



— Coleção - Informação, Tecnologia e Inovação

HIPÁTIA: MODELO DE PRESERVAÇÃO PARA REPOSITÓRIOS ARQUIVÍSTICOS DIGITAIS CONFIÁVEIS

A N O 2 0 2 2



PRESIDÊNCIA DA REPÚBLICA

Jair Messias Bolsonaro
Presidente da República

Hamilton Mourão
Vice-Presidente da República

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES

Paulo César Rezende de Carvalho Alvim
Ministro da Ciência, Tecnologia e Inovações

INSTITUTO BRASILEIRO DE INFORMAÇÃO
EM CIÊNCIA E TECNOLOGIA

Cecília Leite Oliveira
Diretora

Reginaldo de Araújo Silva
Coordenador de Administração - COADM

Gustavo Saldanha
Coordenador de Ensino e Pesquisa, Ciência e Tecnologia da Informação - COEPE

José Luis dos Santos Nascimento
Coordenador de Planejamento, Acompanhamento e Avaliação - COPAV

Marcel Garcia de Souza
Coordenador-Geral de Pesquisa e Desenvolvimento de Novos Produtos - CGNP

Bianca Amaro de Melo
Coordenadora-Geral de Pesquisa e Manutenção de Produtos Consolidados - CGPC

Tiago Emmanuel Nunes Braga
Coordenador-Geral de Tecnologias de Informação e Informática - CGTI



Tiago Emmanuel Nunes Braga
Miguel Ángel Márdero Arellano
Organizadores

Hipátia: MODELO DE PRESERVAÇÃO PARA REPOSITÓRIOS ARQUIVÍSTICOS DIGITAIS CONFIÁVEIS

(Informação, Tecnologia e Inovação ; v. 1)

Coleção organizada por:
Tiago Emmanuel Nunes Braga
Gustavo Silva Saldanha

Brasília
2022

© 2022 Instituto Brasileiro de Informação em Ciência e Tecnologia

Esta obra é licenciada sob licença Creative Commons Attribution 4.0 (CC-BY 4.0), sendo permitida a reprodução parcial ou total, desde que mencionada a fonte.

Os autores são responsáveis pela apresentação dos fatos contidos e opiniões expressas nesta obra.

Comitê científico

Neide Alves Dias de Sordi

Murilo Bastos da Cunha

Miquel Termens

Autores

Alexandre Faria de Oliveira

Cristiano Menezes

Marcos Sigismundo da Silva

Maria de Fátima Duarte Tavares

Miguel Ángel Márdero Arellano

Milton Shintaku

Tatiana Canelhas Pignataro

Tiago Emmanuel Nunes Braga

Revisão gramatical e ortográfica

Poliana Mendes Martins

Normalização

Larissa Alves

Graziela Gomes

Design gráfico, diagramação e capa

Juliana Colem

Alisson Eugênio

Coleção organizada por: Tiago Emmanuel Nunes Braga e Gustavo Saldanha

Dados Internacionais de Catalogação na Publicação (CIP)

H667 Hipátia: modelo de preservação para repositórios arquivísticos digitais confiáveis / Tiago Emmanuel Nunes Braga; Miguel Ángel Márdero Arellano (orgs.); Alexandre Faria de Oliveira ... [et al.]. – Brasília, DF: Ibict, 2022.

101 p.

(Informação, Tecnologia e Inovação ; 1)

ISBN 978-65-89167-50-1

DOI 10.22477/9786589167501

1. Preservação digital. 2. Repositório arquivístico digital confiável – RDC-Arq. I. Braga, Tiago Emmanuel Nunes (org.). II. Márdero Arellano, Miguel Ángel (org.). III. Título

CDU 930.251:004

Ficha catalográfica elaborada por Larissa de Araújo Alves – CRB1/3462



www.gov.br/ibict/pt-br

Ibict – Brasília

Setor de Autarquias Sul (SAUS) Quadra 05 Lote 06, Bloco H

– 5º andar

CEP 70.070-912, Brasília, DF

Ibict – Rio de Janeiro

Rua Lauro Müller, 455 – Botafogo

CEP 22.290-160, Rio de Janeiro, RJ

Como citar:

BRAGA, Tiago Emmanuel Nunes; MÁRDERO ARELLANO, Miguel Ángel. **Hipátia**: modelo de preservação para repositórios arquivísticos digitais confiáveis. Brasília: Ibict, 2022.

SUMÁRIO

Prefácio

Tiago Emmanuel Nunes Braga

Apresentação

Cecília Leite Oliveira

Capítulo 1: Preservação de documentos em formato digital **10**

Miguel Ángel Márdero Arellano

Capítulo 2: O Ibict e a preservação da memória científica **26**

Maria de Fátima Duarte Tavares e Miguel Ángel Márdero Arellano

Capítulo 3: Os principais normativos relacionados ao RDCArq **40**

Tatiana Canelhas Pignataro, Cristiano Menezes e Tiago Emmanuel Nunes Braga

Capítulo 4: O modelo Hipátia: a proposta do Ibict para a preservação digital arquivística **52**

Tiago Emmanuel Nunes Braga

Capítulo 5: Arquitetura do modelo Hipátia **66**

Milton Shintaku e Alexandre Faria de Oliveira

Capítulo 6: Casos de sucesso de implementação do Modelo RDCArq e Hipátia **79**

Alexandre Farias de Oliveira, Milton Shintaku e Marcos Sigismundo da Silva

Capítulo 7: A visão do Ibict para um ecossistema nacional de preservação digital **91**

Alexandre Farias de Oliveira, Miguel Ángel Márdero Arellano e Tiago Emmanuel Nunes Braga



APRESENTAÇÃO

Ciente do potencial transformador inerente à informação, o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) se propõe, desde a década de 50 do século XX, a adotar uma postura de vanguarda no que diz respeito ao estudo e proposição de soluções associadas às temáticas vinculadas às variadas etapas do ciclo informacional. A reflexão sobre os aspectos que envolvem desde a criação de uma informação até a preservação desta para a posteridade é um ponto chave para o Instituto, que contribui com o desenvolvimento da infraestrutura informacional do Brasil.

Ao acreditar na pesquisa e na potência da informação, o Ibict, em parceria com o Tribunal de Justiça do Distrito Federal e Territórios (TJDFT), mergulhou em um trabalho de investigação sobre a preservação digital, a interoperabilidade entre sistemas de gestão documental e os Repositórios Arquivísticos Digitais Confiáveis (RDC-Arq). Dentre os frutos obtidos a partir do empenho de uma equipe de pesquisadores, obtivemos o Hipátia: modelo capaz de se adaptar a qualquer sistema que faça gestão de processos e objetos digitais, e que é apresentado no decorrer desta obra.

Estruturado em sete capítulos, o presente livro reúne os principais pontos a serem considerados para a compreensão do Modelo Hipátia: preservação de documentos digitais; a atuação do Ibict para a preservação da memória científica; a legislação e os normativos essenciais quando se trata de RDCArq; os componentes e arquitetura do Hipátia; relato dos casos de implementação; e, uma abordagem sobre um ecossistema nacional de preservação digital a partir do olhar do Ibict.

Ao relatar a possibilidade que a pesquisa fornece de viabilizar a atuação pioneira na busca de soluções possíveis e inovadoras para problemas informacionais levantados nos mais variados contextos, *o livro Hipátia: modelo de preservação para repositórios arquivísticos digitais confiáveis* constitui o primeiro volume da coleção Informação, Tecnologia e Inovação, que trata de assuntos pertinentes e indispensáveis para a atualidade e para a sociedade.

Desse modo, esperamos que este livro permita valiosas contribuições para todos os seus leitores e registre um dos passos dados na caminhada futura de grandes avanços ainda possíveis em relação às temáticas abordadas.

Cecília Leite Oliveira

Diretora do Instituto Brasileiro de Informação em Ciência e Tecnologia

PREFÁCIO

Quando o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) foi convidado pelo Tribunal de Justiça do Distrito Federal e Territórios (TJDFT) para propor e desenvolver uma pesquisa sobre os Repositórios Arquivísticos Digitais Confiáveis (RDC-Arq), se deparou com uma demanda informacional que tinha abrangência nacional e potencial de contribuição para diversos ramos da Administração Pública. À medida que a pesquisa obteve avanços, variados aspectos passaram a ser incorporados ao entendimento da temática investigada e a interpretação sobre requisitos inicialmente estabelecidos pela resolução 43/2015 do Conarq foi acrescida de novos aspectos informacionais, levantados pela competente equipe de gestão documental do TJDFT. Além das demandas informacionais, o referido Tribunal também levantou demandas computacionais relevantes para a impulsionar a evolução da pesquisa. A proposta inicial do Ibict era construir um barramento que pudesse, posteriormente, também ser utilizado em outros cenários e por outros órgãos públicos. Todavia, diante dos desafios identificados e da complexidade encontrada ao se avançar na pesquisa, percebeu-se que muito mais do que um barramento, era necessário estruturar um modelo capaz de apoiar a implementação dos sistemas utilizados no RDC-Arq. Dessa forma, surgiu o Hipátia: a proposta do Ibict para a preservação digital arquivística.

Este livro buscou contar um pouco da história da preservação digital no âmbito do Ibict, bem como situar os leitores sobre as propostas conceituais e as estratégias adotadas na consolidação do modelo Hipátia. Sabe-se que, apesar de inovador, o modelo Hipátia ainda não está finalizado. Logo, este livro compreende o ponto de partida desta construção, que está sendo feita de forma colaborativa por diversas instituições nacionais.

Boa leitura!

Tiago Emmanuel Nunes Braga



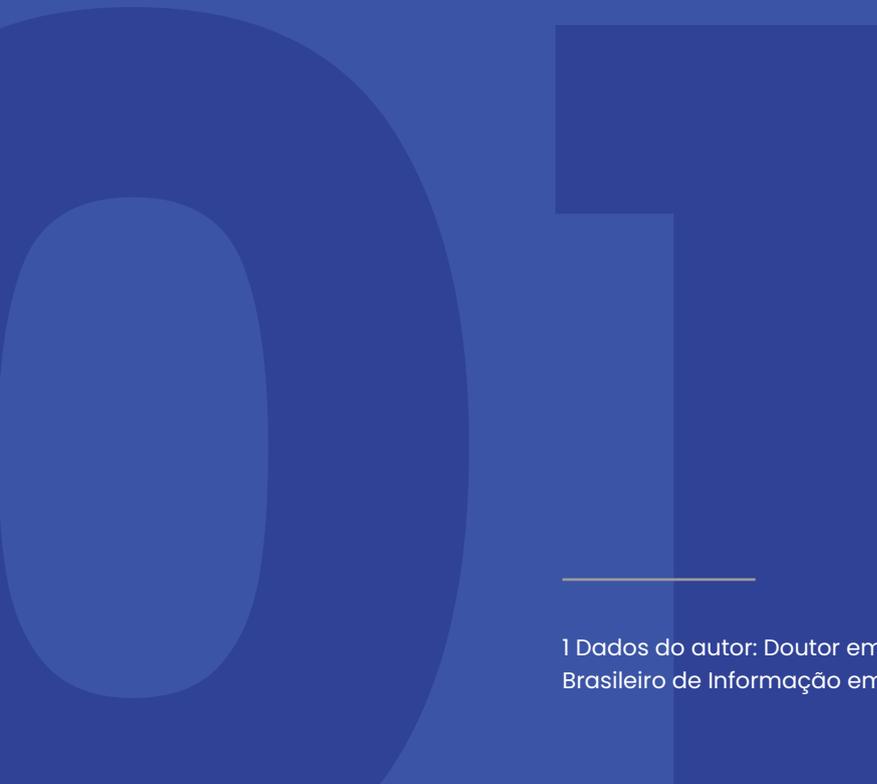


Capítulo 01

PRESERVAÇÃO DE DOCUMENTOS EM FORMATO DIGITAL

**PRESERVATION OF DOCUMENTS IN
DIGITAL FORMAT**

Miguel Ángel Márdero Arellano¹



¹ Dados do autor: Doutor em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, miguel@ibict.br.



A informação digital não sobreviverá acidentalmente.

BEAGRIE, 2006

Resumo: Examina-se a degradação dos objetos digitais e a demora na aplicação de estratégias para garantir sua preservação, assim como, a importância do emprego de técnicas focadas na manutenção das informações contidas nesses objetos. Neste capítulo são analisadas ações de preservação que em seu conjunto formam sistemas de informação integrados beneficiando a preparação e gestão dos conteúdos e coleções em cadeias de transferências entre repositórios e locais de arquivamento confiáveis. O papel da pesquisa nas áreas da Ciência da Informação é mencionado na especificação dos atributos dos registros e dados digitais que dão suporte à presunção de autenticidade. São apresentadas algumas respostas sistêmicas para a padronização da prática de preservação digital que priorizam o uso da infraestrutura governamental e a aplicação de um financiamento contínuo. A abordagem para os sistemas de preservação integrados e distribuídos em rede é recomendado para o tratamento dos objetos digitais, especificamente a sua vulnerabilidade física e lógica, parcerias e contratos e, no recrutamento de pessoal devidamente qualificado.

Palavras-Chave: Preservação digital. Objeto digital. Preservação digital distribuída. Autenticidade.

Abstract: It examines the degradation of digital objects and the delay in applying strategies to ensure their preservation, as well as the importance of employing techniques focused on maintaining the information contained in these objects. This chapter analyzes preservation actions that together form integrated information systems benefiting the preparation and management of contents and collections in transfer chains between repositories and reliable archiving locations. The role of research in the areas of Information Science is mentioned in the specification of the attributes of digital records and data that support the presumption of authenticity. Some systemic responses to the standardization of digital preservation practice are presented that prioritize the use of government infrastructure and the application of continuous funding. The approach to integrated and distributed network preservation systems is recommended for the treatment of digital objects, specifically their physical and logical vulnerability, partnerships and contracts, and in the recruitment of suitably qualified personnel.

Keywords: Digital preservation. Digital object. Distributed digital preservation. Authenticity.

◆ 1 INTRODUÇÃO

Todos os grupos humanos adotam tecnologias para prolongar a prática de crenças e tradições, e objetos produzidos com a tecnologia adotada no momento podem, também, causar mudanças nessas práticas. O governo, a administração pública e privada, a pesquisa científica e tecnológica e a expressão cultural dependem cada vez mais de documentos digitais, não disponíveis em outro formato, para o exercício de suas atividades. A digitalização em massa vem cobrando pesados investimentos em tecnologia da informação, trazendo à tona uma crescente debilidade estrutural dos sistemas eletrônicos de informação, que os incapacita de assegurar a preservação em longo prazo e o acesso contínuo às informações geradas em um contexto de rápido avanço tecnológico. Começam a surgir artefatos digitais que permitirão a identificação das características dos objetos digitais, produtos da necessidade de preservar o significado e a usabilidade da informação codificada neles.

O tema da preservação dos documentos digitais está cada vez mais presente entre os criadores e gerenciadores de coleções digitais. Para aqueles que começam a enfrentar a possibilidade de não poder recuperar informações produzidas há 5 ou 10 anos, é vital contar com estratégias que garantam a persistência desses dados. Para bibliotecários, arquivistas, museólogos e o pessoal de tecnologia da informação, a preservação digital é percebida como um conjunto de atividades desenvolvidas para conservar e prolongar a vida da informação digital, protegendo-a da obsolescência tecnológica, da perda e das falhas de equipamentos. Segundo a Associação Americana de Bibliotecários, “a preservação digital combina políticas, estratégias e ações que garantem o acesso ao conteúdo digital através do tempo” (ALA, 2007). As técnicas de preservação que eram conhecidas pelas bibliotecas e pelos centros de informação se modificaram e encontraram novas práticas em um contexto de rede de computadores, no qual a informação não fica apenas em um só lugar. A preservação da informação em formato digital precisa de um conjunto de práticas técnicas e de gerenciamento que mudam constantemente.

Desde as primeiras definições da preservação digital, a parte de planejamento, alocação de recursos e aplicação de métodos e tecnologias sempre estiveram presentes (HEDSTROM, 1997). Já neste século, seu universo instrumental e o papel dos mecanismos que permitem o armazenamento em repositórios digitais começou a receber mais atenção/ estar em foco (CASTRO, 2018). A preservação dos documentos continua a ser determinada pela capacidade de o objeto informacional servir às utilizações que lhe são imputadas e às suas atribuições, as quais garantem que ele continue a ser satisfatório para utilizações posteriores. Na era da informação digital, a ênfase está na geração e/ou aquisição de material digital, deixando em segundo plano a preservação e o acesso, a longo prazo, aos acervos eletrônicos existentes, quer dizer, a aplicação de estratégias de preservação digital deveria estar presente, pois sem elas não existiria nenhuma garantia de acesso, de confiabilidade e de integridade dos documentos a longo prazo (MÁRDERO ARELLANO; LEITE, 2009).

O domínio da preservação digital se estende para incluir o ciclo de vida completo dos objetos digitais, todos os processos relacionados à ingestão, armazenamento, gerenciamento e fornecimento de acesso ao objeto em questão. Os registros digitais são documentos compostos que consistem em objetos com dependência de intermediários tecnológicos necessários para torná-los visíveis para humanos. São mais facilmente alterados e corrompidos do que materiais analógicos, e as alterações, quando ocorrem, são mais difíceis de detectar. Muitas vezes, dada a complexidade das partes inter-relacionadas que compõem um registro digital, criadores e preservadores acham difícil estabelecer

a identidade e a integridade de um registro, duas características pelas quais a autenticidade é determinada (AUTHENTICITY TASK FORCE, 2002). Esta dificuldade em estabelecer sua autenticidade afeta negativamente o valor dos registros, pois como Ross (2002, p. 7) afirma: “objetos digitais que carecem de autenticidade têm valor limitado como evidência ou como um recurso de informação.”

Para que o objetivo da preservação digital possa ser atingido, os objetos digitais devem ser compreendidos e gerenciados em vários níveis: como um objeto físico, como uma codificação lógica, como objetos conceituais ou possuidores de significado para os humanos e como um conjunto de elementos essenciais, que devem ser preservados para oferecer aos futuros usuários a essência do objeto (UNESCO, 2003).

♦ 2 O OBJETO DA PRESERVAÇÃO

Os objetos digitais seriam aqueles que foram criados em computador, podendo ser originais ou versões depois de terem sido convertidas ou digitalizadas. Pela sua natureza multifacetada, a produção e disseminação dos objetos digitais têm sido ampla e geral, assim como as informações do seu contexto registrado e preservado para que, dessa maneira, futuros usuários possam entender o ambiente tecnológico no qual ele foi criado. Para Thibodeau (2002), objeto digital é o objeto de informação, de qualquer tipo de informação ou de qualquer formato que está expresso em forma digital. Esse raciocínio foi corroborado por Ludwing (2010), que apontou para os três diferentes aspectos visados pela preservação digital: a preservação de *bitstream*, que lida com o objeto físico e garante a estabilidade da cadeia bits, conquanto o objeto físico possa mudar; a preservação de conteúdo, que tenta manter o conteúdo estável, conquanto o formato do objeto lógico possa mudar; e a curadoria de dados, que permite complementações e mudanças no objeto conceitual com o propósito de assegurar a usabilidade a longo prazo.

Entre as ações gerenciais e as técnicas exigidas dentro do planejamento da preservação digital, a alocação de recursos e a aplicação de métodos e tecnologias de preservação, combinadas com políticas e estratégias para garantir o acesso a conteúdo reformatado e “nativo digital” (*born-digital*), levam sempre em consideração a obsolescência tecnológica dos objetos digitais, a fragilidade dos suportes e a sua interpretação a longo prazo (MUHANNA, 2018; CONARQ, 2020). A preservação digital não é apenas um problema técnico, e, sim, “um componente de uma ampla agregação de serviços, políticas e partes interessadas interconectados que, juntos, constituem um ambiente digital” (LAVOIE; DEMPSEY, 2004). As condições básicas à preservação digital seriam, então, a adoção desses métodos e dessas tecnologias que integrariam a preservação física, lógica e intelectual dos objetos digitais.

A preservação física dos objetos digitais está centrada no conteúdo armazenado em mídia magnética e em discos óticos. Assim como ocorre com o suporte físico da informação, em papel, a superfície metálica magnetizada se desintegra ou pode se tornar irrecuperável. Existem, ademais, os efeitos da temperatura, da umidade, do nível de poluição do ar e das ameaças biológicas; os danos provocados pelo uso indevido e regular, as catástrofes naturais e a obsolescência tecnológica. Para Duranti (2010), diferentemente dos documentos tradicionais, os digitais podem ser armazenados em diferentes partes do suporte e eles não existem como objetos físicos até o momento em que são gerados intencionalmente, pois os elementos do documento digital podem se manifestar de múltiplas formas. Uma das dificuldades da aplicação prática das técnicas de preservação digital é a capacidade que possuem as ferramentas tecnológicas de tratar os diferentes objetos digitais adicionais de um único objeto digital.

O dado e a mídia que suportam a informação devem possuir um nível de funcionalidade representacional que permita a sua reprodução a qualquer momento que a instituição mantenedora precisar recuperar o dado. Cada mídia pode armazenar uma sequência de bits de forma diferente, segundo as propriedades físicas da mídia. O *bitstream* precisa, então, ser interpretado, uma vez que toda sequência significativa de bits pode representar qualquer coisa. Os objetos digitais são salvos como coleções de bits, representando documentos específicos, significativos apenas para o programa que os criou. A preservação digital não passa tanto pela preservação do suporte físico (como acontece com a preservação tradicional), mas, sim, por garantir que a informação nele contida permaneça acessível e autêntica ao longo do tempo (FERREIRA, 2011).

Já na preservação da natureza lógica dos objetos digitais, são procurados/buscados, na tecnologia, formatos atualizados para inserção dos dados (correio eletrônico, material de áudio e audiovisual, material em rede etc.), novos *software* e hardware que mantenham vigentes seus bits para conservar sua capacidade de leitura. Na publicação digital, a preservação lógica está associada à necessidade de garantir a conversão dos formatos originais que passaram a ser obsoletos ou de custosa manutenção.

No caso da preservação intelectual, o foco são os mecanismos que garantem a integridade e a autenticidade da informação nos documentos eletrônicos. A preservação do objeto físico e lógico não garante a obtenção do objeto conceitual. É necessário preservar uma forma de traduzir o objeto lógico para o conceitual (YAMAOKA; GAUTHIER, 2013, p. 85,). A preservação da informação descritiva dos materiais digitais é de crucial importância, devido, principalmente, à capacidade de o objeto digital ser passível de modificação no seu desenho (leiaute), apresentação ou interação no formato de publicação. Com isso, a perda do conteúdo intelectual original pode ser declarada inaceitável pelo autor.

Em “Teoria dos Objetos Digitais”, Kallinikos, Aaltonen e Marton (2010) argumentam que estes artefatos possuem algumas características próprias, dentre as quais as de serem editáveis, por permitirem modificações contínuas e sistemáticas; interativos, com funções incorporadas para oferecer caminhos aos agentes; abertos, no sentido de serem acessíveis e modificáveis por um programa e distribuídos, pela possibilidade de disponibilização, em diversas fontes e instituições.

Um fator relevante da natureza intelectual dos objetos digitais é que ela envolve a preservação da propriedade intelectual (IPR) – *Intellectual Property Rights*, que tem um significado mais argúvel do que na mídia tradicional. O IPR deve considerar não apenas o conteúdo, mas também qualquer ação relacionada ao *software* (cópias, encapsulamento de conteúdo, emulação de *software*, migração de conteúdo) que envolva atividades que podem infringir permissões específicas daqueles que mantêm os direitos. Para Bullock (1999), o IPR é uma das principais barreiras que interferem na preservação dos objetos digitais e sua adoção pode interferir nas informações contidas, como funcionalidade, aparência e conteúdo. Por esse motivo, Lavoie e Gartner (2005) recomendam documentar os direitos associados aos objetos arquivados, para que os processos de preservação estejam coordenados com as restrições impostas aos objetos.

◆ 3 A PESQUISA DA PRESERVAÇÃO DIGITAL

Desde suas primeiras conceptualizações, a preservação digital levou em consideração, além da interpretação dos dados, todas as ações que permitiram sua produção e acesso. A pesquisa da preservação digital investiga a natureza de objetos digitais, incluindo registros e dados e os atributos que podem apoiar a sua confiabilidade, a integridade e a presunção de sua autenticidade (GRÁCIO, 2011). Embora muita pesquisa tenha sido e continua a ser conduzida na proteção da autenticidade, no contexto de requisitos para preservação digital, os meios atuais de avaliação da autenticidade para os profissionais da informação ainda não oferecem medidas quantificáveis, e modelos generalizáveis que podem reduzir o problema. Para Moore (2008), as ações de preservação são uma forma de comunicação com o futuro e o grande desafio é como incorporar novas tecnologias de forma eficaz, ao mesmo tempo em que se resguardem as propriedades de preservação, como autenticidade, integridade e cadeia de custódia.

A pesquisa da preservação tem tido foco na natureza dos objetos digitais, no desenvolvimento de padrões, estruturas e sistemas de repositório, em definir e usar/compartilhar esquemas de metadados, nas tecnologias de preservação e nos formatos de arquivo. Todos eles compartilham um objetivo comum: o de preservar objetos digitais confiáveis (ROGERS, 2015). Novas questões surgem a partir desse objetivo, colocando a preservação digital não apenas como um processo isolado, mas também como componente de um conjunto agregado de serviços, políticas e especialistas que constituem o contexto do ciclo de vida da informação digital.

A preservação digital inicia-se no agrupamento da perspectiva que se tinha dos requisitos associados com as atividades tradicionais nas áreas de conservação e de preservação dos objetos informacionais. Os objetos digitais não podem ser deixados em formatos obsoletos para serem transferidos, depois de longos períodos de negligência, para repositórios digitais. O desafio é muito mais um problema social e institucional do que técnico, porque, principalmente para a preservação digital, depende-se de instituições que passam por mudanças de direção, missão, administração e fontes de financiamento (TERMENS, 2009; SASTRE, 2015).

Ao analisar o ciclo da informação (geração, tratamento, preservação, conservação e, finalmente, difusão da informação), percebe-se que a preservação e a conservação necessitam ser revistas no âmbito da informação digital. Até pouco tempo, entendia-se a preservação no sentido de conservar e prevenir os documentos do risco de deterioração. É importante analisar a preservação a partir da necessidade de assegurar o acesso e recuperação da informação como fundamentos para a pesquisa. Aliado a isso, deve-se considerar o crescente uso dos computadores e a quantidade de informações disponíveis em meio eletrônico (algumas, inclusive, somente neste meio).

Muitos materiais publicados digitalmente são produto de serviços de informação disponibilizados por organizações que adotam alguma infraestrutura tecnológica. Essas instituições levam em consideração aspectos legais e culturais que afetam a oferta desses serviços orientados a atender as necessidades de determinados usuários. Mas, muitas vezes, essa preocupação não é a mesma que têm os produtores das novas tecnologias. É fato que, na preservação de documentos digitais, assim como na de documentos em papel, é necessária a adoção de ferramentas que protejam e garantam a sua manutenção. Essas ferramentas deverão servir para reparar e restaurar registros protegidos, prevendo os danos e reduzindo os riscos dos efeitos naturais (preservação prospectiva), ou para restaurar os documentos já danificados (preservação retrospectiva).

Na área da Ciência da Informação, o uso da tecnologia digital, que toma o lugar dos tradicionais meios de conservação e preservação, trouxe consigo a preocupação com as normas para o uso das técnicas digitais e sua prontidão na tarefa da preservação a longo prazo (CHEPESUIK, 1997). Os especialistas da área que trabalham com informação em formatos digitais elaboram normas necessárias para armazenar e compartilhar, de maneira adequada, esses materiais, assim como buscam a formulação de políticas institucionais de preservação.

Os primeiros estudos desenvolvidos na área da preservação digital tiveram origem no seio da Arquivologia. Entretanto, considerando as diferenças existentes no que tange à preservação de acervos bibliográficos e à preservação de documentos arquivísticos ao longo do tempo, a Biblioteconomia e a Ciência da Informação passaram a dedicar-se, também, a este campo. Isto pelo fato de as bibliotecas, os arquivos e os museus serem responsáveis por manter coleções digitais para uso permanente, protegendo-as de ameaças, ou salvando-as e reparando-as para compensar seus impactos. Os sistemas de bibliotecas e de arquivos atuais envolvem um grande esforço de gerenciamento de objetos digitais, que vai além das tarefas tradicionais dessas instituições patrimoniais (aquisição, seleção, classificação, arquivamento etc.), na construção da interoperabilidade de acervos digitais (arquiteturas, metadados, formatos padrão), o que é possível por meio de sistemas relacionados e desenvolvidos para propósitos e comunidades específicas.

As bibliotecas digitais foram as primeiras a serem consideradas o caminho mais adequado para a preservação dos recursos de informação (HILDRETH, 1996). Posteriormente, os repositórios arquivísticos foram considerados, dentro das instituições, locais confiáveis capazes de armazenar, migrar e dar acesso a coleções digitais. O treinamento de especialistas e técnicos também começou a ser contemplado nos custos das atividades desses repositórios, produzindo melhora na aplicação efetiva das estratégias (CHAPMAN, 2004).

Segundo Becker *et al.* (2009), existe um constante crescimento na oferta de ferramentas, teorias e modelos para a preservação digital. Entretanto, apresenta-se uma carência de estudos que auxiliem os gestores na tomada de decisão a respeito da seleção das ferramentas e sobre como implementar os modelos propostos de forma concreta. A responsabilidade de adotar medidas preventivas e corretivas, tendo como objetivo minimizar a ação do tempo sobre o suporte físico da informação e assegurar sua disponibilidade, foi atribuída às instituições arquivísticas, as quais teriam a capacidade de compreender os limites e os significados dos documentos (autenticidade, capacidade probatória, integridade das informações, contexto de produção, manutenção etc.), dando ênfase às tarefas que essas organizações devem observar para lidar com a guarda permanente dos objetos digitais autênticos.

A gestão dos documentos eletrônicos constitui uma área de pesquisa que tem produzido contribuições relevantes relacionadas com a preservação de documentos digitais de “valor permanente” para instituições públicas. Como componentes do documento digital arquivístico, os metadados requerem os mesmos procedimentos de preservação, visto que o contextualizam e colaboram para a manutenção de suas características arquivísticas (SANTOS, 2012).

