Tendências da Pesquisa Brasileira em Ciência da Informação

O MODELO HIPÁTIA: A PROPOSTA DO IBICT PARA IMPLEMENTAÇÃO DE RDC-ARQ NO BRASIL[1]

HIPÁTIA MODEL: IBICT'S PROPOSAL TO IMPLEMENT TDAR IN BRAZIL

Tiago Emmanuel Nunes Braga[2]

Resumo: A preservação arquivística é uma necessidade institucional e social. Nos últimos anos, diversas demandas relacionadas à implementação de repositórios arquivísticos digitais confiáveis foram normalizadas por meio de resoluções, decretos e diretrizes. Todavia, implementar este tipo de repositório não é uma tarefa trivial, tendo em vista a grande quantidade de requisitos que devem ser atendidos. Este trabalho descreve um projeto de pesquisa executado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia a partir de uma demanda do Tribunal de Justiça do Distrito Federal e Territórios. São descritos os passos realizados para se implementar um repositório arquivístico digital confiável que fosse capaz de preservar o diário de justiça eletrônico. As etapas de identificação dos principais atores, levantamento arquivístico, levantamento computacional, estruturação de uma proposta informacional e prova do modelo em uma estrutura organizacional foram detalhadas e os recursos informacionais que as apoiam foram descritos. Como resultado, este trabalho apresenta o modelo Hipátia, primeira experiência brasileira de implementação de um repositório confiável que atende aos requisitos apresentados pelo Conselho Nacional de Arquivos na manutenção da cadeia de custódia. Neste artigo foram incorporados os aprofundamentos obtidos a partir da continuação da pesquisa junto aos parceiros do projeto.

Palavras-Chave: Hipátia. RDC-Arq. Preservação digital.

Abstract: *Archival preservation is an important institutional and social requirement. In recent years, several demands related to the implementation of trusted digital archival repositories have been normalized through norms, decrees, and guidelines. However, implementing this type of repository is not a trivial task given the large number of requirements that must be met. This paper describes a research project executed by the Brazilian Institute of Information on Science and Technology based on the demands of the Court of Justice of the Federal District and Territories. The steps taken to implement a trusted digital archival repository that would preserve the electronic justice diary are described. The stages of the stakeholder identification, archival survey, computational survey, structuring of an informational proposal, and application of the model on the organizational environment were detailed, and the informational resources that support them were described. As a result, this paper presents the Hypatia model, the first Brazilian experience of implementing a trusted repository that meets the requirements presented by the National Council of Archives related*

[1] Texto ampliado a partir do artigo submetido, avaliado, aprovado, apresentado e premiado no XXII ENANCIB.

[2] Doutor em Ciência da Informação (UnB). Pesquisador do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). E-mail: tiagobraga@ibict.br. Orcid: https://orcid.org/0000-0001-6332-7965.

to the custody chain. This paper deepens on the research results as the research process keeps moving forward.

Keywords: *Hipátia. TDAR. Digital Preservation.*

1 INTRODUÇÃO

A necessidade de se preservar o conhecimento existente hoje em dia, a fim de que a sociedade tenha condições de poder acessá-lo no futuro, é tema de várias pesquisas no Brasil e no mundo. O estudo de Chowdhury (2010) acerca da pesquisa sobre preservação reforça o entendimento de que a principal missão das ações de preservação é garantir que as próximas gerações tenham condições de acessar o conhecimento que foi produzido pela sociedade atual. A tarefa de garantir a preservação dos objetos digitais é urgente, dado o enorme número de conteúdos produzidos diariamente, cerca de 40 trilhões de *gigabytes* produzidos no último ano, segundo a revista Exame (2021). Ela também é desafiadora, tendo em vista a acelerada transformação das práticas digitais, que acabam tornando formatos de manipulação de arquivos. Padrões computacionais que antes eram unanimidade, acabam ao par de anos e até mesmo meses, se tornando em estruturas obsoletas, que não possuem os suportes necessários para serem apresentados.

A preservação digital requer a realização de procedimentos complexos e que envolvem todos os aspectos da informação. Além da preservação dos atributos físicos da mídia que guarda a informação digital, é necessário se ater a outras características, como a atualização dos formatos e a manutenção do contexto. Essas ações garantirão que o item a ser preservado possa ser semanticamente compreendido quando for novamente acessado. O modelo de referência DELOS (Candela *et al.*, 2008) destaca três aspectos principais relacionados à preservação digital:

> Manter fisicamente intacta a instância da entidade digital frente a deterioração da mídia física de armazenamento e os sinais gravados nela. Garantir que a sintaxe da entidade digital (sua codificação e formato) pode ser interpretada e que cada

> instância subsequente (acesso, renderização, manipulação etc) é idêntica à instância inicial (com relação ao comportamento, incluindo aparência e sentimento, ou funcionalidade etc). Garantir que o significado semântico da entidade digital é acessível ao longo do espaço e tempo em face das mudanças tecnológicas e culturais (Candela *et al.*, 2008, p. 54, tradução nossa).

Logo, a preservação é um processo sistêmico, que precisa ser realizado a partir da incorporação de várias abordagens. A fim de que todas as perspectivas necessárias sejam consideradas, as instituições usualmente priorizam o tipo de conteúdo a ser preservado. Dessa forma, os esforços são centrados na preservação de um tipo específico de conteúdo, possibilitando a realização de estudos aprofundados para que as especificidades de cada objeto digital sejam analisadas e absorvidas pelas práticas de preservação.

Com vistas a essa especialização no projeto de preservação, várias instituições definiram tipologias específicas para iniciar seus projetos de preservação. O Bureau Europeu de Associações de Biblioteconomia, Informação e Documentação definiu duas prioridades no decorrer das discussões acerca da temática (Eblida, 2010): a preservação das heranças culturais e a preservação das informações científicas. No Brasil, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) também se ocupou em um primeiro momento em promover a preservação das informações científicas (Márdero Arellano, 2008).

