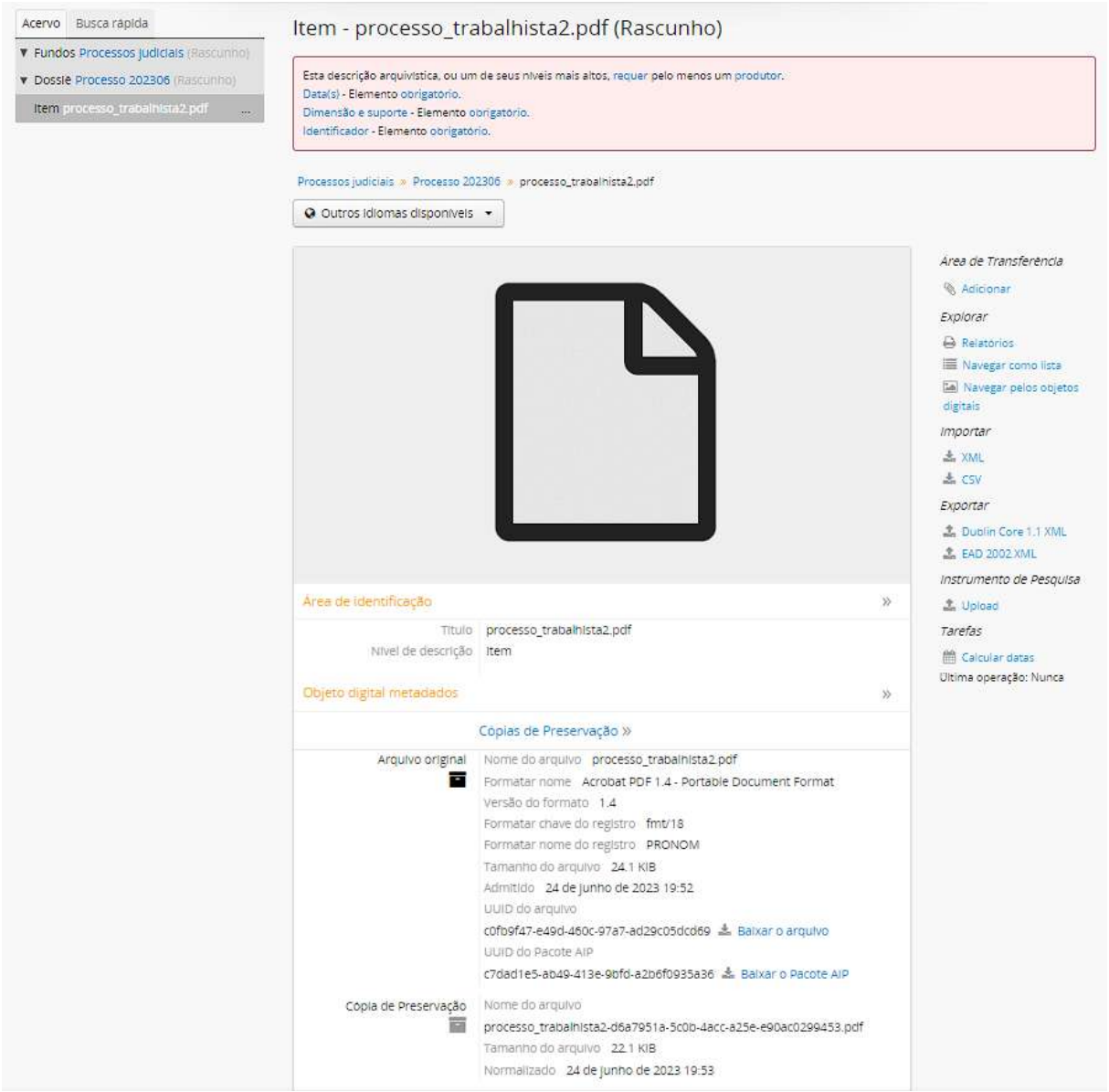
O PDF do subdiretório “access”, diferentemente do original contido na raiz do diretório “processo”, passou por um processo de edição manual prévia na qual foram suprimidos dados pessoais do texto. Essa supressão foi realizada de duas maneiras: na imagem e na camada textual do PDF, de modo a impedir a leitura humana e também a leitura e a indexação por máquina. Propositalmente, uma sequência de caracteres que simula um número de Cadastro de Pessoa Física (CPF) foi deixado no processo sem qualquer tipo de ofuscação na derivada. As duas versões do documento estão justapostas na Figura 3.

Figura 3. Matriz (esquerda) e derivada de acesso (direita) de arquivos em PDF denominados “processo_trabalhista.pdf”

Fonte: documentos elaborados pelo autor.

Designou-se, manualmente, na aba “Transferir” [*Transfer*] do Archivematica, o “nome da transferência” como “derivada2”. Seguiu-se as orientações de configuração do Quadro 4, logo, ignorou-se o exame de conteúdo (*bulk extractor*). A normalização selecionada foi a opção para preservação, pois se o sistema identifica a presença de derivada de acesso, oferece além desta opção, apenas a alternativa de não normalizar ou de rejeitar o SIP. Inseriu-se manualmente metadados descritivos (título, data e descrição), que poderiam constar em uma planilha na transferência a fim de automatizar o registro dessa informação. A resposta dada à transcrição do conteúdo do SIP foi positiva. Armazenou-se o AIP e o DIP nos respectivos locais padrão [*default location*]. Realizou-se o *upload* do DIP para o AtoM.

Figura 4. Resultado do upload de DIP com derivada de acesso em ambiente de testes do AtoM



Fonte: captura de tela de instância de testes do AtoM 2.7.1, realizada pelo autor.

No *upload* do DIP com base na "estratégia de disseminação com transferência de derivada" observou-se os seguintes resultados:

1. A derivada de acesso do subdiretório "access" foi corretamente identificada e o *upload* para o AtoM realizado com sucesso.

2. Metadados descritivos, inseridos manualmente durante o processo de admissão [*ingest*] no Archivematica, deram origem a uma descrição de nível "dossiê". A derivada de acesso "processo_trabalhista2.pdf" foi vinculada como descrição do tipo "item" a este "dossiê".

3. Os dados contidos no arquivo *rights.csv,* que seguiu o modelo do Quadro 1, foram identificados pelo Archivematica e inseridos nos metadados METS do AIP, conforme a Figura 5. Os metadados PREMIS de Direitos também estão no arquivo de metadados METS do DIP produzido.

4. Na configuração padrão, o DIP é salvo no diretório "/tmp" do servidor do AtoM e lá permanece sem ser excluído.

5. A matriz do "processo_trabalhista2.pdf", que contém dados com restrições de acesso, não é acessível pelo AtoM, exceto se estiver configurado o acesso direto de administrador ao *storage service* do Archivematica. Este é um recurso disponível a partir da versão 2.7 do AtoM. Nesta hipótese, um administrador pode fazer o *download* do AIP e do arquivo original diretamente da interface do AtoM. Demais usuários do AtoM terão acesso somente à derivada de acesso previamente editada.

6. Na presente estratégia de disseminação, o dado da derivada de acesso que permaneceu aberto na imagem e na camada de texto do PDF (uma sequência de caracteres simulando um CPF) não foi indexada pelo sistema de busca do AtoM.

Figura 5. Metadados PREMIS Rights em arquivo de metadados METS

```
<mets:rightsMD ID="rightsMD_1">
  <mets:mdWrap MDTYPE="PREMIS:RIGHTS">
    <mets:xmlData>
      <premis:rightsStatement xmlns:premis="http://www.loc.gov/premis/v3" xsi:schemaLocation=
      "http://www.loc.gov/premis/v3 http://www.loc.gov/standards/premis/v3/premis.xsd">
        <premis:rightsStatementIdentifier>
          <premis:rightsStatementIdentifierType>UUID</premis:rightsStatementIdentifierType>
          <premis:rightsStatementIdentifierValue>22f3d007-1ed6-4951-870b-6c56af66e2e3
          </premis:rightsStatementIdentifierValue>
        </premis:rightsStatementIdentifier>
        <premis:rightsBasis>Statute</premis:rightsBasis>
        <premis:statuteInformation>
          <premis:statuteJurisdiction>BR</premis:statuteJurisdiction>
          <premis:statuteCitation>Lei nº 13.709, de 14 de agosto de 2018 - Lei Geral de Proteção de
          Dados Pessoais (LGPD)</premis:statuteCitation>
          <premis:statuteInformationDeterminationDate>2023-06-19
          </premis:statuteInformationDeterminationDate>
          <premis:statuteNote>Registro de aplicação da LGPD obrigatória no repositório, conforme
          determinação do Comitê Gestor de Proteção de Dados Pessoais (CGPD) do Tribunal, criado pela
          Resolução CNJ nº 363, de 12 de janeiro de 2021.</premis:statuteNote>
          <premis:statuteApplicableDates>
            <premis:startDate>2022-09-18</premis:startDate>
          </premis:statuteApplicableDates>
        </premis:statuteInformation>
        <premis:rightsGranted>
          <premis:act>disseminate</premis:act>
          <premis:restriction>Disallow</premis:restriction>
          <premis:termOfRestriction>
            <premis:startDate>2023-06-19</premis:startDate>
          </premis:termOfRestriction>
        </premis:rightsGranted>
        <premis:linkingObjectIdentifier>
          <premis:linkingObjectIdentifierType>UUID</premis:linkingObjectIdentifierType>
          <premis:linkingObjectIdentifierValue>c0fb9f47-e49d-460c-97a7-ad29c05dcd69
          </premis:linkingObjectIdentifierValue>
        </premis:linkingObjectIdentifier>
      </premis:rightsStatement>
    </mets:xmlData>
  </mets:mdWrap>
</mets:rightsMD>
```

Fonte: captura de tela de arquivo XML contendo metadados METS produzidos pelo autor com instância de testes do Archivematica 1.13.2.

Denominou-se a segunda estratégia de “disseminação somente de metadados”. Como o nome sugere, nesta estratégia de disseminação propõe-se que não seja enviado à plataforma de disseminação (AtoM) representações digitais do “processo judicial”, apenas metadados.

Os objetivos dessa estratégia são: registrar de modo automatizado informação sobre Direitos no repositório por meio da inclusão de arquivo com metadados e seus respectivos valores no padrão PREMIS; transferir apenas metadados do ambiente de preservação para a plataforma de disseminação.

O Quadro 5 identifica quais as funcionalidades do Quadro 3 são utilizadas na “estratégia de disseminação somente de metadados” e justifica o porquê dessas escolhas.

Quadro 5. Teste da "estratégia de disseminação somente de metadados"

Recurso ou funcionalidade	Uso no teste da estratégia?	Justificativa
1. Importar metadados de Direitos em arquivo CSV	Sim	Documentar de modo automatizado informação sobre Direitos.
2. Adicionar manualmente metadados de direitos (*Rights*)	Não	Esta estratégia não tem como um de seus objetivos a inserção manual de informação sobre Direitos. Realiza-se este tipo de registro por meio de arquivo CSV incluso na transferência para o repositório.
3. Transferir material com derivadas de acesso	Não	Não é objetivo desta estratégia produzir ou transferir representações digitais.
4. Transferir material com matrizes secundárias [*mezzanine files* / *service files*]	Não	Não é objetivo desta estratégia produzir ou transferir representações digitais.
5. Examinar conteúdo (*examine contents*) no processamento de transferência	Não	Esta estratégia não tem como um de seus objetivos utilizar a ferramenta *bulk extractor* para analisar dados pessoais porque não envia representações digitais para disseminação.
6. Transcrever arquivos (*Transcribe files OCR*) no processamento de transferência	Sim	Verificar se há transcrição de PDF.
7. Arquivo de configuração de transferência	Não	Esta estratégia não tem como objetivo regular a automatização de processos de transferência e admissão no repositório, embora, em tese, possa ser utilizada com alto grau de automatização.
8. Fazer o *upload* somente de metadados para o AtoM [*Metadata-only upload to AtoM*]	Sim	Enviar apenas metadados para a plataforma de disseminação (AtoM), restringindo dados pessoais às representações digitais do AIP.
9. Criptografar AIP	Não	Esta estratégia não tem como objetivo testar a segurança de dados pessoais armazenados no ambiente de preservação.