◆ 4 AS PRÁTICAS DA PRESERVAÇÃO DIGITAL

A preservação digital é a parte mais longa e também a última do ciclo de gerenciamento de objetos digitais. Ela permite o emprego de mecanismos que viabilizam o armazenamento em repositórios de objetos digitais e que garantem a autenticidade e a perenidade dos seus conteúdos. São necessários, além de procedimentos de manutenção e recuperação de dados (no caso de perdas acidentais para resguardar a mídia e seu conteúdo), também estratégias e procedimentos para manter acessibilidade e autenticidade através do tempo, podendo requerer colaboração entre diferentes financiadoras e boa prática de licenciamento, de metadados e de documentação, antes de aplicar ações técnicas.

Nas últimas três décadas, as práticas relacionadas com a preservação digital estão baseadas na ideia de garantir a longevidade dos arquivos, e elas dependem, principalmente, da solução tecnológica adotada e dos custos que as envolvem. Mas, ainda hoje, ainda é evidente a ausência de conhecimento sobre as estratégias de preservação digital e o que isso poderá significar diante da necessidade de garantir a longevidade dos arquivos digitais. Os atuais detentores de acervos digitais percebem a imperiosa necessidade de contar com mecanismos que garantam a preservação de seus documentos em formato digital. Especificamente, essa preocupação parte das comunidades responsáveis pelas bibliotecas e pelos arquivos, nas quais o desenvolvimento de padrões e de mecanismo legais para lidar com arquivos eletrônicos precisa de estratégias metodológicas bem definidas.

As iniciativas se multiplicam e novas soluções são testadas por instituições detentoras de acervos de especial relevância para o desenvolvimento científico e tecnológico de seus países de origem. Muitos projetos e iniciativas têm conseguido estabelecer os benefícios de algumas metodologias e estratégias. A conclusão a que essas iniciativas chegam é que devem ser usados padrões e converterem-se os documentos nos formatos livres, para que eles sejam acessados após a obsolescência dos equipamentos e dos programas informáticos em que foram criados.

Por causa da natureza interdisciplinar, do escopo abrangente e do custo da preservação digital, os projetos são frequentemente realizados por alianças ou consórcios nacionais e internacionais de universidades, bibliotecas e arquivos, agências governamentais, empresas e indústrias. Cada um deles é definido por sua perspectiva epistêmica e propósitos particulares. No entanto, cooperação e colaboração são constantes em toda a comunidade da área. Pesquisadores em várias partes do mundo estão desenvolvendo modelos do que seria a infraestrutura para a preservação, a longo prazo, de informação em formato digital e muitas iniciativas propostas pelos grandes centros de pesquisa continuam sendo utópicas, porém, proporcionam uma base para a discussão sobre a implantação de soluções futuras em relação ao problema. Entre elas, está o uso de padrões abertos que permite seu estudo e sua conversão para novos padrões.

Os principais métodos recomendados para a preservação dos objetos digitais podem ser agrupados em dois tipos: estruturais e operacionais. Os estruturais tratam dos investimentos iniciais por parte das instituições que estão se preparando para implementar algum processo de preservação e que adotam ou adaptam um dos modelos de metadados existentes ou seu próprio esquema. As atividades operacionais são as medidas concretas aplicadas aos objetos digitais, aquelas que englobariam os novos requisitos de preservação, migração de suporte, atualização da mídia (preservação física), conversão dos formatos, emulação (preservação lógica) a preservação do conteúdo (intelectual).

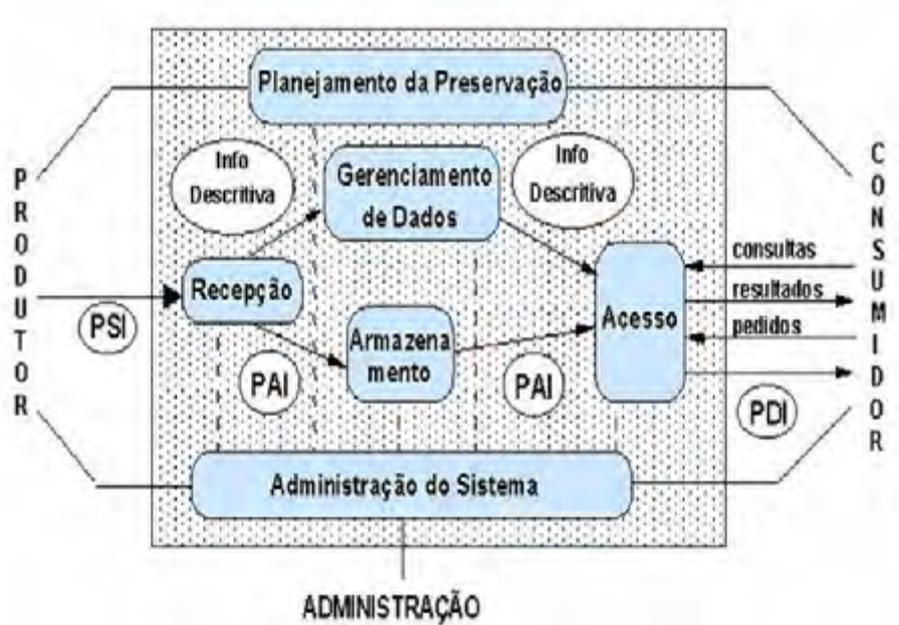
As tarefas da preservação digital são permanentes e devem ser pensadas a longo prazo desde a criação do próprio documento. Por este motivo, é importante tomar como referência padrões internacionais e nacionais existentes, tais como: o modelo OAIS, o sistema de metadados PREMIS e o esquema METS. As coleções que servem de informações estruturadas e padronizadas em metadados tornam-se mais vivíveis em redes de alcance global, o que permite que seus recursos informacionais sejam identificados, localizados, selecionados por critérios de relevância e distinguidos por diferenças e similaridades.

Desde a década de 1990, a importância dos metadados foi reconhecida (HURLEY, 1995) como sendo a chave para a criação de registros digitais confiáveis, parte importante em todas as estratégias estruturais e operacionais de preservação digital, uma vez que elas estão baseadas na conservação de *software* e *hardware*, emulação ou migração, como um meio de garantir a autenticidade, registrar o gerenciamento de direitos e coleções de dados e de interagir com recursos de busca (GILLILAND; MCKEMMISH, 2012). Metadados são afirmações legíveis por máquinas e por humanos a respeito de recursos de informação que permitem o controle físico, intelectual e técnico sobre esses recursos. Os responsáveis pelos acervos digitais criam e anexam e, em seguida, mantêm e preservam metadados, automática e/ou manualmente, por serem elementos ou estruturas de organização da informação que, associados a um objeto digital, classificam, categorizam e descrevem essa informação.

Uma forma especializada de administrar metadados que podem ser usados como um meio de estocar a informação técnica (e que suporta a preservação dos objetos digitais) é utilizar metadados de preservação. Eles visam a apoiar e a facilitar a retenção, a longo prazo, da informação digital (FORMENTON, *et al.*, 2017), codificando e rastreando a identidade e a integridade do material ao longo do tempo por meio de mudanças tecnológicas. Para definir estruturas padronizadas de metadados e determinar a sua utilização nos sistemas eletrônicos de gestão de informação, com o propósito de gerir a preservação e a acessibilidade dos documentos digitais, foi necessária a adoção de um modelo de alto nível para sistemas de preservação digital, o Open Archival Information System (OAIS). Ele é referência em sistemas de preservação digital, abordando todos os aspectos da preservação de longo prazo da informação digital: ingestão, armazenamento em arquivo, gerenciamento de dados, acesso, disseminação e migração para novas mídias e formas.

Desenvolvido em 2002 pelo Comitê Consultivo para Sistemas de Dados Espaciais, o OAIS é hoje uma norma ISO aprovada (ISO 14721: 2003) e passou por várias revisões, a mais recente em 2012 (CCSDS 2012). Um arquivo nos termos do OAIS significa considerar os arquivos como organizações de pessoas e sistemas, que aceitaram a responsabilidade de preservar a informação e torná-la disponível, além de melhorar a comunicação e a produtividade entre diferentes comunidades. No modelo, atuam quatro entidades: produtores, consumidores, administração e o arquivo propriamente dito. No esquema do OAIS, está incluído um modelo de informação para inserção dos metadados de preservação (Figura 1), a seguir.

Figura 1 - Modelo funcional OAIS



Fonte: ABNT (2007).

As entidades funcionais gerenciam o fluxo de informação entre as entidades que fazem parte do ambiente OAIS e identificam os componentes funcionais dos arquivos que estão relacionados com a preservação dos objetos digitais. Cada entidade é responsável por algum processo do sistema no ambiente OAIS, ou seja, o produtor pode ser considerado a pessoa ou sistema cliente que fornece a informação; o consumidor é a pessoa ou sistema cliente que quer obter o material arquivado ou a versão transformada do mesmo; a comunidade-alvo refere-se aos usuários potenciais (MÁRDERO ARELLANO, 2008).

Porém, algumas decisões independem do Modelo OAIS. Como é o caso do tipo de material a ser armazenado, sobre o local do armazenamento das informações descritivas, de relacionamento entre objetos e se outros componentes do gerenciamento devem ou não ser incluídos. O modelo apresenta desafios para as instituições, assim como pode ser considerado uma estratégia de preservação de longo prazo da informação digital. Ele provê conceitos importantes e a adequação necessária dos componentes que podem chegar a ser desenvolvidos no futuro. Em uma adaptação do OAIS especificamente direcionado para documentos arquivísticos digitais, as funções do programa de preservação de documentos arquivísticos digitais devem seguir o modelo de Cadeia de Preservação do InterPARES 2 (2008).

O Dicionário PREMIS (LIBRARY OF CONGRESS, 2008) é usado para definir o conjunto básico de unidades semânticas que os repositórios digitais devem conhecer ao realizar funções de preservação. Para ser considerado um repositório de preservação, é preciso ser ou estar em conformidade com a “responsabilidade exclusiva ou como parte de múltiplas responsabilidades que realiza a preservação de longo prazo de objetos digitais sob sua custódia”. As tarefas de preservação podem variar de um repositório para outro, mas elas devem garantir a viabilidade e a sustentabilidade dos objetos digitais. Além disso, devem assegurar, também, que objetos digitais no repositório não sejam alterados sem nenhuma documentação prévia. Os principais usos do PREMIS são direcionados ao *design* do repositório e à sua avaliação e troca de pacotes de informação entre repositórios. Também, muitos dos metadados PREMIS são projetados para serem fornecidos por um aplicativo de *software*.

Já o formato padronizado para transmissão de metadados, projetado para envio de arquivos, imagens e objetos multimídia em um repositório digital que utiliza a estrutura XML é o Metadata Encoding & Transmission Standard METS (LIBRARY OF CONGRESS, 2007). O padrão é mantido no Gabinete de Desenvolvimento de Redes e Padrões MARC da Biblioteca do Congresso, e consiste na codificação de metadados necessários para o gerenciamento de objetos digitais e na troca entre repositórios digitais.

Em 2014, no âmbito do projeto SCAPE, da União Europeia, ficou amplamente documentada a tendência de as instituições limitarem-se em elaborar políticas de preservação digital, não avançando para as políticas de procedimento (SIERMAN *et al.*, 2014). A expectativa é que as instituições possam definir políticas de gerenciamento que garantam a confiabilidade; definam regras; apliquem as políticas; estabeleçam capacidades que implementem as funções de preservação necessárias; e definam metadados de preservação, para, assim, capturar informações reais sobre a aplicação das funções de preservação. Dessa forma, os metadados de preservação poderão ser consultados para analisar se os critérios de avaliação foram satisfeitos, uma vez que muitos repositórios fazem uso indevido do modelo OAIS, tendo em vista que, nem sempre, eles foram auditados e certificados para verificação de tal conformidade (LAVOIE, 2014).

A comunidade de preservação vem desenvolvendo critérios de avaliação para validar a confiabilidade dos repositórios digitais. Um repositório capaz de autenticar os critérios de avaliação pode ser considerado confiável e, portanto, seria um ambiente razoável para a preservação de dados em longo prazo. Desde 2007, quando um conjunto inicial de critérios de avaliação foi proposto pelo Research Libraries Group and the National Archives and Records Administration (RLG; NARA, 2007), gestores de sistemas de preservação reconhecem que um ambiente de preservação deve ser quantificado em termos de operações de preservação reais (que são realizadas) e de políticas de gestão (que controlam a execução dos processos de preservação).

A importância da automação de todo o processo de preservação digital é resultado da crescente produção de informações a serem preservadas pelas instituições em seus repositórios digitais. Para Térmens e Ribera (2009), os sistemas de preservação digital só podem funcionar com eficiência (ou seja, com custos razoáveis e confiabilidade fidedigna), se puderem normalizar completamente seus processos e automatizar os procedimentos de ingestão. Dessa maneira, os repositórios digitais poderão integrar ferramentas de preservação digital capazes de realizar as tarefas relacionadas à identificação, validação, caracterização, arquivamento e acesso de longo prazo a documentos em formato digital.

◆ 5 CONCLUSÃO

Devido à rápida degradação física, a obsolescência tecnológica, complexidade e os custos envolvidos na preservação, os objetos digitais estão em uma posição de alto risco, tendo em vista a fragilidade e demora na aplicação de estratégias para garantir preservação em um longo período de tempo. A vantagem está no fato de, no meio digital, a destruição do suporte não representar perda de informação, desde que tenham sido empregadas técnicas de preservação neste objeto digital. A preservação digital não é sobre um sistema particular, ou uma série de ações de preservação. Trata-se de preparar conteúdo e coleções em uma grande cadeia de transferências (OWENS, 2018).

Os gestores de repositórios digitais ainda possuem pouca ou nenhuma experiência em relação às práticas de preservação digital. Enquanto isso, a pesquisa continua a investigar a natureza dos objetos digitais, incluindo registros, dados, e os atributos que podem apoiar a presunção da autenticidade. Também continua a procura por respostas sistêmicas que possam ser padronizadas na prática, como infraestrutura governamental e financeira. É imperiosa a necessidade de preservar sistemas complexos de forma integrada e distribuída em rede.

◆ REFERÊNCIAS

ABNT. Associação Brasileira de Normas Técnicas. **NBR 15472**: modelo de referência para um sistema aberto de arquivamento de informação. Rio de Janeiro: ABNT, 2007.

AMERICAN LIBRARY ASSOCIATION. **Definitions of digital preservation**. Chicago: American Library Association. 2007. Disponível em: <http://www.ala.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.pdf>. Acesso em: 1 ago. 2021.

AUTHENTICITY TASK FORCE. Appendix 2 – requirements for assessing and maintaining the authenticity of electronic records. In: L. DURANTI, L. (ed.). **The long-term preservation of authentic electronic records**: Findings of the InterPARES project. Vancouver, BC: InterPARES, 2002. Disponível em: http://www.interpares.org/display_file.cfm?doc=ipl_authenticity_requirements.pdf Acesso em: 1 ago. 2021.

BEAGRIE, N. Digital curation for science, digital libraries, and individuals. **International Journal of Digital Curation**, [s.l.], v. 1, n. 1, 2006. Disponível em: <http://www.ijdc.net/article/view/6> Acesso em: 1 ago. 2021.

BECKER, C. *et al.* Systematic planning for digital preservation: evaluating potential strategies and building preservation plans. **International Journal on Digital Libraries**, [s.l.], v. 10, n. 4, p. 133-157, 2009. Disponível em: <https://doi-org.ez74.periodicos.capes.gov.br/10.1007/s00799-009-0057-1>. Acesso em: 28 jul. 2021.

BULLOCK, A. **Preservation of digital information**: issues and current status. Ottawa: National Library of Canada, 1999. Disponível em: <http://www.referenciasarquivisticas.fci.unb.br:8080/jspui/handle/123456789/3537>. Acesso em: 28 jul. 2005.

CASTRO, R. M. Objetos digitais: da maleabilidade do não-finito à uma ontologia. **Pragmatizes**: Revista Latino-Americana de Estudos em Cultura, [s.l.], n. 16, p. 31-43, 3 jun. 2019.

CHAPMAN, S. Counting the costs of digital preservation: is repository storage affordable? **Journal of Digital Information**, [s.l.], v. 4, n. 2, 2004. Disponível em: <https://journals.tdl.org/jodi/index.php/jodi/article/view/jodi-113>. Acesso em: 30 jul. 2021.

CHEPESUIK, R. The future is here: America's libraries go digital. **American Libraries**, v. 2, n. 1, p. 47-49, 1997.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Câmara Técnica de Documentos Eletrônicos (CTDE). **Glossário Documentos Arquivísticos Digitais**. 2020. 8ª versão. Disponível em: http://antigo.conarq.gov.br/images/ctde/Glossario/glosctde_2020_08_07.pdf Acesso em: 20 mai. 2021.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS). **Reference Model for an Open Archival Information System (OAIS): recommended Practice**, issue 2. Washington, D.C.: CCSDS/NASA, 2012. 135 p. Disponível em: <http://www.ccsds.org>. Acesso em: 20 jul. 2021.

DURANTI, L. Concept and principles for the management of electronic records, or records management theory is archival diplomatics. **Record Management Journal**, [s.l.], v. 20, n. 1, p. 78-95, 2010.

FERREIRA, C. A. S. **Preservação da informação digital**: uma perspectiva orientada para as bibliotecas. 2011. 155f. Dissertação (Mestrado em Informação, Comunicação e Novos Media) – Universidade de Coimbra, Coimbra, 2011.

FORMENTON, D. *et al.* Os padrões de metadados como recursos tecnológicos para a garantia da preservação digital. **Biblios**, Pittsburgh, n. 68, p. 82-95, 2017. Disponível em: <http://www.scielo.org.pe/pdf/biblios/n68/a06n68.pdf> Acesso em: 20 jul. 2020.

GILLILAND, A.; MCKEMMISH, S. Recordkeeping Metadata, the Archival Multivers, and Societal Grand Challenges. *In*: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2012, p. 106-113. **Proceedings [...]** Kuching, Sarawak, Malaysia. Disponível em: <http://dcevents.dublincore.org/IntConf/dc-2012/paper/view/108/66>. Acesso em: 25 jul. 2021.

GRÁCIO, J. C. A. **Preservação digital na gestão da informação**: um modelo processual para as instituições de ensino superior. 2011. 223 f. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Marília, 2011.

HEDSTROM, M. Digital preservation: a time bomb for digital libraries. **Computers and the humanities**, v. 31, n. 3, p. 189-202, 1997. Disponível em: https://www.researchgate.net/publication/30843346_Digital_Preservation_A_Time_Bomb_for_Digital_Libraries. Acesso em: 11 jul. 2021.

HILDRETH, C. R. Preserving what we really want to access, the message, not the medium: challenges and opportunities in the digital age. *In*: HELAL, A. H.; WEISS, J. W. (ed.). INTERNATIONAL ESSEN SYMPOSIUM, 18., 1995, Essen. **Proceeding [...]** Essen: Germany University, 1996. p. 76-95.

HURLEY, C. Ambient Functions: Abandoned Children to Zoos. **Archivaria**, v. 40. 1995. Disponível em: <https://archivaria.ca/index.php/archivaria/article/view/12095>. Acesso em: 31 jul. 2021.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 14721:2012**: Space data and information transfer systems: Open archival information system (OAIS). Reference model. Genebra: ISO, 2012.

INTERPARES 2 PROJECT (The). **Experiential, Interactive and Dynamic Records**. 2008. Disponível em: <http://www.interpares.org/ip2/book.cfm> . Acesso em: 18 jun. 2021.

KALLINIKOS, J.; AALTONEN, A.; MARTON, A. A theory of digital objects. **First Monday**, [s.l.], v. 15, n. 6, 2010. Disponível em: <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3033/2564>

Acesso em: 1 ago. 2021.

LAVOIE, B. **The Open Archival Information System (OAIS) Reference Model**: introductory guide. 2. ed. York: DPC, 2014. Disponível em: <http://dx.doi.org/10.7207/twr14-02>. Acesso em: 1. Ago. 2021.

LAVOIE, B; DEMPSEY, L. Thirteen ways of looking at - digital preservation. **D-Lib Magazine**, v. 10, n. 7/8, jul./ago., 2004. Disponível em: <http://www.dlib.org/dlib/july04/lavoie/07lavoie.html> Acesso em: 28 jun. 2021.

LAVOIE, B.; GARTNER, R. **Preservation Metadata**. Dublin, Ohio: OCLC, September 2005. Disponível em: <https://www.dpconline.org/docs/technology-watch-reports/894-dpctw13-03/file>. Acesso em: 30 jul. 2021.

LIBRARY OF CONGRESS. **Metadata Encoding and Transmission Standard**. 2007. Disponível em: <http://www.loc.gov/standards/mets> Acesso em: 1 ago, 2021.

- LIBRARY OF CONGRESS. **Data dictionary for preservation metadata:** PREMIS. Version 2.2. (2012). Disponível em: <http://www.loc.gov/standards/premis/v2/premis-dd-2-2.pdf> Acesso em: 28 jul. 2021.
- LUDWIG, J. **About the complexity of a digital preservation theory and different types of complex digital objects.** Germany: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2010. Disponível em: <http://www.dagstuhl.de/Materials/Files/10/10291/10291.LudwigJens.ExtAbstract.pdf>. Acesso em: 22 mai. 2021.
- MÁRDERO ARELLANO, M. A. **Crítérios para a preservação digital da informação científica.** 2008. 354 f. Tese (Doutorado) – Universidade de Brasília, Departamento de Ciência da Informação, 2008. Disponível em: <http://eprints.rclis.org/12649/>. Acesso em: 28 jul. 2021.
- MÁRDERO ARELLANO, M. A.; LEITE, F. C. L. Acesso aberto à informação científica e o problema da preservação digital. **Biblios**, Brasília, n. 35, 2009. Disponível em: <http://repositorio.unb.br/handle/10482/4937> Acesso em: 1 ago. 2021.
- MOORE, R. Towards a Theory of Digital Preservation. **International Journal of Digital Curation**, [s.l.], v. 3, n. 1, 2008. Disponível em: <http://www.ijdc.net/article/view/63> Acesso em: 30 jul. 2021.
- MUHANNA, E. What Does “Born Digital” Mean? **International Journal of Middle East Studies**, [s.l.], v. 50, n. 1, Feb. 2018, p. 110–112.
- OWENS, T. **The theory and craft of digital preservation.** Baltimore, Maryland : Johns Hopkins University Press, 2018.
- RESEARCH LIBRARIES GROUP & NATIONAL ARCHIVES AND RECORDS ADMINISTRATION. **Trustworthy repositories audit & certification:** criteria and checklist. 2007. Disponível em: https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf Acesso em: 1 ago. 2021.
- ROGERS, C. **Virtual authenticity:** authenticity of digital records from theory to practice. Electronic Theses and Dissertations, Vancouver: University of British Columbia, 2015. Disponível em: <http://dx.doi.org/10.14288/1.0166169>. Acesso em: 28 jul. 2021.
- ROSS, S. Position paper on integrity and authenticity of digital cultural heritage objects. In: **DigiCULT: Integrity and Authenticity of Digital Cultural Heritage Objects**, August, 2002. p. 7–8. Disponível em: <https://www.digicult.info/downloads/html/7/7.html> . Acesso em: 24 jul. 2021.
- SANTOS, V. B. dos. Preservação de documentos arquivísticos digitais. **Ciência da Informação**, Brasília, v. 41, n. 1, 2014. Disponível em: <http://revista.ibict.br/ciinf/article/view/1357>. Acesso em 28 jul. 2021.
- SASTRE NATIVIDAD, G. **Preservación y conservación de documentos digitales** [en línea]. In: ArchivPost. Salamanca: Asociación de Archiveros de Castilla y León, 2015. Disponível em: <http://www.acal.es/index.php/archivpost-a-fondo>. Acesso em: 29 jul. 2021.
- SIERMAN, B. *et al.* **Catalogue of preservation policy elements.** 2014. Disponível em: http://scape-project.eu/wp-content/uploads/2014/02/SCAPE_D13.2_KB_V1.0.pdf Acesso em: 1 ago. 2021.
- TÉRMENS, M. Investigación y desarrollo en preservación digital: un balance internacional. **El Profesional de la Información**, v. 18, n. 6, p. 613–624, 2009. Disponível em: <http://www.elprofesionaldelainformacion.com/contenidos/2009/noviembre/04.html>. Acesso em: 1 ago. 2021.

TÉRMENS, M; RIBERA, M. El control de los formatos en la preservación digital. *In*: JORNADAS ESPAÑOLAS DE DOCUMENTACIÓN, 11, 2009, Zaragoza. **Anais [...]** Zaragoza: FESABID, 2009. p. 139-146.

THIBODEAU, K. Overview of technological approaches to digital preservation and challenges in coming years the state of digital preservation: an international perspective. **Proceedings [...]** Washington: CLIR and Library of Congress, 2002. Disponível em: <http://www.clir.org/pubs/reports/pub107/pub107.pdf#page=10>. Acesso em: 22 maio 2021.

UNESCO. **Directrices para la preservación del patrimonio digital**. 2003. Disponível em: <http://unesdoc.unesco.org/images/0013/001300/130071s.pdf>. Acesso em: 1 ago. 2021.

YAMAOKA, E. J.; GAUTHIER, F. O. Objetos digitais: em busca da precisão conceitual. **Informação & Informação**, v. 18, n. 2, p. 77-97, 2013. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/viewFile/16162/13077>. Acesso em: 26 jul. 2021.

COMO CITAR:

MÁRDERO ARELLANO, Miguel Ángel. Preservação de documentos em formato digital. *In*: BRAGA, Tiago Emmanuel Nunes; MÁRDERO ARELLANO, Miguel Ángel. (org.). **Hipátia**: modelo de preservação para repositórios arquivísticos digitais confiáveis. Brasília: Ibict, 2022. p. 10-25. (Informação, Tecnologia e Inovação, v. 1). DOI 10.22477/9786589167501.cap1 .

Capítulo 02

O IBICT E A PRESERVAÇÃO DA MEMÓRIA CIENTÍFICA

IBICT AND THE PRESERVATION OF SCIENTIFIC MEMORY

Maria de Fátima Duarte Tavares¹
Miguel Ángel Márdero Arellano²

¹ Dados da autora: Doutora em História Social, Instituto Brasileiro de Informação em Ciência e Tecnologia, fatimatavares@ibict.br

² Dados do autor: Doutor em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, miguel@ibict.br

RESUMO

Resumo: A ampliação do acesso à comunicação científica tem sido garantida por repositórios digitais, estruturados pelas instituições produtoras de conhecimento para dar maior visibilidade à sua própria produção intelectual. No entanto, menor atenção tem sido dada à questão da memória científica, considerando-se a dimensão da longa duração e o impacto da obsolescência tecnológica sobre os objetos digitais não preservados. A partir da instauração da Rede Cariniana, coordenada pelo Ibict, a comunidade científica no Brasil dispõe de uma infraestrutura de suporte nacional e um canal de interlocução dedicado à promoção da salvaguarda do patrimônio científico digital. A preservação dos acervos científicos impõe um olhar sobre as práticas do tempo presente, que dependem da gestão documental e de políticas institucionais sobre uma memória cada vez mais imersa nos fluxos de informação digitais.

Palavras-Chave: Acervos científicos. Longo prazo. Preservação digital. Rede Cariniana

Abstract: The expansion of access to scientific communication has been guaranteed by digital repositories, structured by knowledge-producing institutions to give greater visibility to their own intellectual production. However, less attention has been given to the issue of scientific memory, considering the dimension of long duration and the impact of technological obsolescence on unpreserved digital objects. Since the establishment of the Cariniana Network, coordinated by Ibict, the scientific community in Brazil has a national support infrastructure and a dialogue channel dedicated to promoting the safeguarding of the digital scientific heritage. The preservation of scientific collections imposes a look at the practices of the present time, which depend on document management and institutional policies on a memory increasingly immersed in digital information flows.

Keywords: Scientific collections. Long term. Digital preservation. Cariniana Network.

♦ 1 A HISTORICIDADE DA ERA DIGITAL E OS LUGARES DE MEMÓRIA DA CIÊNCIA

As tecnologias digitais, integradas a um contexto global de mudanças sociais, econômicas e culturais, afetam diretamente o mundo da ciência, em suas práticas e processos, incluindo a comunicação científica. A vivência social, por outro lado, está saturada pela instantaneidade e disponibilidade de informação massiva, a qual está associada a outras formas de transmissão e de construção da memória social, anteriormente caracterizada por relações presenciais de convívio e pela multiplicidade de referências locais.

A relação que as sociedades fazem com o passado está vinculada à percepção social do tempo, que é também histórica e mutável. A problemática da memória, em sua multiplicidade como processo social, teve suas principais questões tratadas por M. Halbwachs, na primeira metade do século XX, em obra póstuma, publicada em 1950, onde o autor apresenta a relevância dos espaços e das instâncias de sociabilidade para a transmissão e apropriação da memória coletiva (HALBWACHS, 2006). Porém, Le Goff (1994), refletindo sobre as relações entre memória e história, mostra que a memória, transmitida pela escrita e instituída pelo poder do Estado, está, desde as mais antigas civilizações humanas, associada às cidades, portanto, à materialidade do desenvolvimento urbano (LE GOFF, 1994, p. 63, 432, 433).

Até um período recente, a atenção privilegiada dada à memória no mundo ocidental foi elaborada na materialização de museus, como já estudado por Pierre Nora e Andreas Huyssen (NORA, 1984; HUYSSSEN, 2000). Portanto, uma memória que está espacialmente inscrita como forma de se manter um vínculo com um passado, transformado pelas práticas modernizadoras da urbanização, da industrialização e dos sistemas técnicos, entre eles a ferrovia, a telefonia, o rádio, a fotografia e a televisão (LE GOFF, 1994). No século XXI, a dimensão da memória está impregnada de novos problemas, conforme debatem os historiadores que creditam às tecnologias digitais a crise da história.

É, no sentido da não relevância da proximidade espacial para a realização das práticas socialmente dependentes das tecnologias digitais, que se revela uma descontinuidade, tanto simbólica quanto do tempo vivido atual, voltado para as trocas comunicacionais não mais mediadas por outras instâncias geradoras de autoridade ou de validação.

No entanto, a comunicação científica, mesmo imersa nos fluxos de informação digitais, resulta de um universo de práticas institucionalizadas para a produção do conhecimento científico, que reveste de valorização diferenciada os meios digitais de acesso à informação de que dispõe, contribuindo para a construção da relação entre informação, memória científica e memória coletiva.

