

REFLEXÕES SOBRE A IMPLANTAÇÃO DO REPOSITÓRIO DIGITAL ARQUIVÍSTICO CONFIÁVEL DA PINACOTECA DE SÃO PAULO

Eliane Barbosa¹

ebarbosa@pinacoteca.org.br

Isabel Cristina Ayres da Silva Maringeli²

imaringelli@pinacoteca.org.br

Resumo

A Pinacoteca de São Paulo deu início à elaboração de sua Política de Preservação Digital em 2016 (a primeira versão foi publicada em 2017 e uma versão atualizada foi finalizada em 2019 e está disponível no website institucional). A partir do diagnóstico realizado para elaboração da Política, foi possível identificar uma quantidade expressiva de objetos digitais e, em alguns casos, foram identificados documentos descontextualizados devido à perda dos metadados descritivos. Outro fato apontado na ocasião foi a constatação de que alguns arquivos estavam inacessíveis por estarem corrompidos ou por terem sido salvos em formatos desconhecidos ou que dependem de softwares que não existem mais ou deixaram de ser adquiridos pela instituição. A Política fornece diretrizes e recomendações para guarda segura dos documentos digitais de guarda permanente e foi a base para a elaboração de um Projeto Piloto para implantação de um Repositório Arquivístico Digital Confiável (RDC-Arq). O Projeto foi submetido à apreciação do Comitê Gestor de Preservação Digital interno (formado por membros dos departamentos a seguir: Diretoria, Tecnologia da Informação, Pesquisa e Curadoria, Acervo Museológico, Comunicação, Conservação e Restauro e Centro de Documentação e Memória - Cedoc). O Comitê Interno e a Diretoria da Pinacoteca aprovaram o Projeto e foi iniciado um estudo comparativo que culminou com a escolha do software Archivematica. Este artigo descreve as ações realizadas no contexto do Projeto Piloto, realizado no período de agosto a dezembro de 2019, bem como as ações realizadas a partir da publicação da Política de Preservação Digital (atualização do quadro de arranjo e elaboração de um manual de recolhimento de objetos digitais). Após a instalação do Archivematica e realização dos treinamentos para instalação e operacionalização, foi dado início à submissão dos documentos selecionados para essa fase inicial. Durante esse processo foi possível perceber que o maior desafio em relação à preservação digital na Pinacoteca será a realização das etapas do tratamento documental. Nos deparamos com uma massa documental digital acumulada por anos e formada por cerca de 700.000 arquivos que equivalem a 4,9 TB. Foram identificadas cerca de 600 extensões de arquivos, muitas delas desconhecidas ou com erros de digitação, além de arquivos em duplicidade. Antes de 2016 as ações foram pautadas pela recomendação de que um storage com capacidade suficiente de armazenamento poderia garantir a preservação dos documentos. Porém constatou-se que os documentos digitais necessitam ser verificados no que diz respeito à sua integridade, além de serem analisados, classificados, padronizados e submetidos aos princípios da arquivística, antes de sua submissão para guarda permanente. Somados a isso existem ainda os documentos que necessitam de estratégias de migração para serem acessados (o acervo conta com mais de 5.000 mídias em CD e DVD, parte inacessível devido à falha nas mídias de transporte). O Projeto Piloto reforçou a necessidade de adoção de procedimentos técnicos e diretrizes consolidadas para práticas de preservação digital em longo prazo. Objetos digitais como bancos de dados, e-mails e obras de arte, bem como os documentos híbridos devem ser cuidadosamente avaliados para não perderem seu contexto. A Pinacoteca não possui um sistema de gestão documental, portanto a cadeia de custódia também se apresenta como outro desafio a ser superado. Conclui-se que o planejamento das etapas que precedem o uso do Archivematica é essencial para garantir a preservação dos documentos, bem como torna-se

¹ Pinacoteca de São Paulo, São Paulo/SP, Brasil.

² ECA/USP, Pinacoteca de São Paulo, São Paulo/SP, Brasil.

imprescindível que a instituição tenha uma política de preservação digital bem definida e consolidada.

