

SEÇÃO 12

ESTUDOS DE CASO

MÓDULO 7

Gestão e preservação de documentos digitais

SEÇÃO 12

Estudos de caso

Adaptação do Arquivo Nacional da Costa Rica

Versão 1, 2024

Este curso foi traduzido e adaptado pela Direção Geral do Arquivo Nacional da Costa Rica, em colaboração com a Seção de Arquivologia da Universidade da Costa Rica, a partir do material original de 2011 da Associação Internacional de Arquivos Francófonos, disponível online no Portal Internacional Arquivístico Francófono. Esclarece-se que podem existir variações em relação ao conteúdo original. Para acessar o material em francês, visite <https://www.piaf-archives.org/se-former/module-7-gestion-et-archivage-des-documents-numeriques>.



Conteúdo

Capítulo 1. Objetivo da seção	4
1.1. Introdução.....	4
Capítulo 2. Contexto global	4
2.1. Internacional.....	4
2.2. Europa.....	9
2.3. Estados Unidos	11
Capítulo 3. Experiências na França.....	11
3.1. A experiência de uma grande organização científica: o Centro Nacional de Estudos Espaciais (CNES).....	11
3.2. A experiência dos serviços públicos de arquivos franceses.....	17
Bibliografia	31

Capítulo 1. Objetivo da seção

1.1. Introdução

Nesta seção serão apresentados vários exemplos de implementação de plataformas de arquivo eletrônico em vários setores: científico, patrimonial, institucional (arquivos, biblioteca, etc.).

No segundo capítulo, será apresentada uma breve incursão em projetos e ferramentas globais que evidenciam a prática generalizada de criação e implementação de arquivos eletrônicos a nível internacional. Mencionam-se apenas algumas dessas conquistas, sem entrar em detalhes, mas com a possibilidade de acessá-las individualmente por meio dos links. Pretende-se ilustrar com isso as atividades em torno do arquivo digital e destacar a necessidade de uma cooperação saudável dentro da globalização da comunicação, ainda mais essencial do que na gestão tradicional de intercâmbios profissionais.

O terceiro capítulo trata da retroalimentação de dois tipos de experiências que tiveram lugar na França:

- por um lado, a análise das múltiplas dificuldades encontradas no passado na gestão, sustentabilidade e disponibilidade da informação digital,
- por outro lado, o exame das primeiras lições resultantes do estabelecimento de organizações e plataformas de arquivo digital em diversos campos, da aplicação do modelo de referência OAIS e outros padrões aplicáveis ao campo do arquivo digital.



NOTA

Nesta seção, o texto não é identificado com caracteres em negrito, não são destacados os termos do glossário que já foram assimilados, nem são propostos temas “complementares”. Portanto, assume-se um nível mais elevado de conhecimento por parte daqueles que consultam este módulo.

Se tiver dificuldades de compreensão, não hesite em retomar o estudo das seções anteriores, desta vez recorrendo aos “complementos”.

Capítulo 2. Contexto global

Em nível internacional, existem muitas iniciativas, programas de estudo e pesquisa, conquistas e projetos de instituições públicas e empresas privadas.

2.1. Internacional¹

Apresenta-se uma breve descrição de projetos, instrumentos, ferramentas e aplicações que oferecem serviços de caracterização de conteúdo, reconhecimento automático de formatos, comparação automática de objetos digitais,

¹ Este item é atualizado para o contexto global de 2020-2021, a partir do Trabalho Final de Graduação para obtenção do título de Mestre em Administração Universitária, intitulado “Modelo de Preservação de Documentos Digitais na Administração Universitária – Estudo de Caso: Universidade Nacional” por María Gabriela Castillo Solano e Raquel Umaña Alpízar (2018).

identificação e extração de metadados, validação da integridade dos documentos digitais, bem como outras funcionalidades de preservação digital que foram desenvolvidas no âmbito internacional.

Apache Tika (Content analysis tool kit)

O kit de ferramentas de Apache Tika™ detecta o texto e extrai metadados de objetos digitais, serve para a indexação de motores de busca, análise de conteúdo e tradução (retirado de <https://tika.apache.org/>).

Arca

Arca é um repositório digital voltado para a preservação dos documentos ao longo do tempo, desenvolvido pela empresa Business Integrators Systems Limitada. Implementa o modelo OAIS e as normas de boas práticas arquivísticas, que garante o valor legal dos seus documentos sob custódia e, também, que no futuro estes serão acessíveis, representáveis e legíveis (retirado de <http://www.bis.co.cr/Products/Arca>).