A noção de lugares de memória de Pierre Nora (1984) é aqui revisitada para tratar das transformações introduzidas pelas tecnologias digitais na ampliação do acesso à informação científica e na temporalidade de sua disponibilidade. Estas dependem, porém, da prévia definição das instituições e do comprometimento social para com a máxima disponibilização pública de seus estoques de informação, no dizer de A. Barreto (2002). Os serviços de acesso digital, integrados a fluxos de comunicação, partem da criação/emissão da informação que será entregue a um receptor, porém, esses serviços devem sua existência aos centros geradores e mantenedores de comunidades científicas (BARRETO, 2002).

A equiparação de F. Sayão entre a estruturação e armazenamento do saber científico em bases de dados e a construção de uma metáfora sobre a memória científica (SAYÃO, 1996) também deve ser considerada. Essa é uma possibilidade de explorar, como um lugar de memória, a noção de estoque informacional informatizado e disponibilizado em instituições produtoras de informação qualificada e de caráter cumulativo. Nos termos de Nora, não é o objeto ou o suporte que são relevantes, mas, sim, a inscrição da memória que se realiza como referente do passado, mantido e administrado no presente (NORA, 1984, p. xxxvii, xli).

Dessa forma, quais são os traços históricos de entendimento para discutir a questão da memória científica na era digital?

A partir das questões e parâmetros colocados pelas instituições produtoras dos saberes científicos é que acontecem os processos de organização, sistematização, armazenamento e disponibilização ao público interessado. A dimensão facilitadora do acesso à produção científica tem uma história e um tempo marcados, que extrapola o quadro de interconectividade e de múltiplos fluxos de informação propiciados pelas tecnologias digitais. Isso acontece porque há um compromisso ético com a publicização digital da produção científica, local ou internacional, nas instituições de pesquisa no Brasil, em adesão aos movimentos internacionais de acesso livre.

Há outro lado da questão, associada ao acesso à informação científica, que está ligada à historicidade do nosso tempo, ditado pela inovação continuada geradora de um efeito de substituição tecnológica acelerada. Portanto, não é suficiente a informação científica estar disponível hoje nos meios oferecidos pelas tecnologias digitais, mas que ela esteja disponível pelo tempo previamente determinado por instâncias decisórias dedicadas à questão da memória nos centros de pesquisa.

As práticas científicas obedecem a um ciclo de desenvolvimento, no qual uma das etapas é a publicação de resultados em diferentes meios, como parte da comunicação científica. Neste circuito, a digitalização vem ganhando, cada vez mais, a dominância dos processos. Entretanto, é notável que duas áreas de conhecimento se coloquem de forma distinta e complementar diante do problema da memória na era digital: a história e a ciência da informação. A primeira tem o papel de retirar suas fontes de acervos para estudo científico de problemas do passado, sejam essas digitais ou não, e colocá-las em circulação novamente na memória social. Já a ciência da informação está voltada para o processo que permite traçar a completude do circuito informacional, nos estudos da comunicação científica e no âmbito do acesso aberto, para que a informação esteja disponível e chegue ao receptor, que livremente a selecionará conforme sua demanda temporal e espacial, independente da fragilidade dos objetos digitais. Nesses termos, a etapa de reflexão sobre a memória atribuída à preservação digital será uma condição necessária.

♦ 2 O FAZER DA HISTÓRIA E A MEMÓRIA DIGITAL DA CIÊNCIA

A história, como área de conhecimento e em sua relação com a memória, depende, fundamentalmente, da disponibilidade das fontes sistematizadas em instituições de memória, sejam estas relacionadas a registros textuais, imagéticos, cartográficos ou da cultura material, com exceção da história oral, que constitui diretamente, no tempo presente, sua incursão pelos vestígios da memória social (LE GOFF, 1994).

No tempo do domínio digital, o acesso *on-line* aos materiais digitalizados e disponibilizados em arquivos, bibliotecas e museus introduz uma série de inovações na atividade do historiador, com repercussão no conjunto de suas práticas (FRIDLUND, 2020). Porém, a maior facilidade de acesso

conjuga-se à incerteza sobre as condições do trabalho do historiador em um futuro próximo, se mantida a descontinuidade de políticas de memória que não atentem para o problema da permanência dos registros digitais, em especial os de origem científica. Essa descontinuidade pode configurar, ainda, uma temporalidade institucional, centrada no presente, deslocada dos registros de longo prazo e, potencialmente, não histórica, em paralelo à existência de lugares de memória, conforme define Nora (1984).

Não há dúvida quanto à agilidade do trabalho de investigação que o acesso *on-line* aos acervos documentais possibilita ao reduzir o tempo de pesquisa. Porém, a facilidade de acesso também leva a ampliar o número de fontes e períodos de investigação, cujos resultados exigem reavaliação de métodos, técnicas ou procedimentos de análise do material disponibilizado, com a incorporação de ferramentas de processamento computacional, já em associação interdisciplinar com outras áreas de conhecimento, como vem sendo apresentado na área denominada de Humanidades Digitais (MANOVICH, 2017).

Do ponto de vista da relação entre acesso à informação e sociedade, o historiador social ou da cultura, como agente que questiona o passado a partir do presente ou, no movimento inverso, o presente a partir do passado para observar os processos de mudança, como propôs Marc Bloch, (LE GOFF, 1994, p. 14, 23, 24, 223) pode, em suas práticas, melhor expressar a dimensão de historicidade do nosso tempo. Na investigação histórica da atualidade está em questão a dominância da instantaneidade das práticas cotidianas, sem mediações ou espaços de interlocução que não sejam informacionais, somada à reduzida perspectiva de armazenamento público desses dados e de sua disponibilidade a médio e longo prazo. A exceção de algumas poucas experiências de arquivamento da web, a exemplo do *arquivo.pt*, em língua portuguesa, aponta um caminho que depende de políticas públicas para sua implementação.

No entanto, a relação entre o meio digital e as transformações sociais e culturais associadas, que também afetam o campo de produção dos saberes científicos, coloca novas questões para o trabalho do historiador e dizem respeito aos problemas levantados e ao processo de análise do volume imensurável de dados publicados e em circulação nas redes sociais, nas demais mídias veiculadas em tecnologias digitais e nos fluxos globais da web, que podem ser trabalhadas para uma história do tempo presente, mas de imprevisível disponibilidade para análise futura. Tendo em conta, porém, que a história, como área de conhecimento, se debruça e se detém sobre os vestígios disponíveis, mas também sobre os registros promotores do deslocamento entre memória social e o seu esquecimento, entre o que é preservado como documento ou não (LE GOFF, 1994, p. 544, 545).

Essa problemática da exponencial imersão no tempo presente e a capacidade de produzir, com apoio das tecnologias digitais, análises de problemas que se tornam apreensíveis pela escala acumulada de dados e interações não previstas identificadas pela computação científica, é a vertente mais promissora de diversas ciências, na biologia, na medicina, na biotecnologia e nas ciências sociais, além de aplicações nas temáticas de desastres naturais e *smart cities* (PIRANI, 2005; MANOVICH, 2017; WILLIAMS, 2017; FUCHS, CHANDLER, 2019). Porém, esse caráter cumulativo das ciências, que estruturam seu processo de conhecimento em inteligência artificial e em práticas de modelação computacional inovadoras, com grande infraestrutura em bases de dados, depende de lógicas compartilhadas de tratamento da investigação em equipes distribuídas mundo afora. Nesse ambiente colaborativo, têm sido geradas soluções para o arquivamento digital em ciência e tecnologia, contemplando, também, acesso a dados primários, que podem ser de uso múltiplo e público (WILLIAMS, 2017).

Distante desses domínios promissores das ciências, com maior aproximação no que diz respeito às tecnologias digitais, nota-se, também, que as demais ciências não estão isentas dessa relação, tendo em vista, prioritariamente, o contexto próprio referente à comunicação científica, dadas as condições de infraestrutura tecnológica e de tratamento da informação necessários à sua efetivação, nomeadamente nos canais formais de disseminação dos saberes, que são, atualmente e em sua maioria, digitais.

Assim, a questão da memória digital se coloca tanto para definir estratégias associadas aos processos de investigação e à respectiva guarda dos dados resultantes, nos domínios científicos associados ao *big data*, quanto nas estruturas de informação vinculadas à disseminação da informação científica, que atendem a diversidade de áreas de conhecimento. Os diversos serviços de informação sustentados por universidades públicas brasileiras têm sido responsáveis por garantir acesso aberto a um volume expressivo de documentação científica do país, mas pouca atenção tem sido dada ao problema da memória e sua estratégia conexa de atuação em preservação digital.

Desse ponto de vista, a história da produção científica no Brasil, como processo, corre o risco de não garantir às gerações futuras o seu legado, que depende de práticas intensivas, sistemáticas e seletivas de arquivamento digital, tendo em conta o planejamento de longo prazo. Sem acervos digitais organizados e sistematizados com previsão de proteção a longo tempo, não haverá a institucionalização da memória, mas a institucionalização do esquecimento, resultante da vulnerabilidade aos riscos do apagamento digital.

3 GESTÃO DOCUMENTAL E INFORMAÇÃO CIENTÍFICA NA ERA DIGITAL

A informação científica na atualidade está cada vez mais dependente, para sua efetiva circulação, da hegemonia dos meios digitais. Pode-se estruturar essa abrangência do digital, com adaptação da proposição de Barreto (2002), em três domínios ou universos informacionais, componentes da inteligibilidade oriunda do o campo de saber da ciência da informação: o que abrange o aparato tecnológico de armazenamento e disponibilização – vinculado à infraestrutura de serviços de informação institucionais; os conteúdos informacionais – elaborados por sujeitos inseridos em sistemas de saberes, materialmente representados em objetos digitais; e o aparato disponível de interconectividade – condicionado à estrutura de redes tecnológicas globalizadas, para a disseminação de conteúdos em canais direcionados ou em fluxos de informação não previstos, que levam esse material informacional a um receptor, situado em qualquer contexto temporal e espacial (BARRETO, 2002).

Considerando-se o esquema acima descrito, convém destacar que a formulação ou criação da informação científica decorre de atividades controladas de pesquisas, fundamentadas teórica e metodologicamente, e cujos resultados são validados nas respectivas instituições de pesquisa ou nos círculos de saber pertinentes, a exemplo da revisão por pares. Assim, a informação científica é constitutiva de um processo amplo de comunicação do processamento de informações e do conhecimento produzido, reconhecido por mecanismos de avaliação consensuais. (VALÉRIO; PINHEIRO, 2008). Essa comunicação científica, de caráter formal e pública, destinada à interlocução com a comunidade científica, tende a uma aproximação com a divulgação científica, no sentido de extrapolação das fronteiras do ambiente especializado, dada a ampliação do público de interesse e a facilidade de acesso à informação no quadro de relações constituído pela interconectividade digital das redes (BARRETO, 2002; VALEIRO; PINHEIRO, 2008).

Dado o caráter público da comunicação científica e o contexto de demandas do movimento pelo acesso aberto às publicações científicas (além do interesse das instituições em ampliar a visibilidade de sua produção), foram instituídas novas práticas de organização, armazenamento e disponibilização, que marcam um novo patamar na história do acesso à informação científica, com destaque em duas frentes já amplamente tratadas desde a *Declaração de Bethesda e a Declaração de Berlin*, de 2003 (PEREIRA, 2014; RIOS *et al.*, 2019). Outro movimento de amplificação do acesso à informação científica está em curso na perspectiva de tornar mais visíveis resultados, dados não processados e metadados, na chamada ciência aberta (VALEIRO; PINHEIRO, 2008; PINHEIRO, 2014). Do primeiro movimento, decorre a ampliação de periódicos científicos eletrônicos de acesso aberto, processados e disponibilizados em grande número no Brasil, por meio do Open Journal Systems (OJS), e os serviços de informação (repositórios institucionais) que, gradativa e cumulativamente, disseminam a produção digital acadêmica brasileira, a qual pode ser acompanhada em consulta à Biblioteca Brasileira Digital de Teses e Dissertações, do Instituto Brasileiro de Informação em Ciência e Tecnologia (BDTD/Ibict).

Porém, o processo de gestão da informação científica apresenta uma lacuna, que diz respeito à disponibilidade de forma perene, tanto de objetos natos digitais, como de imensos acervos não digitalizados, restringindo o acesso público, das atuais e futuras gerações, em um contexto cada vez mais determinado pelas tecnologias digitais, apesar da assertiva pública e das recomendações nacionais e internacionais (CONARQ, 2005).

O reconhecimento da relevância do patrimônio cultural de caráter científico, no domínio digital, cujos critérios de validação pertencem à história dos respectivos saberes, necessita de garantias de preservação em longo tempo, que contemplem a autenticidade e a confiabilidade dos objetos e de seus registros de identificação permanentes, em contraponto ao volume incomensurável de informação, de reprodução e de compartilhamento em meios digitais não confiáveis e à própria condição de obsolescência tecnológica a que estão sujeitos os objetos digitais e seus serviços de suporte (CONARQ, 2005).

Assim, no contexto atual, merece atenção a questão da gestão institucional da documentação científica, que, em conformidade com os ditames da dimensão pública da comunicação científica, incluiria o acesso amplo e aberto à informação e ao conhecimento produzido associado, além da necessidade de criação de políticas e serviços de preservação digital (MÁRDERO ARELLANO, 2008).

Esse aspecto abrangente do foco na gestão documental, como já tratado por Bernardes (2018), aponta para uma conjunção de ações e de políticas que necessitam estar presentes desde a origem da produção da informação científica, já que isso permitirá estabelecer dinâmicas de curadoria, de caráter arquivístico, para o conjunto de dados e informações processadas antes mesmo da criação do objeto digital como produto final. A memória da pesquisa, com seu percurso metodológico, poderia, dessa forma, ser resguardada e estar pronta para eventual e futura recuperação, tanto dos dados primários ou intermediários quanto de processos de investigação, o que contempla parcialmente o debate em torno da ciência aberta.

Os procedimentos para arquivamento de objetos digitais e de seu conteúdo integral em acervos institucionais permitem a associação de múltiplos tipos documentais, que podem compreender não só materiais digitais resultantes de processamento computacional e registros de observação direta (fotografias e ilustrações científicas), quanto inserções de imagens em movimento e arquivos multimídia, desde que haja a descrição e formulação estruturada de metadados, conforme áreas de

conhecimento. O problema da propriedade intelectual também necessita ser definido de acordo com os graus de disponibilidade de acesso indicados pelo autor no autoarquivamento. Outra questão, relacionada à operacionalidade, é a de garantir a compatibilidade com outros serviços com padrões de interoperabilidade. Mas, do ponto de vista da memória científica, sempre seletiva, resta, ainda, uma questão de fundo sobre o longo prazo, que envolve o planejamento de estratégias e práticas de arquivamento digital de menor custo, tanto dedicado à recuperação da informação descritiva do documento científico, quanto à proteção do conteúdo integral, que já conta, no Brasil, com um serviço de modelo distribuído, da Rede Cariniana (Ibict), destinado à preservação digital.³

♦ 4 O IBICT E A INICIATIVA DE PRESERVAÇÃO DIGITAL LOCKSS ALLIANCE

A preservação digital é uma atividade formal na área da biblioteconomia e da arquivologia, onde estudos de casos, pesquisas sobre modelos, e desenvolvimento de ferramentas têm surgido com o propósito de solucionar e atender perdas de informações e normalização dos dados digitais em áreas da gestão documental. Em muitos países, as experiências na implementação e nas ações de preservação digital têm despertado interesse de instituições patrimoniais, as quais procuram formalizar estratégias de preservação por meio de políticas e de planos de ação em diversos contextos de infraestrutura, direcionadas para coleções digitais de diferentes índoles e cenários.

A área de trabalho da preservação digital é ainda pouco conhecida, em parte pela sua curta existência e pelo reduzido número de especialistas no mundo. Com tudo isso, as práticas de digitalização e de criação de repositórios convergem em ações de preservação digital nos espaços em que são formalizadas as normas, ferramentas e sistemas de arquivamento digital. Mas, como atividades pontuais, a junção de políticas, de planos de ação e de estratégias operacionais não é sempre plenamente aplicada (TÉRMENS, 2009).

Para Cabero (2019), as atividades de preservação digital devem ser observadas na forma como são concebidas e na sua atuação como finalidade ativa dentro da curadoria e da gestão de acervos digitais, assim como os perfis específicos segundo as necessidades da comunidade de usuários e do cenário local. Uma interpretação clara da preservação digital pode permitir a sua adoção na cultura organizacional com atividade de gestão e com planejamento estratégico. Em virtude da quantidade de informações produzidas pelas áreas científica e acadêmica, a pesquisa e o desenvolvimento de redes de preservação digital nesta área se tornaram relevantes.

Desde sua fundação, o Ibict tem a missão de organizar a produção científica no Brasil. Portanto, a preservação dessa informação em formato digital é mais uma das suas atribuições. A transição do formato impresso para o digital, com as teses e dissertações e com os periódicos eletrônicos, favoreceu a implementação de uma rede de preservação digital dentro do Instituto. Desde 2002, o Ibict demonstrou interesse por sistemas de informação que servissem para a preservação digital da produção científica nacional. Já em 2003, o Instituto reuniu vários especialistas para discutir o tema da preservação digital e a criação de um comitê. O comitê foi composto por representantes do Ibict, Biblioteca Nacional, Programa Sociedade da Informação, Bireme, Capes, Abec, Unesco e especialistas convidados. Foram criados vários GTs, entre os quais estava o de Certificação Digital. Entretanto, a

3 Rede Brasileira de Serviços de Preservação Digital Cariniana <http://cariniana.ibict.br>

iniciativa não teve prosseguimento devido às mudanças administrativas e de diretoria, mas o Instituto continuou interessado em identificar as tendências e apontar as atividades de preservação digital que pudessem ser aplicadas no contexto nacional de C&T.

Pesquisas em iniciativas reconhecidas pelo uso de soluções tecnológicas de acesso aberto amplamente adotadas em projetos internacionais de preservação digital continuaram sendo feitas no Instituto, identificando ferramentas que passaram pela verificação do gerenciamento e controle de requisitos funcionais para o funcionamento em rede de preservação distribuída. Este modelo de preservação digital está baseado em estudos de instituições de pesquisa, que vêm tendo como foco a necessidade de disponibilizar, de forma permanente, materiais autênticos e patrimoniais, observando padrões internacionais.

Desde 2010, equipes do Ibict iniciaram pesquisas e testes para implementação de uma rede nacional de preservação digital, propondo um projeto estruturante com o mesmo título dentro do Instituto (Ibict, Plano Diretor 2011-2015)⁴. No desenvolvimento do projeto, foi indispensável a realização de pesquisas em organizações internacionais que possuíam projetos de preservação digital. Durante o período de estudo inicial, foi possível realizar visitas técnicas à Universidade do Minho, em Portugal; ao projeto Data-Pass, da Harvard University, e à Aliança LOCKSS, da Stanford University Libraries.

Em 2012, o trabalho de prospecção e de soluções tecnológicas e de iniciativas internacionais no Ibict concretizou a proposta de uma política de preservação digital e a criação de um protótipo do sistema na etapa de teste das funcionalidades desejadas. A política de preservação digital para o gerenciamento de uma Rede Nacional de Preservação Digital foi elaborada tendo como base uma lista de requisitos funcionais para garantir a validade dos registros e a sua permanência, seguindo o modelo de referência OAIS. A proposta indicava as diretrizes que o Ibict adotou relacionadas à administração da Rede⁵.

Outro resultado da prospecção foi a proposta de solução tecnológica e metodológica, LOCKSS (Lots of Copies Keep Stuff Safe), de autoria da Stanford University, que o Instituto indicaria às instituições parceiras do projeto. A escolha da ferramenta LOCKSS foi realizada após análise de sua estrutura, requisitos e fluxos de dados. A definição das condições para a realização das parcerias que possibilitaram a implementação da rede piloto, dirigida à preservação de periódicos que utilizam a ferramenta OJS/SEER, foi resultado das atividades da primeira equipe de bolsistas PCI (Programa de Capacitação Institucional). Os acordos elaborados definiram a interação entre as equipes que gerenciariam essa e as outras sub-redes.

A Cariniana está composta de sub-redes estruturadas pensando na necessidade da preservação do tipo de documento e de seu valor quanto às ligações que ele possui, sejam eles periódicos científicos, livros eletrônicos, teses, dissertações, dados de pesquisa (independentemente do formato), ou base de dados e multimídia incluída. A garantia da existência de um avançado conjunto de serviços de preservação foi considerada na hora de se criarem espaços para armazenamento de documentos digitais em várias instituições.

4 Em 2011, a preservação digital passou a ser o único projeto estruturante do instituto no PDU (2011-2015), tendo como metas as seguintes: (63) desenvolver política de preservação digital da rede nacional; (64) criar a infraestrutura tecnológica necessária, (65) implantar a rede nacional e, (66) integrar a rede nacional à iniciativa LOCKSS.

5 Política de Rede Brasileira de Preservação Digital Cariniana. <https://cariniana.ibict.br/index.php/institucional/documentacao/446-politica-da-rede-brasileira-de-servicos-de-preservacao-digital>

A operação de aquisição da licença LOCKSS ofereceu a participação do Ibict nas atividades de desenvolvimento e transferência de informações dentro da Aliança LOCKSS Internacional, assim como o apoio técnico para o funcionamento da Rede Nacional. O preço praticado pela Stanford University Libraries mostrou a sua intenção favorável de cooperação na viabilização da Rede Nacional, tratando a aquisição da licença como sendo para uma única rede local (6 nós de armazenamento ou caixas LOCKSS), sem nenhuma outra cobrança para outras instituições participantes que comporiam a Rede Brasileira de Serviços de Preservação Digital, Cariniana.

Inicialmente, devido à sua amplitude, o projeto envolveu atividades com seis instituições/universidades brasileiras, com o apoio de seus respectivos centros de informação e de suas editoras. A implementação da primeira etapa da rede de serviços de preservação digital disponível no Ibict incluiu periódicos científicos de acesso livre das instituições parceiras, iniciando o desenvolvimento de uma rede de acervos preservados que permite a livre adesão e a integração de conteúdo da memória institucional de forma consorciada e federada.

Em 2014, com apoio da FINEP, a aquisição da licença LOCKSS para uso na construção da Rede Cariniana foi concretizada, e a adesão à Aliança LOCKSS resultou na apreciação crítica da necessidade da aplicação da dita ferramenta no período relativo a 12 meses, com extensão para 24 meses (2014 a 2016). O Ministério de Ciência, Tecnologia e Inovações permitiu a extensão da adesão, seguindo um Processo de inexigibilidade de licitação (Lei 8.666 Art. 25 Parágrafo I), pois a Stanford University Libraries era a única instituição a disponibilizar ferramenta de preservação digital distribuída que observava os requisitos funcionais para construção de redes colaborativas com nós de armazenamento nacional.

O *software* LOCKSS (Lots of Copies Keeps Stuff Safe), criado em 1998 na Stanford University, é a única ferramenta que replica o armazenamento de dados em vários locais, criando cópias do mesmo conteúdo, comparando-as entre si, podendo ser validadas e reparadas automaticamente. O LOCKSS é uma ferramenta que possui escalabilidade alta no armazenamento de objetos digitais, com capacidade de salvaguarda de milhões de objetos digitais em nós de armazenamento distribuídos dentro do país. Os *plugins* acoplados ao sistema dão subsídio para armazenar e recuperar todo tipo de material digital existente (LOCKSS, 2008).

Desde 2015, a Rede Brasileira de Serviços de Preservação Digital faz parte de um seleto grupo de agências de arquivamento de periódicos eletrônicos, o The Keepers Registry, um serviço online operado pelo Centro do International Standard Serial Number (ISSN), que atua como um monitor global nas disposições de arquivamento para recursos contínuos. A Cariniana é a rede que mais preserva periódicos que não estão em nenhuma outra agência, promovendo, assim, a visibilidade e a preservação que o The Keepers concede, uma vez que há rigorosos critérios e requisitos para garantir a participação das instituições.⁶

No Brasil, o *software* LOCKSS continua servindo para estabelecer a arquitetura de sub-redes colaborativas para tratamento e preservação de publicações técnico-científicas-eletrônicas. O reconhecimento da importância da participação da Aliança LOCKSS esteve baseado no pressuposto de que era necessário desenvolver redes colaborativas nacionais que assegurassem não só as migrações, mas também o contexto, a estrutura e a acessibilidade dos documentos digitais produzidos no país, contribuindo para a salvaguarda do patrimônio nacional. A aceitação de uma

⁶ The Keepers Registry <https://keepers.issn.org/>

solução prática de preservação digital como o LOCKSS, adotado por instituições de pesquisa na Europa, Ásia e América do Norte, significou estar em consonância com outras iniciativas que vêm comprovando o valor científico do sistema como padrão internacional de preservação digital.

◆ 5 CONCLUSÕES

Há uma mudança significativa nos processos de comunicação e de construção social da memória associada às tecnologias digitais que têm implicações na formação da consciência histórica do nosso tempo. Os ambientes institucionais que tratam de ampliar o acesso à informação qualificada também devem instituir as práticas para garantir a democratização da memória coletiva, considerando, em especial, aquela vinculada à circulação da informação científica, como parte do processo de apropriação por atuais e futuras gerações, do conhecimento científico acumulado em centros de pesquisa brasileiros.

Um novo papel está reservado às instituições públicas e privadas para assegurar a preservação da memória em meios digitais, em um contexto de apropriação dispersa dos conteúdos localmente produzidos e de redefinição dos padrões, valores e ordenamentos socioculturais que, no passado, vinculavam memória e identidade à dimensão espacial e à experiência do que era vivido em situações de intersubjetividade.

A preservação digital como prática institucional está, ainda, vinculada às experiências relativamente restritas no Brasil. A perspectiva de expansão de procedimentos e de institucionalização de políticas nesse campo envolve a estruturação em redes de arquivos ou de bibliotecas que contribuam para disseminar as tecnologias e os serviços de suporte, promover o compartilhamento de acervos e ampliar as possibilidades de intercâmbio entre pesquisadores dedicados às mesmas temáticas.

No Brasil, as áreas que lidam com informação e com atividades de desenvolvimento científico não contam com canais oficiais para a identificação, comunicação e avaliação da preservação digital de longa duração. A organização desses canais não pode desconsiderar sua dimensão internacional, pois, no momento em que se escolhe determinado conjunto de critérios de preservação, este deve ser integrado aos outros sistemas que adotem critérios semelhantes.

O Ibict é uma instituição com ampla experiência na prospecção, absorção e customização de sistemas baseados em *software* livre, no modelo "Open Access". Na última década, o Instituto tem reconhecido a importância do estudo de soluções apropriadas para a preservação das informações em C&T em formato digital. Essa constatação levou, em 2012, a uma proposta de aquisição, desenvolvimento e implantação de uma infraestrutura técnica e operacional que integre o Instituto com outras instituições parceiras em uma rede nacional. Com a rede Cariniana, o Ibict trouxe para si o papel de promotor e de intermediador na construção dessa infraestrutura distribuída para diagnóstico e preservação de importantes acervos digitais da memória científica e tecnológica do país.

O modelo distribuído de rede de preservação digital no Brasil já é identificado com aqueles desenvolvidos pelas iniciativas observadas na comunidade internacional durante a última década, a saber: a formulação de políticas que envolvam o uso de tecnologias e de padrões testados e adaptados em projetos de coleções digitais. A rede de preservação distribuída pode salvaguardar a memória científica e permitir sua incorporação em iniciativas internacionais já existentes (MÁRDERO ARELLANO, 2012).

Além da criação de infraestrutura tecnológica, resta o desafio de investigação sobre a relação entre memória institucional e tecnologia, que continua sendo a preservação de todas as características de um objeto digital, o que pode resultar em uma grande quantidade de recursos sendo gastos na preservação de elementos irrelevantes para o significado arquivístico do registro. Para evitar isso, os arquivistas precisam determinar quais elementos de um registro digital são essenciais para que ele retenha significado, tendo em vista a garantia de sua autenticidade.

Determinar a essência dos registros, ou seja, o conjunto de propriedades do objeto digital e a sua caracterização como documento, não é uma ciência e está aberta a subjetividades e interpretação arquivística, mas é essencial para um programa de preservação eficaz e responsável. Focalizar a pesquisa nos componentes básicos de um registro permite declarar abertamente as responsabilidades e os requisitos de arquivamento, indicando a trajetória na cadeia de custódia que levará à preservação desse registro. Isso significa que os pesquisadores, no futuro, poderão ter acesso às decisões de arquivamento que foram feitas sobre a natureza de um registro quando ele foi preservado.

Os desafios da preservação digital afetam as instituições arquivísticas em todo o mundo. Além das bibliotecas, os arquivos desenvolvem padrões de manutenção de registros de seus acervos documentais e estão empenhados em fornecer soluções inovadoras para os problemas de preservação digital. Espera-se que as soluções que estão sendo implementadas no campo da Ciência e Tecnologia sejam valiosas para instituições de outras vertentes de atuação e que sirvam para preservar milhões de registros que serão de valor significativo para gerações futuras.

◆ REFERÊNCIAS

BARRETO, A. A. O tempo e o espaço da ciência da informação. **Transinformação**, v. 14, n. 1, p. 17-24, 2002. Disponível em: <http://ridi.ibict.br/handle/123456789/179>. Acesso em: 18 maio 2021.

BERNARDES, I. P. Os desafios da gestão e preservação de documentos arquivísticos digitais. **Revista de arquivos**, São Paulo, ano II, n. 6, p. 190-207, 2018. Disponível em: http://www.arquivoestado.sp.gov.br/revista_do_arquivo/06/especial_seminario_apresentacao_01.php#inicio_artigo. Acesso em: 29 jul. 2021.

CABERO, M. M. Preservación digital: retos, experiencias y oportunidades. **Revista Ibero-Americana De Ciência Da Informação, Brasília**, v. 12, n.1, p.120-144, 2018. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/10523/18569>. Acesso em: 1 ago. 2021.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Carta para a preservação do patrimônio arquivístico digital**. Brasil: Unesco, 2005. Disponível em: <http://www.gov.br>. Acesso em: 10 ago. 2021.

FRIDLUND, M. **Digital history** 1.5: a middle way between normal and paradigmatic digital historical research. In: FRIDLUND, M.; OLIVA, M.; PAJU, P. (ed). *Digital histories: emergent approaches within the new digital history*. p. 69-87. Helsinki: Helsinki University Press, 2020. Disponível em: <https://doi.org/10.33134/HUP-5-4>. Acesso em: 30 abr.2021.

