

MODELO HIPÁTIA: integração com o *software DSpace*¹

HYPATIA MODEL: integration with DSpace software

Alexandre Faria de Oliveira*

Raul Carlos Neres**

Marilete da Silva Pereira***

RESUMO

Estudo que objetiva apresentar a discussão sobre integração do Modelo Hipátia com o *software DSpace*. Fundamenta-se em uma metodologia que possui características práticas, com aspectos que se assemelham aos tratados na ciência da computação e em estudos focados nos padrões de metadados utilizados no *DSpace*, bem como ajustes na camada informatizada do Modelo Hipátia, realizados em estrita conformidade com os princípios arquivísticos de preservação digital, assegurando a integridade da cadeia de custódia. Resulta na demonstração do fluxo das configurações e adaptações necessárias no Modelo Hipátia para viabilizar a utilização do protocolo nativo *Open Archives InitDSpaceiative - Object Reuse and Exchange* do *DSpace*, com intuito de coletar, empacotar, preservar e disseminar os objetos digitais. Conclui que a principal contribuição deste trabalho é a simplificação do processo de integração de ferramentas ao Modelo Hipátia, por meio do uso de um protocolo já incorporado ao *DSpace*. Essa abordagem facilita a adoção do Modelo Hipátia por instituições que já utilizam o *DSpace* como repositório digital, uma vez que não requer a implementação de protocolos adicionais ou extensas modificações no sistema.

Palavras-chave: Modelo Hipátia; OAI-ORE; DSpace; Interoperabilidade; Preservação digital.

ABSTRACT

Study that aims to present the discussion on the integration of the Hipátia Model with the DSpace software. It is based on a methodology that has practical characteristics, with aspects that resemble those treated in computer science and on studies focused on the metadata standards used in DSpace, as well as adjustments to the computerized layer of the Hipátia Model, carried out in strict accordance with the archival principles of digital

¹Artigo atualizado, oriundo da comunicação oral realizada no X Congresso Nacional de Arquivologia (CNA) em Salvador, Bahia, entre 14 e 18 de outubro de 2024.

* Coordenador de Governança em Tecnologias para Informação e Comunicação - Ibict. Mestrando em Gestão Estratégica de Organizações. alexandreoliveira@ibict.br

** Pesquisador - Ibict. Pós-graduado em Pós em Gestão de Tecnologias da Informação. raulcarlos@ibict.br

*** Pesquisadora - Ibict. Mestranda em Ciência da Informação. mariletepereira@ibict.br



preservation, ensuring the integrity of the chain of custody. It results in demonstrating the flow of necessary configurations and adaptations in the Hipátia Model to enable the use of DSpace's native Open Archives Initiative - Object Reuse and Exchange protocol, with the aim of collecting, packaging, preserving and disseminating digital objects. It concludes that the main contribution of this work is the simplification of the process of integrating tools into the Hipátia Model, through the use of a protocol already incorporated into DSpace. This approach facilitates the adoption of the Hipatia Model by institutions that already use DSpace as a digital repository, as it does not require the implementation of additional protocols or extensive modifications to the system.

Keywords: Hypatia Model. OAI-ORE. DSpace. Interoperability. Digital preservation.

1 INTRODUÇÃO

Atualmente, vivemos em um ambiente cada vez mais tecnológico, caracterizado pela produção, armazenamento e compartilhamento massivo de dados e informações em formato eletrônico. As organizações, independentemente de seu setor de atuação, enfrentam o desafio de lidar com uma verdadeira avalanche de informações digitais.

Para alguns pesquisadores, esse desafio se torna ainda mais complexo devido à natureza sensível e valiosa das informações geradas. Essas informações incluem pesquisas científicas, dados experimentais, relatórios técnicos e documentos científicos, que são fundamentais para o avanço do conhecimento e da inovação tecnológica.