Assim como as informações científicas e a herança cultural, outros tipos de preservação digital foram sendo mapeados e demandados pela sociedade, a exemplo dos arquivos multimídia, das páginas de internet e dos conteúdos arquivísticos. No caso da preservação arquivística, ela foi promovida em grande parte pela publicação da resolução nº 39/2014 do Conselho Nacional de Arquivo (CONARQ) e sua atualização, a resolução 43/2015 (CONARQ, 2015), quando foram estipulados os requisitos para a criação de Repositórios Arquivísticos Digitais Confiáveis (RDC-Arq).

Assim como o CONARQ, outras importantes instituições iniciaram a discussão para normatizar o uso dos RDC-Arq. Uma dessas instituições é o Conselho Nacional

de Justiça (CNJ), que publicou a resolução nº 324/2020 (CNJ, 2020). Nela, é estabelecido o Programa Nacional de Gestão Documental e Memória do Poder Judiciário (Proname), bem como são descritas as diretrizes e normas da gestão da memória e da gestão documental.

No que tange à implementação do RDC-Arq, há uma indicação explícita no normativo do CNJ a respeito da forma como ela deve ser sistematizada:

> Art. 34. Para fins de preservação digital, os órgãos do Poder Judiciário adotarão repositório arquivístico digital confiável (RDC-Arq), desenvolvido como *software* livre, gratuito e de código aberto, projetado para manter os dados em padrões de preservação digital e o acesso em longo prazo (CNJ, 2020).

Além das dificuldades inerentes às diretrizes apontadas pelos documentos do CONARQ e do CNJ, um dos grandes desafios na implementação dos RDC-Arq paira sobre a manutenção ininterrupta da cadeira de custódia, conforme aponta Flores, Rocco e Santos (2016). Este tópico é primordial para o sucesso de um RDC-Arq, já que apenas a certeza da autenticidade do objeto digital é capaz de garantir a confiabilidade inerente a este tipo de solução.

Visando atender às normativas citadas anteriormente, bem como sua própria normativa, o Tribunal de Justiça do Distrito Federal e Territórios (TJDFT) iniciou discussões internas relacionadas à implantação de RDC-Arq. As indicações sobre como deveria ser a gestão documental do TJDFT são estabelecidas na resolução nº 2/2018. Dentre outras indicações, detalhou-se como seria o escopo de utilização do RDC-Arq no Tribunal (TJDFT, 2023).

Diante dos desafios encontrados para o estabelecimento de ações práticas de implementação do RDC-Arq, o TJDFT entendeu por bem convidar o IBICT para apoiar a iniciativa. Tal convite culminou com a realização de um Termo de Execução Descentralizada (TED) entre as duas instituições e que tinha como objetivo:

> Apresentar alternativas de soluções para implantação de um repositório arquivístico digital confiável - RDC - Arq, integrado com solução de gerenciamento documental, tornando possível a elaboração e disseminação de modelo tecnológico de mapeamento, preservação e disseminação das informações associadas (TJDFT, 2018).

A execução da pesquisa por parte do IBICT, fruto da parceria, resultou na implantação do primeiro RDC-Arq do Brasil. Seu grande diferencial é a garantia de manutenção da cadeia de custódia, algo inédito em um ambiente de produção. O RDC-Arq foi implementado com o sistema Diário de Justiça Eletrônico (DJe) do TJDFT. Ainda como fruto da parceria, foi estruturado o modelo Hipátia, a proposta do IBICT para preservação digital arquivística, que tem sido utilizado por diversas instituições nacionais, a exemplo do Arquivo Nacional, do Tribunal de Justiça de Minas Gerais, do Tribunal Regional do Trabalho da 4ª região, dentre outras.

Este trabalho buscou apresentar a pesquisa executada pelo IBICT no âmbito da parceria com o TJDFT, bem como discorrer sobre as bases informacionais que foram consideradas durante a construção do modelo Hipátia.

2 METODOLOGIA

A pesquisa foi realizada a partir de uma perspectiva qualitativa e teve caráter aplicado. A metodologia utilizada foi a proposta durante a formalização da parceria entre o TJDFT e o IBICT e está explicitada no TED. Ela consiste em modelos de análise documental e experimentação e objetiva implementar um RDC-Arq. Os trabalhos foram organizados em quatro metas:

- Levantamento de informações;
- Implantação de repositório;
- Estudo de integração de fluxo de dados; e
- Transferência de tecnologia e disseminação do conhecimento.

Em cada uma delas foram realizados diversos procedimentos metodológicos,

conforme apresentado a seguir.

Visando alcançar as metas propostas, foi realizada inicialmente a análise documental de toda a legislação pertinente à temática. A fim de se aprofundar no entendimento das demandas internas do TJDFT, optou-se pela realização de entrevistas com atores-chave para a temática na instituição, tal como realização de reuniões regulares de acompanhamento, quando foram validadas as estratégias montadas.

Outro procedimento metodológico se apoiou na análise comparativa de programas que poderiam atender os requisitos apontados pelo CONARQ e CNJ. Neste processo, diversos programas de apoio foram instalados. Suas principais características foram sistematizadas de modo a identificar as configurações adequadas para implementação do RDC-Arq. O ambiente que seria preservado, DJe, foi também analisado e seus dados e metadados foram mapeados. Esta etapa visou estabelecer as bases para a proposição de um dicionário de dados. A pesquisa também demandou uma análise compreensiva da infraestrutura de tecnologia da informação do TJDFT de forma a identificar como poderia se dar a coleta e a transferência de dados entre as etapas de implantação do RDC-Arq.

Com os levantamentos realizados anteriormente, foi então proposto um modelo de implementação do RDC-Arq: o Hipátia. Este modelo possuía como requisito a existência de um barramento de extração e preparação, que foi desenvolvido no decorrer da pesquisa. Também previa a incorporação de diversas configurações entre os programas escolhidos para apoiar a implementação do RDC-Arq. Como última etapa da pesquisa, o modelo Hipátia foi sistematizado e disponibilizado para que pudesse ser aplicado em outras instituições.