Archivemática

É uma aplicação de código aberto projetada com o objetivo de preservar a longo prazo os conteúdos digitais, de modo que sua autenticidade e confiabilidade perdurem com o passar do tempo. Archivemática está projetado de acordo com requerimentos do modelo funcional ISO-OAIS (retirado de <https://www.archivematica.org/>).

Catalogue of Criteria for Trusted Digital Repositories

O Nestor Working Group, autores de Catalogue of Criteria for Trusted Digital Repositories, identificaram as condições para poder criar depósitos e arquivos digitais seguros, auditáveis e certificáveis (retirado de http://files.dnb.de/_nestor/materialien/nestor_mat_08-eng.pdf).

C3po (Clever, Crafty Content Profiling of Objects)

É uma ferramenta de criação de perfis de conteúdo para análise de preservação, utiliza dados meta extraídos de arquivos de uma coleção digital como entrada para gerar um perfil do conjunto de conteúdo. Foi projetado de maneira que seja possível integrar facilmente diferentes formatos de metadados provenientes de distintas ferramentas. Atualmente, suporta metadados FITS e metadados Apache TIKa (retirado de <http://peshkira.github.io/c3po/>).

DAITSS

“Dark Archive in the Sunshine State”: desenvolvido por Florida Centre for Library Automation, é um software livre para repositórios digitais, utiliza o esquema de metadados METS e está baseado em OAIS.

DAN

“Documentos de Archivo en la Nube”: é um projeto desenvolvido pela Universidade da Colúmbia Britânica, Canadá, no contexto do projeto InterPARES, com o objetivo de investigar os benefícios e riscos de armazenar documentos de arquivo na *Nuvem*, identificando e examinando os diversos fatores administrativos, operacionais, legais e técnicos relacionados ao armazenamento e à gestão de documentos de arquivo, por meio da terceirização com *Provedores de Serviços na Nuvem* (PSN).

DRAMBORA

O grupo de trabalho RLG-NARA junto com o Center for Research Libraries (CRL), o Digital Curation Centre (DCC) e o Digital PreservationEurope (DPE), desenvolveram o *Digital Repository Audit Method Based on Risk Assessment* (DRAMBORA), uma ferramenta gratuita que permite avaliar um depósito digital, determinando suas capacidades, pontos fracos e pontos fortes, baseada na análise de riscos (retirado de <http://www.repositoryaudit.eu/>).

DROID (Digital Record Object Identification)

É uma ferramenta de identificação de formato de arquivo, que analisa agrupamentos de arquivos por formatos e fornece relatórios para determinar se é necessário aplicar ações de migração de formato. É a primeira de uma série planejada de ferramentas desenvolvidas pelos Arquivos Nacionais do Reino Unido, sob a égide de seu serviço de registro técnico de PRONOM (retirado de <http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>).

DSpace

É um pacote de *software* de repositório de código aberto utilizado para criar repositórios de acesso aberto para conteúdo digital publicado. Desenvolvido em 2002 por HP e pelo Instituto Tecnológico de Massachussets (MIT). Em 2007 HP e MIT formaram a Fundação DSpace e em 2009 em conjunto com a organização de Fedora Commons se uniram para criar Dura Space (retirado de <http://www.dspace.org/>)².

DuraCloud

Dura Cloud é um serviço de armazenamento de Dura Space que permite controlar onde e como se conservam os objetos digitais na *nuvem*. Em 2011, a Universidade de Harvard realizou um projeto piloto utilizando esta solução (retirado de <http://duracloud.org/>).

E-journal Archiving

Com fundos da Fundação Andrew W. Mellon, Harvard realizou um estudo para a construção de um repositório para revistas eletrônicas. Como continuidade, o LDI financiou um projeto colaborativo com a Biblioteca Nacional de Medicina (NLM) para produzir uma definição de tipo de documento XML (DTD, na sigla em inglês). O DTD está projetado para aumentar a facilidade de intercâmbio entre editores e arquivos para o conteúdo da revista eletrônica em nível de artigo. DTD conta com uma estrutura modular que permite a personalização e que deveria ser um objetivo fácil de transformação a partir de conteúdo XML ou SGML codificado

ERA

“Electronic Records Archives”: ERA permite realizar a transferência e preservação de documentos digitais ao NARA, a fim de que esta última garanta a integridade dos documentos digitais e o acesso por parte dos usuários (retirado de <https://www.archives.gov/era>).

FEDORA

Flexible Extensible Digital Object Repositor e Architecture: desenvolvido pela Universidade de Cornell em 1997, é um sistema de repositório de código aberto para a gestão e difusão de conteúdos digitais, é administrado pela organização sem fins lucrativos Dura Space (retirado de <http://fedorarepository.org/>).

FIDO (Format Identification for Digital Objects):

É uma ferramenta de CLI de identificação de formato simples para objetos digitais que utiliza assinaturas PRONOM convertidas em expressões regulares. A funcionalidade de FIDO é similar a DROID (retirado de <http://openpreservation.org/technology/products/fido/>).

² DuraSpace é uma organização sem fins lucrativos que é criada em 2009, quando são fundadas Fedora Commons e a Fundação DSpace. Cabe destacar que esse portfólio de tecnologias de código aberto é desenvolvido por bibliotecários, arquivistas, tecnólogos e pesquisadores.

FITS (File Information Tool Set)

É uma iniciativa da Universidade de Harvard, que desenvolveu um conjunto de ferramentas de informação de objetos digitais que identifica, valida e extrai os metadados técnicos de diferentes formatos de arquivo (<https://projects.iq.harvard.edu/fits/home>).

Functional Requirements for Evidence in Recordkeeping

The Functional Requirements for Evidence in Recordkeeping é um projeto da Universidade de Pittsburg, que com o apoio da National Historical Publications and Records Commission, desenvolveram um conjunto de requisitos funcionais de gestão documental para o design e implementação de sistemas eletrônicos de informação (retirado de <http://www.archimuse.com/papers/nhprc>).

GDFR (Global Digital Format Registry)

É um projeto que já não está ativo, foi uma iniciativa desenvolvida por Harvard University Library, cujo objetivo era construir um registro de formatos de preservação digital.

InSPECT

O projeto InSPECT (Investigating the Significant Properties of Electronic Content Over Time) tem se dedicado à análise dos elementos ou características essenciais (propriedades significativas) para a conservação de um objeto digital (retirado de <https://searchworks.stanford.edu/view/9520222>).

InterPARES

O *Projeto Internacional para a Investigação em Documentos de Arquivo Autênticos e Permanentes em Sistemas Eletrônicos* (InterPARES), que teve início em 1999, sob a responsabilidade da doutora Luciana Duranti como diretora e é um dos principais projetos em nível internacional. Está constituído por 4 etapas, a última é *InterPARES-Trust*, que se concentra nos fatores introduzidos pela gestão de documentos de arquivo e dados, em consonância com o projeto DAN (retirado de <http://www.interpares.org/>).

Jhove

JSTOR/ Harvard Object Validation Environment: é uma ferramenta que é utilizado para a detecção e validação de formatos, e até mesmo para a extração automática de metadados técnicos. Incorporou o DROID em sua estrutura para ampliar os formatos que valida (retirado de <https://jhove.openpreservation.org/>).

LOCKSS

“Lots of Copies Keep Stuff Safe”: desenvolvido pela Universidade de Stanford, é um software que comprova a integridade dos documentos, mantém múltiplas cópias em locais distintos e realiza comparações periódicas entre eles para assegurar a autenticidade dos objetos digitais, sendo utilizado principalmente para revistas eletrônicas (<https://www.lockss.org/>).

PLANETS (Preservation and Long-term Access through Networked Services)

Este projeto nasce em 2006 com o objetivo de oferecer soluções para a preservação a longo prazo e garantir o acesso aos documentos de instituições europeias. PLANETS inclui critérios de *TRAC* (estabelece pautas para medir o cumprimento do modelo OAIS), e *NESTOR* que são utilizados automaticamente com a ferramenta *Plato*.

Plato

É uma ferramenta de planejamento que apoia as decisões de preservação de objetos digitais, por meio de serviços de caracterização de conteúdo, ações de preservação e comparação automática de objetos em uma arquitetura orientada a serviços, e integra DROID, JHOVE y FITS. Além disso, o MiniMEE é um serviço de migração de formatos incluído no Plato.