FUCHS, C.; CHANDLER, D. Introdução. In: FUCHS, C.; CHANDLER, D. **Digital objects, digital subjects: interdisciplinary perspectives on capitalism, labour and politics in the age of big data**. London: University of Westminster Press, 2019. p. 1-20. Disponível em: <http://www.uwestminsterpress.co.uk/site/books/10.16997/book29>. Acesso em: 4 maio 2021.

HALBWACHS, M. **A memória coletiva**. São Paulo: Centauro Editora, 2006.

HUYSEN, A. **Seduzidos pela memória: arquitetura, monumentos e mídia**. Rio de Janeiro: Aeroplano, 2000.

LE GOFF, J. **História e memória**. Tradução Bernardo Leitão et.al. 3.ed. Campinas, SP: Editora da UNICAMP, 1994.

LOTS OF COPIES KEEP STUFF SAFE. About LOCKSS. In: ____. LOCKSS. [s.l.],

2008. Disponível em: http://www.lockss.org/lockss/About_LOCKSS. Acesso em: 11 ago. 2021.

MANOVICH, L. Cultural analytics, social computing and digital humanities. In: SCHÄFER, M T.; ES, K. V. **The Datafied Society: studying culture through data**. Amsterdam: Amsterdam University Press, 2017. p. 55-68. Disponível em: <https://www.aup.nl/en/book/9789462987173/the-datafied-society>. Acesso em: 28 abr. 2021.

MÁRDERO ARELLANO, M. **A. Critérios para a preservação digital da informação científica**. 2008. 354f. Tese (Doutorado) – Programa de Pós-Graduação em Ciência da Informação – Universidade de Brasília, Brasília, 2008. Disponível em: <https://repositorio.unb.br/handle/10482/1518>. Acesso em: 27 jul. 2022.

MÁRDERO ARELLANO, M. A. Cariniana: uma rede nacional de preservação digital. **Ciência da Informação, Brasília**, v. 41, n. 1, p. 83-91, jan./abr 2012. Disponível em: <http://revista.ibict.br/ciinf/article/view/1354>. Acesso em: 30 mai. 2021.

NORA, P. Entre mémoire et histoire: la problématique des lieux. In: NORA, Pierre. (dir.) **Les lieux de mémoire. La république**. [S.l.]: Gallimard, 1984.

PINHEIRO, L.V. R. Do acesso livre à ciência aberta: conceitos e implicações na comunicação científica. **RECIIS**, v. 8, n. 2, p. 153-165, 2014. Disponível em: <http://www.reciis.icict.fiocruz.br>. Acesso em: 30 mai. 2021.

PIRANI, J. R. **Sistemática**: tendências e desenvolvimento, incluindo impedimentos para o avanço do conhecimento na área. CGEE: 2005. Disponível em: <http://www.cgee.org.br>. Acesso em: 28 mai. 2021.

RIOS, F.P.; LUCAS, E. R. O.; AMORIM, I. S. Manifestos do movimento de acesso aberto: análises de domínio a partir de periódicos brasileiros. **Revista Brasileira de Biblioteconomia e Documentação**. v. 1, n. 1, p. 148-169, 2019. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/1152>. Acesso em: 10 ago. 2021.

SAYÃO, F. Bases de dados: a metáfora da memória científica. **Ciência da Informação**. Brasília, v. 25, n. 3, p. 314-318, 1996. Disponível em: <http://revista.ibict.br/ciinf/article/view/629>. Acesso em: 29 jul. 2021.

TÉRMENS I GRAELLS, M. Preservación digital en 2009. **Anuario ThinkEPI**, [S.l.], v. 4, p. 224-230, 2010.

VALEIRO, P. M.; PINHEIRO, L.V. R. Da comunicação científica à divulgação. **Transinformação**, v. 20, n. 2, p. 159-169, 2008. DOI 10.1590/S0103-37862008000200004. Disponível em: <https://www.scielo.br/j/tinf/a/jXWgggxBhXfsT57JDVbghp/?lang=pt>. Acesso em: 29 jul. 2022.

WILLIAMS, E. *et al.* Image data resource: a bioimage data integration and public platform. **Nature methods**, v. 14, n. 8, 2017. p. 775-782. Disponível em: <http://www.nature.com/naturemethods>. Acesso em: 29 abr. 2021.

COMO CITAR:

TAVARES, Maria de Fátima Duarte; MÁRDERO ARELLANO, Miguel Ángel. O Ibict e a preservação da memória científica. In: BRAGA, Tiago Emmanuel Nunes; MÁRDERO ARELLANO, Miguel Ángel. (org.). **Hipátia**: modelo de preservação para repositórios arquivísticos digitais confiáveis. Brasília: Ibict, 2022. p. 26-39. (Informação, Tecnologia e Inovação, v. 1). DOI 10.22477/9786589167501.cap2.

Capítulo 03

OS PRINCIPAIS NORMATIVOS RELACIONADOS AO RDC-ARQ

THE MAIN NORMATIVES RELATED TO THE TDAR

Tatiana Canelhas Pignataro¹

Cristiano Menezes²

Tiago Emmanuel Nunes Braga³

¹ Dados da autora: Arquivista e cientista da computação, pesquisadora do Instituto Brasileiro de Informação em Ciência e Tecnologia, tatianacanelhas@ibict.br

² Dados do autor: Especialista, arquivista do Tribunal de Justiça do Distrito Federal e Territórios, cristiano.alvares@tjdft.jus.br

³ Dados do autor: Doutor em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, tiagobraga@ibict.br

RESUMO

Resumo: Este estudo exploratório e descritivo faz um levantamento dos principais normativos relacionados aos Repositórios Arquivísticos Digitais Confiáveis, em especial no âmbito das instituições jurídicas. Ele identifica cinco tipos de normativos, entre os principais a serem considerados: normas ISO, normas arquivísticas, normas legais, modelos, relatórios e sistemas. Observou-se que diversas normas legais foram publicadas nos últimos anos, o que impactou as instituições que planejam implementar repositórios digitais confiáveis. Como conclusão, observou-se que entre os normativos analisados há indicação da necessidade de atuação conjunta entre os departamentos de gestão documental e tecnologia da informação. Por fim, restou evidenciado que o levantamento realizado não é exaustivo, mas que serve como um orientador para instituições que estejam se movimentando para atender à legislação vigente.

Palavras-Chave: RDC-Arq. Normas de preservação digital. Gestão documental

Abstract: This exploratory and descriptive study surveys the main normatives related to the Trusted Digital Archival Repositories, in particular in the ambit of juridical institutions. It identifies five types of normatives between the mains to be considered: ISO standards, archival standards, legal standards, role models, reports and systems. It was observed that many legal standards were published in the last few years, which impacted the institutions that plan to implement trusted digital repositories. In conclusion it was observed that between the analyzed normatives there is an indication to the need for joint action between the documental management and technology of information. Lastly, it was evidenced that the survey made is not exhaustive, but that serves like an advisor to the institutions that are moving to comply with the legislation in force.

Keywords: TDAR. Digital Preservation Standards. Documental Management.

◆ 1 INTRODUÇÃO

O tema preservação digital há muito é objeto de estudo em diversas áreas que se preocupam com a manutenção da memória humana produzida e armazenada nos meios digitais. Especialmente no Poder Judiciário, essa preocupação tem tomado cada vez mais espaço nas discussões estratégicas, principalmente a partir de meados dos anos 2010, quando a produção de documentos jurídicos nato-digitais e a digitalização daqueles mantidos até então em formato físico se intensificou. Esse processo de intensificação nas discussões sobre o tema se justifica pela adoção de sistemas produtores de documentos nato-digitais e pela necessidade de garantir a integridade, a autenticidade e o acesso a longo prazo a documentos e processos judiciais mantidos em meio digital.

Uma das iniciativas pioneiras foi conduzida pelo Tribunal de Justiça do Distrito Federal e dos Territórios (TJDFT), em 2018, ao estabelecer a *Política de Gestão Documental de Processos e Documentos em Meio Digital* daquela instituição, por meio da Resolução TJDFT 2/2018. Nessa norma, nota-se, além da preocupação com os aspectos de produção, tramitação e migração de suporte, a demanda pela adoção do modelo de repositório arquivístico digital confiável (RDC-Arq) (TJDFT, 2018). O movimento do TJDFT em direção à adoção do RDC-Arq está alinhado com os diversos normativos do Conselho Nacional de Arquivos (CONARQ). Esta tendência pela adoção do RDC-Arq tem sido fortemente defendida tanto nacional quanto internacionalmente.

O conceito da CCDA, segundo Flores *et al.* (2016), pressupõe uma linha ininterrupta desde a criação do documento até sua destinação final, com garantia de acesso permanente, autêntico e confiável. Em outras palavras, caberia ao ambiente de disseminação que compõe o RDC-ARQ garantir à sociedade futura o acesso integral aos objetos digitais que estão sob a guarda da sociedade atual, o que só é possível mediante a certificação de que estes objetos não foram modificados após o seu envio para preservação digital.

Outro aspecto do normativo do TJDFT é a previsão de que a adoção da solução de RDC-Arq deverá ser desenvolvida com base em *software* livre, gratuito e de código aberto, além de estar em conformidade com normas e modelos de preservação consolidados e referenciados em normas e orientações do CONARQ (TJDFT, 2018). Tal medida vem ao encontro da necessidade de ser possível validar e verificar todo o processo de preservação, garantindo a rastreabilidade documental a fim de promover maior segurança ao verificar sua autenticidade. Um outro aspecto colocado pela resolução do TJDFT é a previsão de integração da solução do RDC-Arq com os sistemas produtores de documentos arquivísticos (TJDFT, 2018). Estes deverão estar capacitados a proceder com o tratamento arquivístico dos documentos, sem que haja interrupção da CCDA. Preferencialmente, entende-se que esse processo deva ser automatizado, diminuindo a possibilidade de falhas humanas e habilitando o rastreamento de todas as ações executadas durante a implementação do RDC-Arq.

A norma implementada pelo TJDFT é destacada neste artigo, pois, das ações decorrentes da política estabelecida, teve origem o primeiro projeto de implantação integrada do RDC-Arq que se tem notícia no Poder Judiciário brasileiro, em uma parceria entre o TJDFT e o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) (TJDFT, 2021).

Outra norma pioneira que merece destaque no âmbito do Poder Judiciário é a *Resolução N.º 11/2019*, publicada pelo Superior Tribunal de Justiça (STJ), que, de forma complementar à norma do TJDFT, abordou aspectos técnicos referentes aos metadados descritivos, além da previsão da elaboração de um Plano de Preservação de Documentos Digitais (STJ, 2019).

O surgimento de legislações e normas que abordam o tema da preservação digital arquivística (e, em especial, os RDC-Arq) colocam em evidência o problema de pesquisa que se pretende responder neste estudo: qual o normativo a ser considerado em projetos que busquem implementar o RDC-Arq? A abordagem metodológica adotada para responder a este problema foi exploratória e descritiva. Os autores fizeram um mapeamento das legislações e normas que abordavam ou tinham relação com a temática do RDC-Arq e mapearam seus principais aspectos e tipos a fim de oferecer uma coletânea abrangente capaz de suportar as ações e projetos que envolvam a construção e implementação de RDC-Arq. O ponto de partida do estudo foi o projeto de pesquisa *Estudo para Implantação de Repositório Arquivístico Digital Confiável (RDC-Arq) no âmbito do Tribunal de Justiça do Distrito Federal e Territórios*, uma parceria do TJDF e do Ibict para promoção e aprofundamento da temática e que resultou na estruturação do modelo Hipátia de preservação digital.

♦ 2 SISTEMATIZAÇÃO DAS NORMAS E LEGISLAÇÕES APLICADAS AO TEMA

Após a identificação dos normativos afetos ao tema de preservação digital arquivística e RDC-Arq, foi realizada uma sistematização de todos os documentos analisados neste estudo. O **Quadro 1** a seguir, apresenta a listagem de documentos, uma descrição sucinta e o *link* para *download*, quando aplicável, do tipo de documento e o ano de sua publicação.

Quadro 1 – Legislações aplicáveis ao contexto do RDCArq

Título	Descrição	Tipo	Ano
PRONOM	É um sistema de informações <i>on-line</i> sobre formatos de arquivos de dados e seus produtos de <i>software</i> de suporte. Originalmente desenvolvido para apoiar a adesão e preservação a longo prazo dos registros eletrônicos mantidos pelo Arquivo Nacional, o PRONOM agora está sendo disponibilizado como um recurso para qualquer pessoa que necessite de acesso a esse tipo de informação. (https://www.nationalarchives.gov.uk/PRONOM/Default.aspx#)	Sistema	Primeira versão em 2002, desenvolvimento contínuo desde então.
Relatório Trusted Digital Repositories: Attributes and Responsibilities (An RLG-OCLC Report)	Relatório que descreve os atributos e responsabilidades inerentes a um repositório digital confiável. (https://www.oclc.org/content/dam/research/activities/trustedrepositories.pdf)	Relatório	2002

<p>Resolução CONARQ N° 24</p>	<p>Estabelece diretrizes para a transferência e recolhimento de documentos arquivísticos digitais para instituições arquivísticas públicas. (https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes-do-conarq/resolucao-no-24-de-3-de-agosto-de-2006)</p>	<p>Norma legal</p>	<p>2006</p>
<p>Resolução CONARQ N° 37</p>	<p>Aprova as Diretrizes para a Presunção de Autenticidade de Documentos Arquivísticos Digitais (https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes-do-conarq/resolucao-no-37-de-19-de-dezembro-de-2012)</p>	<p>Norma legal</p>	<p>2012</p>
<p>Resolução CONARQ N° 39</p>	<p>Estabelece diretrizes para a implementação de repositórios arquivísticos digitais confiáveis para o arquivamento e manutenção de documentos arquivísticos digitais em suas fases corrente, intermediária e permanente, dos órgãos e entidades integrantes do Sistema Nacional de Arquivos - SINAR. [Redação dada pela Resolução nº 43 de 04 de setembro de 2015] (https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes-do-conarq/resolucao-no-39-de-29-de-abril-de-2014)</p>	<p>Norma legal</p>	<p>2014</p>
<p>Resolução CONARQ N° 43</p>	<p>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. (https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes-do-conarq/resolucao-no-43-de-04-de-setembro-de-2015)</p>	<p>Norma legal</p>	<p>2015</p>
<p>Portaria CONARQ N° 272</p>	<p>Dispõe sobre a aprovação de Planos de Destinação de Documentos pelo Arquivo Nacional. (https://www.gov.br/conarq/pt-br/legislacao-arquivistica/portarias-federais/portaria-no-272-de-9-de-novembro-de-2020)</p>	<p>Norma legal</p>	<p>2020</p>

Resolução CNJ N° 324	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. (artigos 34 e 40) (https://atos.cnj.jus.br/atos/detalhar/3376)	Norma legal	2020
Portaria CNJ N° 295	Institui o Manual de Gestão Documental do Poder Judiciário e o Manual de Gestão de Memória do Poder Judiciário. (https://atos.cnj.jus.br/atos/detalhar/3646)	Norma legal	2020
Resolução STJ/GP N.º 11	Institui a Política de Preservação Digital do Superior Tribunal de Justiça. (https://bdjur.stj.jus.br/jspui/bitstream/2011/131788/Res_11_2019_PRE.pdf)	Norma legal	2019
Resolução TJDF N.º 02	Estabelece a Política de Gestão Documental de Processos e Documentos em Meio Digital do TJDF (https://www.tjdft.jus.br/publicacoes/publicacoes-oficiais/resolucoes-do-pleno/2018/resolucao-2-de-27-03-2018)	Norma legal	2018
ISO 14721: 2012	ISO 14721: 2012 Space data and information transfer systems – Open archival information system (OAIS) – Reference model Sistemas espaciais de dados e informações - Modelo de referência para um sistema aberto de arquivamento de informação (SAAI) Publicado no Brasil pela ABNT NBR 15472:2021 (https://www.normas.com.br/visualizar/abnt-nbr-nm/26223/abnt-nbr15472-sistemas-espaciais-de-dados-e-informacoes-modelo-de-referencia-para-um-sistema-aberto-de-arquivamento-de-informacao-saai)	Norma ISO	2012

<p>ISO 16363:2012</p>	<p>ISO 16363:2012 Space data and information transfer systems – Audit and certification of trustworthy digital repositories</p> <p>Define e recomenda práticas para avaliar os repositórios digitais confiáveis. Pode ser utilizado como uma base para certificação.</p> <p>(https://www.iso.org/standard/56510.html)</p>	<p>Norma ISO</p>	<p>2012</p>
<p>ISO 16919: 2014</p>	<p>ISO 16919: 2014 Space data and information transfer systems – Requirements for bodies providing audit and certification of candidate trustworthy digital repositories</p> <p>Fornece as adições específicas requeridas para certificação e auditoria de sistemas candidatos a se tornarem repositórios digitais confiáveis.</p> <p>(https://www.iso.org/standard/57950.html)</p>	<p>Norma ISO</p>	<p>2014</p>
<p>ISAD (G)</p>	<p>Norma Geral Internacional de Descrição Arquivística. Estabelece diretrizes gerais para a preparação de descrições arquivísticas. Deve ser usada em conjunção com as normas nacionais existentes ou como base para a sua criação.</p> <p>(https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/isad_g_2001.pdf)</p>	<p>Norma arquivística</p>	<p>2000</p>
<p>NOBRADE</p>	<p>Norma brasileira de descrição arquivística. Consiste na adaptação das normas internacionais à realidade brasileira, incorporando preocupações que o Comitê de Normas de Descrição do Conselho Internacional de Arquivos (CDS/CIA) considerava importantes, porém, de foro nacional.</p> <p>(http://www.siga.arquivonacional.gov.br/images/publicacoes/nobrade.pdf)</p>	<p>Norma arquivística</p>	<p>2006</p>

OAI-PMH	<p>O Open Archives Initiative Protocol for Metadata Harvesting é um protocolo desenvolvido pela Open Archives Initiative que define um mecanismo para coleta de registros de metadados em repositórios.</p> <p>(https://www.openarchives.org/pmh/)</p>	Modelo	2002
METS	<p>Metadata Encoding & Transmission Standard é um XML Schema elaborado sob o patrocínio da Digital Library Federation e mantido pela Biblioteca do Congresso Norte Americano. Ele permite a criação de uma rica estrutura capaz de registrar não apenas os múltiplos tipos de metadados usados para descrever o acervo de uma biblioteca digital, como também os próprios objetos digitais, seja para gerenciá-los, seja para permitir o intercâmbio entre diferentes instituições.</p> <p>(http://www.loc.gov/standards/mets/)</p>	Modelo	2018
EAD	<p>Encoded Archival Description é um padrão para a codificação de instrumentos de descrição documental arquivísticos utilizando a linguagem XML mantida pela Biblioteca do Congresso e pela Society of American Archivists.</p> <p>(https://www.loc.gov/ead/)</p>	Modelo	2019
ePING	<p>Os Padrões de Interoperabilidade de Governo Eletrônico definem um conjunto mínimo de premissas, políticas e especificações técnicas que regulamentam a utilização da Tecnologia de Informação e Comunicação (TIC) na interoperabilidade de serviços de Governo Eletrônico, estabelecendo as condições de interação com os demais Poderes e esferas de governo e com a sociedade em geral. (http://eping.governoeletronico.gov.br/)</p>	Modelo	2018

PREMIS	O Dicionário de Dados PREMIS para Metadados de Preservação é o padrão internacional de metadados para apoiar a preservação de objetos digitais e garantir sua usabilidade em longo prazo. A implementação do padrão consiste no Dicionário de Dados, um esquema XML e documentação de apoio. (https://www.loc.gov/standards/premis/v3/)	Modelo	2015
MoReq-Jus	Modelo de Requisitos para Sistemas Informatizados de Gestão de Processos e Documentos do Poder Judiciário (https://atos.cnj.jus.br/atos/detalhar/78) Download em (https://www.cnj.jus.br/wp-content/uploads/2011/01/manualmoreq.pdf)	Modelo	2009
Manual de Gestão Documental do Poder Judiciário	Tem o objetivo de especificar as diretrizes, os princípios e as normas de Gestão Documental. Constitui material de consulta e de orientação para o planejamento, implementação e execução do tema nos diversos órgãos do Poder Judiciário. (https://www.cnj.jus.br/wp-content/uploads/2021/04/Manual_de_Gestao_Documental_v16-04-2021.pdf)	Modelo	2021
Manual de Gestão de Memória do Poder Judiciário	Tem o objetivo de especificar as diretrizes, os princípios e as normas de Gestão de Memória. Constitui material de consulta e de orientação para o planejamento, a implementação e a execução relacionados a ela nos órgãos do Poder Judiciário. (https://www.cnj.jus.br/wp-content/uploads/2021/02/Manual_de_Gestao_de_Memoria.pdf)	Modelo	2021

O **Quadro 1** apresenta os diversos documentos que devem ser considerados ao se propor um projeto de implementação de um RDC-Arq, em especial quando se trata de instituições do campo jurídico. Os documentos analisados foram divididos nos seguintes tipos: norma ISO, norma arquivística, norma legal, modelo, relatório e sistema. Enquanto as normas ISO, normas arquivísticas e modelos já vêm sendo trabalhadas há muito e se mantêm atualizadas, observa-se que nos últimos anos houve um grande movimento de definições de normas legais acerca do tema.

Desde 2018, foram lançadas cinco normas legais que fomentaram a discussão sobre os RDC-Arq no âmbito do judiciário. Os documentos foram publicados por instituições que direcionam (ou, pelo menos, influenciam) o sistema jurisdicional brasileiro. Iniciando com a publicação do TJDFT, em

2018, e do STJ, em 2019, bem como duas do Conselho Nacional de Arquivos (CNJ), no ano de 2020. A criação de tais normativos indicam que estas instituições perceberam a importância da preservação digital e a fragilidade das infraestruturas atuais quando se analisam as características propostas nos modelos mais atuais.

Outro dado importante obtido pela análise do quadro anterior diz respeito à quantidade de modelos existentes, principalmente em nível internacional, o que indica a preocupação em se estabelecer parâmetros que garantam a interoperabilidade da preservação digital. Destaca-se também a atuação do judiciário em estabelecer seus próprios modelos de gestão documental e de memória, o que direciona as ações deste tipo de instituição.

Por fim, observa-se que há um movimento contínuo de atualização e manutenção dos documentos levantados, o que indica que a comunidade preocupada com a preservação digital arquivística se mantém ativa e colaborativa.

◆ 3 O IMPACTO DOS NORMATIVOS NAS INSTITUIÇÕES NACIONAIS

Este movimento observado tanto em nível nacional quanto internacional de estabelecimento de normativas que tratam da temática da preservação digital arquivística fomentou uma reação das instituições, principalmente públicas, que começaram a buscar alternativas para implementar os requisitos estabelecidos. Dentre este normativo, sem dúvida, duas publicações provocaram uma maior reação da comunidade: a resolução CONARQ Nº 43/2015 e a resolução CNJ Nº 324/2020.

A resolução CONARQ Nº 43/2015 estabeleceu o conceito de repositórios digitais e repositórios arquivísticos digitais, bem como os requisitos para que estes possam ser considerados confiáveis (CONARQ, 2015). Além disso, esta resolução estabelece os requisitos conceituais para que uma solução seja considerada um RDC-Arq. O documento diz que “os requisitos estão organizados em três conjuntos: infraestrutura organizacional; gerenciamento do documento digital; e tecnologia, infraestrutura técnica e segurança” (CONARQ, 2015, p. 18). A iniciativa do CONARQ obteve um avanço ao estabelecer suas diretrizes baseadas em normas nacionais e internacionais (FLORES; PRADEBON; CÉ, 2020) e, com isso, definir de forma clara a diferença entre os tipos de repositórios. O grande desafio estrutural apresentado pela resolução Nº 43 do CONARQ é a necessidade de atuação conjunta das equipes de gestão documental e de tecnologia da informação, sem a qual não é possível uma implementação bem-sucedida do RDC-Arq.

No que diz respeito à resolução Nº 324/2020 do CNJ, há a indicação de que as instituições do poder judiciário deveriam se movimentar no sentido de suportar o *Programa Nacional de Gestão Documental e Memória do Poder Judiciário* - PRONAME (CNJ, 2020). Além disso, esta resolução faz a importante definição de que a preservação digital no âmbito do poder judiciário deverá ser implementada em ambiente RDC-Arq. Por fim, a resolução ainda estabelece prazos para as instituições do poder judiciário se adaptarem à demanda, o que funciona como um motivador para que estas instituições iniciem de forma acelerada a busca por alternativas que as permitam atender aos requisitos postos.

Assim como estas duas resoluções, outras normativas também trouxeram à tona a necessidade de adaptações na estrutura institucional das organizações, em especial do poder judiciário, e promoveram um movimento de valorização do processo de preservação digital.

◆ 4 CONSIDERAÇÕES FINAIS

O recorte atribuído a esta pesquisa exploratória e descritiva se deve ao projeto de pesquisa construído na parceria entre o TJDFT e o Ibict, mas entende-se que o levantamento poderá ser uma fonte rica de consulta para qualquer projeto de preservação digital. As principais referências legais e de padrões foram levantadas e sistematizadas de forma a orientar a realização de projetos de implementação de RDC-Arq.

Uma das observações importantes advindas da análise documental realizada diz respeito à percepção de que as normas apresentadas demandam, em geral, a integração das áreas de Gestão Documental e Tecnologia da Informação das instituições. Apenas com esforços conjuntos dessas duas áreas é possível viabilizar a implantação da tecnologia de repositório arquivístico digital confiável.

Por fim, ressalta-se que o levantamento, embora não seja exaustivo, traz as principais referências no contexto da preservação digital arquivística, em especial no que diz respeito à implementação de repositórios arquivísticos digitais confiáveis no âmbito das instituições do sistema judiciário.

◆ REFERÊNCIAS

BRASIL. SUPERIOR TRIBUNAL DE JUSTIÇA (STJ). **Resolução STJ/GP n. 11 de 25 de junho de 2019. 2019.** Disponível em: https://bdjur.stj.jus.br/jspui/bitstream/2011/131788/Res_11_2019_PRE.pdf. Acesso em: 2 out. 2021.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Resolução no 43, de 04 de setembro de 2015:** diretrizes para a implementação de repositórios arquivísticos digitais confiáveis - RDC-Arq. 2015. Disponível em: <https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes-do-conarq/resolucao-no-43-de-04-de-setembro-de-2015>. Acesso em: 4 ago. 2021.

CONSELHO NACIONAL DE JUSTIÇA (CNJ). **Resolução no 324 de 30 de junho de 2020.** 2020. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3376>. Acesso em: 4 ago. 2021.

DISTRITO FEDERAL (Brasil). TRIBUNAL DE JUSTIÇA DO DISTRITO FEDERAL E TERRITÓRIOS (TJDFT). **RDC-Arq.** Brasília: Tribunal de Justiça do Distrito Federal e dos Territórios, 2021.

DISTRITO FEDERAL (Brasil). TRIBUNAL DE JUSTIÇA DO DISTRITO FEDERAL E TERRITÓRIOS (TJDFT). **Resolução 2 de 27 de março de 2018. 2018.** Disponível em: <https://www.tjdft.jus.br/publicacoes/publicacoes-oficiais/resolucoes-do-pleno/2018/resolucao-2-de-27-03-2018>. Acesso em: 14 jul. 2021

FLORES, D.; PRADEBON, D. S.; CÉ, G. Análise do conhecimento teórico-metodológico da preservação digital sob a ótica da OAIS, SAAI, ISO 14721 e NBR 15472. **Brazilian Journal of Information Science: research trends**, Marília, v. 11, n. 4, p. 73, 25 set. 2020. Disponível em: <https://doi.org/10.36311/1981-1640.2017.v11n4.11.p73>. Acesso em: 8 ago. 2021.

FLORES, D.; ROCCO, B.; SANTOS, H. Cadeia de custódia para documentos arquivísticos digitais. **Acervo**, Rio de Janeiro, v. 29, n. 2, p. 117-132, 2016.

COMO CITAR:

PIGNATARO, Tatiana Canelhas; MENEZES, Cristiano; BRAGA, Tiago Emmanuel Nunes. Os principais normativos relacionados ao RDCArq. *In*: BRAGA, Tiago Emmanuel Nunes; MÁRDERO ARELLANO, Miguel Ángel. (org.). **Hipátia**: modelo de preservação para repositórios arquivísticos digitais confiáveis. Brasília: Ibict, 2022. p. 42-53. (Informação, Tecnologia e Inovação, v. 1). DOI 10.22477/9786589167501.cap3.

Capítulo 04

O MODELO HIPÁTIA: A PROPOSTA DO IBICT PARA A PRESERVAÇÃO DIGITAL ARQUIVÍSTICA

THE HIPÁTIA ROLE MODEL: THE IBICT PROPOSAL FOR ARCHIVAL DIGITAL PRESERVATION

Tiago Emmanuel Nunes Braga¹

¹ Dados do autor: Doutor em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, tiagobraga@ibict.br.

◆ 1 INTRODUÇÃO

A preservação digital é um tema extremamente relevante na sociedade atual. A capacidade de manter preservada a memória nacional, de instituições e de pessoas é fundamental para que possamos ser capazes de compreendermos a forma como a história de nossa sociedade foi construída. A demanda por estabelecer mecanismos capazes de cumprir esta meta ambiciosa de manutenção do conhecimento humano ao longo do tempo fez com que o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) contemplasse o assunto em seu Plano Diretor (PDU) válido entre os anos de 2006 e 2010 (Ibict, 2006) e o mantivesse presente na publicação de todos os planos publicados posteriormente.

Já no ano de 2006 foram estabelecidas diversas metas relacionadas à temática de preservação digital. Tal construção estava condizente com os esforços realizados pelo Instituto desde o ano de 2002, quando começou a promover o aprofundamento teórico sobre o tema. Um dos frutos advindos desse esforço do Ibict foi a definição de padrões e modelos a serem utilizados no processo de preservação digital, a exemplo do OAIS (MÁRDERO ARELLANO, 2004), que posteriormente, seriam adotados pela comunidade nacional que pesquisa o tema. Eventualmente as ações empreendidas pelo Ibict geraram outros resultados, como a criação, em 2013, da Rede Cariniana, responsável, dentre outras coisas, por organizar as ações que estavam sendo realizadas por diferentes instituições nacionais.