A preservação digital desses arquivos é essencial para garantir que as descobertas e os dados científicos permaneçam acessíveis e utilizáveis a longo prazo. Isso envolve a implementação de estratégias de arquivamento e backup, bem como a utilização de formatos de arquivo que assegurem a integridade e a legibilidade dos dados a longo prazo.

Desde a sua criação, o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) tem a missão de organizar e promover a disseminação da produção científica brasileira. Uma de suas principais responsabilidades é a preservação dessa informação em formato digital, uma tarefa especialmente relevante no atual cenário permeado pelas tecnologias.

Em 2018, a parceria estabelecida entre o Ibict e o Tribunal de Justiça do Distrito Federal e Territórios (TJDFT) resultou na criação do Modelo Hipátia, destinado à



preservação digital, foi cuidadosamente desenhado não apenas para atender às diretrizes das normativas atuais, mas também para garantir a interoperabilidade entre diferentes sistemas de informação. Além de ter como objetivo principal facilitar uma implementação eficiente e segura de Repositórios Arquivísticos Digitais Confiáveis (RDC-Arq), promovendo assim a preservação e o acesso confiável à informação digital (Braga; Márdero Arellano, 2022).

Estudos mais recentes sobre a capacidade de interoperabilidade do Modelo Hipátia em integrar qualquer sistema produtor vêm sendo realizados pelo Ibict, baseando-se em conceitos, paradigmas e técnicas de integração sistêmica, Shintaku *et al.* (2023) apresentaram, de forma detalhada, as modificações requeridas no software *DSpace* para habilitar o uso do protocolo *Open Archives Initiative - Object Reuse and Exchange* (OAI-ORE), visando a integração efetiva com o Modelo Hipátia.

Diante do exposto, este estudo objetiva apresentar a discussão sobre integração do Modelo Hipátia com o *software DSpace*.

2 SOFTWARE DSPACE E PROTOCOLO OAI-ORE

O *DSpace* é um software livre, de código aberto, desenvolvido originalmente para a Biblioteca do Massachusetts Institute of Technology (MIT) pela *Hewlett Packard* (HP). O *DSpace* adota o Protocolo de Metadados de Iniciativa de Arquivos Abertos (OAI-PMH), que permite a coleta de metadados de seus repositórios por serviços agregadores e motores de busca, facilitando a visibilidade e acessibilidade do conteúdo digital. Ademais, o *DSpace* também utiliza a implementação do OAI-ORE permite a descrição e intercâmbio de recursos agregados na web, melhorando a integração com outros sistemas e serviços digitais (Silveira *et al.*, 2018; Van de Sompel; Lagoze, 2009).

O OAI-ORE é baseado em várias tecnologias e conceitos fundamentais que permitem a interoperabilidade e a reutilização de recursos digitais. De acordo com Lagoze *et al.* (2008, p.2), o " [...] OAI-ORE explora a arquitetura da web e os princípios do Linked Data para criar uma infraestrutura que permita a descrição e a troca de agregações de recursos". Nesse contexto, a arquitetura da web é a base sobre a qual o OAI-ORE é construído. O uso de Uniform Resource Identifiers (URIs) para identificar recursos e



Hypertext Transfer Protocol (HTTP) para a transferência de dados são princípios fundamentais que garantem a interoperabilidade. Como Berners-Lee et al. (2006, p.5) afirmam, "a arquitetura da web permite a integração de informações de maneira global e escalável".

O protocolo OAI-ORE permite recuperar metadados de itens, coleções e comunidades no *DSpace*. Para isso, basta enviar uma requisição `getRecord` com o identificador do recurso desejado e o formato de metadados desejado (ex: XML). O *DSpace* então responde com os metadados nesse formato, além dos próprios objetos digitais, se solicitados. A especificação do formato de saída e, principalmente, o fornecimento do identificador correto são essenciais para o sucesso da requisição. A estrutura hierárquica do *DSpace* é representada no quadro 1.