Palavras-chave: 1. Archivemática. 2. Repositório Arquivístico Digital Confiável (RDC-Arq). 3. Política de Preservação Digital 4. Arquivos de Museu. 5. Preservação Digital.

1 Introdução

A obsolescência tecnológica caracteriza-se pela subordinação do documento digital a um contexto tecnológico que o torne compreensível para os humanos e que o torne, portanto, acessível (FERREIRA, 2006). Para Santos e Flores o ciclo de obsolescência se dá em etapas que não se alteram: depois do surgimento de uma nova tecnologia, se segue a adesão ou rejeição a ela, seu declínio até que se torne obsoleta com o surgimento de uma nova tecnologia.

Nesse contexto, evidencia-se a fragilidade do documento digital perante a preservação em longo prazo. A preservação de documentos arquivísticos deve ser pautada a partir dos conceitos de patrimônio digital, já discutidos no âmbito da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO). A instituição, por meio do *UNESCO Charter on the Preservation of Digital Heritage*, tem abordado essas questões há um tempo, e algumas recomendações foram adotadas pelo Conselho Nacional de Arquivos (Conarq).

O entendimento de que os documentos digitais são patrimônio a ser preservado é essencial para que o planejamento de ações de preservação digital em um arquivo permanente possa ser realizado.

Na Pinacoteca de São Paulo a preservação digital adquiriu uma discussão e foco mais amplos a partir da constatação de que documentos digitais de guarda permanente se tornaram obsoletos e inacessíveis. Tal fato impulsionou os estudos e ações internas que serão descritos nesse artigo. Inicialmente relataremos o processo de elaboração de nossa Política de Preservação Digital. Depois comentaremos alguns aspectos conceituais sobre o Repositório Digital Arquivístico Confiável (RDC-Arq) e finalmente descreveremos as etapas do projeto piloto realizado.

2 Política de preservação digital

A Política de Preservação Digital da Pinacoteca de São Paulo foi publicada inicialmente em 2017, sendo atualizada em 2019. Trata-se de documento que fundamentou as ações realizadas na Pinacoteca, pois, além de se constituir por um conjunto de regras e princípios, “Uma política de preservação digital fornece a estrutura para ação e planejamento para garantir a manutenção e preservação de longo prazo dos documentos de uma organização. (INTERPARES; ICA, 2012, p. 10).

Para (GRÁCIO, 2010, p. 10) “Uma política de preservação digital deve estabelecer diretrizes que garantam a qualidade e autenticidade da informação, de forma que possa ser acessada e recuperada no futuro”.

Nesse contexto, elaborou-se um diagnóstico para mapear as práticas e estruturas existentes de modo que se pudesse ter uma visão panorâmica do estágio atual e do que se deseja para o futuro, por meio da elaboração de estratégias de preservação.

O documento da Pinacoteca, disponível para consulta no website institucional, tem por objetivos: estabelecer diretrizes para preservação do patrimônio digital da Pinacoteca; proporcionar confiabilidade, autenticidade e acessibilidade a longo prazo e estabelecer diretrizes e estratégias de preservação (PINACOTECA, 2019, p. 7).

De acordo com Luz e Maringelli:

Nas estratégias operacionais foram definidos os padrões a serem adotados de forma macro na estratégia e na política de preservação digital, além da definição de metadados de preservação digital e montagem de infraestrutura. Faz parte da estratégia principal para a instituição, estabelecido a partir de então, o uso de um repositório digital confiável arquivístico (RDC-Arq). (LUZ; MARINGELLI, 2016, p. 194).

Neste artigo abordaremos uma das principais estratégias adotadas pela Pinacoteca para preservação de seus documentos digitais: a implementação de um RDC-Arq, cujos conceitos são apresentados no próximo tópico. Cabe ressaltar que a Política propôs a constituição de um Comitê Gestor interno, formado por diversas áreas do Museu, responsável pelos processos decisórios que envolvem o patrimônio digital institucional. Tal modelo de gestão se mostrou bastante produtivo, pois a diversidade de conhecimentos das áreas externas ao Cedoc, portanto fora do âmbito da arquivística,

proporcionaram debates enriquecedores sobre quais os caminhos seriam viáveis para o museu. Após diversos estudos e análises o Comitê Gestor optou por utilizar o software Archivematica.