Há diversos tipos de objetos digitais passíveis de se aproveitarem das vantagens da preservação digital. Documentos científicos, páginas da internet, arquivos multimídia e documentos arquivísticos são alguns exemplos. O foco inicial do Ibict centrou-se na preservação de publicações científicas, mas, já em 2006, a preservação arquivística foi inserida como uma meta de seu PDU. A meta 2 do objetivo específico 1 estabelecia que o Instituto pretendia “elaborar e propor, até o final de 2006, recomendação e normas de preservação de arquivos digitais” (Ibict, 2006, p. 23).

A discussão sobre preservação digital arquivística tem avançado muito no Brasil. Há diversas discussões acerca da especificidade da preservação arquivística em relação à preservação de objetos digitais provenientes de outras estruturas. Santos (2012) sintetiza as principais características atribuídas a um documento arquivístico: fixidez, organicidade, naturalidade, unicidade, autenticidade e imparcialidade. Este autor argumentou que, dadas as características distintas dos documentos digitais arquivísticos, era necessário que a preocupação com a preservação fosse evidenciada durante todo o ciclo documental.

Um agente fundamental na promoção da discussão sobre preservação arquivística foi a publicação da Resolução nº 39/2014 do Conarq, atualizada posteriormente pela resolução nº 43/2015 (CONARQ, 2015). Neste documento há o estabelecimento de que “a preservação dos documentos arquivísticos digitais, nas fases corrente, intermediária e permanente, deve estar associada a um repositório digital confiável” (CONARQ, 2015, p. 4). A definição, por parte do Conselho Nacional de Arquivo (CONARQ), gerou reflexo nas discussões até então empreendidas em outras estruturas, como o Conselho Nacional de Justiça (CNJ), o Conselho Superior de Justiça do Trabalho (CSJT) e o próprio Arquivo Nacional (AN).

Tendo em vista a discussão pregressa empreendida pelo Ibict, resolveu-se estabelecer um modelo nacional capaz de implementar as diretrizes definidas para os Repositórios Arquivísticos Digitais Confiáveis (RDCArq). Este processo foi aprofundado por meio do projeto de pesquisa *Estudo para Implantação de Repositório Arquivístico Digital Confiável (RDC-Arq)*, no âmbito do Tribunal de

Justiça do Distrito Federal e Territórios. Este projeto foi realizado em parceria com o Tribunal de Justiça do Distrito Federal e Territórios (TJDFT) e culminou na formatação do Modelo Hipátia, a proposta do Ibict para a implementação de repositórios arquivísticos digitais confiáveis.

Este artigo é um estudo de caso que se propõe a apresentar o modelo Hipátia de preservação arquivística digital. O objetivo é estruturar, de forma sistêmica, as etapas para implementação do modelo, de forma que outras instituições possam adotá-lo e, com isso, otimizar a difícil tarefa de preservar o conhecimento humano produzido e armazenado de forma digital.

♦ 2 A CONSTRUÇÃO DO MODELO DE PRESERVAÇÃO ARQUIVÍSTICA DIGITAL HIPÁTIA

A opção por construir um modelo que compreenda o processo de preservação digital arquivística passou pela identificação da demanda proveniente de uma preocupação natural da sociedade por manter resguardada a sua memória. Tal movimento promoveu o aprofundamento dos estudos e pesquisas relacionadas, o que de certa forma, resultou na proposição de ações práticas por parte das instituições. Tendo sido constatada esta tendência, coube ao Ibict a iniciativa de procurar parceiros que estivessem dispostos a avançar na identificação e formatação dos fatores e requisitos basilares e conceituais dos RDCArq.

A promoção do modelo partiu do levantamento teórico, passou pela identificação e caracterização dos requisitos informacionais, computacionais e conceituais e culminou com a proposição de um modelo de implementação do RDCArq integrado ao sistema produtor documental, que pode ser implementado em cinco etapas: preparação arquivística; preparação computacional; extração e preparação dos objetos digitais; preservação; e disseminação. Estas etapas são detalhadas a seguir.

2.1 SUPORTE TEÓRICO

O suporte teórico para proposição de um modelo de preservação digital arquivística deve ser construído apoiado pelo aspecto conceitual e pelo arcabouço normativo. Tanto a área arquivística quanto a de preservação possuem produções teóricas nacionais e internacionais capazes de orientar quanto à dinâmica necessária para sistematização do modelo. Além disso, tanto as normativas trazidas pelo CONARQ quanto pelo CNJ servem para instruir quanto aos requisitos a serem seguidos na construção do modelo. Uma análise mais abrangente deste *corpus* conceitual foi realizada nos três primeiros capítulos deste livro e a síntese é apresentada a seguir.

2.2 PRÉ-REQUISITOS CONCEITUAIS

Foram mapeados uma série de requisitos conceituais a serem considerados durante o processo de concepção do modelo de preservação digital. Estes requisitos, identificados durante a fase de análise do referencial bibliográfico, caracterizam as principais condições a serem contempladas pelo modelo.

Considerando-se a diversidade de sistemas que se propõe a fazer a gestão de documentos em formato digital, estabeleceu-se que o modelo a ser proposto deveria possibilitar sua implementação em ambientes informacionais diversos, independente do conjunto de *software* e hardware utilizado pela instituição que almeja aplicá-lo. Esta opção por um modelo com características tão abrangentes,

embora demandem maior esforço no processo de concepção do modelo, traz também vantagens significativas ao definir como característica do modelo primário a interoperabilidade entre tecnologias e ambientes informacionais.

Um ambiente interoperável, ou seja, que se adapta às distintas configurações, acontece apenas se consideradas premissas abertas, sejam relacionadas ao *software* ou aos padrões adotados. Uma explicação detalhada dos aspectos relacionados às arquiteturas computacionais e informacionais do modelo de preservação arquivística digital concebido no âmbito do Ibict será apresentada no próximo capítulo. Por ora, é preciso apenas qualificar que tanto os padrões quanto as tecnologias consideradas para contemplar a proposta do modelo devem ser abertas e livres, de forma a permitir que nenhuma restrição financeira relacionada à aquisição de tecnologias ou padrões possa ser empecilho para a implementação da proposta.

Além dessa definição acerca dos padrões e tecnologias, outro importante requisito identificado durante a análise negocial dos aspectos a serem contemplados pelo modelo diz respeito à possibilidade de personalização do processo de preservação. Embora exista todo um normativo a ser considerado, bem como diversos guias e orientações de boas práticas, uma vez que se tem a intenção de permitir a aplicação do modelo para qualquer sistema informacional que faça a gestão de documentos digitais, há também a necessidade de se permitir que sejam feitas personalizações no processo de preservação a fim de que as peculiaridades de cada instituição sejam atendidas. Essa personalização é importante para garantir que aspectos basilares do projeto de preservação digital apoiados pelo modelo sejam contemplados e que garantam o acesso futuro ao conteúdo preservador.

Estes dois requisitos formam a espinha dorsal do modelo, mas outros aspectos também devem ser considerados, principalmente aqueles relativos às definições estabelecidas pelo CONARQ (2015). Flores, Rocco e Santos (2016) destacam que além dos aspectos conceituais, o processo de implementação de um RDCArq deve contemplar a garantia da preservação da cadeia de custódia. Para isso, é necessário entender que a implementação do modelo terá que possibilitar a atuação de outros atores além da instituição arquivística, como a de certificadores, de auditores e de curadores. Tais aspectos também foram analisados durante a concepção do modelo, que é apresentado na sequência deste artigo.

2.3 O MODELO HIPÁTIA

Tendo em vista as definições conceituais propostas para o modelo de preservação digital arquivística, foi definida a criação do Modelo Hipátia. O nome do modelo se refere à cientista que contribuiu com a estruturação da biblioteca de Alexandria (CUNHA *et al.*, 2014) e que, por muito tempo, foi considerada como a única mulher cientista da qual se tinha conhecimento. Hipátia não constituiu matrimônio e a ela é atribuída a declaração de que era “casada com a verdade” (GOMES, 2018), focando seu tempo e esforço no aprofundamento da Matemática, Física, Filosofia e Astronomia. Acredita-se que sua recusa em se converter ao cristianismo teria provocado seu assassinato (CUNHA *et al.*, 2014) e a conseqüente perda de grande parte de suas descobertas. Apesar de muito do seu legado ter sido considerado perdido por séculos, eventualmente algumas de suas contribuições foram resgatadas e seu papel revelado para a sociedade moderna. A história de Hipátia e de como seu legado chegou aos dias atuais serviu de inspiração para a construção do modelo proposto pelo Ibict (BRAGA, 2021).

A aplicação do modelo Hipátia de preservação digital é estruturada em cinco fases: preparação arquivística, preparação computacional, extração de objetos digitais, preservação, disseminação. Estas fases podem acontecer de forma síncrona, no caso da preparação arquivística e computacional, ou de forma sequencial, no caso das etapas de extração de objetos digitais, preservação e disseminação. Visando uma melhor compreensão do modelo, ele é apresentado na figura 1, de forma sequencial.

Figura 1 – As cinco etapas do modelo Hipátia



Fonte: elaboração própria

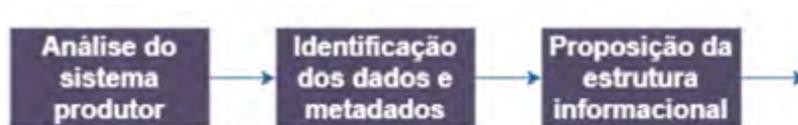
As fases de preparação arquivística e computacional compreendem o processo de identificação e análise do ambiente informacional e estrutural no qual a preservação acontecerá. Na fase de extração dos objetos digitais são sistematizados os conteúdos informacionais a serem preservados, tanto do ponto de vista da informação em si quanto dos metadados necessários para sua recuperação posterior. A fase de preservação é aquela em que os objetos são enviados para o repositório de guarda e, finalmente, a etapa de disseminação considera a disponibilização do conteúdo preservado para acesso por parte do público. Estas cinco fases são descritas mais detalhadamente nos tópicos a seguir.

2.3.1 Preparação arquivística

A preparação arquivística é a primeira fase da aplicação do modelo Hipátia e aquela que define as diretrizes a serem adotadas durante todo o processo. Nesta etapa é analisado o ambiente a ser preservado de forma que possam ser identificados todos os objetos digitais que farão parte dos pacotes a serem futuramente preservados. A etapa de preparação arquivística pode ser considerada a mais importante, tendo em mente o papel do arquivista no processo de preservação, como colocado por Fernal e Vechiato (2013). Logo, de suas definições dependem todas as outras etapas.

Dentre as ações a serem realizadas nesta etapa de aplicação do modelo destacam-se: definição dos dados a serem preservados, definição dos metadados que irão compor o pacote de preservação, proposição do dicionário de dados a ser utilizado no processo automatizado de preservação, definição da estrutura do objeto digital a ser preservado por meio do pacote OAIS, avaliação documental a partir da análise da tabela de temporalidade e dos critérios arquivísticos, e estudos que possam apoiar o restante da pesquisa. Estas ações da primeira etapa de implementação do modelo são resumidas em três grandes blocos, conforme mostra a figura 2.

Figura 2 – principais ações da etapa um: preparação arquivística



Fonte: elaboração própria

Ao final desta fase, os produtos gerados serão utilizados para possibilitar a extração dos objetos digitais do sistema produtor visando seu envio para o ambiente de preservação.

2.3.2 Preparação computacional

Assim como a preparação arquivística, a preparação computacional também se concentra nos requisitos a serem seguidos ao se implementar o modelo de preservação Hipátia. Ela pode acontecer concomitante com a preparação arquivística ou de forma sequencial. Mas, destaca-se que, embora a preparação computacional tenha como principal função estabelecer as diretrizes para a instalação do modelo no parque tecnológico da instituição que o está implantando, os estudos e as preparações realizadas neste momento devem também atender aos requisitos estabelecidos pela primeira etapa, da preparação arquivística.

Durante esta etapa são realizados os seguintes processos: identificação da estrutura lógica dos servidores, análise da arquitetura de rede adotada pela instituição, configuração das permissões de acesso para os sistemas e usuários, e instalação dos três sistemas computacionais adotados pelo modelo. A figura 3 apresenta as principais ações realizadas nesta fase.

Figura 3 – principais ações da etapa dois: preparação computacional



Fonte: elaboração própria

A preparação computacional se pauta pela mesma fonte que a preparação arquivística, ou seja, baseia-se na análise do sistema produtor de objetos informacionais digitais. Com os resultados dessa análise, é estabelecida uma proposta de extração dos dados e metadados que alimentarão o modelo de preservação digital. Além disso, a análise da infraestrutura da instituição fornecerá as informações necessárias para o planejamento de um ambiente de preservação que garanta a *performance* necessária ao mesmo tempo em que ratifica os critérios de segurança estabelecidos para a rede computacional. O resultado dessa análise é, em última instância, a instalação e configuração dos três sistemas computacionais que compõem o modelo Hipátia: BarraPres, Archivematica e AtoM.

Embora a espinha dorsal do processo de preservação seja composta por profissionais da informação, em especial o arquivista, a fase de preparação computacional depende essencialmente

do conhecimento da área da computação. Sendo assim, para que o sucesso do modelo seja alcançado, os profissionais dos dois campos devem atuar não apenas em parceria, mas de forma conjunta, complementando os saberes um do outro.

Os resultados desta fase, em conjunto com os resultados da fase de preparação arquivística, possibilitará que a próxima fase, de extração de objetos digitais, ocorra com sucesso.

2.3.3 Extração de objetos digitais

Nesta etapa é que são realizadas as conexões entre o ambiente de produção de objetos digitais e o ambiente de preservação. Sendo assim, esta etapa é a responsável por garantir a interoperabilidade do modelo Hipátia, tornando-se imprescindível sua personalização a cada novo projeto de implementação do RDCArq. Para possibilitar esta conexão entre os sistemas é possível utilizar algumas estratégias, a saber:

- a) utilização de APIs e *webservices*;
- b) conexão direta com o banco de dados;
- c) acesso à estrutura de arquivos do sistema operacional.

Há a recomendação de que, sempre que possível, sejam utilizados APIs e *webservices* para realizar a conexão entre os sistemas, mas, como o processo de preservação muitas vezes acontece com sistemas legados, ou, ainda, que não possuem manutenções evolutivas, eventualmente é preciso utilizar outros métodos para a conexão entre os sistemas. Em alguns casos, inclusive, pode-se propor um processo de conexão mista, em que mais de uma estratégia é utilizada simultaneamente a fim de extrair do sistema produtor todos os objetos informacionais previstos para preservação.

Um aspecto importante desta etapa é a criação do dicionário de dados de preservação a partir das propostas advindas das fases anteriores. Segundo a proposta PREMIS (2015), um dicionário de dados aplicado ao contexto da preservação deve suportar os objetos digitais a serem preservados de forma que garanta durante o processo: viabilidade, renderização, compreensibilidade, autenticidade e identidade. Ele também deve contemplar as informações que serão necessárias para realizar a preservação de longo prazo (PREMIS, 2015). Outra questão que deve ser incorporada pelo dicionário de dados, segundo o PREMIS, é a definição de metadados que possibilitem a automatização do processo de criação, gerenciamento e uso dos objetos digitais. Por fim, o dicionário de dados não deve fazer suposições tecnológicas, ou seja, deve garantir as informações necessárias para a preservação, independente da tecnologia a ser utilizada na próxima fase de preservação (PREMIS, 2015). Estes quatro grandes aspectos apresentados no relatório PREMIS devem ser considerados no processo de criação do dicionário de dados do modelo Hipátia. Uma vez que isso tenha sido feito, será possível garantir que o fluxo de preservação avance de forma fluida.

Para proceder com a extração dos objetos digitais, o modelo Hipátia prevê a utilização do pacote *BagIt*, um arquivo em formato de empacotamento que segue uma estrutura hierárquica para armazenamento e transmissão de conteúdo digital (KUNZE *et al.*, 2018). Esta proposta de empacotamento do conteúdo digital possui dois conceitos-chave: forte garantia de segurança e acesso direto aos arquivos.

Na prática, a garantia de segurança quer dizer que é possível aplicar ao pacote *BagIt* criptografia por meio de algoritmos Hash, o que permite que o mesmo receba melhoria em sua segurança sem comprometer a compatibilidade com as versões anteriores. Este aspecto de segurança do pacote garante que o conteúdo não será corrompido, mas não é projetado para ser uma ferramenta contra ataques feitos de forma ativa (KUNZE *et al.*, 2018), o que deve ser tratado por outras estratégias do modelo de preservação. Já o acesso direto foi pensado para permitir acesso ao conteúdo de maneira hierárquica, conforme mostra a figura 4. Dessa forma, é possível acessar os objetos digitais e seus metadados de maneira integral ou, caso não exista o interesse em interpretar todo o pacote, de forma parcial (KUNZE *et al.*, 2018). A implementação do pacote *BagIt* aumenta as possibilidades do Hipátia ao permitir a integração de diversos sistemas de preservação do modelo. Ele também garante a qualidade da preservação e do envio de dados entre os sistemas sem que a cadeia de custódia dos objetos digitais seja comprometida.

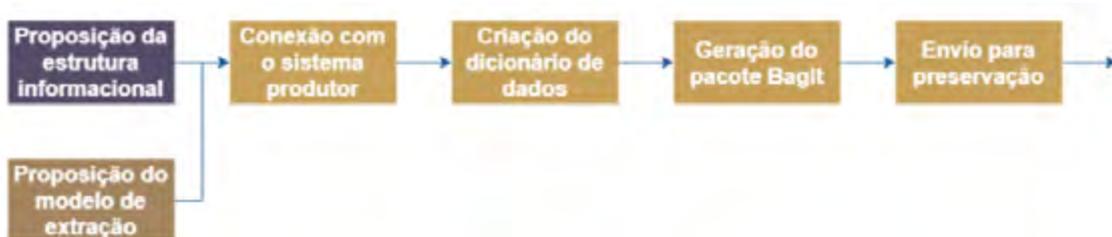
Figura 4 - estrutura hierárquica do pacote *BagIt*

```
<base directory>/
|
+-- bagit.txt
|
+-- manifest-<algorithm>.txt
|
+-- [additional tag files]
|
+-- data/
|   |
|   +-- [payload files]
|
+-- [tag directories]/
|   |
|   +-- [tag files]
```

Fonte: Kunze *et al.*, 2018

A etapa de extração de objetos digitais pode ser inicializada de três formas: em intervalos de tempo, sob demanda ou por eventos. Estas três configurações permitem que a instituição que está implementando o modelo Hipátia tenha autonomia sobre o início do processo de preservação. No formato intervalo de tempo, também conhecido como método passivo, é realizada uma configuração no BarraPres para que ele faça o processo de extração dos objetos digitais em intervalos de tempo predeterminados. Neste caso, após o acesso ao sistema produtor por parte do barramento, há a extração dos dados e metadados, a geração do pacote e o envio para preservação sem que seja necessária nenhuma intervenção humana. No formato sob demanda, o barramento BarraPres é acionado ativamente por uma pessoa ou sistema. Um parâmetro informando qual conjunto de dados e metadados deve ser preservado é enviado ao barramento, que inicia as ações previstas para a etapa. Por fim, no formato por eventos, o sistema produtor dispara automaticamente uma chamada sempre que determinados requisitos são alcançados. Esta chamada é armazenada em um controlador de eventos que está sendo monitorado pelo BarraPres. Quando a chamada é identificada, mais uma vez, são iniciadas as ações da etapa. Um resumo de todas as ações desencadeadas pela etapa de extração de objetos digitais é apresentado na figura 5.

Figura 5 – principais ações da etapa três: extração de objetos digitais



Fonte: elaboração própria

Este conjunto de ações, provenientes do resultado das duas etapas anteriores, culminam com a geração do pacote *BagIt* e finalizam a interação com o sistema produtor. A última ação prevista para a etapa de extração dos objetos digitais é o envio do pacote *BagIt* para a etapa de preservação, quando se inicia o processo de armazenamento. Nesta etapa também é gerado um *log* de controle a fim de verificar os passos realizados e, eventualmente, é possível configurar uma mensagem de status para que o sistema produtor de objetos digitais apresente um retorno a seus usuários.

2.3.4 Preservação

Na etapa de preservação é que os objetos digitais são, de fato, preservados em um repositório específico para esta função. A fim de obter sucesso nesta etapa, o modelo Hipátia propõe, atualmente, a utilização do *software* Archivematica como ambiente de preservação. Uma das vantagens dessa solução é sua arquitetura robusta e comunidade ativa, o que tende a possibilitar o aprimoramento da ferramenta ao longo do tempo. O Archivematica implementa o modelo OAIS de preservação com seus três subtipos de pacotes de informação: *Submission Information Package* (SIP), *Archival Information Package* (AIP) e *Dissemination Information Package* (DIP). Este modelo de empacotamento é entendido como o melhor para o processo de preservação de dados e informações, tendo em vista suas diversas possibilidades (ARELLANO, 2004). A figura 6 exemplifica as entidades funcionais do modelo OAIS.

Figura 6 – Modelo OAIS: entidades funcionais



Fonte: ISO, 2012

Embora o esquema acima apresente um fluxo relativamente simples entre os três subtipos de pacote, a ISO 14721:2012 (2012) estabelece relações complexas entre eles. Essas relações definem a estrutura dos subtipos de pacote da seguinte forma:

- d) Um SIP pode gerar um AIP;
- e) Um SIP pode gerar muitos AIP;
- f) Diversos SIP podem gerar um AIP;
- g) Muitos SIP podem gerar muitos AIP;
- h) Um DIP pode conter parte de um AIP;
- i) Um DIP pode conter todo o conteúdo de um AIP.

Enquanto o pacote de transferência inicial é submetido pelo produtor, ou seja, neste caso, pelo BarraPres, o SIP e o AIP são gerados internamente pelo Archivematica, assim como o DIP. Neste último caso, visando a disseminação dos objetos digitais preservados, o pacote DIP é disponibilizado por meio de um sistema consumidor. A geração destes pacotes de informação ocorre por meio dos micros serviços executados pelo *software* Archivematica. As principais ações realizadas na etapa de preservação são sintetizadas na figura 7 a seguir.

Figura 7 - principais ações da etapa quatro: preservação



Fonte: elaboração própria

Ao final da etapa de preservação, aqueles objetos digitais que podem ser acessados pelo usuário são disponibilizados para a etapa de disseminação de forma automatizada a partir dos parâmetros estabelecidos na etapa de preparação arquivística.

2.3.5 Disseminação

Por fim, o modelo Hipátia de implementação do RDCArq prevê a disseminação, para o público, dos objetos digitais preservados. Deve-se fazer uma análise dos objetos digitais à luz da política institucional e da legislação vigente, contemplando, inclusive, a *Lei Geral de Proteção de Dados* (LGPD). Após esta análise, aqueles objetos digitais destacados para disseminação serão

enviados automaticamente do ambiente de preservação para o ambiente de acesso, quando são disponibilizados para o público. Existe apenas uma grande ação prevista para esta etapa do modelo, como pode ser observado na figura 8 a seguir.

Figura 8 – principais ações da etapa cinco: disseminação



Fonte: elaboração própria

A indicação atualmente adotada pelo modelo Hipátia é pela utilização do sistema AtoM, que é um sistema robusto de descrição arquivística pautada em normas internacionais e que possui ambientes e configurações condizentes com o RDCArq.

2.4 A IMPLEMENTAÇÃO DO MODELO HIPÁTIA

Estas etapas propostas para o modelo Hipátia de preservação arquivística digital confiável foram sistematizadas e aprimoradas a partir da experiência de implementação do modelo em instituições parceiras. Todavia, elas não são estáticas, ou seja, o modelo ainda está em evolução, tendo em vista a diversidade de ambientes nos quais ele se propõe a ser aplicável. Uma visão sistêmica do modelo e suas principais ações é apresentada na figura 9 a seguir.

Figura 9 – visão geral das ações propostas pelo modelo Hipátia



Fonte: elaboração própria

Embora já esteja sistematizado, a implementação do modelo implica em uma atuação conjunta de diversas áreas da instituição, principalmente aquelas responsáveis pela gestão documental e pela gestão dos ativos computacionais. Estas equipes, ao analisarem o sistema produtor e a infraestrutura, poderão identificar ações estruturantes não previstas inicialmente no modelo, o que é perfeitamente compreensível, dadas as diferentes realidades institucionais existentes. Neste caso, a recomendação é que os profissionais da informação, principalmente os arquivistas e bibliotecários, se pautem pelos requisitos informacionais para adicionar as novas ações e, com isso, adaptar o modelo à realidade identificada pelas equipes.

◆ 3 CONSIDERAÇÕES FINAIS

Uma das conclusões possíveis acerca do modelo Hipátia é que, para sua implementação, torna-se indispensável o apoio da alta direção, tendo em vista que o RDCArq contempla mudanças significativas na cultura informacional da instituição. Além disso, destaca-se que a construção do modelo se deu por meio do estabelecimento de parcerias entre o Ibict e outras organizações que visavam a promoção da pesquisa científica acerca do tema. A primeira parceria estabelecida foi com o TJDFT e, posteriormente, com outras instituições representativas do tema no cenário governamental, como o AN. Disso, resulta que o modelo Hipátia é fruto de uma construção coletiva entre instituições públicas, o que corrobora com sua disponibilização em formato aberto, bem como a opção adotada no modelo de fomentar apenas tecnologias desenvolvidas com *software* livres.

Como dito anteriormente, o modelo Hipátia ainda está em desenvolvimento. Diversos aprimoramentos já foram identificados e estão apresentados como desafios a serem superados pelas instituições que decidirem adotá-lo. Dentre estes recursos, é possível listar a necessidade de otimização do processo de adaptação do dicionário de dados para novos sistemas produtores, a utilização de tecnologias computacionais emergentes, como os *containers*, e a estruturação do código fonte para facilitar a contribuição da comunidade.

Embora ainda não esteja finalizado, o modelo já possui maturidade suficiente para ser utilizado por instituições que possuem a necessidade de implementar o RDCArq. A experiência de implementação no TJDFT e no AN mostrou que é possível obter retornos significativos com o modelo, atendendo à legislação vigente e garantindo um acesso de qualidade para as gerações futuras aos objetos informacionais digitais produzidos no momento presente.

Na década de 1980, a UNESCO já entendia que a informação era uma commodity e deveria ser utilizada para gerar riqueza e desenvolvimento à nação (1981). Dessa forma, entende-se que o modelo Hipátia é uma contribuição do Ibict para a definição de estruturas informacionais que visem o desenvolvimento da sociedade e o avanço no entendimento da preservação da informação enquanto garantia de guarda da riqueza informacional produzida no país.

◆ REFERÊNCIAS

BRAGA, T. E. N. Hipátia: a proposta do Ibict para o RDC-Arq. *In*: SEMINÁRIO INTERNACIONAL DE PRESERVAÇÃO DIGITAL, 5., Brasília, 2021. **Anais [...]** Brasília: Ibict, 2021. Disponível em: <http://eventoscariniana.ibict.br/index.php/sinpred/article/view/71> Acesso em: 4 jul. 2022.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Resolução no 43, de 04 de setembro de 2015**. 2015. Disponível em: https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/conarq_diretrizes_rdc_arq_resolucao_43.pdf. Acesso em: 11 jan. 2022.

CUNHA, M. B.; RITTER PERES, O. M.; GIORDAN, M.; BERTOLDO, R. R.; QUADROS MARQUES, G. de; DUNCKE, A. C. As mulheres na ciência: o interesse das estudantes brasileiras pela carreira científica. **Educación Química**, [s.l.], v. 25, n. 4, p. 407–417, 1 out. 2014. Disponível em: [https://doi.org/10.1016/S0187-893X\(14\)70060-6](https://doi.org/10.1016/S0187-893X(14)70060-6). Acesso em: 25 ago. 2021.

FERNAL, A.; VECHIATO, F. L. Repositórios digitais como ambientes de atuação do arquivista: um estudo dos princípios arquivísticos e da preservação digital nesse contexto. **Informação@Profissões**, v. 2, n. 1, p. 103–122, 2013. Disponível em: <https://doi.org/10.5433/2317-4390.2013v2n1p103>. Acesso em: 15 ago. 2021.

FLORES, D.; ROCCO, B.; SANTOS, H.. Cadeia de custódia para documentos arquivísticos digitais. **Acervo**, [s.l.], v. 29, n. 2, p. 117–132, 2016.

GOMES, V. de S. A Vida de Hipátia de Alexandria. **Mulheres na Matemática**. 2018. Disponível em: <http://mulheresnamatematica.sites.uff.br/hipatia-de-alexandria/>

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (Ibict). **Plano Diretor**: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2006 - 2010. 2006. Disponível em: http://sitehistorico.ibict.br/sobre-o-ibict/regimento-interno-1/copy_of_2010.05.2014.44xPDUIBICT.pdf

INTERNATIONAL STANDARD ORGANIZATION (ISO). **ISO 14721:2012**. Space data and information transfer systems: Open archival information system (OAIS) – Reference model. 2012. .

KUNZE, J. A.; LITTMAN, J.; MADDEN, L.; SCANCELLA, J.; ADAMS, C.. The *BagIt* File Packaging Format (V1.0). **Request for Comments**, n. RFC 8493. [S. l.]: Internet Engineering Task Force, out. 2018. DOI 10.17487/RFC8493. Disponível em: <https://datatracker.ietf.org/doc/rfc8493>. Acesso em: 25 jan. 2022.

MÁRDERO ARELLANO, M. Á. Preservação de Documentos Digitais. **Ciência da Informação**, v. 33, n. 2, 10 dez. 2004. DOI 10.18225/ci.inf..v33i2.1043. Disponível em: <http://revista.ibict.br/ciinf/article/view/1043>. Acesso em: 11 jan. 2022.

PREMIS, Editorial Committee. **Data dictionary for preservation metadata**: PREMIS version 3.0. [S. l.]: Technical report, Mar, 2015.

UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION (UNESCO). **The small and medium industrial enterprises and technological information services**: concepts, insights and experiences. 1981. Disponível em: <http://unesdoc.unesco.org/images/0004/000460/046079EB.pdf>. Acesso em: 9 jul. 2016.

COMO CITAR:

BRAGA, Tiago Emmanuel Nunes. O modelo Hipátia: a proposta do Ibict para a preservação digital arquivística. *In*: BRAGA, Tiago Emmanuel Nunes; MÁRDERO ARELLANO, Miguel Ángel. (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.