3 RESULTADOS E ANÁLISES

A seguir, são descritos os resultados obtidos durante a pesquisa. Eles são divididos em levantamentos arquivísticos, levantamentos computacionais, sistematização do modelo Hipátia e aplicação no ambiente do TJDFT. O primeiro resultado identificado na pesquisa diz respeito à responsabilização pela implementação do RDC-Arq. O CONARQ prevê que as equipes de gestão documental e de tecnologia da informação seriam aquelas responsáveis pela implementação do repositório no âmbito institucional. Embora estas duas equipes, de fato, recebam esta tarefa, ela precisa ser compartilhada também com a equipe responsável pela gestão institucional em alto nível, a fim de garantir os subsídios necessários para o sucesso da empreitada. Assim, a implementação do RDC-Arq é baseada na atuação de três áreas da organização, conforme mostra a figura a seguir.

Figura 1: Responsáveis pelo RDC-Arq.



Fonte: elaborada pelo autor.

Uma vez identificados os principais atores no processo de viabilização da implementação do RDC-Arq, concebeu-se as fases seguintes na estruturação do projeto. Os dois primeiros passos passavam pelo levantamento de informações relacionadas à gestão documental, aqui denominado de levantamentos arquivísticos,

e o levantamento dos aspectos computacionais da instituição. Estas fases são descritas a seguir.

3.1 Levantamentos arquivísticos

Os metadados levantados no âmbito do projeto ao se analisar o sistema DJe podem ser divididos em duas categorias: metadados de gestão e metadados descritivos. A definição desses metadados levou em consideração o MoReq-Jus, instituído pela Resolução nº 91 de 2009 (CNJ, 2009) e o e-ARQ Brasil, instituído pela Resolução nº 25 de 2007 (CONARQ, 2007), recentemente substituída pela Resolução nº 50 de 2022.

Os metadados de gestão foram extraídos do próprio DJe e são compostos, inicialmente, por 22 itens. Um extrato da lista de metadados de gestão mapeados é apresentado no quadro a seguir.

Quadro 1: Metadados de gestão mapeados.

Descrição	Campo CSV	Tabela	Exemplo
Número gerado pelo sistema na hora de sua captura. Geralmente é o campo ID de uma tabela.	numeroSistema	[DB].DiarioEletronico.ID	5620
Identificar de forma unívoca e persistente o documento. Podemos usar no caso do DJe, por exemplo, o ano concatenado com o número do Diário da Justiça.	numeroIdentificador	[DB].DiarioEletronico.Ano + [DB].DiarioEletronico.Numero	201961
[...]	[...]	[...]	[...]
Registro do algoritmo utilizado para o criptografia do cálculo do checksum. Exemplo: SHA, MD5	algoritmoChecksum	Utilizar de codificação para salvar esse dado.	sha-256

Fonte: Ibict (2019b).

Além dos metadados de gestão, foram estipulados os metadados descritivos,

que seguiram dois padrões: ISAD(G) e Dublicore. A utilização dos dois padrões procurou atender a infraestrutura computacional proposta para implementar o RDC-Arq, tendo em vista que alguns sistemas utilizavam o ISAD(G) e outros usavam o Dublicore. Dada a arquitetura informacional que foi desenvolvida para o Hipátia, é possível adicionar qualquer outro padrão de metadados sem maiores transtornos. Os metadados incorporados para os dois padrões foram extraídos do ambiente do DJe ou preenchidos de forma estática, quando aplicável. No primeiro momento foram mapeados 15 metadados seguindo o padrão Dublincore. A lista de metadados inicialmente mapeados foram: dc.title, dc.creator, dc.subject, dc.description, dc.publisher, dc.contributor, dc.date, dc.type, dc.format, dc.identifier, dc.source, dc.language, dc.relation, dc.coverage e dc.rights.

Já no padrão ISAD(G), foram selecionados os seguintes metadados: identifier, title, eventDates, levelOfDescription, extentAndMedium, eventActors, eventActorHistories, archivalHistory, acquisition, scopeAndContent, appraisal, accruals, arrangement, accessConditions, reproductionConditions, language, physicalCharacteristics, findingAids, locationOfOriginals, locationOfCopies, relatedUnitsOfDescription, publicationNote, generalNote, archivistNote, rules e revisionHistory. Estes 26 metadados são uma lista exemplificativa, mas outros metadados foram adicionados no decorrer do projeto. A lista completa dos metadados utilizados está disponível no repositório GIT destinado ao Hipátia[3]. Todos os metadados foram mapeados a partir de entrevistas e reuniões com a equipe do TJDFT.

A tabela de temporalidade do TJDFT foi considerada a fim de pautar a forma como seria feita a integração entre o RDC-Arq e o DJe. Foram levantadas duas

[3] O acesso ao repositório GIT do projeto poderá se dar pelo endereço https://git.ibict.br/cgti/barrapres_/Dje mediante um cadastro prévio. O cadastro precisa ser realizado neste momento por motivo de segurança, mas espera-se que esta exigência seja eliminada em breve.

possibilidades: a primeira, denominada de Integração Ativa, ocorre de forma que o RDC-Arq acesse ativamente o ambiente DJe a fim de verificar se existem novas atualizações e proceder com a preservação dos objetos informacionais. Já a segunda, denominada de Por Demanda, funciona quando o ambiente que está sendo preservado acessa o RDC-Arq, passando parâmetros para iniciar o processo de preservação. Uma vez que todas as edições publicadas no DJe precisam ser preservadas, optou-se pela Integração Ativa, que ocorre a cada 24 horas.

Por fim, nos levantamentos arquivísticos foi definida a estrutura do dicionário de dados do DJe. Este dicionário é responsável por mapear os dados e metadados extraídos do DJe para o RDC-Arq, funcionando como um tradutor das tabelas do sistema para os metadados de gestão e descritivos. Para que cada sistema seja integrado com o Hipátia, seu próprio dicionário de dados precisa ser definido. Este documento é lido pelo RDC-Arq, e então, os metadados são coletados para permitir que o contexto de utilização do objeto informacional que está sendo preservado seja resgatável no futuro. A figura 2 apresenta um extrato do dicionário de dados que foi construído para o DJe.