3 Repositório Digital Arquivístico Confiável

De acordo com o Conarq

um repositório digital arquivístico confiável é definido como: “um ambiente de armazenamento e gerenciamento de materiais digitais. Esse ambiente constitui-se de uma solução informatizada em que os materiais são capturados, armazenados, preservados e acessados. (CONARQ, 2015, p. 9)

Ainda de acordo com a Resolução 43 do Conarq, um repositório digital confiável deve ser “capaz de manter autênticos os materiais digitais, de preservá-los e prover acesso a eles pelo tempo necessário” (CONARQ, 2015, p. 10).

Para que os repositórios digitais possam cumprir com sua função principal, que é possibilitar o acesso a longo prazo dentro dos requisitos de integridade e autenticidade, atributos arquivísticos que visam garantir a confiabilidade das informações, é necessário que este ambiente esteja em conformidade com modelos e padrões.

O modelo Open Archival Information System (OAIS) estabelece diversos critérios para o armazenamento seguro das informações depositadas nos repositórios, entre os quais a inclusão de metadados de descrição, administração e preservação dos objetos digitais dentro de padrões estabelecidos.

Atualmente o Modelo OAIS é considerada pelo Conarq, a norma mais importante da área de preservação digital. Modelo conceitual desenvolvido pelo Consultive Committee for Space Data Systems – CCSDS, que resultou na norma ISO 14721:2003. O OAIS descreve as funções de um repositório digital e os metadados necessários para a preservação.

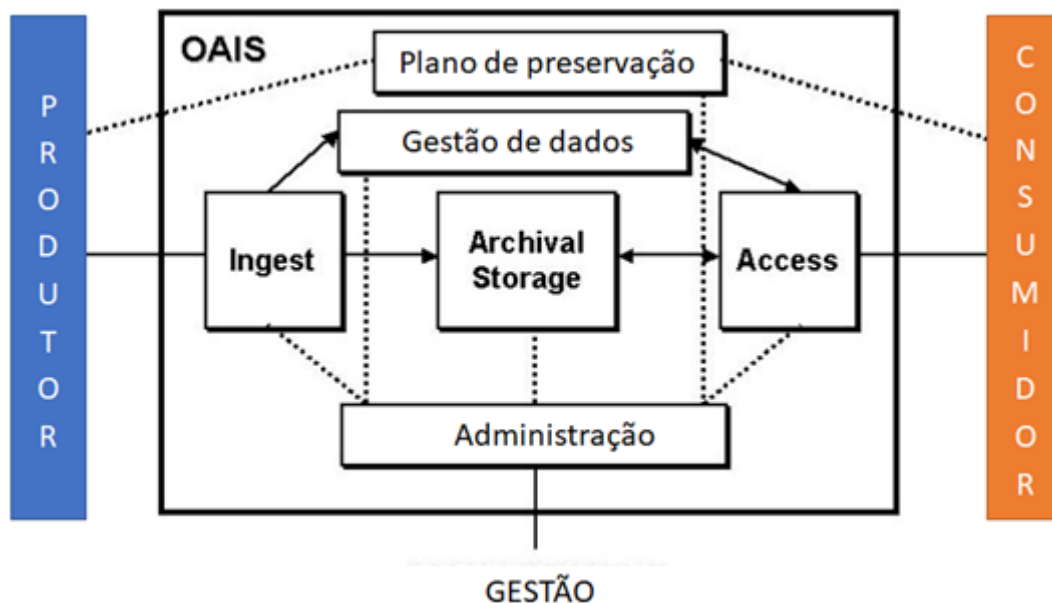


Figura 1 Modelo de Gestão OAIIS -adaptação de: (Rogers 2012)

De acordo com CCSDS, (1998 apud ANGEL ARELLANO, 2004,)

Um arquivo nos termos do OAIIS significa considerar os arquivos como organizações de pessoas e sistemas, que aceitaram a responsabilidade de preservar a informação e torná-la disponível e melhorar a comunicação e produtividade entre diferentes comunidades.