Capítulo 05

ARQUITETURA DO BARRAPRES/HIPÁTIA

THE BARRAPRES/HIPÁTIA MODEL
ARCHITECTURE

Milton Shintaku¹

Alexandre Faria de Oliveira²

¹ Dados do autor: Doutor em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, shintaku@ibict.br

² Dados do autor: Mestrando em Computação, Instituto Brasileiro de Informação em Ciência e Tecnologia, alexandreoliveira@ibict.br

RESUMO

Resumo: A preservação sempre esteve presente na história da humanidade. Com a evolução da tecnologia e o surgimento da informática, surge um novo mundo, no qual documentos estão em formato digital, com toda sua tipologia e possibilidades. Nesse sentido, este trabalho tem por objetivo apresentar a arquitetura do BarraPres, *software* desenvolvido pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), e que compõe o modelo Hipátia, construído por meio de um projeto de pesquisa em parceria com o Tribunal de Justiça do Distrito Federal e Territórios (TJDFT). Este barramento permite que sistemas heterogêneos possam interagir entre si, compondo assim, o ecossistema de preservação Hipátia. A partir da descrição sobre a estrutura e a interoperabilidade do barramento BarraPres/Hipátia, observa-se a eficácia do barramento em relação à integração de vários sistemas, sendo uma opção flexível e robusta para atender às normas e orientações vigentes. Logo, o barramento se transforma em um elemento importante de interface integradora entre sistemas de negócio, como o Sistema Eletrônico de Informações (SEI), Diário da Justiça Eletrônico (DJe) e outros sistemas comumente utilizados na preservação digital, como o Archivematica. A integração de todos esses sistemas é um componente importante do modelo Hipátia.

Palavras-Chave: Preservação digital. BarraPres. Modelo Hipátia. Ecossistemas de informação.

Abstract: Preservation has always been present in the history of mankind. With the evolution of technology and the emergence of information technology, a new world emerges, in which documents are in digital format, with all its typology and possibilities. In this sense, this work aims to present the architecture of the BarraPres model – software developed by Ibict, and which composes the Hipátia Model, built through a research project in partnership with TJDFT. This software allows heterogeneous systems to interact with each other, thus forming a preservation ecosystem. From the description of the structure and interoperability of BarraPres/Hipátia software, it is observed the effectiveness of the software in relation to the integration of various systems, being a flexible and robust option when meeting the current standards and guidelines. Therefore, the software becomes an important element of an integrative interface between business systems, such as SEI, DJe and other systems commonly used in digital preservation, such as Archivematica. The integration of all of these systems is an important component of the Hipátia Model.

Keywords: Digital Preservation. BarraPres. Hipátia Model. Information ecosystems.

◆ 1 INTRODUÇÃO

A preservação sempre esteve presente na história da humanidade. Se nas sociedades primitivas os anciãos eram respeitados por acumular informações, tendo no repasse oral a preservação da cultura, com a evolução, o surgimento da escrita catapultou a preservação a outro nível, no qual, mesmo depois de séculos, tem-se acesso a documentos escritos em diversos suportes, como papiros, tábuas de argila, pergaminhos e papel.

O surgimento da prensa móvel, por sua vez, possibilitou que o conhecimento fosse disseminado de forma mais democrática, e, de certa forma, ajudou a preservá-lo, visto que os livros puderam ser distribuídos em quantidades maiores, e as cópias passaram a compor os acervos de diversas bibliotecas, inclusive as particulares. Com isso, bibliotecas, arquivos e museus passaram a ter a missão de preservar documentos, em grande parte, impressos.

Com a evolução da tecnologia e o surgimento da informática, surge um novo mundo, no qual documentos estão em formato digital, com toda sua tipologia e possibilidades. Com isso, tem-se um novo desafio na preservação desta documentação, que possui características próprias, diferentemente do suporte físico. Nesse sentido, se no mundo físico a preservação é física, no mundo digital a preservação deve ocorrer nos ambientes digital e físico.

Nesse contexto, urge a necessidade de atuar na preservação digital de conteúdos criados, tramitados e armazenados em sistemas informatizados, não sendo uma preocupação recente, visto que Kimley (1995) relata sobre as preocupações a respeito das tecnologias para preservação digital nos anos 90 do século passado. Naquela época, a preservação digital estava sendo definida pela Biblioteca Digital da Califórnia como “as atividades gerenciais necessárias para assegurar a retenção e usabilidade a longo prazo de objetos digitais”, conforme relata Schmidt, Gehring e Nicholson (1995).

Desde então, a preocupação com a preservação digital tem promovido estudos, tanto que Patra, Sahoo e Mohanty (2021), ao avaliarem as publicações sobre o tema, verificaram que a tecnologia mais citada para essa atividade foi o Lot of Copies Keep Stuff Safe (LOCKSS), projeto mantido pela Universidade de Stanford e utilizado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) em sua rede de preservação, denominada Rede Cariniana. Esta iniciativa tem por objetivo “salvaguardar os registros da ciência, tecnologia e do patrimônio cultural do Brasil” (MÁRDERO ARELLANO, 2012, p. 87).

Assim, a Rede Cariniana é uma alternativa para organizações que desejam colecionar, armazenar, preservar e oferecer acesso a seus acervos mais importantes em cópias digitais autorizadas no Brasil. Para a interligação dos diversos sistemas informacionais de preservação digital, identificou-se a necessidade de desenvolver um barramento tecnológico de interoperabilidade, de forma a possibilitar que diferentes programas possam compor um ecossistema de preservação digital de documentos arquivísticos, denominado como Repositório Arquivístico Digital Confiável (RDC-Arq).

Desta forma, o Ibict desenvolveu, em 2018/2019, por meio de um projeto de pesquisa em parceria com o TJDF, o barramento BarraPres, que viria a ser parte do Modelo Hipátia, voltado a ser uma interface que permite que sistemas heterogêneos possam interagir entre si, formando assim um ecossistema de preservação e alcançando os requisitos do RDC-Arq.

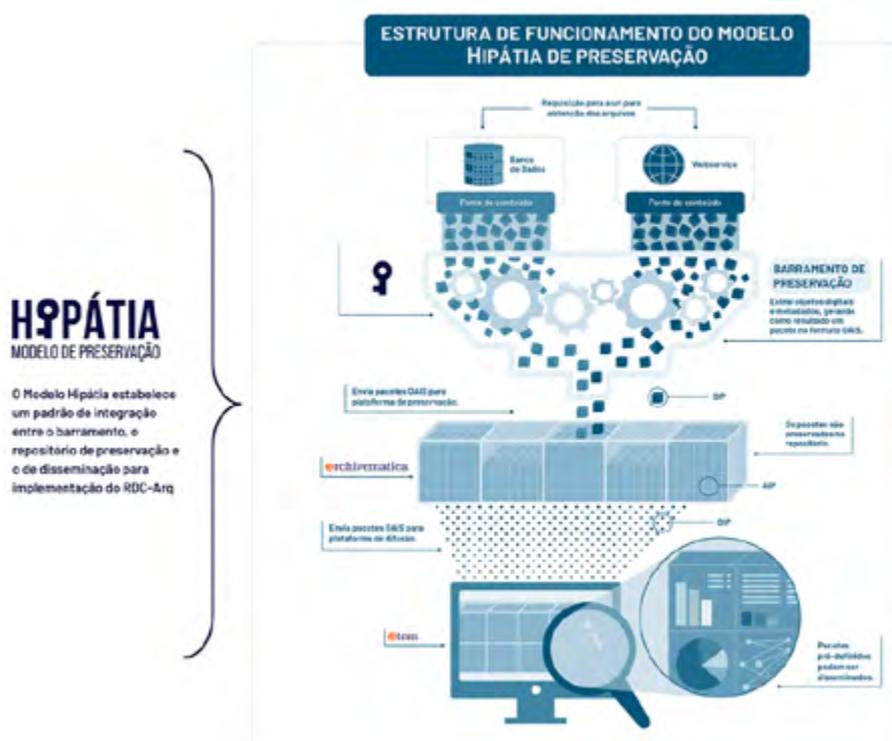
2 BARRAPRES/HIPÁTIA

O nome Hipátia se refere à filósofa Hipátia da biblioteca de Alexandria, do período romano do Egito, reconhecida como a primeira mulher a ser documentada como matemática. Alic (2005) relata que, apesar de terem existido outras mulheres cientistas no mundo antigo, poucas tiveram registros de suas vidas, como Hipátia. Mesmo que muitos dos seus escritos tenham se perdido, há inúmeras referências aos seus trabalhos em matemática e astronomia.

Devido a importância da filósofa Hipátia nas ciências, o Ibict, por meio da sua Coordenação-Geral de Tecnologia da Informação e Informática (CGTI), nomeou o modelo de preservação digital arquivística com o nome desta cientista do mundo antigo. Faz parte desse modelo o BarraPres, um barramento capaz de extrair, aplicar estruturas informacionais e empacotar pacotes de dados. Dessa forma, o BarraPres/Hipátia é um *software* livre, de código aberto, desenvolvido e mantido pelo Ibict, e voltado a integrar sistemas seguindo o modelo Open Archival Information System (OAIS) para preservação.

Conforme a estrutura apresentada na Figura 1, o BarraPres/Hipátia ou somente BarraPres, atua como uma camada de integração entre os sistemas gestores (produtores) de conteúdos e um sistema de preservação. Assim, o BarraPres recebe os conteúdos oriundos dos sistemas fontes, convertendo os conteúdos para pacotes em formato do sistema de preservação no padrão *BagIt*, contendo metadados de conservação baseados na política de preservação definida. O *BagIt* é o formato padronizado de empacotamento utilizado pelo Archivematica. Ele foi criado pela Biblioteca do Congresso Americano sob licença pública e organiza de forma hierárquica os documentos digitais junto aos seus metadados, possibilitando a criação de um pacote de Submissão de Informação, na sigla inglesa de Submission Information Package (SIP), que é enviado para um repositório de preservação, como o Archivematica.

Figura 1 – BarraPres no âmbito do modelo Hipátia



Fonte: Ibict (2022).

O Archivematica recebe esses pacotes compactados, que, no contexto do Hipátia, nomeamos como “Pacotes de Transferência de Informação (Initial Transfer Package – ITP)”. Este pacote é internamente recebido pelo Archivematica que gera um pacote de submissão (Submission Information Package – SIP) e depois realiza automaticamente a migração ou conversão de formatos dos documentos digitais armazenados para preservação e executa todos os micros serviços existentes e pré-configurados pelos curadores do sistema, gerando o Pacote de Informação de Arquivo (AIP). Neste momento, é atribuído ao pacote uma codificação única persistente que garante a integridade e autenticidade do conteúdo armazenado. Após esse processo, o Pacote de Informação para Disseminação (Dissemination Information Package – DIP) pode ser criado, sendo derivado de um ou mais pacotes AIP, contendo uma derivada de acesso dos originais, o que garante a divulgação da informação por meio do envio direto para uma plataforma de acesso. Este poderia ser o caso da plataforma AtoM.

O BarraPres pode integrar vários tipos de sistemas fontes, precisando de configurações e adaptações para cada tipo de sistema. Dessa forma, para os sistemas de entrada, funciona como uma plataforma, na qual adicionam-se configurações de código para atender a sistemas de informação diferentes. Com isso, apresenta flexibilidade necessária para auxiliar as organizações que mantêm vários sistemas de informação e precisam preservar seus dados.

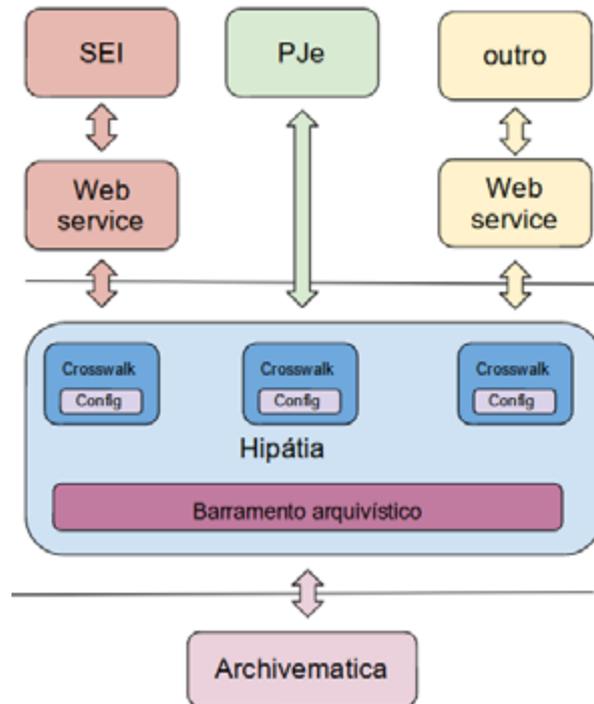
Quanto aos sistemas de preservação, no qual os pacotes devem ser depositados, o mais comum é o Archivematica, que se configura como o repositório digital confiável, conforme as orientações arquivísticas. Assim, o Hipátia converte as informações recolhidas nos sistemas de entrada e as transforma em pacotes *BagIts* a serem depositados no Archivematica. Finalmente, é possível ter um ambiente de acesso acoplado, papel a ser desempenhado pelo AtoM.

◆ 3 ESTRUTURA DO BARRAPRES/HIPÁTIA

O propósito do BarraPres é possibilitar a integração de sistemas que gerenciam documentos digitais, compondo um RDC-Arq. Para tanto, conta com a orientação conceitual de modelos de referência, como o OAIS, e normas brasileiras propostas pelo Conselho Nacional de Arquivo (CONARQ). Com isso, atende aos requisitos de manutenção da cadeia de custódia.

A estrutura do BarraPres é simples, como mostra a **Figura 2**. Ela pode ser dividida em três partes, sendo uma para se conectar a um *Web Service* disponibilizado pelo sistema fonte ou a conexão com o banco de dados, um arquivo de configuração e uma parte que empacota e deposita no repositório de preservação, no caso, o Archivematica. Todo o processo é automatizado sem a intervenção humana.

Figura 2 - Estrutura do Hipátia



Fonte: Elaboração dos autores (2022).

Alguns sistemas utilizam *Web Service* voltados a oferecer coleta de dados, sendo uma porta para extrair informações, assim como no caso do Sistema Eletrônico de Informações (SEI), onde é ofertado *Web Services* nos quais o BarraPres se conecta para extrair os metadados e arquivos. Porém, no caso do SEI, algumas informações também são coletadas diretamente no banco de dados, devido à ausência de todos os metadados de gestão necessários para recolhimento.

Em outros casos, quando o sistema não oferece uma forma de integração, o BarraPres pode extrair os dados diretamente do banco de dados, como é o caso do Diário de Justiça Eletrônico (DJe). Para tanto, é preciso mapear o banco de dados de forma a possibilitar que sejam feitas consultas aos metadados necessários para a preservação. Neste caso, é necessário o conhecimento da estrutura do banco de dados para possibilitar a extração dos metadados.

Independentemente da forma pelo qual os metadados e os arquivos são coletados, o BarraPres promove a flexibilidade e a modularidade, podendo ser utilizado por instituições e órgãos de governo e até usuários de sistemas distintos. Os extratores de metadados e arquivos (*Crosswalk*) executam seus processos com o apoio de um arquivo de configuração, em formato eXtensible Markup Language (XML).

Tanto para obter os metadados pelo *Web Service* ou pelo banco de dados, o BarraPres precisa de informações sobre como deve fazer a conexão. De forma a facilitar esse processo, esses dados são inseridos em arquivos de configuração e lidos pelos processos queries para sua execução.

Por fim, com os metadados e arquivos provenientes do *query*, o BarraPres pode empacotar em formato que alimentará o RDC-Arq. Esse empacotamento é feito a partir de um dicionário de dados estruturado previamente e segue todas as orientações do OAIS, de forma a permitir que o Archivematica proceda com a montagem do SIP. Após esse processo, o pacote é depositado no repositório. É importante ressaltar que as etapas do processo não contam com intervenção humana, mantendo, assim, a cadeia de custódia.

◆ 4 WEB SERVICE DO SISTEMA ELETRÔNICO DE INFORMAÇÃO

O conceito de *Web Service* não é novo, tanto que Tanenbaum (1994) define esse componente de *software* ou *hardware* como um facilitador de intercâmbio de dados entre aplicações. Assim, os *Web Services* tornam-se elementos essenciais para integração entre sistemas, servindo, em muitos casos, como interfaces que possibilitam que dados fluam de um sistema para outro.

Nesse sentido, os *Web Services* podem atuar de várias formas, implementando inúmeras funcionalidades. No âmbito empresarial, por exemplo, Coulouris *et al* (2003) defendem que os *Web Services* promovem a agilidade e eficiência na comunicação, com segurança na troca de dados, sem a necessidade de intervenção humana. Assim, os *Web Services* tornam-se essenciais na formação de Ecossistemas de Informação.

Mesmo que integrem sistemas de informação, os *Web Services* são elementos independentes e que podem possuir tecnologias distintas. Assim, possibilita a compatibilidade entre sistemas, mesmo que utilizem plataformas diferentes. Com isso, dados podem ser enviados e recebidos, mesmo entre sistemas que utilizam padrões diferentes, nos quais os *Web Services* atuam nessa compatibilização e comunicação.

Como componentes dos ecossistemas de informação, os *Web Services* não apenas possibilitam a comunicação, mas também oferecem alguns serviços de normalização, conversão de dados simples ou com alguma complexidade. Mesmo que ainda não haja consenso, grande parte dos *Web Services* oferecem segurança nas suas transações, tanto que são componentes altamente utilizados.

4.1 WEB SERVICE DO SISTEMA ELETRÔNICO DE INFORMAÇÃO

O SEI, desenvolvido pelo Tribunal Regional Federal da 4ª Região (TRF4), se autodefine como uma ferramenta de gestão de documentos e processos de forma eletrônica. Este sistema, amplamente utilizado pelos órgãos de governo, atua principalmente na produção e tramitação de documentos em formato nativos digitais ou digitalizados, atendendo grande parte das atividades de um Sistema Informatizado de Gestão Arquivística de Documentos (SIGAD), mas não atende aos requisitos de preservação.

O SEI é um sistema atual e estruturalmente fechado, ou seja, não é disponibilizado em código livre e tem sua gestão mantida pelo TRF4. De forma a possibilitar a integração com outros sistemas, o SEI oferta um *Web Service* com a possibilidade de coletar uma grande quantidade de informações de forma estruturada.

Entretanto, o *Web Service* não está disponível abertamente. Para se conectar e extrair os dados, é preciso ter um cadastro que associa o sistema (neste caso, o BarraPres) ao *Web Service*, assim como quais os tipos de informações a serem coletadas. As requisições de informações provenientes de sistemas desconhecidos serão rejeitadas de forma a garantir certa segurança e privacidade.

A permissão para acessar o *Web Service* do SEI é feita pelo endereço IP (Internet Protocol) cadastrado. Desta forma, apenas requisições feitas por endereço válidos serão processadas. Em um segundo nível, apenas as requisições dentro do escopo de permissão serão atendidas. Portanto, o processo de validação das requisições do *WebService* (SEI) tem dois níveis: de acesso e de permissão de uso.

O *Web Service* SEI oferece a possibilidade de recuperar uma grande quantidade de informações mantidas por esse sistema de informação. Neste caso, este *Web Service* do SEI atende a requisições e responde no formato JavaScript Object Notation (JSON), amplamente utilizado no intercâmbio de informações entre sistemas de informação. O formato estruturado das informações facilita o seu processamento, visto que atende a questões semânticas e sintáticas.

O BarraPres faz requisições ao SEI por meio do *Web Service* e banco de dados, a fim de obter os metadados e os arquivos necessários para criar os pacotes de preservação. Para tanto, utiliza o arquivo de configuração a fim de obter as informações para a conexão e formação das requisições. Destaca-se que um *Web Service*, como o do SEI, requer requisição estruturada e padronizada para que possa processar e enviar a resposta.

◆ 5 CONEXÃO DIRETA COM BANCO

Quando um sistema de informação não oferta *Web Services* ou outra forma de extração de dados por agente externo, uma das maneiras possíveis de integração é por meio de requisição direta ao banco de dados. Este processo nem sempre é recomendável, por apresentar certos riscos que podem ser atenuados por técnicas e métodos de conexões, as quais garantem alguma segurança.

Segundo Azevedo Filho e Costa (2008), todo acesso direto ao servidor de banco de dados deve ser restrito a computadores cliente específicos para impedir conexões não autorizadas. Dessa forma, é garantida segurança de acesso, sendo uma das premissas primordiais nos acessos remotos aos bancos de dados, tanto que em todos os gerenciadores de banco de dados há validadores de acesso.

Machado (2004), por sua vez, refina a questão do acesso externo a banco de dados, advogando sobre a definição dos privilégios que cada usuário pode ter em relação ao banco de dados. Não apenas a conexão ao banco deve ser avaliada, mas também o que cada usuário que acessa o banco pode fazer. Grande parte dos bancos de dados criam níveis de acesso diferenciados de forma a garantir a segurança.

Bancos de dados estruturados, ou mesmo os que atuam com texto por meio de objetos grandes (large object), utilizam a linguagem de requisição de dados Structured Query Language (SQL) para extrair as informações. Neste caso, muitas implementações de segurança dos bancos de dados estão de acordo com esta linguagem, estruturando-se para garantir restrições de acesso e de requisições.

Entretanto, a recuperação de dados por meio de acesso direto ao banco, mesmo com as questões de segurança, possibilita maiores opções, pois dá ao *crosswalk* acesso a todas as informações do sistema, desde que o usuário tenha essa permissão. Com isso, é possível ter uma maior flexibilidade, visto não estar restrito aos dados que um *WebService* retorna.

Assim, pode ser que, mesmo que um sistema oferte um *Web Service*, o Hipátia tenha que desenvolver *crosswalks* de conexão direta ao banco de dados. Nem sempre os *Web Services* oferecem dados importantes para a preservação, como aqueles voltados para a certificação, importante na garantia de validade.

Por outro lado, esse tipo de *crosswalk* requer conhecimento sobre o banco de dados do sistema, o que pode apresentar maior complexidade. Geralmente os *Web Services* são amplamente documentados com todos os tipos de requisição e respostas. Para que tenha acesso direto ao banco

de dados, é preciso que o desenvolvedor conheça a estrutura do banco de dados para fazer as requisições e, com isso, extrair as informações necessárias para o BarraPres.

5.1 DIÁRIO DE JUSTIÇA DO TRIBUNAL DE JUSTIÇA DO DISTRITO FEDERAL E TERRITÓRIOS

No caso do projeto de pesquisa que possibilitou a implantação do BarraPres no Diário de Justiça do TJDF, adotou-se a conexão direta ao banco de dados, devido às particularidades do sistema do Diário de Justiça à época. Portanto, o BarraPres processa uma função que é responsável por realizar uma tarefa a qual vai desde a leitura do arquivo XML, passando por conexão direta com o banco DJe, até o envio para o Archivematica de forma automatizada.

É importante destacar que o processo de conexão direta com o banco de dados pode conter muitas ameaças de segurança devido às vulnerabilidades de sistemas operacionais, serviços e aplicativos que são amplamente publicados e conhecidos por todos. Assim, as melhores práticas de desenvolvimento de *software* recomendam a não utilização de conexão direta ao banco de dados, sendo importante definir uma política de segurança de dados formais que regem como e quando usuários privilegiados podem acessar sistemas de banco de dados.

No caso do *Web Service* do SEI, as requisições e respostas são padronizadas e fixas, tendo um conjunto básico para a extração de dados. No Diário do TJDF, contudo, pode-se extrair uma quantidade maior de dados, por meio de requisições diretas ao banco de dados, utilizando as Queries, que são estruturas de requisição em linguagem SQL. Posteriormente, o *Crosswalk* manipula os metadados e arquivos da mesma forma, independentemente da sua origem e conexão direta ao banco ou *Web Service*.

◆ 6 CROSSWALK

O termo *Crosswalk*, em sua essência, significa faixa de pedestre, ferramenta em que é controlado o fluxo de pessoas e automóveis, por meio de sinalizações. Esse conceito inicial foi apropriado a comunicação de dados, ganhando um novo significado, pois, como relatam Arora e Shah (2009), neste contexto, *Crosswalk* também pode ser chamado de mapeamento de tag ou de translação de metadados, servindo para denominar processos pelos quais campos em um padrão de metadados são mapeados e convertidos para outro campo em outro padrão.

Assim, o *crosswalk* é amplamente utilizado na interoperabilidade entre sistemas, principalmente para sistemas que não compartilham o mesmo padrão de metadados, ou mesmo para converter dados oriundos de banco de dados com o intuito de formar metadados em algum esquema. O seu uso prático é defendido por vários estudos, tanto que Bountori e Gergatsoulis (2009) defendem que a técnica de mapeamento *Crosswalk* é apropriada para automatização de conversão de esquemas de metadados conhecidos.

Assim, a técnica de *Crosswalk* é tanto metodológica quanto ferramenta de atuação na pesquisa, sendo extremamente útil no BarraPres, no mapeamento dos metadados para preservação. Nesse sentido, atua na conversão ou criação de um arquivo no formato XML como os metadados requeridos pelo modelo OAIS, necessários para a formação do pacote de depósito no Archivematica.

6.1 CONFIGURAÇÃO DO CROSSWALK DO BARRAPRES

O *Crosswalk*, processo pelo qual o BarraPres se conecta aos sistemas externos e coleta informações e arquivos, deve atuar em *Web Services* ou diretamente no Banco. Para tanto, requer uma série de informações sobre como fazer a conexão e como extrair os dados. Na estrutura do BarraPres, cada sistema externo requer um *Crosswalk*. Dessa forma, foram desenvolvidos *crosswalk* para o SEI, PJE, e assim por diante. Este *Crosswalk* também pode ser denominado como um dicionário de dados.

Para facilitar o desenvolvimento dos *Crosswalks*, um arquivo de configuração no formato XML foi implementado, com o intuito de conter as informações de conexão, sejam elas em *WebService* ou em banco de dados, e sobre os metadados a serem coletados. Dessa forma, o desenvolvimento de *crosswalks* torna-se mais simples e as variáveis são declaradas no arquivo de configuração.

Assim, cada *Crosswalk* possui estruturas semelhantes, com a maior diferença ocorrendo no arquivo de configuração, indicando como se devem obter os dados e arquivos. Cabe ressaltar que a saída do *Crosswalk* é padrão, seguindo a estrutura *BagIt*, que é formada por diretórios e arquivos XML.

◆ 7 BARRAMENTO ARQUIVÍSTICO

Os *Crosswalks* acessam os sistemas de origem e extraem os metadados e arquivos, formando um diretório com todas as informações necessárias para formação de pacotes de preservação. Essa estrutura de diretórios é padrão, independentemente do tipo de acesso, se por *Web Service* ou por conexão direta com o banco de dados. Assim, cabe ao barramento processar esse diretório, de forma a montar o pacote a ser depositado no Archivematica.

Com isso, o barramento tem três grandes tarefas a serem executadas: empacotamento, controle e depósito. Atualmente, o barramento está preparado para depositar apenas no Archivematica, seguindo todas as normas e orientações desta ferramenta. Entretanto, por estar preparado para seguir padrões rígidos da arquivologia, torna-se fácil ser utilizado em outras ferramentas que sigam as mesmas normas.

O empacotamento é um processo de compactar o diretório e metadados para formação de um pacote que segue o modelo de referência OAIS. Este processo gera o pacote no formato padrão de depósito do Archivematica. Para tanto, requer que os metadados e diretórios de entrada estejam no padrão necessário para o empacotamento.

O controle tem duas grandes atuações, sendo o primeiro gerar o checksum, que oferta um número a ser validado no Archivematica de forma a garantir que não houve alteração durante o depósito. Outro ponto de controle é indicar a ordem dos arquivos para o empacotamento.

Por fim, o depósito no Archivematica pede, primeiramente, a conexão entre o BarraPres e o Archivematica. Este processo requer validação do Archivematica ao pedido de acesso. Com a conexão feita, é possível iniciar o depósito propriamente dito. Este processo é crítico, pois se trata de uma transferência de pacotes do BarraPres ao Archivematica, no qual a validação também é efetuada no sistema destino.

Assim, o Barramento recebe as informações oriundas dos *Crosswalks* e deposita no Archivematica, executando todas as transformações necessárias, ofertando formas de controle. Todo o processo ocorre sem a intervenção humana, o que garante a cadeia de custódia.

8 CONSIDERAÇÕES FINAIS

O impacto do digital no mundo é visível, mesmo que o processo de informatização ainda esteja em curso. Historicamente, não faz muito tempo que os computadores começaram a ser utilizados na oferta de serviços, mas seus impactos são amplamente vistos na sociedade. Dos primeiros computadores de grande porte aos *smartphones*, cada vez mais a informática faz parte da vida cotidiana.

Com isso, algumas atividades humanas precisam ser adaptadas ao novo cenário em que o digital está presente. Nesse sentido, é possível destacar atividades como os da biblioteca, arquivo e museu, que tem por objetivo ser intermediador da informação. Tanto que Rubi, Euclides e Santos (2006) discutem os desafios dos profissionais de informação no cenário atual em que o digital se faz presente, requerendo mudanças de postura.

Atualmente, grande parte das instituições e órgãos de governo têm fomentado a diminuição da tramitação de papel, iniciativa com o apoio de toda a sociedade, visto o impacto que isso provoca no meio ambiente. Entretanto, essa implementação apresenta um desafio a várias atividades, com destaque para a informática, administração, contabilidade, arquivo, entre outras.

Especificamente nas atividades de arquivo, a digitalização afetou significativamente os seus processos, fundamentados nos documentos físicos. Com a presença cada vez maior dos documentos nativos digitais ou digitalizados, as atividades do arquivo que é voltado à preservação passaram a ser afetadas. Assim, novas normas, modelos e orientações que têm como foco a preservação digital passaram a fazer parte do cotidiano do arquivista.

Dessa maneira, todas as tarefas voltadas ao tratamento documental e executadas nos arquivos precisam ser repensadas, pois foram elaboradas para o formato físico. Com isso, estudos estão sendo realizados e baseiam-se no uso de RDC-Arq como solução para que os SIGADs possam ser totalmente implementados, visto que algumas iniciativas têm atuado principalmente na criação e tramitação de documentos digitais.

Nesse contexto, o modelo Hipátia e seu barramento BarraPres mostram-se eficazes na integração de vários sistemas ao Archivematica e ao AtoM para a preservação digital e disseminação, sendo uma opção flexível e robusta ao atender as normas e orientações vigentes. O BarraPres mantém a cadeia de custódia, tratando as informações de forma a gerar pacotes de informação arquivística (*Archival Information Package - AIP*).