Figura 2: Dicionário de dados do DJe

```
<mapping>
    <temporality field="DiarioEletronico.DataPublicacao">+2</temporality>
    <filename format="object/"></filename>
    <relation table="tjdf_dje.DiarioEletronico">
        <reference column="ID">tjdf_dje.DatasDePublicacao.ID</reference>
    </relation>

    <metadatas table="tjdf_dje.DiarioEletronico">
        <nomePasta>'dje_' Ano '_' Numero</nomePasta>
        <nomeArquivo>'DJ' Numero '_' Ano</nomeArquivo>
        <filename>'objects/DJ' Numero '_' Ano '.pdf'</filename>

        <!--<"Gestão">-->
        <numeroSistema>ID</numeroSistema>
        <numeroIdentificador>Ano Numero</numeroIdentificador>
        <responsavelPubLogin>Login</responsavelPubLogin>
        <dataCriacao>DataHoraDeModificacao</dataCriacao>
        <dataPublicacao>DataDePublicacao</dataPublicacao>
        <pdf>Ano Numero</pdf>
        <anoDiario>Ano</anoDiario>
        <numeroDiario>Numero</numeroDiario>

        <!--<"Dublin Core">-->
        <dc.title>'Edição_nº_' Numero '/' Ano</dc.title>
        <dc.date>DataDePublicacao</dc.date>
        <dc.identifier>Ano Numero</dc.identifier>

        <!--<"ISAD(G)">-->
        <isad.title>'Edição_nº_' Numero '/' Ano</isad.title>
        <isad.eventDates>DataDePublicacao</isad.eventDates>
        <isad.identifier>Ano Numero</isad.identifier>
        <isad.sources>'https://pesquisadje-api.tjdft.jus.br/v1/diarios/pdf/' Ano '/' Numero '.pdf'</isad.sources>
    </metadatas>
```

Fonte: Ibict (c2021).

Visando promover a interoperabilidade, o dicionário de dados do DJe foi construído na linguagem textual XML, largamente utilizada.

3.2 Levantamentos computacionais

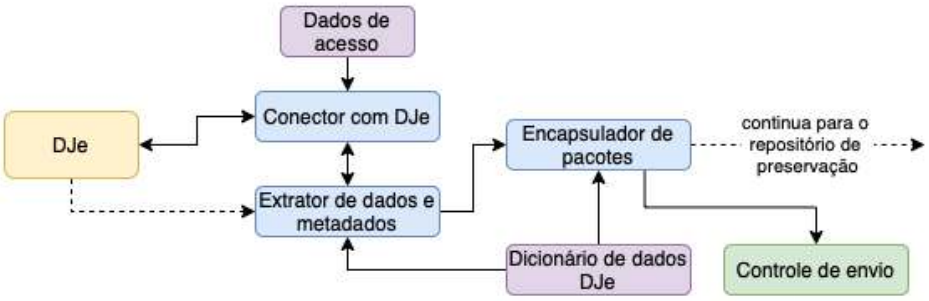
O levantamento computacional considerou a infraestrutura tecnológica adotada pelo TJDFT e as opções de programas consolidados que poderiam ser consideradas para implementação do RDC-Arq. Foram mapeados dois programas já existentes: o Archivematica e o AtoM. Todavia, não foi possível identificar um programa capaz de conectar-se ao DJe e realizar a extração e preparação do conteúdo. Para isso, foi desenvolvido um programa denominado BarraPres. Além da parte de programas, foram apreciadas questões referentes à infraestrutura de rede do TJDFT. Isso se deu pela necessidade de garantir o funcionamento seguro da solução RDC-Arq.

Com relação aos programas que fazem parte do modelo, passa-se a descrever como se deu a utilização de cada um deles.

3.2.1 Barrapres

O BarraPres foi inteiramente desenvolvido no IBICT. Apoiado na linguagem Python, ele possui duas funções principais: extrair os dados e metadados do DJe e prepará-los para envio ao Archivematica. Para tanto, o BarraPres foi estruturado a partir de uma arquitetura simplificada, permitindo que adaptações sejam realizadas de forma otimizada. Entre os componentes que são incorporados no BarraPres estão: conector com o sistema produtor, dicionário de dados do sistema produtor e estrutura de empacotamento. A figura 3 apresenta o esquema de funcionamento do BarraPres.

Figura 3: Esquema de funcionamento do BarraPres.



Fonte: elaborado pelo autor.

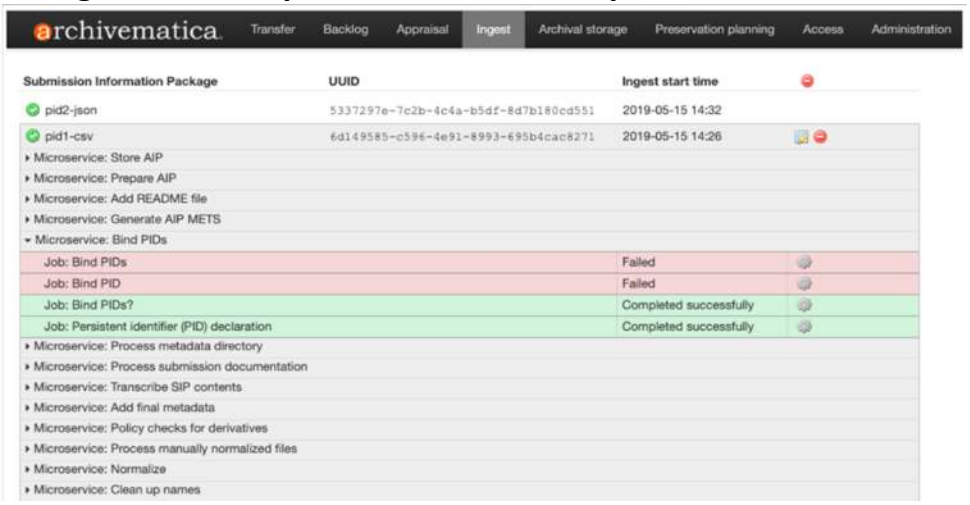
Os itens marcados na cor roxa são os que eventualmente deverão ser alterados caso o BarraPres seja utilizado junto ao DJe de outra instituição. Enquanto o módulo de Dados de Acesso recebe os dados de conexão, o módulo Dicionário de dados DJe recebe os metadados que devem ser preservados na nova instância. Entende-se que cada instituição priorizará um conjunto de metadados personalizado para além daqueles obrigatórios já definidos.