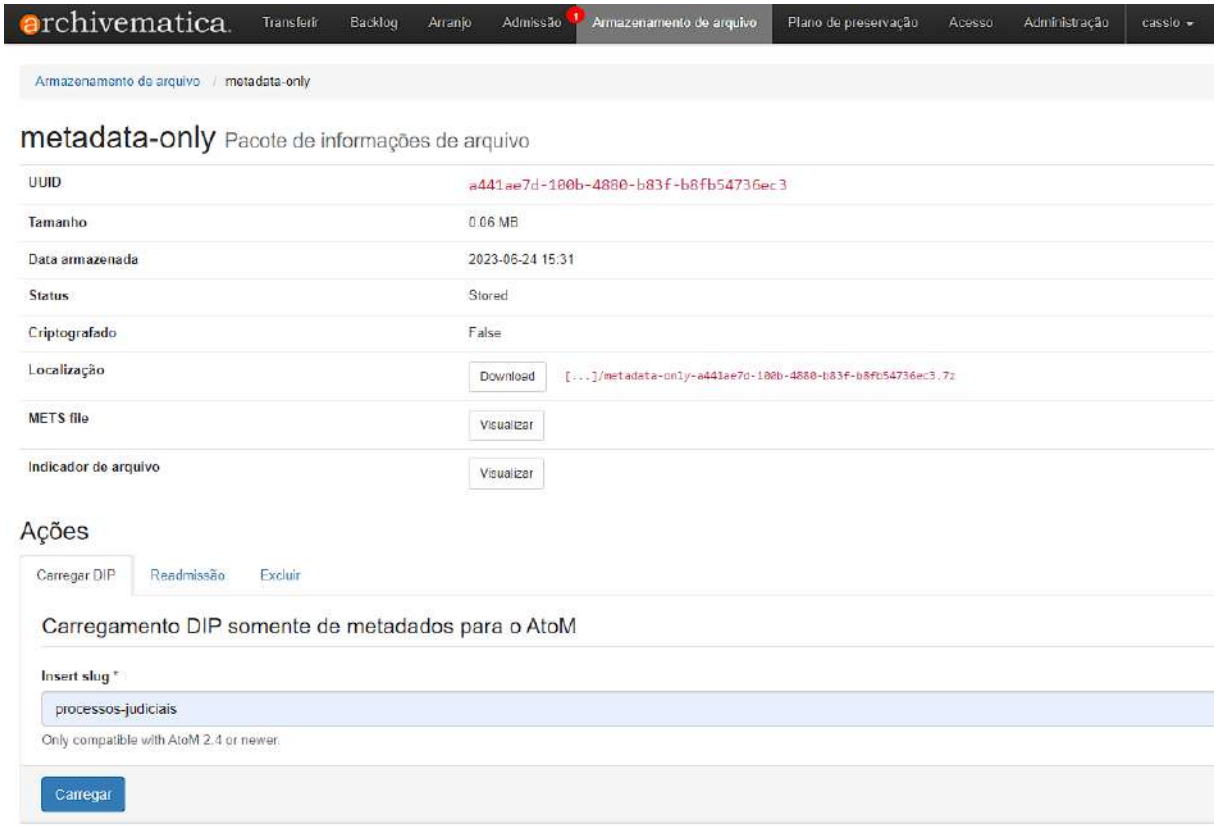
Fonte: elaborado pelo autor com base em testes realizados em instância do *software* Archivematica 1.13.2 e na sua documentação oficial.

No teste da "estratégia de disseminação somente de metadados" utilizou-se na transferência para o Archivematica um diretório "processo_judicial" contendo: um arquivo "processo_trabalhista.pdf" e um subdiretório "metadata". Para

importar metadados de Direitos (primeira funcionalidade do Quadro 5), é necessário incluir um arquivo *rights. csv* na transferência para o Archivematica dentro do subdiretório "metadata". Esse arquivo CSV seguiu o modelo do Quadro 1, analisado na seção 1 do relatório.

Designou-se, manualmente, na aba "Transferir" [*Transfer*] do Archivematica, o "nome da transferência" como " *metadata-only* ". Seguiu-se as orientações de configuração do Quadro 5, logo, os conteúdos não foram examinados com o *bulk extractor,* a normalização foi apenas para preservação e adicionou-se metadados descritivos manualmente (título, data e descrição). A resposta dada foi positiva à transcrição do conteúdo do SIP. Armazenou-se o AIP no local padrão [*default location*]. Como a resposta manual ao microsserviço de normalização foi apenas de preservação (mas não de disseminação), o Archivematica não acionou o microsserviço de carregar o DIP para o AtoM. Contudo, após a conclusão da admissão, o AIP ficou disponível na aba "armazenamento de arquivo" do Archivematica. Ao abri-lo, foi possível fazer manualmente o *upload* dos metadados inserindo a *slug* de uma simulação de fundo arquivístico "Processos judiciais" (que já existia no AtoM), conf orme a Figura 6.

Figura 6. *Upload* somente de metadados do AIP para o AtoM



Fonte: captura de tela, realizada pelo autor, de instância de testes do Archivematica 1.13.2.

O *upload* somente de metadados do pacote de informação para uma instância do AtoM demonstrou os seguintes resultados, de acordo com análise da Figura 7:

1. A indicação do *slug* "processos-judiciais" no *upload,* como ilustrado na Figura 4, que correspondia ao fundo "Processos judiciais" já existente no AtoM, resultou na criação de uma descrição de nível "dossiê", vinculada ao fundo "Processos judiciais" e a uma descrição de nível "item". O sistema atribuiu ao título do dossiê o nome dado à transferência ("metadata-only") e, ao título do item, o nome do arquivo transferido.

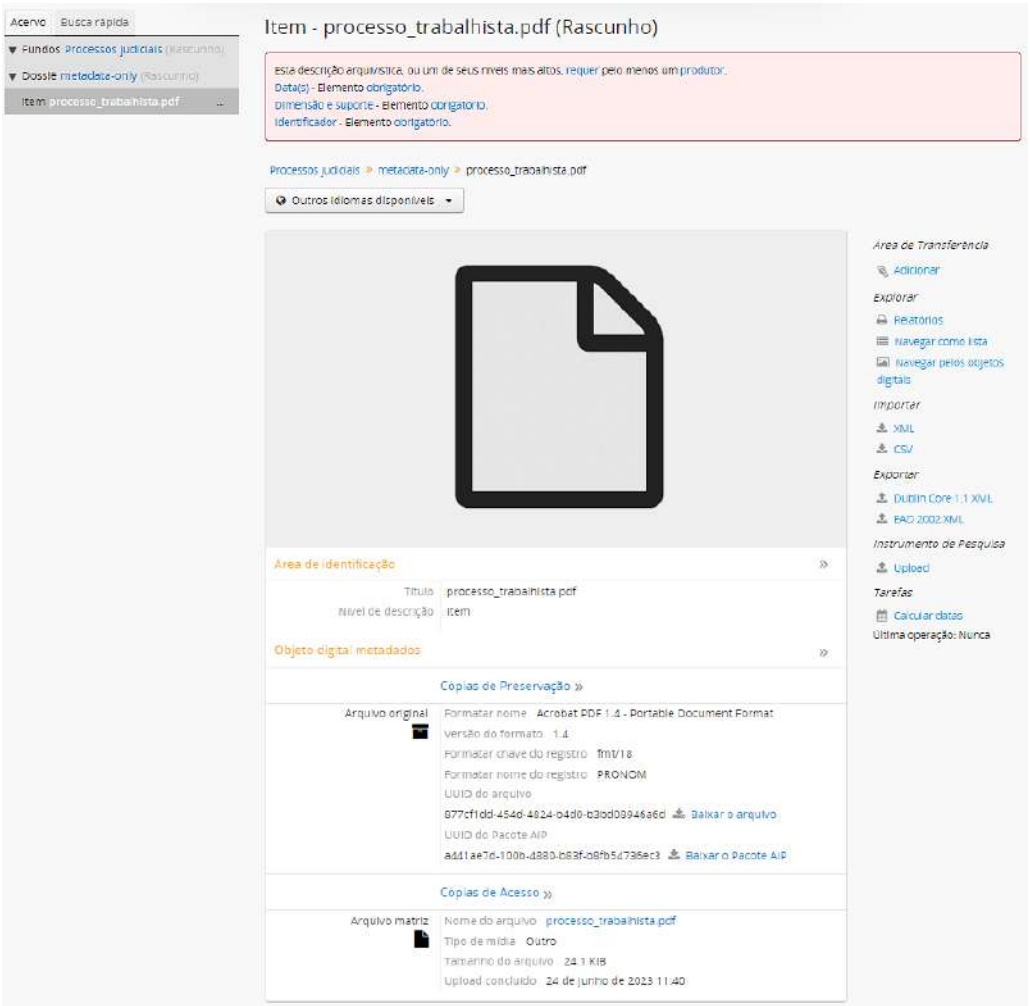
2. Os metadados descritivos inseridos no processo de admissão no Archivematica não foram utilizados pelo AtoM na descrição arquivística do item. Este resultado atesta a falta de suporte ao *upload* de metadados descritivos na opção de " *upload* somente de metadados" do Archivematica.

3. Metadados do objeto digital foram transferidos corretamente para o AtoM.

4. Ao listar os objetos digitais do AtoM, o "processo_trabalhista.pdf" está indexado. Contudo, o objeto digital não é acessível pela descrição no AtoM se não estiver configurado o acesso direto ao *storage service* do Archivematica (recurso disponível a partir da versão 2.7 do AtoM). A produção de um DIP para armazenamento (e não para *upload*) não afeta a manifestação dos metadados do objeto digital, que permanecem como os da Figura 7.

5. Os dados contidos no arquivo *rights.csv* foram identificados pelo Archivematica e inseridos nos metadados METS do AIP, como representa a Figura 8. Se for produzido um DIP para armazenamento no *storage service,* estes metadados PREMIS de Direitos também estarão no DIP.

Figura 7. Resultado do *upload* somente de metadados em ambiente de testes do AtoM



Fonte: captura de tela, realizada pelo autor, de instância de testes do AtoM 2.7.1.

Figura 8. Metadados PREMIS *Rights* em METS

```
<mets:rightsMD ID="rightsMD_1">
  <mets:mdWrap MDTYPE="PREMIS:RIGHTS">
    <mets:xmlData>
      <premis:rightsStatement xmlns:premis="http://www.loc.gov/premis/v3" xsi:schemaLocation=
      "http://www.loc.gov/premis/v3 http://www.loc.gov/standards/premis/v3/premis.xsd">
        <premis:rightsStatementIdentifier>
          <premis:rightsStatementIdentifierType>UUID</premis:rightsStatementIdentifierType>
          <premis:rightsStatementIdentifierValue>a6b5b168-2795-4239-9e93-b6fa5db97c6a
          </premis:rightsStatementIdentifierValue>
        </premis:rightsStatementIdentifier>
        <premis:rightsBasis>Statute</premis:rightsBasis>
        <premis:statuteInformation>
          <premis:statuteJurisdiction>BR</premis:statuteJurisdiction>
          <premis:statuteCitation>Lei nº 13.709, de 14 de agosto de 2018 - Lei Geral de Proteção de
          Dados Pessoais (LGPD)</premis:statuteCitation>
          <premis:statuteInformationDeterminationDate>2023-06-19
          </premis:statuteInformationDeterminationDate>
          <premis:statuteNote>Registro de aplicação da LGPD obrigatória no repositório, conforme
          determinação do Comitê Gestor de Proteção de Dados Pessoais (CGPD) do Tribunal, criado pela
          Resolução CNJ nº 363, de 12 de janeiro de 2021.</premis:statuteNote>
          <premis:statuteApplicableDates>
            <premis:startDate>2022-09-18</premis:startDate>
          </premis:statuteApplicableDates>
        </premis:statuteInformation>
        <premis:rightsGranted>
          <premis:act>disseminate</premis:act>
          <premis:restriction>Disallow</premis:restriction>
          <premis:termOfRestriction>
            <premis:startDate>2023-06-19</premis:startDate>
          </premis:termOfRestriction>
        </premis:rightsGranted>
        <premis:linkingObjectIdentifier>
          <premis:linkingObjectIdentifierType>UUID</premis:linkingObjectIdentifierType>
          <premis:linkingObjectIdentifierValue>877cf1dd-454d-4824-b4d0-b3bd08946a6d
          </premis:linkingObjectIdentifierValue>
        </premis:linkingObjectIdentifier>
      </premis:rightsStatement>
    </mets:xmlData>
  </mets:mdWrap>
</mets:rightsMD>
```

Fonte: captura de tela de arquivo XML contendo metadados METS produzidos pelo autor com instância de testes do Archivematica 1.13.2

Em relação aos itens 1 e 2 da estratégia de *upload* somente de metadados, a documentação do Archivematica 1.13.2 registra que o *workflow* do *upload* somente de metadados não suporta metadados descritivos incluídos por meio de arquivo CSV ou por meio da interface do usuário. Além disso, SIPs contendo níveis de descrição para arranjo no AtoM também são ignorados neste *workflow* (ARTEFACTUAL SYSTEMS, *s. d.*). Destaca-se que estas limitações afetam a capacidade desses *softwares* de prover acesso à informação de modo eficiente observando eventuais restrições de acesso a dados contidos em representações digitais.

Observa-se que, na instância do AtoM utilizada nos testes das duas estratégias, habilitou-se o acesso direto ao AIP armazenado no *storage service* do Archivematica por parte de usuários **administradores** do AtoM. Por esse motivo, nas Figuras 7 e 4 há *links* para *download* do arquivo original e do AIP completo.

Não se observou quaisquer implicações dos valores documentados nos metadados PREMIS de Direitos em relação às tarefas e aos microsserviços executados pelo Archivematica nas duas estratégias de disseminação. Além disso, não se identificou funcionalidade para importar esses metadados de Direitos para o AtoM, o que, teoricamente, poderia automatizar restrições de acesso no AtoM a partir de metadados incluídos previamente na transferência para o repositório. Localizou-se no fórum "ICA-AtoM Users" do Google Groups uma manifestação de representante oficial da Artefactual Systems confirmando a inexistência dessas funcionalidades. De acordo com ele, "[...] suporte para importação de metadados de direitos provenientes do Archivematica ou de qualquer outra fonte não está atualmente no nosso mapa de desenvolvimento" (GILLEAN, 2021, sem paginação).

3.3.2 Testes com AtoM: restrições de exibição de objetos digitais

No Quadro 3, o item 1 "Metadados de Direitos [*Rights*]: apagar, explorar, **exibir,** disseminar, migrar, modificar, replicar" refere-se à funcionalidade do AtoM de aplicar ações que regulam o acesso de usuários a documentos digitais com base em Direitos. Conforme destacado na seção anterior, **em testes realizados, observou-se que a informação sobre Direitos admitida pelo repositório por meio do Archivematica não é herdada pelo AtoM mesmo com a integração entre ambos os *softwares* devidamente configurada.**

O AtoM também tem funcionalidades para configuração de Direitos acionáveis. Entretanto, este estudo localizou significativas limitações nesse âmbito do desenvolvimento do AtoM como *software*. No fórum "ICA-AtoM users" do Google Groups, representante oficial da Artefactual Systems explicou, em 2022, o funcionamento dessas funcionalidades de restrições baseadas em direitos no AtoM

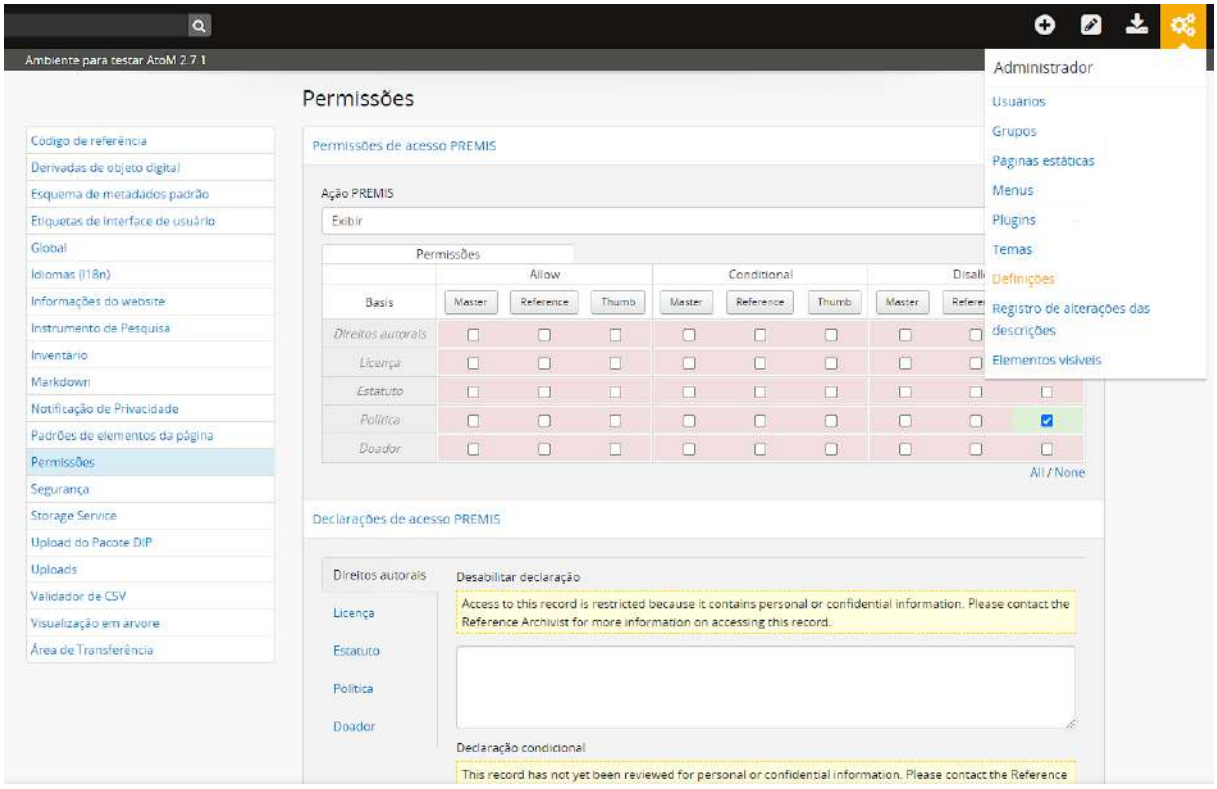
(GILLEAN, 2022, sem paginação). Reproduz-se abaixo, o trecho mais relevante para a problemática em análise:

> [...] Os direitos PREMIS acionáveis no AtoM só podem ser configurados atualmente para controlar a visibilidade de objetos digitais. É útil se você deseja que seus usuários finais saibam que há conteúdo digital disponível, mas você não quer deixar esse conteúdo diretamente disponível (devido a acordos, *copyright*, políticas etc). Todavia, como você deve ter percebido, qualquer um que puder logar será capaz de visualizar os objetos digitais originais. Isto é semelhante ao comportamento utilizado no módulo de Elementos Visíveis, que pode ocultar alguns campos de metadados em alguns

> templates, mas somente para usuários públicos. Observe que para Direitos de PREMIS, somente usuários autenticados podem ver as declarações de Direitos que forem adicionadas aos documentos arquivísticos - usuários públicos verão somente as declarações de acesso que você configurar para aparecerem no lugar de derivadas de objetos digitais. [...]

Testes realizados com o AtoM 2.7.1, em função do presente estudo, entre os meses de abril e junho de 2023, corroboram a mensagem de Gillean. As funcionalidades a que ele se refere estão disponíveis no AtoM 2.7 por meio de definições estabelecidas em duas páginas: a de configuração geral de permissões do AtoM e a de Direitos da respectiva descrição arquivística. A configuração de permissões é acessível por meio do menu do "Administrador", na opção "Definições" [*Settings*] e "Permissões" no menu da esquerda, como mostra a Figura 9.

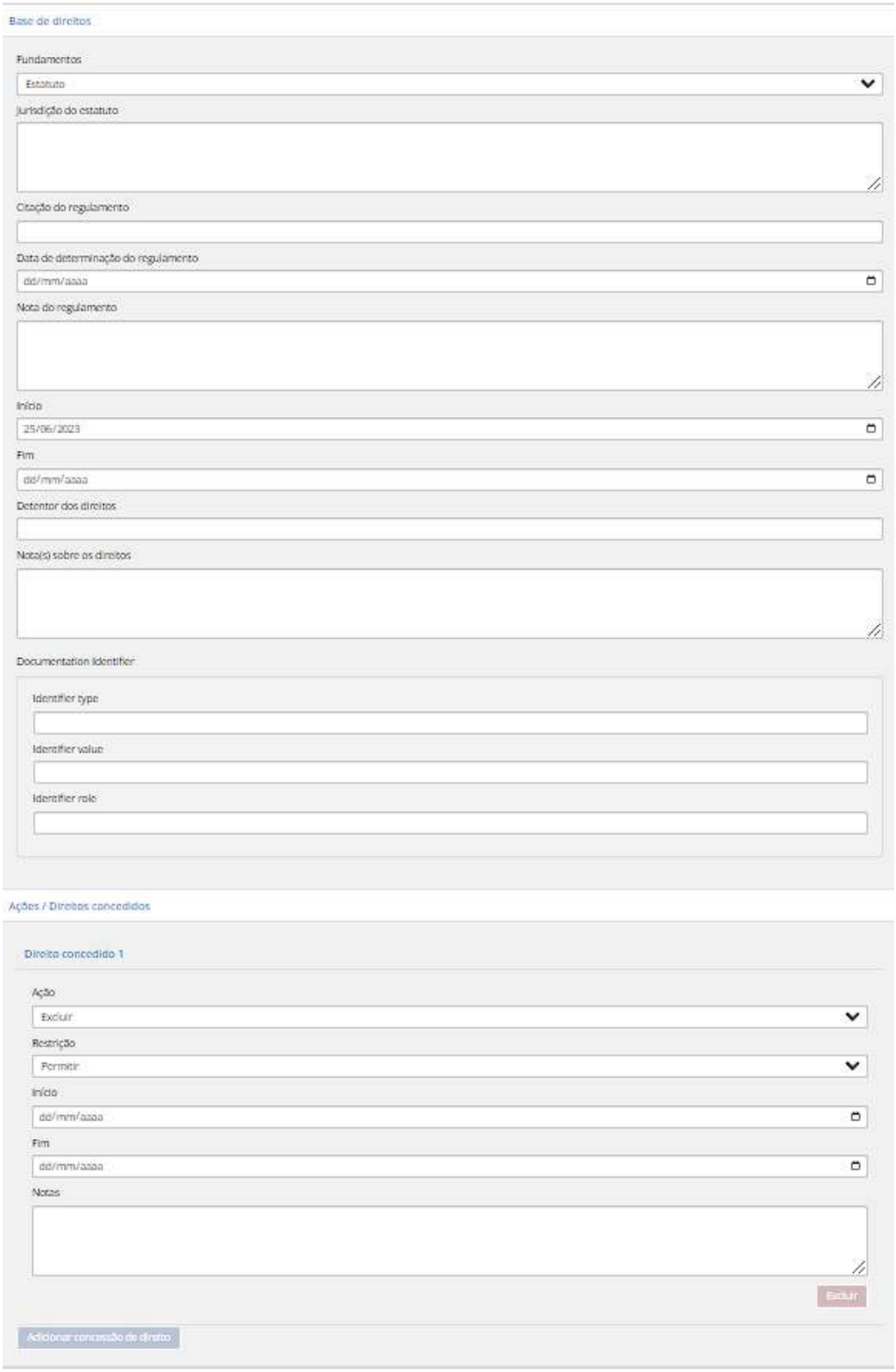
Figura 9. Permissões no AtoM 2.7.1



Fonte: captura de tela, realizada pelo autor, de instância de testes do AtoM 2.7.1.

O acesso à configuração de direitos de uma descrição arquivística específica ocorre por meio da opção "Criar novos direitos" no menu aberto pelo botão "Mais", na parte inferior da página da descrição. A Figura 10 representa essa página.

Figura 10. Página de "criar novos direitos" para uma descrição do AtoM 2.7.1



Fonte: captura de tela, realizada pelo autor, de instância de testes do AtoM 2.7.1.

Nos testes realizados e na análise da documentação, **chegou-se à conclusão de que o desenvolvimento dessas funcionalidades no *software* AtoM não estão concluídas,** pois a interação entre as permissões e os direitos das respectivas descrições é responsiva apenas naquele sentido destacado por Gillean (2022). Isto é, serve para restringir ou não a visualização de um objeto digital para um grupo de usuários e inserir uma mensagem. Ademais, inserir manualmente informações de Direitos na plataforma de acesso desconsiderando que há metadados PREMIS com o mesmo tipo de informação já nos metadados do AIP e do DIP exige repetir informações com a mesma finalidade no âmbito do repositório.

O uso de mensagens de direitos para substituir o acesso a representações digitais pode ser configurado com algumas limitações. Entretanto, não estão suficientemente claras as implicações de segurança ao disponibilizar em uma mesma instância do AtoM documentos de acesso ostensivo e irrestrito junto com outros que têm restrições de acesso. Este questionamento é particularmente importante em um contexto no qual as restrições de acesso a objetos digitais no AtoM dependem de funcionalidades que estão apenas parcialmente concluídas, que é o caso do sistema de permissões baseado em Direitos do AtoM.

3.4 Recomendações

A fim de proteger dados pessoais e outros dados que tenham restrições de acesso contidos em documentos arquivados permanentemente, recomenda-se a implementação do modelo Hipátia considerando as possibilidades do Quadro 6. Sugere-se essa lista de possibilidades com o objetivo de que aquelas consideradas como mais adequadas possam, eventualmente, serem empregadas por um RDC-Arq.

Quadro 6. Aspectos positivos e negativos de possibilidades para apoiar restrições de acesso a dados no modelo Hipátia

Possibilidade de implementação	Aspecto(s) positivo(s)	Aspecto(s) negativo(s)
Documentação de Direitos com metadados PREMIS	Proporciona ao repositório documentação permanente sobre Direitos que pode, futuramente, orientar permissões e restrições por meio de ações automáticas. A informação sobre Direitos fica registrada nos metadados METS do AIP e do DIP. O DIP pode ficar armazenado também no servidor do AtoM.	Atualmente, AtoM e Archivematica não utilizam os metadados PREMIS de Direitos fornecidos na transferência para orientar ações automáticas no repositório. AtoM não incorpora em suas funcionalidades a informação sobre Direitos fornecida por *uploads* de DIP do Archivematica.
Estratégia de disseminação baseada na transferência prévia de derivadas de acesso	Garante acesso a uma derivada de acesso do documento arquivístico a todos os usuários. Os usuários administradores podem acessar o documento original submetido ao repositório por meio da configuração de acesso ao *storage service* do Archivematica no AtoM, disponível a partir da versão 2.7.	Requer que a instituição implemente um *workflow* baseado em métodos e tecnologias com confiabilidade comprovada para ocultação de dados a fim de produzir **previamente** derivadas de acesso para inclusão nos SIPs.
Estratégia de disseminação baseada no *upload* somente de metadados	Garante que não sejam enviados documentos que tenham restrições de acesso ao ambiente de disseminação. Todos os usuários da plataforma de acesso podem visualizar que existe um documento no ambiente de preservação que está vinculado àquela respectiva descrição. Os usuários administradores podem acessar o documento original submetido ao repositório por meio da configuração de acesso ao *storage service* do Archivematica no AtoM, disponível a partir da versão 2.7.	Usuários que não tiverem permissões especiais ou repositório que optem por não permitir acesso ao *storage service* do Archivematica pelo AtoM não poderão acessar representações digitais por meio do AtoM. O Archivematica não suporta o envio de metadados descritivos e de uma estrutura de arranjo nesse tipo de *upload*, o que pode dificultar a automatização de transferências.

Pipelines e instâncias de Archivematica/AtoM para uso interno	Permite que a instituição crie ambientes totalmente projetados para determinados perfis de usuários. Por exemplo, AtoM e seu respectivo servidor configurados exclusivamente para o acesso de usuários internos da instituição responsável pelo repositório.	A instituição precisa gerir virtualizações de recursos computacionais e diferentes instalações Archivematica e AtoM com distintas configurações.
Configurações de Direitos no AtoM	Disponibiliza mensagens textuais personalizadas sobre Direitos no ambiente de acesso. Essas mensagens podem substituir o acesso às representações digitais para usuários anônimos (sem *login*).	O desenvolvimento das funcionalidades do *software* não parece concluído. Não é uma boa prática repetir na plataforma de acesso a inserção manual de informação sobre Direitos que já está registrada em metadados contidos no AIP e no
Configurações de Direitos no AtoM	Disponibiliza mensagens textuais personalizadas sobre Direitos no ambiente de acesso. Essas mensagens podem substituir o acesso às representações digitais para usuários anônimos (sem *login*).	O desenvolvimento das funcionalidades do *software* não parece concluído. Não é uma boa prática repetir na plataforma de acesso a inserção manual de informação sobre Direitos que já está registrada em metadados contidos no AIP e no DIP.

Fonte: elaborado pelo autor.

A análise do Quadro 6 evidencia que, das quatro possibilidades elencadas, apenas as duas estratégias de disseminação são mutuamente excludentes se estiverem no mesmo ambiente e com a mesma finalidade. Isso ocorre porque entre essas estratégias está implícita a decisão de enviar ou não representações digitais para a plataforma de acesso. Contudo, se a instituição optar por implementar o modelo Hipátia com mais de uma instância do AtoM, uma dessas instâncias poderá receber representações digitais contendo dados com restrições de acesso e a outra não. Esta é a alternativa sugerida pela linha "Pipelines e instâncias de Archivematica/AtoM para uso interno" do Quadro 6.

Os testes realizados não contemplaram as potencialidades da ferramenta *bulk extractor* utilizada no microsserviço de exame de conteúdo do Archivematica. Essa análise extrapolaria os objetivos do presente relatório. A *bulk extractor* e, principalmente, ferramentas que façam uso de IA permanecem como possíveis hipóteses futuras de uso na identificação de conteúdos a serem protegidos em documentos digitais.

4.
Considerações finais

No estudo que resultou na produção deste relatório técnico, buscou-se verificar as potencialidades do padrão de metadados PREMIS e dos *softwares* Archivematica e Atom, utilizados no modelo Hipátia, com foco no apoio ao acesso à informação do repositório arquivístico observando hipóteses de sigilo e de proteção de dados.

Analisou-se o padrão de metadados PREMIS sob a perspectiva da entidade *Rights.* Elaborou-se um modelo para documentar declarações de Direitos (*rightsStatement*) com metadados PREMIS utilizando o *software* Archivematica. Identificou-se funcionalidades dos *softwares* Atom e Archivematica com potencial para auxiliar no acesso à informação observando hipóteses de sigilo e de proteção de dados. Analisou-se essas funcionalidades. Realizou-se testes baseados em funcionalidades identificadas. No Archivematica, esses testes seguiram dois roteiros elaborados de "estratégias de disseminação" e, no Atom, concentraram-se nas possibilidades de restrições de acesso. Procedeu-se à análise dos resultados dos testes realizados e, com isso, foi possível gerar um balanço das potencialidades das tecnologias do modelo Hipátia para as finalidades analisadas. Por fim, realizou-se recomendações com base nas análises e nos testes realizados.

Conclui-se que o padrão de metadados PREMIS tem potencial ainda não explorado pelas soluções de *softwares* no que diz respeito à questão de Direitos. Se, por um lado, é possível documentar informação sobre Direitos com as tecnologias de *softwares* atualmente utilizadas pelo modelo Hipátia, por outro, essas tecnologias não são capazes de tirar o máximo de proveito da automatização de permissões e de restrições possibilitadas pelo padrão PREMIS. A incapacidade do *software* Atom, em sua versão atual (2.7), de implementar automaticamente restrições de acesso definidas ou enviadas para o Archivematica pode dificultar soluções. Outrossim, o próprio sistema de permissões de Direitos do Atom mostra-se incompleto.

Apesar desses fatores, considera-se a opção pelo padrão PREMIS nessas tecnologias de *software* uma decisão acertada de desenvolvimento. Ademais, a despeito das limitações identificadas, existem possibilidades que podem ser exploradas na implementação do repositório, conforme tratado na seção 3.4. O desenvolvimento dessas estratégias depende das decisões tomadas no planejamento do repositório digital, que devem estar alinhadas com políticas e planos de preservação digital no âmbito do Poder Judiciário.

5. Referências

Referências

ARTEFACTUAL SYSTEMS. **Archivematica user manual version 1.13.2.** [*s. d.*]. Disponível em: https://archivematica.org/pt-br/docs/archivematica-1.13/user--manual/. Acesso em: 24 jun. 2023.

BRAGA, T. E. N; MÁRDERO ARELLANO, M. Á. (org.). **Hipátia: modelo de preservação para repositórios arquivísticos digitais confiáveis.** Brasília, DF: Ibict, 2022. Disponível em: https://doi.org/10.22477/9786589167501.

BRAGA, T. E. N.; HOLANDA, A. P.; CANELHAS PIGNATARO, T. Resolução RDC-Arq Conarq: uma análise dos novos requisitos informacionais propostos. **Revista Brasileira de Preservação Digital,** v. 3, p. e022004, 2022. Disponível em:

https://doi.org/10.20396/rebpred.v3i00.16583.