Com a tendência atual pela formação de ecossistemas de informação, o BarraPres torna-se um elemento importante ao atuar como interface integradora entre sistemas de negócio, como o SEI, DJe e outros com sistemas de preservação digital, como o Archivematica e AtoM. Com o modelo Hipátia é possível adicionar serviços de preservação aos ofertados pelos outros sistemas da instituição ou órgão público por meio da implementação de RDC-Arq, o que é um grande avanço em se tratando de gestão documental digital.

◆ REFERÊNCIAS

ALIC, M. **El legado de Hipatia**: historia de las mujeres en la ciencia desde la Antigüedad hasta fines del siglo XIX. 2. ed. México, D.F.: Siglo XXI, 2005.

AZEVEDO FILHO, E.; COSTA, M. C. L. Segurança em Servidores com banco de dados Microsoft SQL Server: meios de proteção contra invasões. In: SIMPÓSIO DE EXCELÊNCIA EM GESTÃO E TECNOLOGIA, 5., 2008, Resende, RJ. **Anais [...]**. Resende, RJ: AEDB, 2008. Disponível em: https://www.aedb.br/seget/arquivos/artigos08/261_261_Seguranca.pdf. Acesso em: 16 ago. 2021.

BOUNTOURI, L.; GERGATSOULIS, M. Interoperability between archival and bibliographic metadata: an EAD to MODS *crosswalk*. **Journal of Library Metadata**, [s.l.], v. 9, n. 1-2, p. 98-133, 2009.

COULOURIS, G. F.; DOLLIMORE, J.; KINDBERG, T.; BLAIR, G. **Distributed systems**: concepts and *design*. 5. ed. Boston: Addison-Wesley, 2012.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). **O modelo Hipátia**. Disponível em: <https://hipatia.ibict.br/o-modelo/>. Acesso em: 3 jun. 2022.

KLIMLEY, S. Digital preservation: the promise vs. the reality. In: MEETING OF THE GEOSCIENCE INFORMATION SOCIETY, 30., 1995, Luisiana, EUA. **Proceedings [...]**. Alexandria, US: American Geological Institute, 1995. p. 5-10. Disponível em: <https://oaktrust.library.tamu.edu/handle/1969.1/156095>. Acesso em: 3 jun. 2022.

MACHADO, F. N. R. **Banco de Dados**: projeto e implementação. São Paulo: Erica, 2004.

MÁRDERO ARELLANO, M. Á. Cariniana: uma rede nacional de preservação digital. **Ciência da Informação**, v. 41, n. 1, 2012. DOI: 10.18225/ci.inf..v41i1.1354. Disponível em: <https://revista.ibict.br/ciinf/article/view/1354>. Acesso em: 3 jun. 2022.

PATRA, S.; SAHOO, J.; MOHANTY, B. Research on Digital Preservation: an empirical analysis. **Library Philosophy and Practice (e-journal)**, 4693, 2021. Disponível em: <https://digitalcommons.unl.edu/libphilprac/4693/>. Acesso em: 3 jun. 2022.

RUBI, M. P.; EUCLIDES, M. L.; SANTOS, J. C. dos. Profissional da informação: aspectos de formação, atuação profissional e marketing para o mercado de trabalho. **Informação & Sociedade**, João Pessoa, v. 16, n. 1, 2006. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/443>. Acesso em: 3 jun. 2022.

SCHMIDT, L.; GHERING, C.; NICHOLSON, S. Notes on operations digital curation planning at Michigan State University. **Library Resources & Technical Services**, v. 55, n. 2, p. 104-118, 2011. Disponível em: <https://journals.ala.org/index.php/lrts/article/view/5227/6354>. Acesso em: 3 jun. 2022.

SHAH, V.; ARORA, J.. **Metadata Crosswalks**. Ahmedabad: INFLIBNET Centre, 2009. (Topics on cutting-edge technology in LIS, v. 16). Disponível em: <http://ir.inflibnet.ac.in/handle/1944/1927>. Acesso em: 3 jun. 2022.

TANENBAUM, A. S. **Distributed Operating Systems**. Nova Iorque: Prentice Hall, 1994. 648 p.

COMO CITAR:

SHINTAKU, Milton; OLIVEIRA, Alexandre Faria de. Arquitetura do modelo Hipátia. *In*: BRAGA, Tiago Emmanuel Nunes; MÁRDERO ARELLANO, Miguel Ángel. (org.). **Hipátia**: modelo de preservação para repositórios arquivísticos digitais confiáveis. Brasília: Ibict, 2022. p. 66-78. (Informação, Tecnologia e Inovação, v. 1). DOI 10.22477/9786589167501.cap5.

Capítulo 06

CASOS DE SUCESSO DE IMPLEMENTAÇÃO DO MODELO RDCARQ E HIPÁTIA

SUCCESSFUL CASES OF IMPLEMENTATION OF THE TDAR MODEL AND HIPÁTIA

Alexandre Faria de Oliveira¹

Milton Shintaku²

Marcos Sigismundo³

¹ Dados do autor: Mestrando em Computação, Instituto Brasileiro de Informação em Ciência e Tecnologia, alexandreoliveira@ibict.br

² Dados do autor: Doutor em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, shintaku@ibict.br

³ Dados do autor: Mestrando em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, marcossigismundo@ibict.br

RESUMO

Resumo: A atividade de arquivo remonta à Antiguidade, desde o nascimento da escrita. A partir de então, técnicas e métodos foram desenvolvidos para essa prática. Com o surgimento da informática e dos documentos digitais, outros desafios surgiram, refletindo nas práticas de gestão, preservação e acesso de documentos arquivísticos. Este capítulo tem por objetivo apresentar os casos de sucesso de implementação do modelo Repositórios Digitais Arquivísticos Confiáveis (RDC-Arq) e Hipátia. A partir do método de estudo de casos de natureza exploratória e explicativa, investiga-se o caso de sucesso da implementação no Tribunal de Justiça do Distrito Federal e Territórios e no Arquivo Nacional, por meio da observação direta e participativa, no âmbito dos projetos de pesquisa firmados entre o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) e ambas as instituições. Conclui-se que o *software* Hipátia é eficaz em relação à integração de vários sistemas, tornando-se uma opção flexível e robusta ao atender as normas e orientações vigentes.

Palavras-Chave: Preservação digital. *Software* Hipátia. Ecossistemas de informação.

Abstract: Archival activity dates back to antiquity, from the birth of writing. Since then, techniques and methods have been developed for this practice. With the emergence of information technology and digital documents, other challenges have arisen, reflecting on management practices, preservation and access to archival documents. This chapter aims to present successful cases of implementation of the Trusted Digital Archival Repositories (TDAR) and Hipátia model. Based on the case study method of an exploratory and explanatory nature, the case of successful implementation in the Court of Justice of the Federal District and Territories and in the National Archives is investigated, through direct and participatory observation, within the scope of research projects signed between the Brazilian Institute of Information in Science and Technology (Ibict) and both institutions. It is concluded that the Hipátia software is effective in relation to the integration of various systems, making it a flexible and robust option to meet current standards and guidelines.

Keywords: Digital Preservation. Hipátia *Software*. Information eco systems.

♦ 1 INTRODUÇÃO

A atividade de arquivo remonta à Antiguidade, desde o nascimento da escrita, na medida em que grande parte dos registros antigos tinha relação com religião e administração, atividades ligadas ao funcionamento do Estado, requerendo ações voltadas à preservação dos documentos. Desde então, técnicas e métodos foram desenvolvidas para essa prática. Entretanto, com o surgimento da informática e dos documentos digitais, outros desafios surgiram e foi necessário que os conhecimentos arquivísticos precisassem ser adaptados ao novo contexto.

Nesse novo cenário, Santos e Flores (2015) relatam que a digitalização proporcionou significativas mudanças nos conceitos e na concepção de documentos arquivísticos, refletindo nas práticas de gestão, preservação e acesso. Tanto que os autores defendem que, para a preservação a longo prazo, os Repositórios Digitais Arquivísticos Confiáveis (RDC-Arq) são a alternativa mais viável mundialmente, salientando a necessidade da realização de auditorias periódicas para a sua certificação e a criação de políticas de preservação digital que contemplem os requisitos arquivísticos e diplomáticos.

Sendo assim, o RDC-Arq apresenta-se como a opção válida para a preservação arquivística de documentos digitais. Lehmkuhl, Macedo e Silva (2018) defendem que a ferramenta mais apropriada para construir esses sistemas de informação é o Archivematica, por questões qualitativas, em comparação com outros *softwares* livres. Conforme o estudo, o Archivematica atende todos os principais requisitos para preservar, de forma confiável, os documentos arquivísticos.

A escolha do *software* livre Archivematica, em grande parte, dá-se pelo atendimento ao modelo, desenvolvido pela Organização Internacional de Normalização (ISO), denominada de *Open Archival Information System* (OAIS), visto que essa norma serviu de base conceitual para o desenvolvimento da ferramenta, como relatam Van Garderen *et al.* (2016). Fontana, Flores, Nora e Machado (2014) defendem o uso do Archivematica como uma estratégia para preservação, de forma a diminuir a possibilidade de inacessibilidade aos documentos digitais, na medida em que possibilita ao administrador manter os documentos atualizados tecnologicamente garantindo, assim, que os documentos digitais permaneçam autênticos, acessíveis e utilizáveis por um prazo indeterminado.

♦ 2 O HIPÁTIA E ARCHIVEMATICA

O uso de Repositórios Digitais Arquivísticos Confiáveis (RDC-Arq) torna-se uma necessidade aos órgãos e instituições, tendo em vista a necessidade de preservação da sua documentação, com a implementação de sistemas de tramitação de documentos digitais, como o Sistema Eletrônico de Informações (SEI), em grande parte dos órgãos públicos, por exemplo. Com isso, evidencia-se a necessidade de estudos que visem a integração de sistemas distintos, como os de tramitação de documentos digitais e o RDC-Arq, de forma a ter um ecossistema compatível com as premissas do SIGAD.

Nesse contexto, se vislumbra o desenvolvimento de interfaces que possibilitem a integração de tecnologias distintas, mas que mantenham as características de ambos. Neste caso, um grande orientador é o modelo OAIS e as normas brasileiras determinadas pelo Conselho Nacional de Arquivo (CONARQ). Assim, esse aparato de regras e modelos amparam o funcionamento de interfaces que integram sistemas de arquivo.

Os sistemas que tramitam documentos digitais dependem do órgão ou instituição usuária, pois é uma escolha que tem a ver com as suas características, selecionando o que melhor lhe atende. A partir da informatização, esses sistemas têm evoluído, desde os primeiros controles de acompanhamento de protocolo criados ainda no século passado. Entretanto, o maior desenvolvimento deveu-se ao uso cada vez maior de documentação nato digital ou digitalizado.

Quanto ao RDC-Arq, nota-se cada vez mais o uso do Archivematica, por atender a todos os requisitos necessários para a preservação digital, seguindo as normas internacionais arquivísticas. Entretanto, para manter a cadeia de custódia, todo o processo de transferência dos documentos e seus metadados do sistema de tramitação de documento para o Archivematica deve ser automatizado, o que agrega desafios, levando em conta que os sistemas a serem integrados podem utilizar tecnologias distintas.

Assim, torna-se clara a necessidade de interfaces que possibilitem a integração de sistemas distintos, utilizando ou não tecnologias diferentes. É nesse contexto que se situa o Hipátia (**Figura 1**), o qual possibilita a integração de sistemas que tramitam documentos com o RDC-Arq, neste caso o Archivematica, seguindo todas as orientações do modelo OAIS, de forma a manter a cadeia de custódia.

Figura 1 - Contexto de inserção do Hipátia para tramitação de documentos digitais



Fonte: Elaboração dos autores (2022)

O Hipátia foi totalmente desenhado para integrar com o Archivematica, mas permite vários sistemas produtores, como o SEI e o Processo Judicial eletrônico (PJe) utilizado no Tribunal de Justiça do Distrito Federal de Territórios (TJDFT). Com isso, ele atende a todos os requisitos do sistema, que, por sua vez, está de acordo com as normas internacionais. Assim, pode ser utilizado por vários órgãos ou instituições, independente do seu sistema de tramitação de documentos.

Nesse sentido, o Hipátia recebe a requisição, vindo do sistema de tramitação de documento para o arquivamento de um documento no Archivematica. Com essa requisição, o Hipátia coleta os metadados diretamente no banco de dados do sistema de tramitação de documentos e, com isso, transfere automaticamente o documento do sistema fonte para o Archivematica junto com os metadados, sem a intervenção humana, que poderia comprometer o processo.

Destaca-se a flexibilidade ofertada pelo Hipátia, no qual vários sistemas podem se integrar ao Archivematica para preservar seus documentos em formato digital. Com isso, possibilita que todos os sistemas que atuam com gestão de documentação digital possam utilizar o Archivematica como repositório confiável para preservação, desde que forneça os requisitos requeridos, como os metadados e capacidade de transferência dos arquivos.

♦ 3 CASOS DE SUCESSO NO USO DO HIPÁTIA

O desenvolvimento de novas ferramentas visa o atendimento a necessidades com pouca ou nenhuma oferta de soluções automatizadas que as atendam. Com isso, pode-se considerar que apenas com o uso, como uma forma de validação, uma ferramenta tem uma finalização de uma versão. Como é de conhecimento, um *software* vive em constante evolução, no qual o versionamento são marcos de etapas. Assim, o ciclo de desenvolvimento se completa, pelo menos, em uma etapa da evolução.

No caso específico do Hipátia, como validadores, o projeto de desenvolvimento contou com a colaboração de duas instituições, com sistemas fontes de documentos distintos, o Tribunal de Justiça do Distrito Federal e Territórios (TJDFT) e o Arquivo Nacional (AN). Com isso, expandiu-se os testes e ajustes para atendimento das necessidades, levando em consideração as diferenças entre as instituições e seus sistemas de origem de documentos na medida em que o sistema alvo é o mesmo, o Archivematica.

Em ambos os casos, o Hipátia se mostrou eficaz em sua tarefa de integrar o sistema fonte de documentos ao sistema de preservação, de forma automatizada, sem a interferência humana, mantendo a cadeia de custódia. Assim, pode-se contar as experiências da implementação do Hipátia no TJDFT e AN como casos de sucesso, mesmo que possam ser considerados como estudos iniciais, visto o potencial proposto pela ferramenta.

3.1 TRIBUNAL DE JUSTIÇA DO DISTRITO FEDERAL E TERRITÓRIOS

O Tribunal de Justiça do Distrito Federal e Territórios (TJDFT) é o órgão do poder judiciário brasileiro que tem raízes ainda quando a capital residia na cidade do Rio de Janeiro e do primeiro tribunal instalado em Salvador (primeira capital do Brasil). Entretanto, da forma que se reconhece atualmente, o TJDFT nasce na transferência da capital para Brasília, por meio da Lei nº3.752 de 14 de abril de 1960, dispendo sobre a Organização judiciária do Distrito Federal.

Assim, a história do TJDFT e, por conseguinte, sua documentação, tem uma trajetória que acompanha a cidade de Brasília, e requer grande parte de cuidados especiais de preservação, com garantias de integridade e autenticidade. Seguindo a evolução tecnológica, o TJDFT aderiu às novas tecnologias, que possuem grande apelo digital, com oferta de serviços por meio da internet, produção e transmissão de documentos digitais, entre outros.

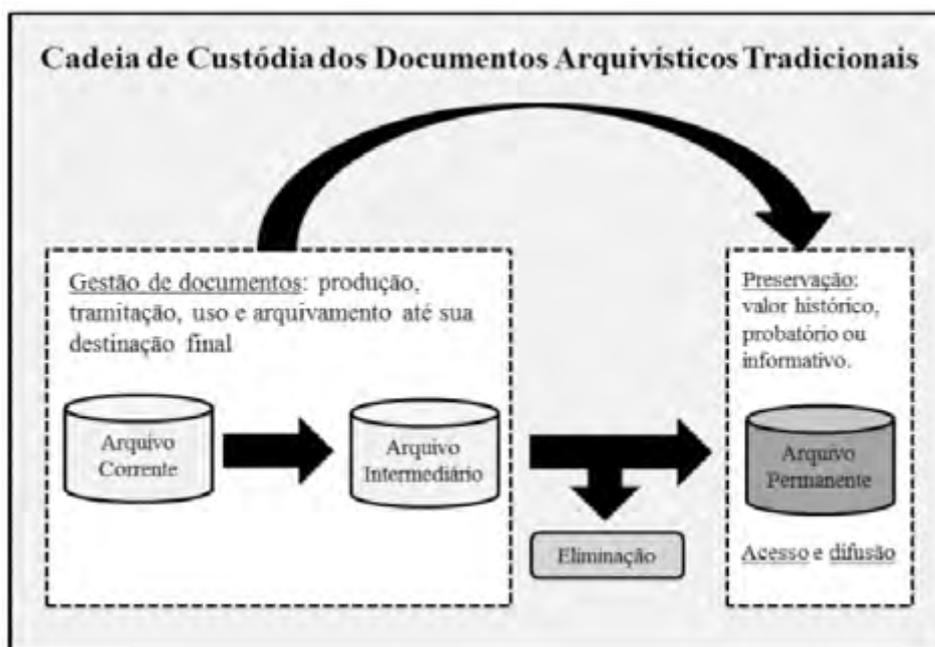
A crescente demanda pela gestão dos processos e documentos em meio digital na administração pública fez o TJDFT perceber a necessidade de se criar uma política integrada de padronização, classificação, avaliação, descrição e preservação dos documentos judiciais nesse ambiente. Desta maneira, em 2018 foi editada pelo Tribunal a Resolução 2, de 27 de março de 2018, que estabelecia a *Política de Gestão Documental de Processos e Documentos em Meio Digital do TJDFT*.

Neste mesmo ano, devido às necessidades identificadas na gestão documental em formato digital ocorridas no TJDF, articulou-se com o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) a possibilidade de fomento a um projeto de pesquisa voltado a estudos para criação de alternativas de soluções para implantação de um Repositório Digital Confiável Arquivístico (RDC-Arq). Este projeto desenvolveu estudos voltados à integração de um sistema fonte de documentos digitais com um sistema de preservação digital, atendendo a todos os requisitos arquivísticos necessários para integridade e autenticidade.

A escolha do Ibict como instituição parceira no projeto de pesquisa deu-se pela expertise do instituto com a preservação digital, que, desde 2002, tem atuado neste tema, no qual a Rede Cariniana é destaque, preservando periódicos científicos em formato digital. Mesmo que o projeto de pesquisa tenha distinções com o modelo de preservação digital da Rede Cariniana, por tratarem de objetos digitais de características muito diferentes, o Ibict é a referência nacional no tema.

Assim, o projeto de pesquisa ampara-se em base conceitual composta por elementos oriundos da Arquivologia e preservação digital, juntamente com as especificidades da documentação do TJDF, na implementação de um RDC-Arq que atendesse ao tribunal. Nesse sentido, os estudos voltaram-se para os critérios necessários para a implementação do RDC-Arq, sendo que uma das questões primordiais tem relação com a preservação digital, garantindo a cadeia de custódia documental. Como entendimento para o projeto, Flores (2016) declara que a cadeia de custódia compreende um ciclo que contempla o Arranjo, Descrição, Digitalização, Difusão e Acesso de Documentos de caráter corrente, intermediário e permanente em SIGAD'S e Repositórios Digitais confiáveis Arquivísticos respectivamente, como apresentado na **Figura 2**.

Figura 2 - Cadeia de custódia



Fonte: Flores; Rocco; Santos (2021).

O fluxo documental, no qual embasa-se o estudo, repousa no processo de gestão documental, em que um documento é produzido de forma digital, ou digitalizado, tramita neste formato até o arquivamento (que ocorre em arquivo corrente, quando ainda está em uso), passa por arquivo intermediário até que chega ao ponto de passar para o arquivo permanente ou descarte. Para os

documentos digitais, o RDC-Arq torna-se um sistema para atuar no arquivo permanente, mantendo a cadeia de custódia e preservando-o em longos períodos de tempo.

Diante do desafio de se implementar um Repositório Digital Confiável Arquivístico para o TJDFT, o Ibict iniciou a pesquisa em torno dos principais desafios de integrar sistemas heterogêneos, sendo um fonte de documentação e o outro de preservação. Assim nasceu a ideia de construir um barramento de preservação que fizesse a automação de coleta, empacotamento e envio para o Archivematica. A estrutura de criação do sistema de automação levou em consideração o Modelo OAIS (*Open Archival Information System*).

O sistema produtor inicial selecionado para o estudo foi o Diário de Justiça Eletrônico (DJe), criado com base na Lei nº 11.419 de 19 de dezembro de 2006, que em seu capítulo II, dispendo sobre a comunicação eletrônica dos atos de processo, permitiram que os tribunais criassem os seus diários de justiça eletrônicos. Entretanto, na mesma lei, dispendo sobre o processo eletrônico, relatam sobre a preservação dos processos.

Com uma equipe multidisciplinar composta por estudiosos da Ibict e TJDFT nas áreas de Tecnologia da Informação e Gestão da Informação definiu-se, inicialmente, quais informações seriam coletadas do DJe e enviadas para preservação no Archivematica. Dessa forma, estabeleceu-se a unidade informacional a ser preservada, da qual é possível levantar as características informacionais.

No que se refere aos sistemas a serem integrados, há uma diferença nos padrões utilizados na representação da informação. O DJe utiliza o Dublin Core, assim como parte dos sistemas voltados à disseminação da informação, enquanto o Archivematica utiliza o padrão General International Standard Archival Description (ISAD (G)), seguindo as orientações arquivísticas. Para realizar a integração, primeiramente foi preciso compatibilizar os dois padrões, como nos modelos de *Crosswalk*.

O RJe oferta uma *Application Programming Interface* (API) para recuperar dados (metadados) em Dublin Core. Entretanto, não oferece todas as informações necessárias para o ISAD (G). Assim, foi preciso desenvolver formas de extração diretamente do banco de dados, por meio de conexão externa. Com isso, é possível compor o registro completo ISAD (G) necessário para a preservação.

Além dos metadados, é preciso extrair os arquivos, muitas vezes armazenados em estrutura de diretórios, para compor o pacote *BagIt*, composto por metadados mais os arquivos. Assim, é necessário identificar os arquivos relacionados aos metadados e extraí-los, mantendo a integridade e repassando para que possa ser montado o pacote *BagIt*.

Foi então desenhada a primeira versão do barramento tecnológico. O sistema estava apto a fazer, de forma agendada e autônoma, a coleta dos metadados diretamente no banco de dados, coletando também os respectivos arquivos dos diários, montando um pacote *BagIt* com toda a estrutura necessária para envio ao Archivematica. A transação era registrada em registros cronológicos (*Logs*) de sucesso e erros. Posteriormente, foi implementada uma nova funcionalidade que tratava os erros de maneira mais rápida, evitando a paralisação da coleta.

Após a exitosa integração entre o DJe e o RDC-Arq, por meio do Hipátia, o TJDFT manteve a parceria de pesquisa para a implementação da solução tecnológica com os demais sistemas da instituição, em especial com o Processo Judicial Eletrônico (PJe). Nesta fase, a pesquisa será feita integrando tecnologicamente os sistemas dentro de novos conceitos como programação orientada a eventos, sistemas de mensagens em tempo real e uma nova plataforma de serviços criada em contêineres, utilizando a tecnologia *OpenShift*®.

Para tanto, foi preciso instalar o Archivematica em um contêiner *OpenShift*[®]. Este processo foi pioneiro, visto não documentado relatando esse processo. O uso de contêineres para gerenciar sistemas tem sido uma tendência e tem como intuito facilitar a manutenção. A tecnologia *OpenShift*[®] umas das mais utilizadas, mas impõe certos desafios às equipes de informática por ser uma novidade em certos casos.

Essa nova fase do projeto viabiliza a gestão documental dos processos judiciais no PJe (PJe-Arq), garantindo a preservação digital daqueles com prazos de guarda longo ou com a destinação final de guarda permanente no ambiente de preservação dos documentos arquivísticos digitais. Com isso, expande o uso do Hipátia para atender a mais de um sistema, funcionando realmente como um barramento integrador de sistemas para preservação.

3.2 ARQUIVO NACIONAL

O Arquivo Nacional (AN) é um órgão público do poder executivo, vinculado ao Ministério da Justiça e Segurança Pública (MPJP), voltado a gerir, preservar e disseminar os documentos produzidos pela administração pública brasileira. Para tanto, tem a sua história ligada ainda ao Império do Brasil, por meio do Regulamento nº 2 de 02 de Janeiro de 1838, com o nome de Arquivo Público do Império.

Com história rica e longínqua, o AN tem um acervo documental único no país, composto por documentos físicos inestimáveis, com textos, fotografias, cartões e outros. O Arquivo Nacional tem papel fundamental na preservação do patrimônio documental brasileiro, abrigando documentos como os originais da Constituição de 1824 e da Lei Áurea. O AN tem a finalidade de guardar, conservar e divulgar os documentos de valor histórico, administrativo ou legal dos órgãos que integram os poderes da União.

Desde sua fundação, temas ligados à história do Estado brasileiro são discutidos no Arquivo Nacional, que possui a memória do país revelada em seu acervo. O conhecimento e o acesso a dados importantes que fazem parte da história, da política e da sociedade brasileira convertem o Arquivo em parte da riqueza nacional.

Acompanhando a evolução tecnológica, grande parte dos órgãos tem tramitado cada vez mais documentos em formato digital, em grande parte impulsionado pelo chamado e-GOV, que, desde o início do século, tem ações de informatização das atividades administrativas no âmbito governamental. Com isso, impõe-se ao AN desafios à adaptação das suas atividades aos documentos no formato digital.

Nesse sentido, o Arquivo Nacional buscou parceria com o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) para aperfeiçoar suas metodologias na gestão de sua informação, como também alcançar o domínio em novas soluções tecnológicas livres que dinamizam o processo de gestão da informação. AN e Ibict tem missões que se complementam, no qual o AN garante acesso aos documentos da administração pública federal, fazendo por meio de infraestrutura informacional, promovida pelo Ibict.

Nesse contexto, o AN e o Ibict firmaram projeto de pesquisa para estudos voltados à implantação de estratégias de preservação digital dos documentos tramitados no Sistema Eletrônico de Informações (SEI). Com isso, desenvolver modelos que possam ser replicados para outros órgãos públicos que façam uso do SEI e desejam implementar a preservação digital, por meio de RDC-Arq, como o Archivematica.

Estudos têm mostrado que o SEI ainda não pode ser considerado um Sistema Informatizado de Gestão Arquivística de Documentos (SIGAD), principalmente por não atender a questão de preservação, como defende Flores (2017). Por isso, alguns especialistas têm definido o SEI como um sistema de tramitação de processos no formato digital, no qual ainda precisa de estudos voltados para a preservação. Tanto que, Nogueira, Costa e Saraiva (2018) consideram o SEI como um sistema que possibilita a produção, edição, assinatura e tramitação de documentos em formato digital e que ainda requer melhorias na gestão de documentos de arquivo.

O SEI é um *software* governamental, não podendo ser considerado completamente como uma ferramenta livre, visto que ainda mantém certo controle de seus desenvolvedores. A história do SEI remonta ao ano de 2009, com o propósito de implementar melhorias nos fluxos de informações e documentos no Tribunal Regional da 4ª Região (TRF4), com sede em Porto Alegre, órgão de 2º grau da justiça federal. Posteriormente, tem sido amplamente adotado por órgãos de governo em todas as esferas e poderes, na medida em que compõe o Processo Eletrônico Nacional, como iniciativa dos órgãos e entidades públicas na adoção de medidas para agilização dos processos administrativos, com o uso de documentação digital.

Um dos principais resultados do projeto de pesquisa firmado entre o AN e Ibict é desenvolver um modelo de integração entre o SEI e o Archivematica, por meio do Hipátia. A integração entre os sistemas vai permitir coletar os processos e seus arquivos do SEI e empacotar e depositar no Archivematica. Todo esse processo deve, no entanto, garantir a cadeia de custódia.

Inicialmente, foi estudado o SEI e quais as formas de interação e interoperabilidade, no qual foi identificado um *web service* voltado a coleta automática de informações, com uso restrito por usuários institucionais validados. Assim, foi feito um extenso levantamento das funcionalidades do *web service*⁴ nativo do sistema, mapeando e definindo locais para coleta dos metadados obrigatórios na preservação. Foram também definidas coletas diretamente no banco de dados do SEI de modo a complementar as informações necessárias.

Uma das grandes funcionalidades implementadas no Hipátia, nesse projeto de pesquisa, foi a habilidade de criar e inserir documentos dentro do sistema SEI. Para atender uma regulamentação interna da instituição, foi necessário criar fluxo de trabalho que permitisse ao operador do sistema validar os processos documentais sob diversas diretrizes que definem os requisitos mínimos para arquivamento do Arquivo Nacional. Caso os requisitos não estejam contemplados, o Hipátia cria documentos dentro do processo, com a maioria dos metadados preenchidos automaticamente, cabendo ao operador do sistema terminar o preenchimento. Esses documentos contêm informações obrigatórias para validar a preservação.

Os estudos revelaram que o *web service* não atendia totalmente às necessidades do Hipátia para a geração do pacote *BagIt*, sendo necessário o desenvolvimento de outras funcionalidades. Com isso, o Hipátia aumentou a oferta de serviços, com implementação de códigos para extrair informações diretamente do banco de dados, no modelo do SEI. Como um dos resultados do projeto, é possível considerar o melhor conhecimento do funcionamento interno do SEI.

4 Web Service é uma solução utilizada na integração de sistemas e na comunicação entre aplicações diferentes. Com esta tecnologia, é possível que novas aplicações possam interagir com aquelas que já existem e que sistemas desenvolvidos em plataformas diferentes sejam compatíveis.

O modelo desenvolvido pelo projeto AN e Ibict para atender a integração do SEI ao Archivematica, por meio do Hipátia, envolve o uso do *web service* nativo do SEI para coletar algumas informações, recuperação de informações direto no banco de dados e extração dos documentos. Os estudos revelaram que as informações ofertadas pelo *web service* não contemplavam todos os metadados necessários para a preservação imposta pelo padrão ISAD (G), o que requereu outras ações do Hipátia. Da mesma forma, a extração dos arquivos do processo precisa de ações do Hipátia.