Quadro 1 - Estrutura hierárquica do *DSpace*

ESTRUTURA	DESCRIÇÃO
Comunidades	Representam grandes temas ou áreas de conhecimento.
Coleções	São conjuntos de documentos pertencentes a uma.
Itens	Consistem em metadados e objetos digitais que compõem uma coleção.

Fonte: Dados da pesquisa, 2024

Para assegurar a preservação adequada desses elementos, é fundamental que o sistema ofereça opções flexíveis de preservação, permitindo:

- Preservar uma comunidade inteira, incluindo suas subcomunidades, coleções e itens.
- Preservar uma coleção específica com todos os seus itens.
- Preservar um único item individualmente.

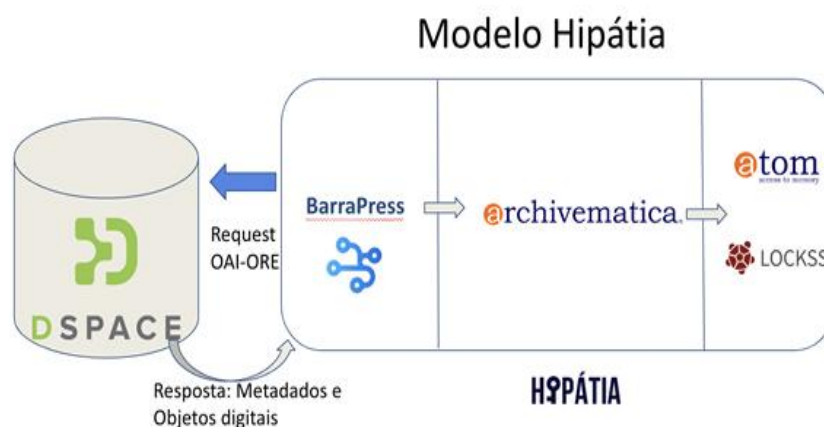
Essa granularidade nas opções de preservação garante que os administradores possam adequar as estratégias de preservação de acordo com as necessidades específicas



de cada nível da hierarquia do *DSpace*.

Nesse contexto, conforme ilustrado na Figura 1, o processo começa no *DSpace* com a indicação do desejo de preservar um determinado elemento ou conjunto de objetos. Em seguida, o processo é iniciado também no Hipátia. Essa interação é essencial para sincronizar as duas ferramentas e permitir o início da transferência dos itens entre elas.

Figura 1 - Integração Modelo Hipátia - *DSpace*



Fonte: Elaborada pelos autores, 2024

Para iniciar o processo de preservação, o Hipátia envia uma requisição ao *DSpace* solicitando o identificador do elemento a ser preservado. O *DSpace*, então, retorna o identificador ao Hipátia. Com o identificador em mãos, o Hipátia envia uma nova requisição, desta vez utilizando o protocolo OAI-ORE com o verbo "*getRecord*", para obter os itens a serem preservados. O *DSpace* responde com os itens solicitados, e estes seguem as fases de preservação digital do Modelo Hipátia.

Segundo Braga(2022), o modelo Hipátia de preservação digital é organizado em cinco fases distintas: preparação arquivística, preparação computacional, extração de objetos digitais, preservação e disseminação (Figura 2).

Figura 2 - Etapas do Modelo Hipátia



Fonte: Adaptado de Braga, 2022.



É importante destacar que o Modelo Hipátia assegura a cadeia de custódia ao seguir rigorosamente todas as diretrizes arquivísticas, desde a fase de preparação até a disseminação da informação, por meio de um fluxo totalmente informatizado. Além disso, o Hipátia, em sua concepção, baseou-se no modelo *Open Archival Information System* (OAIS²) de preservação digital, que oferece uma estrutura abrangente para a gestão e preservação de arquivos digitais, garantindo que eles permaneçam acessíveis e compreensíveis no futuro.

3 METODOLOGIA

A pesquisa apresenta características práticas, com aspectos que se assemelham aos tratados na ciência da computação. De acordo com Wazlawick (2009), há estudos que se concentram em apresentar inovações na resolução de problemas. Esses estudos investigam novas abordagens e metodologias que possam proporcionar soluções mais eficientes e eficazes para os desafios existentes. A criatividade e a aplicação de conhecimentos interdisciplinares são enfatizadas para desenvolver tecnologias e processos que não apenas solucionem problemas de maneira inovadora, mas também contribuam para o avanço do campo em questão. O autor argumenta, em grande medida, que essas pesquisas são conduzidas em áreas emergentes e esses esforços são essenciais para promover o progresso e a adaptação contínua às mudanças e demandas emergentes na sociedade.

Dessa maneira, este estudo apresenta a discussão de propostas para integrar o Modelo Hipátia ao software *DSpace*, visando a preservação digital dos objetos digitais armazenados no *DSpace*. Em seguida, é apresentada uma automatização computacional seguindo todas as etapas previstas no Modelo Hipátia, que almeja averiguar como preparar o modelo Hipátia para realizar a coleta, empacotamento, preservação e disseminação dos objetos digitais do *DSpace* e inferir se é possível, através do protocolo OAI-ORE, a integrar do Modelo Hipátia ao *DSpace* para a preservação digital de metadados e objetos digitais.

² Modelo de referência criado pela *Consultative Committee for Space Data Systems* para padronizar atividades de preservação digital (De Souza *et al.*, 2012).



Nesse sentido, o estudo foi dividido em quatro etapas:

1. Iniciou-se com a etapa de preparação do ambiente computacional, por meio da instalação dos *softwares* utilizados pelo Modelo Hipátia (BarraPres, *Archivematica* e *AtoM*).
2. Posteriormente realizou-se estudos de integração com o protocolo OAI-ORE/OAI-PMH, permitindo a criação de agregações que podem incluir diversos tipos de objetos digitais, como documentos, imagens, vídeos e metadados.
3. Em seguida, foram realizados estudos sobre os padrões de metadados para descrever e organizar os itens que serão preservados.
4. Logo após, os resultados das etapas 2 e 3 foram aplicados ao contexto da pesquisa.

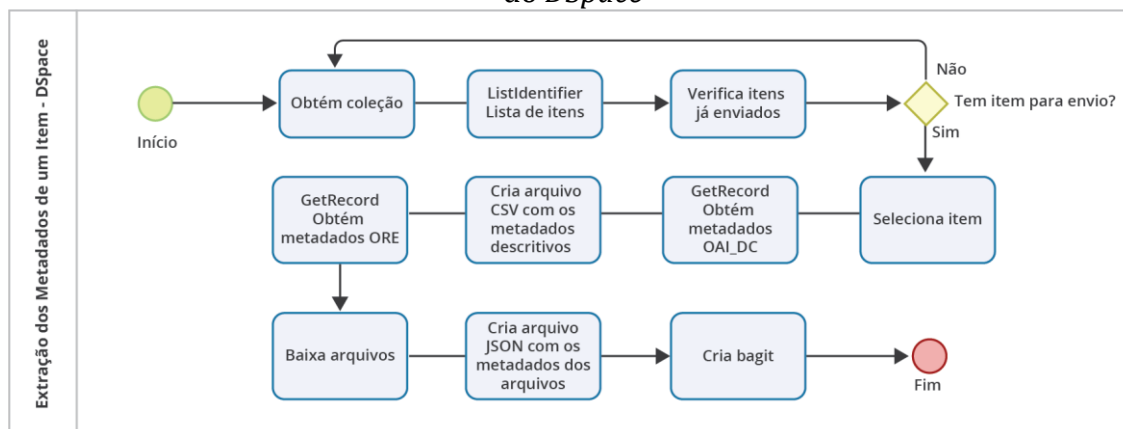
Ressalta-se que para viabilizar essa integração, foi necessário conduzir estudos focados nos padrões de metadados utilizados no *DSpace*, bem como ajustes na camada informatizada do Modelo Hipátia. As modificações implementadas no modelo foram cuidadosamente realizadas em estrita conformidade com os princípios arquivísticos de preservação digital, assegurando a integridade da cadeia de custódia.

4 RESULTADOS

Os resultados deste estudo demonstram as configurações e adaptações necessárias no Modelo Hipátia para viabilizar a utilização do protocolo nativo OAI-ORE do *DSpace*, com intuito de coletar, empacotar, preservar e disseminar os objetos digitais. Diante disso, apresenta-se o fluxo do mecanismo de extração dos metadados para preservação, ilustrado na Figura 3.



Figura 3 - Fluxo do processo de extração dos metadados de um item e de seus arquivos do *DSpace*



Fonte: Elaborado pelos autores, 2024.

Ademais, elencam-se as seguintes atividades realizadas no âmbito desse processo.

1. Realização de uma consulta em *link* onde estão dispostas as coleções que terão seus itens preservados e uma coleção é selecionada;
2. Os itens da coleção são listados através do verbo “*ListIdentifiers*” e dos parâmetros “*MetadataFormats*” e “*set*”. Para esse tipo de consulta o formato utilizado foi o “*OAI_DC*”;
3. Verificação no banco do barramento os itens já preservados e então obtido o identificador do próximo item. Para o caso de coleções com todos os itens preservados é iniciado o mesmo processo com outra coleção;
4. Outra requisição é realizada, através do verbo “*GetRecord*” e dos parâmetros “*MetadataFormats*” e “*identifier*”, buscando os metadados descritivos do item selecionado. Novamente, para esse tipo de consulta o formato utilizado foi o “*OAI_DC*”;
5. Após a extração dos metadados no padrão Dublin Core, com o mesmo verbo e parâmetros da última consulta, porém com formato “*ORE*” definido no “*MetadataFormats*”, os arquivos do item são baixados, bem como seus dados extraídos.

Conforme demonstrado na Figura 3, com os metadados *OAI_DC* é gerado o arquivo “*metadatas.csv*” e com os *ORE*, além do download dos arquivos do item, é criado o arquivo “*ore.json*”. O barramento organiza esses documentos em um pacote que segue o padrão *bagit*, o qual é voltado para empacotar arquivos e diretórios de maneira hierárquica para que os metadados e objetos digitais de um item do *DSpace* possam ser



enviados para o ambiente de preservação. Assim, o pacote bagit é gerado e compactado no formato zip, o processo de extração é finalizado e o próximo passo é o envio para o Archivematica (Figura 4 e 5).

Figura 4 - Estrutura raiz do pacote gerado pelo barramento no padrão bagit

Nome	Data de modificação
data	19/12/2024 18:13
bag-info.txt	15/08/2024 20:54
bagit.txt	15/08/2024 20:54
manifest-sha256.txt	15/08/2024 20:54
tagmanifest-sha256.txt	15/08/2024 20:54

Fonte: Elaborado pelos autores, 2024.

Figura 5 - Arquivos dos metadados do item na estrutura bagit (data/metadada)

Nome	Data de modificação
metadata.csv	15/08/2024 20:54
ore.json	15/08/2024 20:54

Fonte: Elaborado pelos autores, 2024.

A adoção do modelo Hipátia envolveu várias etapas críticas, incluindo a avaliação inicial das coleções, a análise dos tipos documentais do repositório e o estudo aprofundado dos metadados de preservação possíveis presentes no *DSpace*. Cada uma dessas etapas foi cuidadosamente planejada e executada para garantir a integridade e a acessibilidade dos acervos a longo prazo. Como resultado prático, foi possível implementar integralmente todo o fluxo de preservação delineado no modelo Hipátia.

5 CONSIDERAÇÕES FINAIS

Em conclusão, os ajustes realizados no Modelo Hipátia para suportar o protocolo OAI-ORE do *DSpace* representam um avanço significativo na integração de repositórios digitais no Brasil. Essa integração simplificada, utilizando um protocolo nativo, contribui para a disseminação e adoção de práticas de preservação digital de longo prazo, essenciais para garantir a disponibilidade e o acesso contínuo aos recursos digitais armazenados em repositórios como o *DSpace*.



O OAI-ORE é um padrão flexível que facilita a interoperabilidade e a reutilização de recursos digitais na web. Ao ser adotado, o protocolo oferece uma infraestrutura poderosa para a descrição e a troca de agregações de recursos e resulta no desempenho de um papel crucial na promoção da descoberta e reutilização de informações digitais.

A principal contribuição deste trabalho é a simplificação do processo de integração de ferramentas ao Modelo Hipátia, por meio do uso de um protocolo já incorporado ao *DSpace*. Essa abordagem facilita a adoção do Modelo Hipátia por instituições que já utilizam o *DSpace* como repositório digital, uma vez que não requer a implementação de protocolos adicionais ou extensas modificações no sistema.

REFERÊNCIAS

BERNERS-LEE, Tim et al. Creating a Science of the Web. *Science*, v. 313, n. 5788, p. 769-771, 2006. Disponível em: DOI: 10.1126/science.1126902. Acesso em 11 jun. 2024.

BRAGA, T. E. N. O modelo Hipátia: a proposta do Ibict para a preservação digital arquivística. In: BRAGA, T. E. N.; MÁRDERO ARELLANO, M. Á. (org.). **Hipátia: modelo de preservação para repositórios arquivísticos digitais confiáveis**. Brasília: Ibict, 2022. p. 52-65. (Informação, Tecnologia e Inovação, v. 1). DOI 10.22477/9786589167501.cap4. Disponível em: <http://labcotec.ibict.br/omp/index.php/edcotec/catalog/view/livrohipatia/136/943>. Acesso em: 17 jun. 2024.

BRAGA, T. E. N.; MÁRDERO ARELLANO, M. Á. **Hipátia: modelo de preservação para repositórios arquivísticos digitais confiáveis**. Brasília, DF: Ibict, 2022. Disponível em: <http://labcotec.ibict.br/omp/index.php/edcotec/catalog/view/livrohipatia/136/943>. Acesso em: 18 jun. 2024.

DE SOUZA, A. H. L. R. et al. O modelo de referência OAIS e a preservação digital distribuída. **Ciência da Informação**, v. 41, n. 1, 2012. Disponível em: <https://revista.ibict.br/ciinf/article/view/1352>. Acesso em: 15 jun. 2024.

LAGOZE, C. et al. Open Archives Initiative Object Reuse and Exchange (OAI-ORE). **Open Archives Initiative**, 2008.

SHINTAKU, M. et al. Uso do protocolo de interoperabilidade OAI-ORE para preservação digital. **Revista Brasileira de Preservação Digital**, Campinas, SP, v. 4, n. 00, p. e023012, 2023. DOI: 10.20396/rebpred.v4i00.17992. Disponível em: <https://econtents.bc.unicamp.br/inpec/index.php/rebpred/article/view/17992>. Acesso em: 17 jun. 2024.



SILVEIRA, Lucas Ângelo et al. **Guia para a Integração do DSpace com o TemaTres**. Brasília, DF: Ibict, 2020. Disponível em:

https://ridi.ibict.br/bitstream/123456789/1118/1/SILVEIRA_GOMES_MENESES_SHIN_TAKU_%20Integra%c3%a7%c3%a3oDspaceETematres_2020.pdf. Acesso em 10 jun. 2024.

VAN DE SOMPEL, H.; LAGOZE, C. Notes from the interoperability front: a progress report on the Open Archives Initiative. In: **International Conference on Theory and Practice of Digital Libraries**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002. p. 144-157.

WAZLAWICK, Raul Sidnei. **Metodologia de pesquisa para ciência da computação**. Rio de Janeiro: Elsevier, 2009.