Foram vislumbrados três formatos de conexão entre o BarraPres e o sistema produtor: acesso direto ao banco de dados, API Rest, acesso aos diretórios e pastas do sistema operacional. Para o caso do DJe, entendeu-se que o acesso direto ao banco era a melhor opção, tendo em vista a simplicidade do sistema. Sendo assim, foi necessário configurar um usuário com acesso somente leitura à base oficial do DJe do TJDFT. Com isso, o BarraPres pôde gerar os pacotes e enviá-los para o repositório de preservação. No módulo de Controle de Envio são registrados todos os pacotes enviados para preservação, de forma a evitar envios duplicados e a realizar a comunicação em caso de algum erro de execução.

3.2.2 Archivematica

O Archivematica foi escolhido por ser a solução de acesso livre mais compreensiva e completa em termos de preservação arquivística. Ele foi pensado para ser o repositório de preservação do modelo Hipátia, ou seja, o sistema que recebe os pacotes enviados pelo BarraPres. A figura 4 mostra uma das etapas de funcionamento do Archivematica.

Figura 4: Microprocessos realizados pelo Archivematica.

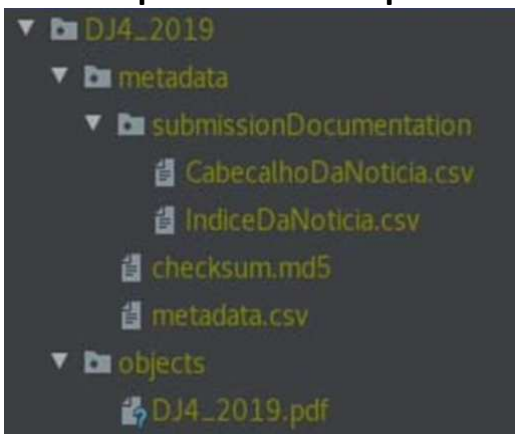


Fonte: Ibict (2019a).

Um dos pontos fundamentais para o sucesso de implementação do RDC-Arq foi a configuração automática para que o BarraPres fosse capaz de submeter pacotes diretamente para o Archivematica, sem a necessidade de interferência humana. Com isso, todo o processo de preservação foi automatizado, bastando alguma atuação humana apenas quando eram encontrados erros que não podiam ser solucionados dinamicamente pelo sistema. O mapeamento de erros do Archivematica é realizado por meio de seus microprocessos de *Ingest*.

Para envio dos pacotes de preservação para o Archivematica foi necessário definir o formato do pacote. Após a análise de todas as possibilidades oferecidas pelo programa, optou-se por um pacote que possibilitasse o envio de forma simplificada ao mesmo tempo que compreendesse todas as informações necessárias para que o pacote pudesse ser acessível no futuro, mantendo sua autenticidade. A figura 5 apresenta a estrutura adotada.

Figura 5: Estrutura do pacote enviado para o Archivematica.



Fonte: Ibict (2019a).

Com o envio do pacote para o Archivematica e a realização bem-sucedida dos microprocessos, os conteúdos preservados já estão prontos para serem enviados para disseminação. Todavia, embora a utilização do AtoM esteja prevista no modelo Hipátia, para este primeiro projeto a opção foi por finalizar o processo com a

verificação de que o pacote encontrava-se preservado no Archivematica. Já existem implementações do modelo Hipátia nas quais a disponibilização do AtoM é contemplada. Todavia, estas implementações não compõem o escopo desta pesquisa.

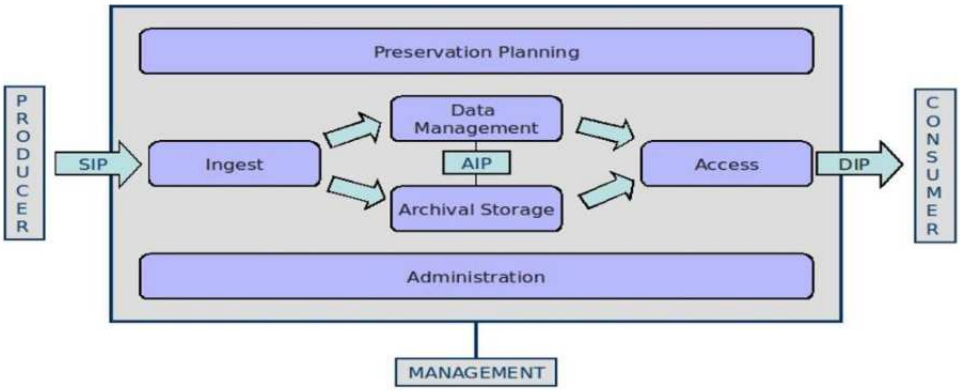
Além das configurações dos programas descritos anteriormente, foram necessários criar usuários locais na rede do TJDFT. Tal demanda era necessária para permitir a realização de ações automatizadas sem a necessidade de interferência humana. Outra opção computacional adotada pelas equipes foi a segmentação da estrutura de rede em diversas redes virtuais (VLAN, na sigla em Inglês). Com isso, objetivou-se dificultar a ação de criminosos que eventualmente tenham como objetivo acessar os ativos informacionais do TJDFT.

3.3 A estruturação do modelo Hipátia

Após os levantamentos arquivísticos e computacionais, foi possível estruturar uma proposta de modelo capaz de apoiar a implementação de RDC-Arq. A estruturação de tal proposta tinha como premissa atender ao modelo de referência Open Archival Information Systeam (OAIS). Entre os principais diferenciais do OAIS está a proposição de separação entre os pacotes de submissão, preservação e disseminação. Aumentando ou diminuindo a complexidade da estrutura do pacote a depender de sua utilização.

A figura 6 apresenta o esquema do modelo de referência OAIS.

Figura 6: Esquema proposto pelo modelo de referência OAIS



Fonte: ISO (2012).

O OAIS prevê três principais atores: o produtor de conteúdo, o gerenciador do processo de preservação e o consumidor de conteúdo. Cabe ao sistema produtor de conteúdo gerar a informação a ser preservada, o que ocorre por meio do pacote de submissão de informação (SIP, na sigla em Inglês). Este pacote passará pelo processo de *Ingest*, tal como mostrado anteriormente na análise do Archivematica, e então passará pelo gerenciamento de dados e pelo armazenamento arquivístico, quando o pacote de arquivamento de informação é gerado (AIP, na sigla em Inglês). Para viabilizar o acesso, é disponibilizado o pacote de disseminação da informação (DIP, na sigla em Inglês). Todo este processo ocorre sob a guarda de um planejamento de preservação e administração.

Ao se propor a estruturação do modelo Hipátia, o OAIS serviu como parâmetro de referência. No caso do Hipátia foram concebidas algumas ações adicionais, que não eram previstas no OAIS, mas que eram primordiais para garantir o sucesso de implementação do RDC-Arq.

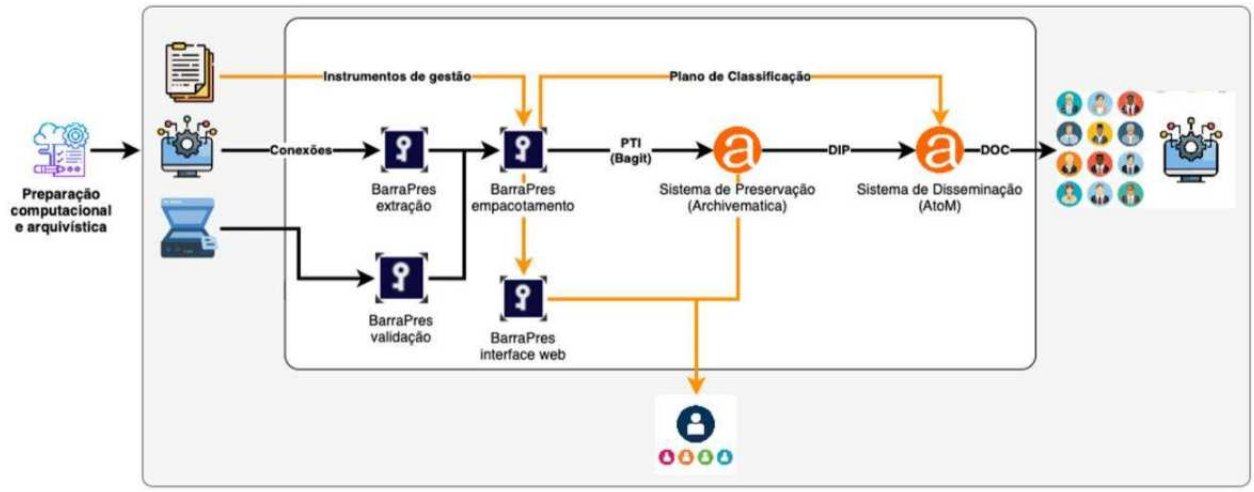
O Hipátia funciona a partir de um sistema produtor de massa documental. Até o momento não foram encontradas restrições com relação à abrangência dos sistemas produtores, o que serve como um indicativo de que o Hipátia pode se adaptar a diversos cenários. Além disso, não é necessário que o sistema produtor seja

um Sistema Informatizado de Gestão Arquivística de Documentos (SIGAD), ou seja, a preparação arquivística pode ser realizada inteiramente no escopo do modelo Hipátia por meio do BarraPres.

Cabe ao BarraPres acessar o sistema a ser preservado, extrair os conteúdos definidos pela equipe de gestão documental, preparar o pacote de preservação e então encaminhar para a fase de preservação. Neste processo são utilizadas algumas estruturas avançadas, como um dicionário de dados, um módulo de conexão e um módulo de empacotamento. Embora não tenha sido o caso descrito sobre o DJe, atualmente o BarraPres gera os pacotes seguindo o padrão BagIt, um avançado mecanismo de organização de documentos, dados e metadados para preservação (Kunze *et al.*, 2018).

Por ser baseado no modelo de referência OAIS, o modelo Hipátia buscou estruturar suas fases propostas em mecanismos que representassem o modelo de referência. Sendo assim, há fluxos que remetem aos pacotes SIP, AIP e DIP. Além disso, há a criação do pacote de transferência inicial (PTI), bem como fluxos que permitem considerar as políticas de preservação existentes na instituição. A figura 7, apresentada a seguir, sintetiza a estrutura de funcionamento do modelo Hipátia.

Figura 7: Estrutura do modelo Hipátia.



Fonte: elaborada pelo autor.

Para a preservação o modelo vislumbra a utilização do software o Archivematica. Eventualmente, este sistema poderá ser substituído por outros sistemas desde que os recursos previstos pela resolução do CONARQ sejam contemplados. Os pacotes AIP e DIP são gerados por este sistema, que realiza uma série de microprocessos a fim de garantir a preservação do conteúdo e do contexto, possibilitando o acesso futuro. O Archivematica é configurado para enviar aqueles pacotes marcados para disseminação diretamente para o AtoM, sem a necessidade de interferência humana neste processo. Além disso, o sistema Archivematica disponibiliza um avançado mecanismo de controle, garantido que os pacotes arquivados mantenham sua autenticidade enquanto estiverem ali armazenados.

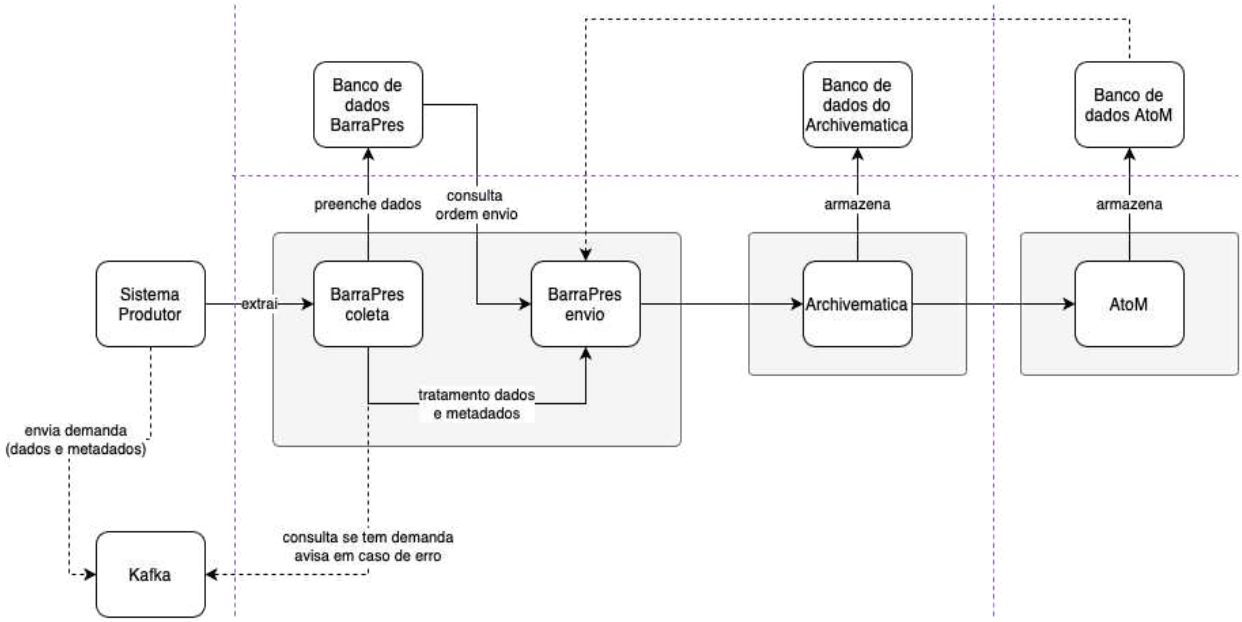
Por fim, os pacotes DIP são encaminhados para o AtoM a fim de que fiquem disponíveis para consumo futuro. O AtoM também poderá ser substituído por outros software, mas a integração entre o ambiente de preservação e de disseminação é fundamental para o sucesso na aplicação do modelo Hipátia. Foram definidas configurações que garantem ao AtoM recuperar a classificação existente no sistema produtor, facilitando a navegação por parte dos usuários. O ambiente também pode ser personalizado de forma a contemplar a identidade visual da instituição que está realizando a implementação do RDC-Arq.

3.4 A prova do modelo

O modelo Hipátia foi aplicado na infraestrutura computacional do TJDFT e atualmente encontra-se no ambiente de produção, ou seja, é entendido como uma das soluções informacionais adotadas por aquele Tribunal. Houve necessidade de se conectar diversos outros sistemas denominados auxiliares, como os bancos de dados, sistemas de eventos, sistemas de *log*, dentre outros. A organização de redes virtuais também permite aprimorar os mecanismos de segurança da informação. A forma

como se deu a implementação do modelo a fim de concretizar a prova da pesquisa é apresentada a seguir.

Figura 8: Organização do ambiente de prova do modelo.



Fonte: elaborada pelo autor.

Para suportar tal aplicação foram desenvolvidos ambientes de apoio que podem ser acessados por qualquer pessoa ou instituição interessada no Hipátia. Dentre as infra estruturas informacionais disponíveis para suporte ao Hipátia destacam-se: o fórum de suporte técnico, utilizado para dirimir quaisquer dúvidas sobre a aplicação do modelo; o site do modelo Hipátia, local em que são reunidas todas as documentações, exemplos e avanços do modelo e o repositório de códigos do modelo Hipátia, onde são disponibilizados os códigos de todas as implementações do Hipátia.

Dessa forma, além de se aplicar o modelo no proposto no âmbito do TJDFT, foram disponibilizadas infra estruturas abertas de apoio ao modelo a fim de fomentar a implementação do Hipátia em outras instituições que tenham demandas similares.

4 CONSIDERAÇÕES FINAIS

O modelo Hipátia consiste na proposta do Ibict para oferecer infra estruturas informacionais voltadas para a preservação arquivística. Sua sistematização somente foi possível por meio da construção da parceria com o TJDFT. Como resultado principal da pesquisa, tem-se a primeira implementação nacional de um RDC-Arq com manutenção da cadeia de custódia dos conteúdos preservados, garantindo a autenticidade destes conteúdos. A forma como o modelo Hipátia foi desenvolvido indica a importância de se estabelecer parcerias entre agentes públicos que possuam interesses comuns. Este tipo de atuação envolvendo órgãos públicos é, sem dúvida, um dos caminhos possíveis frente a um cenário de diminuição do corpo técnico disponível nas instituições.

Um dos indicativos do acerto na estruturação do modelo Hipátia pode ser vislumbrado pelo número de instituições que já estabeleceram parcerias com o IBICT a fim de implementá-lo. Além do TJDFT, que em 2021 renovou a parceria, também houve formalização de parcerias com o Arquivo Nacional e com o Tribunal de Justiça de Minas Gerais. Além dessas parcerias, atualmente a formalização de parcerias entre o IBICT e o Tribunal Regional do Trabalho da 4ª Região, o Superior Tribunal de Justiça, o Arquivo Público do Estado de São Paulo e o Arquivo Público do Estado do Rio Grande do Sul estão em estágio avançado.

No entanto, estima-se que diversos outros projetos apoiados no modelo Hipátia estão sendo executados, tendo em vista que o modelo foi disponibilizado de forma livre e gratuita por meio de uma série de canais. O número expressivo de palestras e reuniões de apresentação demandadas ao IBICT corroboram com esta impressão.

Com a construção deste modelo, o IBICT reafirma sua missão de fornecer infraestruturas informacionais voltadas para o desenvolvimento da ciência e

tecnologia nacional. O entendimento de que a preservação digital é um mecanismo de salvaguarda das informações institucionais, e, neste caso, de incentivo à transparência organizacional, contribui para entender a importância do projeto frente ao movimento que busca recontar a atuação de atores públicos brasileiros diante de eventos históricos consolidados. Há espaço para se pensar em muitos aprimoramentos em relação ao modelo Hipátia, mas o estágio atual em que ele se encontra já pode ser considerado um grande avanço em termos de preservação digital arquivística. As instituições que estão na vanguarda da gestão documental e preservação da memória podem contar com o Hipátia como ponto de partida a fim de obterem sucesso nesta árdua tarefa de preservação do conhecimento humano.

REFERÊNCIAS

CANDELA, Leonardo; CASTELLI, Donatella; FERRO, Nicola; IOANNIDIS, Yannis; KOUTRIKA, Georgia; MEGHINI, Carlo; PAGANO, Pasquale; ROSS, Seamus; SOERGEL, Dagobert; AGOSTI, Maristella; DOBREVA-MCPHERSON, Milena; KATIFORI, Vivi; SCHULDT, Heiko. **The DELOS Digital Library Reference Model**: foundations for digital libraries. [*S. l.*]: DELOS, 2008.

CHOWDHURY, Gobinda. From digital libraries to digital preservation research: the importance of users and context. **Journal of Documentation**, *[s. l.]*, v. 66, n. 2, p. 207-223, 2010. DOI: https://doi.org/10.1108/00220411011023625. Disponível em: https://www.emerald.com/insight/content/doi/10.1108/00220411011023625/full/html. Acesso em: 8 jun. 2022.

CONSELHO NACIONAL DE ARQUIVOS (Brasil). Resolução nº 25, de 27 de abril de 2007. Dispõe sobre a adoção do Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos - e-ARQ Brasil pelos órgãos e entidades integrantes do Sistema Nacional de Arquivos - SINAR. **Diário Oficial da União**: seção 1, Brasília, DF, n. 81, p. 14-15, 27 abr. 2007. Disponível em: http://dspace.mj.gov.br/handle/1/407. Acesso em: 12 jun. 2022.

CONSELHO NACIONAL DE ARQUIVOS (Brasil). **Resolução nº 43, de 04 de setembro de 2015.** Altera a redação da Resolução do CONARQ nº 39, de 29 de abril de 2014,

que estabelece diretrizes para a implementação de repositórios digitais confiáveis para a transferência e recolhimento de documentos arquivísticos digitais para instituições arquivísticas dos órgãos e entidades integrantes do Sistema Nacional de Arquivos - SINAR. [*S. l.*]: CONARQ, 2015. Disponível em: https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes-do-conarq/resolucao-no-43-de-04-de-setembro-de-2015.

CONSELHO NACIONAL DE JUSTIÇA (Brasil). **Resolução nº 91, de 29 de setembro de 2009**. Institui o Modelo de Requisitos para Sistemas Informatizados de Gestão de Processos e Documentos do Poder Judiciário (MoReq-Jus) e disciplina a obrigatoriedade da sua utilização no desenvolvimento e manutenção de sistemas informatizados para as atividades judiciárias e administrativas no âmbito do Poder Judiciário. Brasília, DF: CNJ, 9 out. 2009. Disponível em: https://hdl.handle.net/20.500.12178/4030. Acesso em: 12 jun. 2022.

CONSELHO NACIONAL DE JUSTIÇA (Brasil). **Resolução nº 324, de 30 de junho de 2020**. 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. [*S. l.*]: CNJ, 2020. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3376. Acesso em: 12 jun. 2022.

EBLIDA. Digital agenda for Europe: digital libraries initiative: European Bureau of Library Information and Documentation Associations (EBLIDA). **Eblida**, *[s. l.]*, 2010. Disponível em: http://www.eblida.org/news/digital-agenda-for-europe-digital-libraries-initiative.html. Acesso em: 10 jun. 2022.

EXAME. Temos mais dados do que nunca. Como usá-los a nosso favor? **Exame**, *[s. l.]*, 9 jun. 2021. Disponível em: https://exame.com/carreira/dados-uso-favor/. Acesso em: 10 jun. 2022.

FLORES, Daniel; ROCCO, Brenda; SANTOS, Henrique. Cadeia de custódia para documentos arquivísticos digitais. **Acervo**, Rio de Janeiro, v. 29, n. 2, p. 117-132, 2016. Disponível em: https://revista.arquivonacional.gov.br/index.php/revistaacervo/article/view/717/732. Acesso em: 10 jun. 2022.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Git IBICT**. IBICT, Brasília, DF, c2021. Disponível em: https://git.ibict.br/users/sign_in. Acesso em: 10 jun. 2021.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Relatório Parcial:** estudo para Implantação de Repositório Arquivístico Digital Confiável: RDC-Arq no âmbito do Tribunal de Justiça do Distrito Federal e Territórios-TJDFT. Brasília, DF: IBICT, 2019a.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Relatório parcial**: Meta 2. [*S. l.*]: IBICT, 2019b.

INTERNATIONAL STANDARD ORGANIZATION. **ISO 14721:2012**: Space data and information transfer systems: Open Archival Information Aystem (OAIS): reference model. [*S. l.*]: ISO, 2012.

KUNZE, John A.; LITTMAN, Justin; MADDEN, Liz; SCANCELLA, John; ADAMS, Chris. **The BagIt File Packaging Format (V1.0)**. Request for Comments, [*s. l.]*, n. 8493, p. 1-25, 2018. DOI: 10.17487/RFC8493. Disponível em: https://datatracker.ietf.org/doc/rfc8493. Acesso em: 25 jan. 2022.

MÁRDERO ARELLANO, Miguel Ángel. **Critérios para a preservação digital da informação científica**. 2008. Tese (Doutorado em Ciência da Informação) – Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Universidade de Brasília, Brasília, DF, 2008. Disponível em: https://repositorio.unb.br/handle/10482/1518. Acesso em: 1 mar. 2021.

TRIBUNAL DE JUSTIÇA DO DISTRITO FEDERAL E TERRITÓRIOS (Brasil). RDC-Arq. **TJDFT**, Brasília, DF, 28 fev. 2023. Disponível em: https://www.tjdft.jus.br/institucional/gestao-do-conhecimento/gestao-documental/rdc-arq. Acesso em: 1 mar. 2021.

TRIBUNAL DE JUSTIÇA DO DISTRITO FEDERAL E TERRITÓRIOS (Brasil). **Termo de execução descentralizada entre órgãos da União 002/2018, que entre si celebram o TJDFT e o IBICT.** [*S. l.*]: Imprensa Nacional, 2018.

AGRADECIMENTOS

Agradeço aos órgãos que financiaram essa pesquisa: Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Tribunal de Justiça do Distrito Federal e Territórios (TJDFT), Arquivo Nacional (AN), Tribunal de Justiça de Minas Gerais (TJMG) e a Justiça do Trabalho por meio do Tribunal Regional do Trabalho da 4ª Região (TRT-4ª).

Copyright: Esta obra está licenciada com uma Licença Creative Commons Atribuição 4.0 Internacional.

tpbci@ancib.org @anciboficial @ancib_brasil