Com todas as informações obtidas, conforme o padrão ISAG (G), e os arquivos do processo, o Hipátia gera o pacote *BagIt* conforme as normas arquivísticas, implementadas pelo Archivematica. O processo de geração do pacote *BagIt* e depósito no Archivematica é único e independente do sistema fonte das informações. Assim, modularmente, o projeto AN e Ibict apenas gerou mais uma possibilidade de integração.

Esse projeto de pesquisa deu ao Hipátia uma grande flexibilidade de integração com vários sistemas produtores de informações arquivísticas, dando robustez necessária para validar realmente como um barramento de dados, ao garantir integridade e autenticidade, essa automação proporcionada pelo barramento facilita e agiliza o processo de preservação digital nas instituições. Com isso, o Hipátia pode ser considerado uma plataforma de integração de sistemas voltada à preservação digital, no qual possibilita a integração de diversos sistemas produtores de documentos.

◆ 4 CONSIDERAÇÕES FINAIS

O Barramento Hipátia atua como interface integradora entre sistemas, recebendo, coletando e depositando os pacotes no sistema destino. Com isso, permite que sistemas heterogêneos estabeleçam comunicação, voltados à preservação digital. A flexibilidade ofertada pelo Hipátia, no qual vários sistemas podem se integrar ao Archivematica para preservar seus documentos em formato digital, possibilita que todos os sistemas que atuam com gestão de documentação digital, possam utilizar o Archivematica como repositório confiável para preservação, desde que forneça os requisitos requeridos, como os metadados e capacidade de transferência dos arquivos.

Nesse sentido, o Hipátia é mais uma iniciativa do Ibict, alinhado à sua missão de promover o desenvolvimento de recursos e infraestrutura informacional. Da mesma forma, como um grande incentivador da filosofia aberta, com a promoção pelo código aberto.

Portanto, a avaliação de um *software* livre, por meio do seu uso, representa, de certa forma, o ápice, principalmente se considerar a viabilidade das instituições que a adotam. No caso do Hipátia, o uso em projetos ligados ao TJDFT e AN demonstram a importância da iniciativa do Ibict e a sua vanguarda na preservação digital, sendo uma das instituições precursoras neste tema no Brasil.

A parceria em projetos de pesquisa com o TJDFT e o Arquivo Nacional tem possibilitado o desenvolvimento de um *software* livre de alta qualidade que automatiza uma importante etapa do processo de preservação digital nas instituições. Essa funcionalidade é fundamental para garantir o acesso contínuo a materiais digitais durante o tempo necessário, além dos limites e mudança tecnológica e organizacional.

◆ REFERÊNCIAS

ALIC, M. **El legado de Hipatia**: historia de las mujeres en la ciencia desde la Antigüedad hasta fines del siglo XIX. 2. ed. México, D.F.: Siglo XXI, 2005.

AZEVEDO FILHO, E.; COSTA, M. C. L. Segurança em Servidores com banco de dados Microsoft SQL Server: meios de proteção contra invasões. *In*: SIMPÓSIO DE EXCELÊNCIA EM GESTÃO E TECNOLOGIA, 5., 2008, Resende, RJ. **Anais** [...]. Resende, RJ: AEDB, 2008. Disponível em: https://www.aedb.br/seget/arquivos/artigos08/261_261_Seguranca.pdf. Acesso em: 16 ago. 2021.

BOUNTOURI, L.; GERGATSOULIS, M. Interoperability between archival and bibliographic metadata: an EAD to MODS *crosswalk*. **Journal of Library Metadata**, [s.l.], v. 9, n. 1-2, p. 98-133, 2009.

COULOURIS, G. F.; DOLLIMORE, J.; KINDBERG, T.; BLAIR, G. **Distributed systems**: concepts and *design*. 5. ed. Boston: Addison-Wesley, 2012.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (Ibict). **O modelo Hipátia**. Disponível em: <https://hipatia.ibict.br/o-modelo/>. Acesso em: 3 jun. 2022.

KLIMLEY, S. Digital preservation: the promise vs. the reality. *In*: MEETING OF THE GEOSCIENCE INFORMATION SOCIETY, 30., 1995, Luisiana, EUA. **Proceedings** [...]. Alexandria, US: American Geological Institute, 1995. p. 5-10. Disponível em: <https://oaktrust.library.tamu.edu/handle/1969.1/156095>. Acesso em: 3 jun. 2022.

MACHADO, F. N. R. **Banco de Dados**: projeto e implementação. São Paulo: Erica, 2004.

MÁRDERO ARELLANO, M. Á. Cariniana: uma rede nacional de preservação digital. **Ciência da Informação**, v. 41, n. 1, 2012. DOI: 10.18225/ci.inf..v41i1.1354. Disponível em: <https://revista.ibict.br/ciinf/article/view/1354>. Acesso em: 3 jun. 2022.

PATRA, S.; SAHOO, J.; MOHANTY, B. Research on Digital Preservation: an empirical analysis. **Library Philosophy and Practice (e-journal)**, 4693, 2021. Disponível em: <https://digitalcommons.unl.edu/libphilprac/4693/>. Acesso em: 3 jun. 2022.

RUBI, M. P.; EUCLIDES, M. L.; SANTOS, J. C. dos. Profissional da informação: aspectos de formação, atuação profissional e marketing para o mercado de trabalho. **Informação & Sociedade**, João Pessoa, v. 16, n. 1, 2006. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/443>. Acesso em: 3 jun. 2022.

SCHMIDT, L.; GHERING, C.; NICHOLSON, S. Notes on operations digital curation planning at Michigan State University. **Library Resources & Technical Services**, v. 55, n. 2, p. 104-118, 2011. Disponível em: <https://journals.ala.org/index.php/lrts/article/view/5227/6354>. Acesso em: 3 jun. 2022.

SHAH, V.; ARORA, J.. **Metadata Crosswalks**. Ahmedabad: INFLIBNET Centre, 2009. (Topics on cutting-edge technology in LIS, v. 16). Disponível em: <http://ir.inflibnet.ac.in/handle/1944/1927>. Acesso em: 3 jun. 2022.

TANENBAUM, A. S. **Distributed Operating Systems**. Nova Iorque: Prentice Hall, 1994. 648 p.

COMO CITAR:

OLIVEIRA, Alexandre Faria de; SHINTAKU, Milton; SILVA, Marcos Sigismundo da. Casos de sucesso de implementação do Modelo RDCArq e Hipátia. *In*: BRAGA, Tiago Emmanuel Nunes; MÁRDERO ARELLANO, Miguel Ángel. (org.). **Hipátia**: modelo de preservação para repositórios arquivísticos digitais confiáveis. Brasília: Ibict, 2022. p. 79–90. (Informação, Tecnologia e Inovação, v. 1). DOI 10.22477/9786589167501.cap6.

Capítulo 07

A VISÃO DO IBICT PARA UM ECOSSISTEMA NACIONAL DE PRESERVAÇÃO DIGITAL

THE IBICT'S VISION FOR A NATIONAL DIGITAL PRESERVATION ECOSYSTEM

Alexandre Faria de Oliveira¹

Miguel Ángel Márdero Arellano²

Tiago Emmanuel Nunes Braga³

¹ Dados do autor: Mestrando em Computação, Instituto Brasileiro de Informação em Ciência e Tecnologia, alexandreoliveira@ibict.br

² Dados do autor: Doutor em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, miguel@ibict.br

³ Dados do autor: Doutor em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, tiagobraga@ibict.br

RESUMO

Resumo: O capítulo apresenta a visão de futuro do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) para os projetos de preservação digital e discorre sobre as áreas de suporte à pesquisa na temática. Apresenta também um panorama sobre como o Hipátia será integrado nas iniciativas já existentes e quais caminhos podem ser trilhados a fim de consolidar a atuação do Instituto enquanto agente promotor das ações de Preservação Digital.

Palavras-Chave: Ibict. Preservação Digital. Hipátia.

Abstract: The chapter presents a vision of the future of the Ibict Instituto Brasileiro de Informação em Ciência e Tecnologia - Ibict to the projects in Digital Preservation and discusses the areas of research support on the theme. It also presents an overview of how Hipátia will be integrated into existing initiatives and which paths can be trodden in order to consolidate the Institutes's *performance* as a promoter of Digital Preservation actions.

Keywords: Ibict. Digital Preservation. Hipátia.

◆ 1 INTRODUÇÃO

A preservação dos registros eletrônicos oriundos dos resultados da ciência e do patrimônio histórico e cultural em documentos digitais adquire novas práticas dentro de um contexto de rede de computadores. A informação não está mais em um só local e a gestão desses acervos encontra novos desafios. O tratamento e a organização de bases de dados, repositórios e bibliotecas digitais precisam incorporar estratégias de preservação relacionadas com a sua dependência tecnológica: custos, políticas, acesso e critérios técnicos.

O *design* do ecossistema de preservação digital do Ibict parte do conceito de Arquitetura da Informação Pervasiva proposto por Resmini e Rosati (2011). Ele seria um ecossistema de informação que conecta pessoas, informações e processos que estão em toda parte, um espaço de processos de informação que conectam soluções tecnológicas com o *design* de serviços. No caso da preservação digital, esse ambiente é formado por tecnologias que gerenciam processos de preservação para tipologias específicas de documentos digitais. São sistemas para coleta, submissão e disseminação de dados em repositórios digitais confiáveis, sistemas para arquivamento e de preservação digital distribuída.

Nessa perspectiva, o Ibict se enquadra no fornecimento de diferentes tipos de serviços de preservação digital para bibliotecas, arquivos, pesquisadores e instituições de ensino e pesquisa, disponibilizando uma série de soluções tecnológicas, tais como sistemas de preservação de publicações eletrônicas, de arquivamento de páginas web, de dados de pesquisa e para preservação de todo tipo de documentos arquivísticos. Esse ecossistema de serviços está inserido dentro de um contexto teórico e tecnológico amplo e de um volume crescente de objetos digitais que precisam da aplicação de atividades de gestão relacionadas à preservação durante todo seu ciclo de vida.

Dentro do contexto das atividades do Instituto, a identificação de oportunidades para a criação de um ecossistema de serviços de preservação digital iniciou-se reconhecendo o problema principal: nem tudo poderia ou deveria ser preservado. Sendo assim, decisões qualitativas tiveram que ser tomadas e a prioridade das ações foi definida como a preservação do conteúdo relacionado ao serviço e repasse da ferramenta *Open Journal Systems*, a qual o Instituto havia pioneiramente fomentado junto aos periódicos científicos do Brasil. Como a preservação digital tem muito a ver com o longo prazo, essa escolha poderia não ser reconhecida como boa ou ruim por várias décadas, mas o panorama tecnológico daquele momento exigiu a consideração do curto prazo na definição de prioridades, da mesma forma que outras escolhas são feitas quando há recursos escassos.

Como uma alternativa para organizações que desejam colecionar, armazenar, preservar e oferecer acesso a seus acervos mais importantes em cópias digitais autorizadas no Brasil, o Ibict criou a Rede Brasileira de Serviços de Preservação Digital Cariniana, promovendo, assim, o arquivamento digital em rede. Na sua primeira etapa, com a adesão à Aliança LOCKSS, da Universidade de Stanford, estava estabelecida sua concordância com as normas internacionais já aplicadas em importantes iniciativas de preservação de publicações eletrônicas, assim como a incorporação de atividades de capacitação de profissionais dedicados ao registro de acervos em formato digital ao conjunto das primeiras ações da rede.

A Rede Cariniana compreende doze instituições de ensino e pesquisa brasileiras, as quais compartilham sua infraestrutura local para instalação e uso de ferramentas de preservação digital distribuída, e outras seis instituições afiliadas participam de projetos e do uso das ferramentas

promovidas pela Rede. Além desse apoio, todas as instituições parceiras usufruem dos benefícios do *networking*, do compartilhamento de informações e de experiências em eventos e em cursos. Em sua estrutura organizacional, há o conselho técnico consultivo, que é renovado a cada cinco anos, e é responsável pela identificação de soluções tecnológicas e pela avaliação de coleções digitais e eletrônicas nas quais os membros da rede trabalharão coletivamente.⁴

A crescente demanda por ações de preservação no contexto das instituições parceiras da Rede influenciou nas fases de desenvolvimento do projeto da Cariniana, do fornecimento básico inicial de coleta e armazenamento distribuído de publicações digitais para um serviço orientado ao usuário/parceiro e que participa de uma gama de projetos cada vez maior, contemplando recursos educacionais e uma gama de tipos de materiais digitais crescente.

Com o projeto da Rede Cariniana como um sistema de preservação digital baseado no modelo de rede distribuída, o Ibict acompanha e provê subsídios a outros projetos brasileiros que precisam preservar materiais autênticos e certificados por instituições reconhecidas. As atividades de colaboração entre as instituições parceiras da Rede levaram à concretização de políticas de preservação em instituições de ensino e pesquisa no Brasil, assim como a participação em projetos institucionais de preservação de acervos analógicos desde a etapa de digitalização até seu armazenamento seguro em repositórios digitais confiáveis.

2 O PROGRAMA

A visão do Programa de Preservação Digital Brasileiro (Ibict, 2017) é baseada no estabelecimento de um conjunto de fontes de referência confiáveis que seriam o catalisador para o compartilhamento de informações e boas práticas em um conjunto de sistemas de preservação digital de acesso aberto, bem como em uma comunidade de pesquisa. As instituições usuárias seriam capazes de determinar as melhores e mais adequadas sequências de ações de preservação para atender às suas necessidades e teriam a oportunidade de revisar, selecionar ou mesmo descartar essas opções sempre que preciso.

O Programa compreende sete metas que reforçam a importância do trabalho prospectivo e da aplicação e metodologias baseadas em estudos comparativos de tecnologias abertas, tornando a rede Cariniana o principal centro brasileiro de referência em preservação digital, o que pode ser mensurado por meio das parcerias, publicações, transferência de tecnologias e itens preservados. A previsão é que o Programa continue atuando como um facilitador, conectando as instituições proponentes de projetos a fontes de boas práticas, permitindo-lhes revisar ações de técnicas e de gestão de preservação digital, por meio de seus próprios sistemas, escolhendo importar essas ações em seus próprios ambientes locais. Da mesma forma, as instituições usuárias teriam a capacidade de disponibilizar seus próprios fluxos de trabalho complexos para todos, garantindo que os dados sejam compartilhados com os usuários dos serviços do Programa.

O futuro do Programa de Preservação Digital Brasileiro do Ibict responde por demandas em vários níveis, incluindo o técnico, político e organizacional, todos os quais devem ser considerados dentro de um modelo de ecossistema de serviços de preservação digital. Fundamentalmente, embora o Ibict seja uma instituição federal, a ênfase dos serviços em rede tem uma boa relação custo-benefício e

⁴ Estrutura organizacional da Rede Cariniana
<https://cariniana.ibict.br/index.php/institucional/estrutura-organizacional>

promove a colaboração e a utilização de padrões, o que permitiu que ele integrasse com sucesso aspectos de sustentabilidade de um ecossistema informacional para que a preservação de recursos seja parte integrante de seu sucesso geral.

Figura 1 - Ecossistema de preservação digital do Ibict



Fonte: elaboração própria

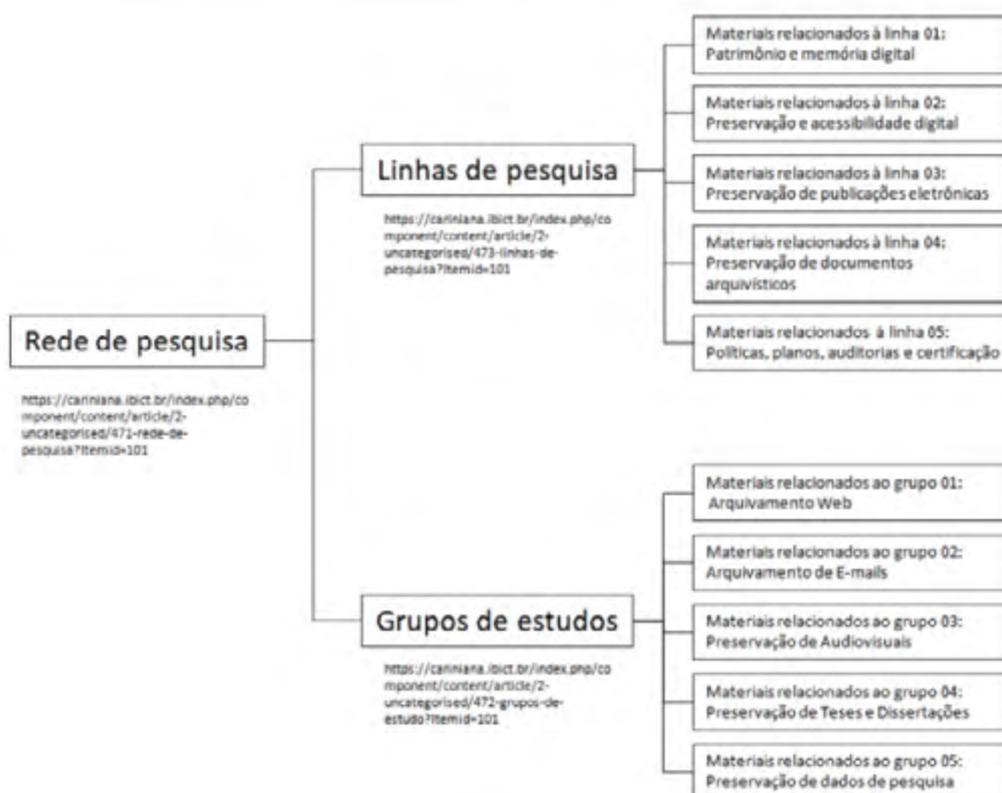
A proposta do Ibict para um ecossistema de preservação digital passa pela promoção de uma rede de colaboração, estruturação de serviços aplicáveis no contexto da preservação, bem como definição de infraestruturas e tecnologias capazes de suportar este movimento informacional que busca preservar o conhecimento a longo prazo. A experiência de décadas promovendo a preservação digital por meio de mecanismos e ações sistematizados junto à comunidade dá respaldo para que essa proposta encontre caminhos para se manter viável em um ambiente que está em constante transformação.

◆ 3 REDE DE PESQUISA CARINIANA

A pesquisa é um dos pilares ativos do Ecossistema de Preservação Digital do Ibict. Um dos vários grupos de pesquisa basilares desta rede é o grupo “Estudos e Práticas de Preservação Digital – Rede Dríade”⁵. Ele concentra seu trabalho em linhas e grupos de estudo baseados em comunicações acadêmicas e dados de pesquisa e projetos individuais de mais de 100 estudantes, professores e pesquisadores do Brasil e do Exterior. De muitas maneiras, incluindo um nível de compartilhamento de esforços na construção de produtos de informação, o compromisso de trabalhar juntos e dentro de um modelo de liderança aberta, o interesse de colaborar para a preservação de materiais digitais tem pavimentado o caminho da Rede Cariniana (TAVARES; FREIRE, 2020).

5 A Rede de Pesquisa Dríade <https://cariniana.ibict.br/index.php/grupos-de-pesquisa>

Figura 2 – Rede de Pesquisa em Dríade



A rede de pesquisa, seja por meio do grupo de pesquisa ou dos eventos científicos que promove, possui um papel primordial na estrutura do ecossistema de preservação proposto pelo Ibict. É ela quem dá o suporte teórico para o avanço da temática, bem como para a definição dos padrões que serão promovidos ou mesmo das tecnologias que serão adotadas nas ações e serviços de preservação. Além disso, a rede Cariniana é responsável pelo suporte organizacional ao ecossistema de preservação digital, pois sistematiza institucionalmente a promoção de suas políticas e ações.

◆ 4 SERVIÇOS

O segundo pilar do ecossistema de preservação digital proposto pelo Ibict está nos serviços que a instituição propõe, oferece ou apoia. Estes serviços representam a materialização de ações pensadas a partir dos padrões e tecnologias fomentadas no pilar anterior. É por meio dos serviços que se busca aplicar os conceitos de preservação digital ao mesmo tempo em que há o atendimento a demandas da sociedade relacionadas à preservação de objetos informacionais digitais.

O primeiro serviço oferecido neste ecossistema foi o de preservação digital de periódicos científicos. Este serviço, ativo há anos, conta com uma grande quantidade de periódicos científicos preservados e indexados por meio de uma estrutura integrada de sistemas. Seu metiê está associado à necessidade de garantir o acesso a longo prazo aos avanços teóricos oriundos da pesquisa científica brasileira. Para que seja viabilizado, este serviço foi estruturado em parceria com diversas instituições, responsáveis pela coleta, tratamento e armazenamento dos objetos digitais destacados para preservação. Ele usa ainda padrões e tecnologias nacionais e internacionais que visam garantir a interoperabilidade entre os distintos ambientes institucionais.

Outros serviços como a preservação de dados de pesquisa, focado nos dados oriundos de pesquisas científicas financiadas ou realizadas por instituições de ensino e pesquisa também foram estruturados a fim de pavimentar a atuação do Ibict na temática. No caso específico da preservação de dados de pesquisa, há interesse em integrar essa iniciativa com outras áreas do Instituto que discutem o acesso aberto ou mesmo a incorporação destes dados em cenários inovadores.

Há ainda o desejo de estruturar serviços capazes de atender a outras áreas da preservação da memória digital, como a preservação de páginas da internet. Neste caso, a ideia inicial é preservar as páginas de governo eletrônico, uma vez que várias comunicações públicas são feitas atualmente por este meio. Diante um cenário de constante atuação tecnológica devido à evolução das tecnologias de tratamento e apresentação de conteúdos na internet, bem como à vulnerabilidade que estes sistemas *on-line* possuem, a preservação deste tipo de conteúdo garantirá o acesso a longo prazo às construções realizadas e disseminadas por este meio. Em uma segunda etapa ainda é estudado a possibilidade de se fazer a preservação de todo o conteúdo hospedado sob a denominação .br. Este passo, no entanto, demanda um planejamento acerca da sustentabilidade financeira da iniciativa.

Discute-se também no Ibict a construção de outros dois serviços no arcabouço da preservação digital. De um lado há a intenção de disponibilizar um sistema para construir Planos de Gestão de Dados (PGD). Este documento contém, entre outras coisas, as informações acerca do cuidado que se terá com os dados coletados e armazenados durante projetos de pesquisa. Com a disponibilização do serviço de geração de PGD, pretende-se que a preocupação com os aspectos de preservação possam ser consideradas desde o início da pesquisa. Há ainda a intenção de aprofundar os estudos acerca da preservação de caixas de correio eletrônico. Este tipo de instrumento de comunicação é o principal utilizado no meio corporativo e por isso entrou no radar dos serviços a serem estruturados no âmbito do Instituto. Neste caso, as preocupações em incorporar os aspectos da Lei Geral de Proteção de Dados Pessoais (LGPD) se fazem extremamente importantes, dado o caráter pessoal deste tipo de instrumento.

Com esta série de serviços sistematizados no contexto do Ibict, têm-se o segundo pilar do ecossistema de preservação digital. Estes serviços são estruturados e disponibilizados de acordo com o marco teórico estabelecido pela rede de pesquisa e desenvolvidos seguindo os requisitos identificados no pilar de infraestrutura e tecnologias.

◆ 5 INFRAESTRUTURAS E TECNOLOGIAS

O Programa conta com um conjunto de ferramentas de preservação digital de código aberto, abordando problemas comuns encontrados nos estágios de ingestão e fluxos de processos de preservação digital. As ferramentas e a sua documentação estão sob a administração da Coordenação de Tecnologia da Informação e Informática (CGTI), são disponibilizados gratuitamente com licenças abertas e podem ser adaptados para atender às necessidades de quase qualquer organização.

As equipes técnicas dos projetos do Ibict fornecem liderança e coordenam o desenvolvimento de serviços por meio de testes e repasse de tecnologias, aproveitando a infraestrutura aberta e seguindo as melhores práticas de acesso aberto.

Dessa forma, a infraestrutura tecnológica do Ecossistema de Preservação Digital representa o terceiro pilar em que faz o uso de algumas ferramentas previamente definidas, baseando-se no objeto digital que será preservado.

Assim, algumas ferramentas de preservação digital foram desenvolvidas e customizadas para atender as necessidades das instituições parceiras, buscando Lockss.

O Lockss foi a primeira ferramenta utilizada pela Rede Cariniana para realizar a preservação de periódicos científicos e teses e dissertações que utilizam como repositório de guarda o DSpace.

O exemplo do modelo RDCArq Hipátia também ilustra algumas experiências que as instituições usuárias do Programa de Preservação do Ibict compartilham no trabalho conjunto para desenvolver programas relacionados à preservação de benefício mútuo. No entanto, embora a necessidade de preservação digital e os benefícios dos serviços compartilhados sejam bem compreendidos pelos usuários desses sistemas, o serviço de suporte do Ibict precisa do trabalho de equipes locais na sua implementação de forma consistente, particularmente em algumas das instituições menores.

Para resolver a lacuna entre o armazenamento de preservação e os serviços de preservação completos, algumas instituições começaram a usar o *software* de preservação digital Archivematica, e o Ibict formou um grupo de trabalho para discutir os resultados dos testes e as experiências e fluxos de trabalho de algumas implementações de produção local. Várias outras instituições da área jurídica estavam interessadas em implementar o Archivematica, mas não tinham infraestrutura local para apoiá-lo.

◆ 6 PRÓXIMOS PASSOS

Até a data atual, o progresso na implementação do Ecossistema de Preservação Digital no Ibict tem variado por várias razões. Desde o início, a Cariniana tem trabalhado com todas as instituições parceiras no treinamento, planejamento de preservação e operacionalização dos serviços e, no geral, o progresso tem sido substancial. Durante o período de implementação inicial, vários aspectos não previstos surgiram. Em primeiro lugar, a demora na formalização dos acordos de cooperação, na avaliação das equipes jurídicas das instituições e no convencimento das equipes técnicas locais de que a preservação de registros organizacionais usando ferramentas de código aberto é consistente com as políticas de segurança das instituições.

Com a adoção de outras soluções de preservação digital, a questão da segurança foi tomando lugar de destaque na estruturação dos serviços. Todas as iniciativas de preservação digital usadas como referência para a construção dos serviços do Programa consideram a segurança como uma parte importante do Modelo de Referência OAIS, mas elas são “informativas” e não “normativas”, o que significa que elas simplesmente apontam problemas de segurança e definem o que um sistema compatível deve fazer para resolvê-los e não estipulam um modelo técnico de segurança específico. Por esse motivo, todas as estratégias de preservação digital recomendadas pelo Programa visam estar em conformidade com o modelo funcional OAIS, implementando controles de acesso específicos no conteúdo sob sua supervisão, mas esses controles também devem ser consistentes com as políticas locais que tratam de segurança, privacidade e retenção de registros.

Outra barreira na implementação de soluções de preservação digital está ligada à necessidade de recursos de infraestrutura tecnológica e à ausência de políticas e planos de preservação digital que definam as prioridades e responsabilidades no início da implementação dos serviços. Mesmo que algumas das instituições usuárias começaram a implementar essas políticas quando os projetos se tornaram oportunidades concretas para concentrar suas prioridades em torno de estratégias operacionais.

Outro empecilho relacionado com os anteriores tem sido o que está ligado com a integração das soluções de preservação digital com sistemas de gestão locais e repositórios institucionais. Algumas dessas plataformas requerem trabalhar com as áreas centrais de TI ou em setores de arquivo ou bibliotecas. A integração com aplicativos externos pode depender das políticas de segurança aplicadas à infraestrutura local. A oportunidade de que as equipes técnicas na gestão da importação e exportação de informações trabalhem juntas é facilitada quando a solução de preservação digital funciona como um serviço hospedado e se investe na sua implementação e integração.

A Coordenação da CGTI continuará encorajando as instituições custodiadoras de informação digital relevantes no Brasil a usarem os serviços do Programa como uma prioridade; eles combinados fornecem um modelo flexível para provisionar a infraestrutura necessária para atender às futuras demandas de preservação digital. Também o Programa continuará explorando o desenvolvimento de serviços a serem compartilhados com toda a comunidade, modelados a partir dessas primeiras experiências com as soluções adotadas até o momento. Uma primeira experiência poderá ser o desenvolvimento de um novo serviço compartilhado de arquivamento de conteúdos da Internet no projeto de arquivamento colaborativo da web do Ibict.

◆ REFERÊNCIAS BIBLIOGRÁFICAS

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (Ibict).

Plano diretor da unidade: PDU. 2018 – 2022. Brasília, DF. 2017.

RESMINI, A.; ROSATI, L. A brief history of information architecture. **Journal of Information Architecture**, [s.l.], v. 3, n. 2, 2012. Disponível em: <https://journalofia.org/volume3/issue2/03-resmini/>. Acesso em: 20 ago. 2021.

TAVARES, A.; FREIRE, I. Preservação digital: estudo exploratório sobre a literatura científica e as redes sociais colaborativas no Brasil. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, Rio de Janeiro, v. 14, n. 3, 2020. DOI <https://doi.org/10.29397/reciis.v14i3.2108>. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/2108>. Acesso em: 28 ago. 2021.

COMO CITAR:

OLIVEIRA, Alexandre Faria de; MÁRDERO ARELLANO, Miguel Ángel; BRAGA, Tiago Emmanuel Nunes. A visão do Ibict para um ecossistema nacional de preservação digital. In: BRAGA, Tiago Emmanuel Nunes; MÁRDERO ARELLANO, Miguel Ángel. (org.). **Hipátia**: modelo de preservação para repositórios arquivísticos digitais confiáveis. Brasília: Ibict, 2022. p. 91-100. (Informação, Tecnologia e Inovação, v. 1). DOI [10.22477/9786589167501.cap7](https://doi.org/10.22477/9786589167501.cap7).

Resumo

Por meio do Hipátia, propõe-se um modelo de preservação digital arquivística. A pesquisa que resultou no modelo foi executada pelo Instituto Brasileiro de Informação em Ciência e Tecnologia com o apoio do Tribunal de Justiça do Distrito Federal e Territórios.

Em sete capítulos, este livro apresenta as primeiras percepções quanto à prospecção, o estudo, a proposição e o desenvolvimento do Hipátia, que é um modelo capaz de se adaptar a qualquer sistema que faça gestão de processos e objetos digitais. Temáticas como a preservação de documentos digitais e os Repositórios Arquivísticos Digitais Confiáveis (RDC-Arq) permeiam todas as discussões apresentadas pelos autores que participaram dos processos que resultaram no Modelo Hipátia.



Apoio:



Organização
das Nações Unidas
para a Educação,
a Ciência e a Cultura

Realização:

