

DOCUMENTO 2: SOBRE O FUNCIONAMENTO E AS CARACTERÍSTICAS DO REPOSITÓRIO LEGATUM – SONUS ET IMAGO

a) COMO FUNCIONA O REPOSITÓRIO LEGATUM?

O Repositório Legatum visa a constituir-se como um ambiente federado, o qual já reúne uma série de instituições públicas parceiras e colaboradoras. Dois ambientes interligados caracterizam o Repositório Legatum, um para o *acesso* e outro para a *preservação digital*. Para o acesso remoto, usa-se o software *Access to Memory (AtoM)* e, para a preservação digital, o software *Archivematica*, e ambos nos permitem aderir ao que a legislação brasileira preconiza com relação ao acesso e preservação de documentos produzidos por instituições públicas. No Repositório Legatum ambos estarão interligados, permitindo-nos realizar preservação digital e gerar versões de acesso confiáveis. Esses softwares são livres e de código aberto, e são reconhecidos pelo Conselho Internacional de Arquivos (ICA, na sigla original em inglês) e pelo Conselho Nacional de Arquivos do Brasil (CONARQ). Atualmente são adotados, por exemplo, pela UNESCO, entre muitas outras importantes instituições em muitos países. O AtoM e o Archivematica têm por base normas arquivísticas internacionais de preservação digital e de acesso remoto arquivísticos.

Assim, no ambiente AtoM do Repositório Legatum, o esquema de metadados adota a Norma Internacional de Descrição Arquivística (ISAD-G), procurando atender também a Norma Brasileira de Descrição Arquivística (NOBRADE), a Norma Internacional para Descrição de Instituições com Acervo Arquivístico (ISDIAH), a Norma Internacional de Registro de Autoridade Arquivística para Entidades Coletivas, Pessoas e Famílias (ISAAR-CPF) e a Norma internacional para descrição de funções (ISDF).

O Archivematica, por sua vez, foi desenvolvido com base na norma ISO 14.721:2012 – *Space data and information transfer systems – Open Archival Information System (OAIS) – Reference model*. Devido à importância que essa norma vem adquirindo em todo o mundo e ao fato de que o software Archivematica adota este modelo, indicaremos aqui um pouco mais de detalhes.

O Modelo de Referência OAIS (traduzido na ABNT NBR 15.472:2007¹ como SAAI – Sistema Aberto de Arquivamento de Informação)

fornece uma estrutura sistemática para o entendimento e implementação de conceitos arquivísticos necessários ao acesso e à preservação de informação digital por longo prazo e para a descrição e comparação de arquiteturas e operações de arquivos existentes e futuros. Descreve funções, processos e métodos para preservação por longo prazo. Desenvolvido pelo Consultative Committee for Space Data Systems (CCSDS), o OAIS foi publicado pela primeira vez em 1999 e desde o

¹ A norma ABNT 15.472:2007 é baseada (praticamente uma versão integral para o português) na norma ISO 14.721:2003, encontrando-se, portanto, atualmente, com uma defasagem de quase 15 anos com relação à ISO, já que em 9 de janeiro de 2018 esta foi revisada e validada.

início dos anos 2000 tem influenciado muitos desenvolvimentos de preservação digital.²

De acordo com a norma NBR 15.472:2007 (p.vi):

O sistema aberto de arquivamento de informação (SAAI) é um arquivo, compreendido nesta Norma como uma organização de pessoas e sistemas, que aceitou a responsabilidade de preservar informação e torná-la disponível a uma comunidade-alvo. Um arquivo SAAI distingue-se de outros usos do termo 'arquivo' por cumprir um conjunto de responsabilidades, como as definidas nesta Norma. O termo 'aberto' em SAAI significa que esta recomendação, bem como futuras recomendações e padrões relacionados, sejam desenvolvidos em fóruns abertos, mas não significa que o acesso ao arquivo seja irrestrito.

Ainda de acordo com a referida norma NBR, o modelo OAIS:

- oferece esquema [estrutura] para amplo entendimento e compreensão dos conceitos arquivísticos necessários para preservação por longo prazo e acesso à informação digital;
- oferece conceitos necessários para que organizações não arquivísticas participem efetivamente da função de preservação;
- oferece esquema [estrutura], incluindo terminologia e conceitos, para descrever e comparar arquiteturas e operações de arquivos existentes e futuros;
- oferece esquema [estrutura] para descrever e comparar diferentes estratégias e técnicas de preservação por longo prazo;
- oferece base para comparar modelos de dados de informação digital preservada por arquivos e para discutir como esses modelos e a informação relacionada podem mudar ao longo do tempo;
- oferece fundamentos que podem ser expandidos por outros esforços dirigidos à preservação por longo prazo de informação que NÃO se encontra em formato digital (por exemplo, meios físicos e amostras físicas); ...

Estamos na Fase 3 de 3 da pesquisa, cuja conclusão está prevista para fevereiro de 2019. A participação ativa e contínua das instituições parceiras e colaboradoras, mesmo após encerrada a pesquisa em execução, é fundamental para a evolução do Repositório Legatum e, conseqüentemente, para a preservação digital de acervos audiovisuais públicos, que cresceram exponencialmente nos últimos anos. A Universidade Federal da Bahia (UFBA), em parceria com o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), pode liderar esta inovação.

b) QUE CONTEÚDOS O REPOSITÓRIO LEGATUM DISPONIBILIZA?

Repositórios que adotam os softwares AtoM e Archivematica comportam qualquer tipo de documento digital em seus ambientes. No entanto, o Repositório Legatum dedica-se aos complexos arquivos audiovisuais digitais (e também iconográficos, textuais, fotográficos, desde que relativos ao acervo audiovisual admitido). Assim, versões digitais nativas ou resultantes de processos de

² Cf. *Digital Preservation Coalition*. Disponível em: <https://dpconline.org/handbook/institutional-strategies/standards-and-best-practice>. Acesso em: 23jun2018

conversão digital de itens constituintes de acervos audiovisuais de natureza institucional, científica, acadêmica e artística podem ser disponibilizados nesta plataforma.

c) QUAL A INFRAESTRUTURA E A PLATAFORMA DO REPOSITÓRIO LEGATUM?

A plataforma do Repositório Legatum para a função de acesso está hospedada em uma máquina virtual fornecida pela Superintendência de Tecnologia da Informação da Universidade Federal da Bahia. O sistema operacional na máquina virtual é a distribuição Linux Debian, versão estável. A aplicação é baseada no software livre AtoM, atualmente na versão 2.4.0. A linguagem de programação é o PHP.

O módulo do RL-SI para a função de Preservação Digital está hospedado em um computador servidor dedicado, fornecido pela Dell e adquirido com recursos de bolsa em produtividade do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). O sistema operacional é a distribuição Linux Ubuntu 14.04, Long Term Support (LTS). A aplicação é baseada no software livre Archivematica, cuja linguagem de programação é o Python.

Os arquivos *master* (originais nato-digitais ou representantes digitais destinados à preservação digital), no momento atual, não ficarão online no servidor em que está instalado o software de preservação digital Archivematica. Serão gravados em fitas do tipo Linear Tape Open (LTO) e geradas três cópias, ficando 01 cópia com a instituição parceira / colaboradora, 01 cópia com uma entidade indicada pela instituição parceira / colaboradora e 01 cópia com a UFBA. Entretanto, estamos estudando as vantagens das infraestruturas *cloud*, inclusive por já serem adotadas pela UFBA internamente. Pensamos, para um futuro cooperativo, nas possibilidades de expandi-la de forma colaborativa com outras destacadas instituições públicas. Adotando-se “nuvens”, reduz-se o trabalho com infraestrutura. Gravar e manter fitas LTO também é lidar com infraestrutura, mas – com a eventual possibilidade da opção de contratação de infraestruturas *cloud* externas, institucionais, públicas, cooperativas, distribuídas – as questões associadas às aquisições de hardware, mídias e atualizações de softwares deixariam de ser uma dificuldade para as instituições parceiras e para a UFBA. A título de ilustração, sabe-se que há opções, neste ano de 2018, com custo de armazenamento de USD 0,01 por GB/mês (ou 1 dólar por Terabyte), mas são necessárias análises mais densas e que considerem outros aspectos relativos às esferas públicas. No entanto, não se pode desprezar a opção *Cloud* de armazenamento, visto que todo o trabalho de infraestrutura é repassado/terceirizado para a empresa que fornece o serviço. Nesse caso, ao invés de manter-se uma infraestrutura dentro da instituição, há apenas configuração, upload e download quando necessário. Por outro lado, a solução da nuvem envolve outras questões, não apenas a relativa aos custos (sobretudo para arquivos digitais de vídeo de alta resolução, que exigem espaço considerável

de armazenamento), como, por exemplo, a dependência a uma empresa, que pode fechar sem aviso prévio. Sendo o Legatum uma iniciativa pública, é algo a se levar em consideração. Mas o fato é que há um campo novo a ser pensado nessa linha, não apenas relativo ao armazenamento, mas uma série de outros recursos, como, por exemplo, inteligência artificial (IA), que permite analisar vídeos automaticamente e indexá-los (o que aparece em cada momento do vídeo). Temos uma nova geração de sistemas de acesso baseadas em IA muito próxima de chegar ao mercado, e não duvidamos de que as IA também possam ajudar na preservação digital muito em breve. Outro aspecto do que vimos pensando, parte da premissa de que não se pode guardar tudo: talvez haja a necessidade de critérios de avaliação e seleção mais rígidos, como, por exemplo, considerar guardar apenas as versões finais das obras e não todo o material criado durante a produção. O difícil seria definir critérios que categorizem os arquivos audiovisuais que separem, por exemplo, obras-primas daquelas mais comuns, ou aquelas que representem idealmente uma época ou técnica que sejam eleitas para preservação. Sobre a definição desses critérios, trata-se de questão que historicamente tem sido um tema complexo para os arquivos audiovisuais. Portanto, tal solução deverá ser estabelecida pelos membros do Grupo de Estudos sobre Cultura, Representação e Informação Digitais (CRIDI), seus parceiros formais e pela UFBA, por meio de um documento que reflita o consenso, em meio às dificuldades que caracterizam até onde se pode ir, considerando-se, sempre, as limitações então vigentes.