Ainda de acordo com (ANGEL ARELLANO, 2004, p. 19):

Por ser um modelo de referência, ele não é uma implementação específica, mas proporciona uma lista de condições do que deve ser considerado no estabelecimento de um projeto de preservação de qualquer tipo de documento, seja digital ou analógico.

O modelo OAIIS descreve as funções de um repositório digital, bem como os metadados necessários para a preservação e acesso aos objetos digitais gerenciados pelo repositório, constituindo desse modo um modelo funcional e um modelo de informação.

O OAIIS contempla desde a criação dos pacotes de informação, considerados na admissão, captura, arquivamento, planejamento, preservação, armazenamento e gerenciamento das informações e de acesso

Uma das importantes características de um repositório digital arquivístico confiável, são os metadados, que de acordo com a *ISO 2308*, norma internacional que trata especificamente dos metadados, e traz em seu conteúdo a importância da inserção

desses dados para preservação dos documentos digitais. Eles podem ser conceituados de diversas formas, entre as quais, a mais comum é: “dados sobre os dados”, ou “informação sobre informação”. A ISO 15489, que trata de gestão de documentos, define metadados como: Informação que descreve o contexto, conteúdo e estrutura do documento arquivístico e da sua gestão ao longo do tempo

Um esquema de metadados de preservação digital inclui três classes conceituais, sendo os metadados descritivos, estruturais e os administrativos.

Descritivos: “descrevem um recurso com o propósito, por exemplo, de descoberta ou identificação. Isso pode incluir elementos como título, resumo, autor e palavras-chave”.

Estruturais: “indicam como objetos compostos são colocados juntos, por exemplo, como é que páginas são ordenadas para formar capítulos”.

Administrativo oferece informações sobre como os arquivos foram criados e armazenados, direitos sobre propriedade intelectual, informações sobre o objeto original a partir do qual o objeto foi derivado, informações sobre a proveniência dos arquivos que compõe o objeto.

3.1 *Archivematica*

O Archivematica é um conjunto integrado de ferramentas de software de código aberto que permite aos usuários processarem objetos digitais desde o ingresso até o acesso, em conformidade com o modelo funcional ISO-OAIS.

O Sistema é flexível e customizável, é compatível com diversos formatos, ou seja, o archivematica contempla todas as características de um RDC-Arq, sendo o principal objetivo deste repositório o armazenamento seguro do acervo digital arquivístico, com a finalidade de possibilitar o acesso as informações por tempo indeterminado.

Após levantamento dos Repositórios Digitais Arquivísticos Confiáveis disponíveis e suas características, e de acordo com critérios pré-estabelecidos pelo Comitê interno e diretorias da Pinacoteca, além de consultoria com especialistas em preservação digital, a instituição implantou o sistema archivematica, para armazenamento de seu acervo digital com objetivo de evitar perdas informacionais e

possibilitar acesso seguro aos objetos digitais de modo íntegro, autêntico e confiável a longo prazo.

Os principais critérios considerados pela Pinacoteca para seleção do Archivematica foram:

- Aplicativo de código aberto
- Está em conformidade com modelo OAIS
- É flexível e customizável
- Garante os metadados por meio do METS, PREMIS e Dublin Core
- É compatível com centenas de formatos
- É indicado por especialistas em preservação digital
- Possibilita pesquisa avançada e gerenciamento de armazenamento
- Tem como objetivo garantir acesso aos objetos digitais no futuro de modo autêntico, íntegro e seguro
- Possui uma comunidade ativa de usuários com objetivo de colaboração entre si.
- Está em constante aprimoramento, de modo constante e ativo por equipe comprometida com as atualizações.

De acordo com as características acima citadas o archivematica foi o repositório selecionado e implementado na Instituição em 2019. A finalidade é assegurar o acesso ao acervo digital da instituição de modo que todas as características arquivísticas e de preservação digitais sejam cumpridas, bem como o acesso a longo prazo seja possível.

4 Projeto Piloto

O acervo digital da Pinacoteca de São Paulo foi acumulado por mais de uma década e tratado sob o ponto de vista do acesso, sem que a questão da preservação digital tenha sido considerada. Quando falamos em massa documental geralmente pensamos em documentos acumulados sem organização. É sabido que algumas experiências de organização englobam a digitalização de documentos como forma de preservação. Essa prática foi adotada no Cedoc até 2015, quando se percebeu que a digitalização de arquivos necessitava de um planejamento para garantir o acesso futuro diante das obsolescências de software e hardware.

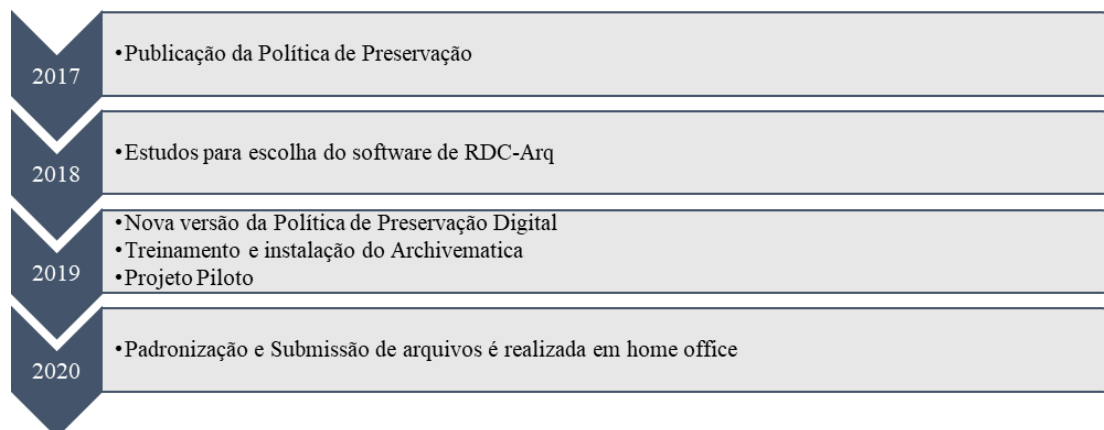


Figura 2 – Linha do tempo da Preservação Digital da Pinacoteca (Fonte: Elaboração das autoras)

Para atender a uma das principais estratégias de preservação indicadas na Política de Preservação Digital citada anteriormente foi apresentado um Projeto Piloto para implementação e testes com o Archivematica.

O projeto foi aprovado pelo Comitê Gestor de Preservação Digital da Pinacoteca (formado por membros dos núcleos: Diretoria, TI, Pesquisa e Curadoria, Acervo Museológico, Comunicação, Conservação e Restauro e Cedoc) e foi realizado no período de 12 de agosto a 30 de dezembro de 2019. O projeto foi realizado nas etapas descritas a seguir:

- Instalação do software Archivematica;
- Treinamento geral para equipe Cedoc;
- Treinamento específico para o Núcleo de Tecnologia da Informação;
- Realização de estudos de aprofundamento nos padrões de metadados de preservação digital e demais recursos utilizados pelo Archivematica;
- Elaboração de fluxograma de tramitação interna de objetos digitais;
- Desenvolvimento de manual de uso do software;
- Atualização da Política de Preservação Digital

A seleção de documentos que fizeram parte do Projeto Piloto foi baseada em demandas de preservação que surgiram durante o processo de elaboração da Política, mais especificamente a partir das reuniões do Comitê Gestor. A seguir indicamos os documentos e respectivos locais de guarda:

- Dossiê de exposição (Fundo Institucional)

- Fotografia de eventos expositivos (Fundo Institucional)
- Planos anuais do MinC (Fundo Institucional)
- Filmagens de cursos e palestras (Fundo Institucional)
- Obras de arte/entrevista INCCA (Núcleo de Conservação e Restauro)
- Atas do COA (Núcleo de Acervo Museológico)

Cabe destacar que alguns documentos foram avaliados nessa etapa, mas não foram contemplados no projeto. Um exemplo são os e-mails. Esse conjunto necessita de uma política de gestão de documentos que compreenda a produção e uso. A gestão de correio eletrônico possui recomendações específicas do Conarq, porém, embora a Pinacoteca possua normas para utilização do correio eletrônico, faz-se necessário um programa de conscientização e uma política para seleção do que deve ser classificado como documento de guarda permanente dado o volume produzido.

Após a realização dos treinamentos e após a seleção do que seria submetido ao Archivemática, foi necessário avaliar as práticas já existentes no Cedoc. A descrição arquivística na Pinacoteca começou a ser realizada de forma sistemática somente a partir de 2015, quando foi realizada a migração dos dados registrados em MS Excel, para o novo banco de dados no sistema Automação de Bibliotecas e Centros de Documentação (ABCD). Nesse processo alguns campos foram modelados de acordo com a Norma Geral Internacional de Descrição Arquivística (ISAD-G), porém outros foram criados de acordo com as necessidades institucionais, sendo campos criados localmente.

Nesse contexto foi necessário fazer uma tabela de equivalência com os campos existentes para que pudéssemos adotar o Dublin Core, padrão de metadados escolhido para registro dos metadados descritivos pois ele “vem ganhando destaque em diversos ambientes informacionais devido a sua simplicidade, flexibilidade e abrangência” (ALVES; SANTOS, 2012, p. 72), o que o torna atrativo para os repositórios em geral.

Outro ponto a ser destacado é que o Archivemática traz uma política de preservação padrão, com indicação de formatos de preservação baseada no Pronom¹, um recurso online que agrega informações sobre os diversos formatos de arquivo conhecidos. Tal diversidade de formatos nos levou a estudos mais aprofundados para ampliação do entendimento sobre as escolhas que poderiam ser feitas.

Em linhas gerais o processo pode ser visualizado nas figuras 3 e 4.

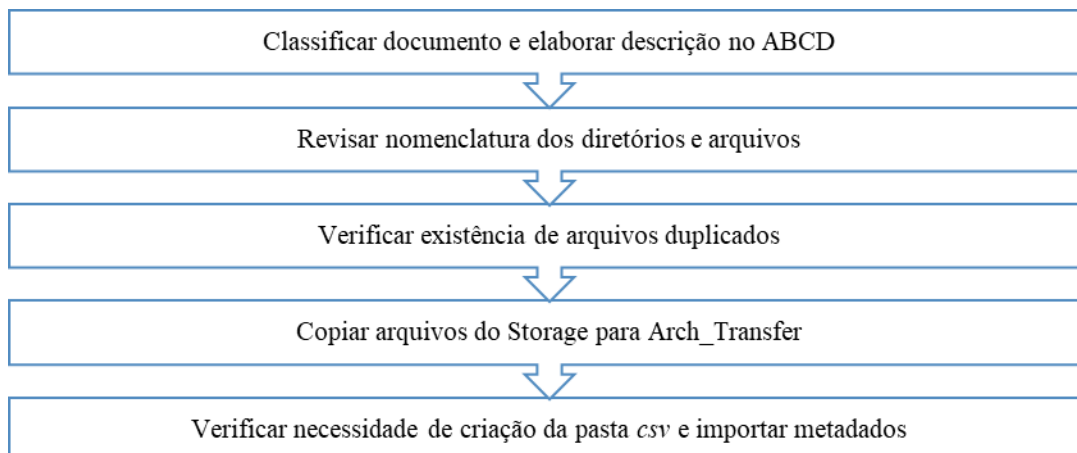


Figura 3 – Etapas da fase pré-repositório (Fonte: Elaboração da Equipe do Cedoc)

A Figura 3 apresenta o processamento operacional das etapas de preparo dos arquivos na fase pré-repositório para submissão ao Archivematica. Para essa etapa foram selecionados documentos que ainda não estavam descritos no banco de dados. A verificação de duplicidade foi necessária devido ao fato de os documentos digitais não passarem, até então, pelo processo de recolhimento habitual.

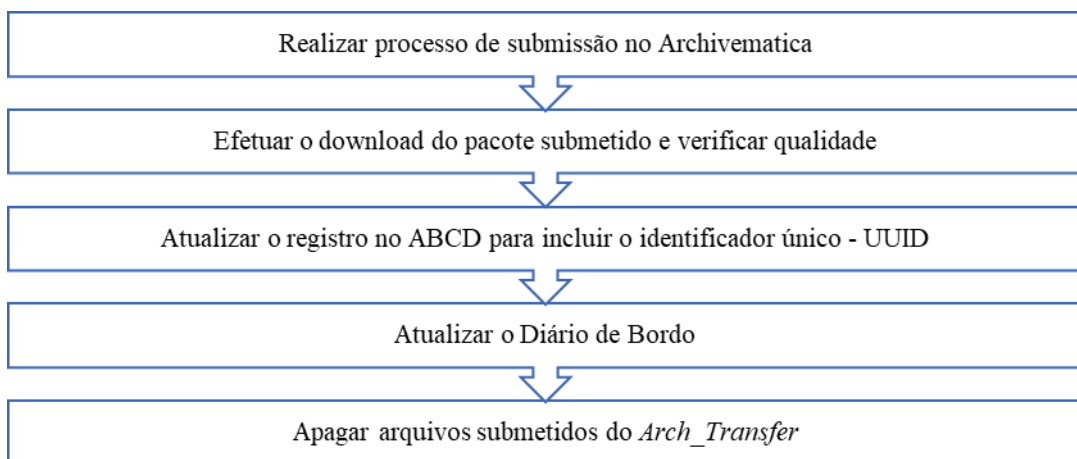


Figura 4 – Etapas do processo de submissão ao Archivematica (Fonte: Elaboração da Equipe do Cedoc)

A figura 4 mostra as etapas do processo de submissão ao repositório. Após a submissão foi realizado o download para verificar se o processo foi concluído de forma satisfatória. Depois procede-se à atualização do metadados descritivos do documento no banco de dados, bem como do diário de bordo onde foram registradas todas as ações.

É importante ressaltar que devido ao fato dos objetos digitais selecionados já se encontrarem no *Storage* institucional, tornou-se inviável verificar a autenticidade em alguns casos. No período do projeto o *Storage* contava com cerca de 748.627 arquivos (6,2 TB). Diante dessas dimensões, e considerando a diversidade de formatos de arquivos nesse contexto, será necessário realizar um estudo de quais formatos são contemplados na conversão automática do Archivematica e quais são de preservação.

Do ponto de vista da Diplomática, a autenticidade não é verificada no momento da produção do documento e nem sempre foi possível identificar a cópia do original, e questões relacionados à análise tipológica de documentos digitais se mostraram mais complexas do que os documentos não digitais, conforme apontado por DELMAS:

A “Diplomática digital” não pode se limitar ao reconhecimento da autenticidade. Para esta, o processo digital deve, como para os arquivistas responsáveis pela conservação perene dos dados, situar-se no ponto de origem, no momento de criação do documento. (Delmas, 2015, p. 50).

O autor defende que não é possível a criação de um documento digital sem a influência de um arquivista diplomata. Tal afirmação pode ser corroborada a partir da experiência relatada nesse artigo. A participação do documentalista não pode ser limitada à organização documental, especialmente no caso dos objetos digitais que possuem estrutura mais complexa e mais dependente de tecnologias externas.

5 Considerações finais

A execução do Projeto Piloto possibilitou a prática com o Archivematica de um lado, e o estudo de conceitos sobre as características do documento digital de outro. A partir do projeto foi possível delinear um plano de ações para depósito dos objetos digitais de guarda permanente no Archivematica.

Cabe destacar que nos deparamos com questionamentos em relação ao tratamento arquivístico dado aos documentos digitais, tanto no acervo retrospectivo quanto no acervo nato-digital, recém produzido. Desde a compreensão do que é o objeto digital, suas características e especificidades até o tratamento arquivístico usual, o qual é embasado em princípios constituídos no meio não digital.

Muitas são as fragilidades encontradas no universo digital, e ainda que o Archivematica e a Política de Preservação Digital se configurem como soluções plausíveis, ainda há muito trabalho a ser feito no campo da organização documental do

acervo digital retrospectivo, e também no que podemos denominar de arquivo corrente, produzido durante o período de pandemia pelo qual sobrevivemos, e onde a Pinacoteca teve uma produção de documentos digitais em grande escala e que dada à rapidez e sobreposição de eventos no mundo virtual, já fazem parte da nossa memória, sendo portanto, objetos de guarda permanente e acesso imediato.

Referências

ALVES, Rachel Cristina Vesú, SANTOS, Plácida Leopoldina Ventura Amorim da Costa. *Metadados no domínio bibliográfico*. Rio de Janeiro: Intertexto, 2012.

ANGEL ARELLANO, Miguel. Preservação de documentos digitais. *Ciência da Informação (on line)*. Brasília v.33, n.2, p.197-217, maio./ago. 2004. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652004000200002. Acesso em: 16 de dez. 2020.

ARCHIVEMATICA. Disponível em: <https://www.archivematica.org/pt-br/>. Acesso em: 28 de nov. 2020.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Câmara Técnica de Documentos Eletrônicos (CTDE). Diretrizes para a implementação de repositórios arquivísticos digitais confiáveis: RDC-Arq 2015. Rio de Janeiro: CONARQ-CTDE, 2015. Disponível em: http://conarq.gov.br/images/publicacoes_textos/diretrizes_rdc_arq.pdf. Acesso em: 28 de nov. de 2020.

DELMAS, Bruno. Por uma Diplomática contemporânea: novas aproximações. In: CAMARGO, Ana Maria de Almeida et. al. (org.). *Dar nome aos documentos: da teoria à prática*. São Paulo: IFHC, 2015.

DURANTI, Luciana; SUDERMAN, Jim; TODD, Malcolm. *A framework of principles for the development of policies, strategies and standards for the long-term preservation of digital records*. Inter pares 2 Project, 2008. Disponível em: [http://www.interpares.org/display_file.cfm?doc=ip2\(pub\)policy_framework_document.pdf](http://www.interpares.org/display_file.cfm?doc=ip2(pub)policy_framework_document.pdf). Acesso em 20 out. 2020.

GRÁCIO, José Carlos Abbud. *Preservação digital na gestão da informação: um modelo processual para as instituições de ensino superior*. São Paulo: Cultura Acadêmica, 2012. Disponível em: <http://hdl.handle.net/11449/113727>. Acesso em 19 dez. 2020.

INTERPARES; ICA. *Digital records pathways: topics in digital preservation*. 2012.

Disponível em:

http://interpares.org/ip3/display_file.cfm?doc=ip3_canada_gs12_module_1_july-2012_DRAFT.pdf. Acesso em 20 de dez. 2020.

LUZ, Charlley dos Santos; MARINGELLI, Isabel Cristina Ayres da Silva. Política de preservação digital: caso Pinacoteca de São Paulo. *Perspect. ciênc. inf. (online)*, Belo Horizonte, v. 23, n. 2, p. 189-200, June 2018. Disponível em:

http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362018000200189&lng=en&nrm=iso . Acesso em 20 dez. 2020.

PINACOTECA DE SÃO PAULO. *Política de preservação digital da Pinacoteca do Estado de São Paulo*. São Paulo: Pinacoteca de São Paulo, 2019. Disponível em:

http://pinacoteca.org.br/wp-content/uploads/2016/11/at_BIBLI_final.pdf. Acesso em 10 nov. 2020.

Notas

1 Disponível em: <https://www.nationalarchives.gov.uk/PRONOM/Default.aspx>.