BRASIL. Conselho Nacional de Arquivos. **Diretrizes para a implementação de Repositórios Arquivísticos Digitais Confiáveis - RDC-Arq.** 2015. Disponível em: https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/conarq_diretrizes_rdc _arq_resolucao_43.pdf. Acesso em: 02 abr. 2023.

BRASIL. Conselho Nacional de Arquivos. **Minuta: diretrizes para a implementação de repositórios arquivísticos digitais confiáveis - RDC-Arq.** [S. l.]: Conarq, 2022. Disponível em:

https://www.gov.br/conarq/pt-br/assuntos/noticias/consulta-publica--proposta-de-atualizac ao-das-diretrizes-para-a-implementacao-de-repositorios-arquivisticos-digitais-confiaveis 2013-rdc-arq/minuta_diretrizes_rdc_arq_2021VF_Consulta.pdf/view. Acesso em: 27 mai. 2023.

CAPLAN, P. **Comprende PREMIS.** Traduction: Louise Fauduet, Clément Oury, Sébastien Peyrard, Jean-Philippe Tramoni. [*S. l.*]: Library Of Congress, 2009a. Disponível em:

https://www.loc.gov/standards/premis/Understanding-PREMIS_french.pdf. Acesso em: 01 mai. 2023.

CAPLAN, P. **Entendendo o PREMIS.** Tradução: Amarílis Montagnolli Gomes Corrêa; Laerte Pereira da Silva Júnior. [*S. l.*]: Library Of Congress, 2021a. Disponível em: https://loc.gov/standards/premis/understandingPREMIS_portuguese_2021.pdf. Acesso em: 01 mai. 2023.

CAPLAN, P. **Entender PREMIS.** Traducción: María Luisa Martínez-Conde. [*S. l.*]: Library Of Congress, 2009b. Disponível em:

https://www.loc.gov/standards/premis/UnderstandingPREMIS_espanol.pdf. Acesso em: 01 mai. 2023.

CAPLAN, P. **Understanding PREMIS.** Disponível em:

https://www.loc.gov/standards/premis/understandingPREMIS_english_2021.pdf. [*S. l.*]: Library of Congress, 2021b. Acesso em: 01 mai. 2023.

CONSELHO NACIONAL DE JUSTIÇA. **Manual de digitalização de documentos do Poder Judiciário.** Brasília: CNJ, 2023. *E-book.* Disponível em:

https://www.cnj.jus.br/wp-content/uploads/2023/03/proname-manual--digitalizacao-15-03- 2023.pdf. Acesso em: 25 jun. 2023.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução nº 363, de 12 de janeiro de 2021.** Estabelece medidas para o processo de adequação à Lei Geral de Proteção de Dados Pessoais a serem adotadas pelos tribunais. Brasília, DF: CNJ, 2021. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3668. Acesso em: 25 jun. 2023.

CCSDS. **Audit and Certification of Trustworthy Digital Repositories.** Recommended practice. CCSDS 652.0-M-1. Consultative Committee for Space Data Systems, 2011. Disponível em: https://public.ccsds.org/pubs/652x0m1.pdf. Acesso em: 02 abr. 2023.

CCSDS. **Reference model for an open archival information system (OAIS).** Recommended practice. CCSDS 650.0-M-2. Consultative Committee for Space Data Systems, 2012. Disponível em: https://public.ccsds.org/pubs/650x0m2.pdf. Acesso em: 02 abr. 2023.

COPYRIGHT. *In*: WEX free legal dictionary and encyclopedia sponsored and hosted by the Legal Information Institute at the Cornell Law School. 2022. Disponível em: https://www.law.cornell.edu/wex/*copyright*. Acesso em: 01 mai. 2023.

COYLE, K. The automation of Rights. **The Journal of Academic**

Librarianship, v. 32, n. 2, 2006, p. 326–329. Disponível em: http://www.kcoyle.net/jal-32-3.html. Acesso em: 01 mai. 2023.

DAPPERT, A. **DePICT: a conceptual model for digital preservation.** 2013. Tese de Doutorado - University of Portsmouth, Portsmouth, 2013. Disponível em: https://researchportal.port.ac.uk/portal/files/5899441/Final_Thesis_Angela_Dappert_pdf a.pdf. Acesso em: 04 jun. 2023.

GARFINKEL, S. L. bulk_extractor. [*s. l.*], 2023. **Github: simsong/bulk_extractor.** Disponível em: https://github.com/simsong/bulk_extractor. Acesso em: 25 jun. 2023.

GILLEAN, D. Right and user permissions. [*s. l.*], 13 abr. 2022. **Google groups: ica-AtoM-users.** Disponível em:

https://groups.google.com/g/ica-AtoM-users/c/DGG4klUrBN8/m/rBf9c-5q7AQAJ. Acesso em: 25 jun. 2023.

GILLEAN, D. Can AtoM import PREMIS:RIGHTS. [*s. l.*], 01 jul. 2021. **Google groups: ica-AtoM-users.** Disponível em:

https://groups.google.com/g/ica-AtoM-users/c/UyfChSs7s5I/m/A6v0aXF-sAAAJ?pli=1. Acesso em: 25 jun. 2023.

LIBRARY OF CONGRESS. **PREMIS: Preservation Metadata Maintenance Activity.** 2022. Disponível em: https://www.loc.gov/standards/premis/. Acesso em: 26 abr. 2023.

LICENSE. *In:* WEX free legal dictionary and encyclopedia sponsored and hosted by the Legal Information Institute at the Cornell Law School. 2020. Disponível em: https://www.law.cornell.edu/wex/license. Acesso em: 01 mai. 2023.

MCLELLAN, E. Implementing Rights metadata for digital preservation. *In* : DAPPERT, A.; GUENTHER, R. S.; PEYRARD, S. (org.). **Digital Preservation Metadata for Practitioners: Implementing PREMIS.** 1ª ed. Cham: Springer International Publishing: Imprint: Springer, 2016. p. 151-160. *Ebook.*

PREMIS EDITORIAL COMMITTEE. **PREMIS Data Dictionary for Preservation Metadata, Version 3.0.** 2015. Disponível em: https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf. Acesso em: 30 abr. 2023.

ROCHA, R. P. da; PIRES, C. de O. Finalidade e Atividades da Curadoria Digital na Perspectiva de sua Implantação em uma Instituição. **Brazilian Journal of Information Science,** v. 14, n. 4, p. e020012, 2020. Disponível em: https://doi.org/10.36311/1940-1640.2020.v14n4.10857

STATUTE. *In*: WEX free legal dictionary and encyclopedia sponsored and hosted by the Legal Information Institute at the Cornell Law School. 2020. Disponível em: https://www.law.cornell.edu/wex/statute. Acesso em: 01 mai. 2023.

HIPÁTIA

MODELO DE PRESERVAÇÃO