Pronom

É um registro de formatos que foi lançado em 2004 pelos Arquivos Nacionais do Reino Unido, com capacidade de reconhecimento de cerca de 150 formatos; possui uma ferramenta Open Source, o Digital Record Object Identification (DROID), para identificação e descrição automática do formato dos arquivos, validação do formato de um conjunto de arquivos, além de permitir ao Pronom a integração com sistemas informáticos mais amplos (retirado de <https://www.nationalarchives.gov.uk/PRONOM/Default.aspx>).

PUIDs (Pronom Unique Identifiers)

Foi criado no âmbito do projeto PRONOM, devido às deficiências dos descritores MIMETypes, pois este distingue entre diferentes versões e subformatos (retirado de http://www.nationalarchives.gov.uk/aboutapps/pronom/pdf/pronom_unique_identifier_scheme.pdf).

RODA “Repository of Authentic Digital Objects”

É uma solução completa de repositório digital que entrega todas as principais unidades funcionais do modelo de referência OAIS. RODA é capaz de gerenciar e fornecer acesso a diversos tipos de conteúdos digitais produzidos por grandes corporações ou instituições públicas. RODA baseia-se em tecnologias de código aberto e suporta grande parte dos padrões existentes, tais como o Sistema de Informação de Arquivos Abertos (OAIS), codificação e transmissão de metadados padrão (METS), Descrição Arquivística Codificada (EAD), Dublin Core (DC) e PREMIS (metadados para a conservação) (retirado de http://roda-community.org/?locale=es_CL#welcome).

SPIRT

O projeto SPIRT e seu esquema de metadados RKMS3 (Record keeping Metadata Schema) publicado em 1999, é um projeto da Universidade Australiana de Monash. O acrônimo SPIRT provem da bolsa de pesquisa que deu nome ao projeto: *Strategic Partnership with Industry- Research & Training*, pesquisa que foi concebida para construir uma estrutura na qual outros padrões de metadados de outros setores pudessem desenvolver sua própria aplicação (retirado de https://www.monash.edu/_data/assets/pdf_file/0003/857550/SPIRT.pdf).

Technical Guidelines for Digitizing Archival Materials for Electronic Access Creation of Production Master Files- Raster Images (2004)

São diretrizes técnicas para digitalizar materiais de arquivo que definem as abordagens para a criação de cópias digitais para facilitar o acesso e a reprodução. Essas diretrizes são a base dos procedimentos utilizados pelo Laboratório de Imagens Digitais do Laboratório de Preservação de Mídias Especiais da NARA (retirado de <https://www.archives.gov/files/preservation/technical/guidelines.pdf>).

3 GuiaAGRkMS Australian Government Recordkeeping Metadata Standard Implementation Guidelines: Exposure Draft de 2010.

Trustworthy Repositories Audit & Certification: Criteria and Checklist

Em 2003, Research Library Group RLG y National Archives and Records Administration NARA criaram um grupo de trabalho para analisar a possibilidade de criar repositórios digitais seguros que pudessem ser auditados, avaliados e certificados. O desafio foi elaborar os critérios de certificação e delinear um processo de certificação aplicável a uma ampla gama de repositórios digitais e arquivos, desde a preservação institucional e acadêmica de repositórios de grandes arquivos de dados e bibliotecas nacionais até os serviços de arquivo digital para terceiros. (RLG-NARA, 2003, p. 1). É a base da Norma ISO 16363:2012 Sistemas de transferência de informação e dados espaciais: auditoria e certificação de repositórios digitais confiáveis.

UDFR (Unified Digital Formats Registry)

A UDFR é uma iniciativa de abril de 2009 para criar um único registro de formatos compartilhados. O UDFR foi desenvolvido no Centro de Curadoria da Universidade da Califórnia (UC3) com fundos da Biblioteca do Congresso (retirado de <http://www.udfr.org/>).

VERS

VERS é o projeto de pesquisa que promoveu o Public Records Office Victoria (PROV) no ano de 1996 e que se tornou um padrão cujo objetivo é garantir a veracidade e a autenticidade dos documentos digitais; descreve uma metodologia para converter registros digitais em formatos de longo prazo conhecidos como objetos encapsulados VERS (retirado de <https://prov.vic.gov.au/recordkeeping-government/vers>).

VidArch

O Projeto VidArch abordou tanto aspectos teóricos quanto práticos do contexto da preservação digital, desenvolvendo uma estrutura de preservação para o contexto de vídeo digital, aplicando-a a duas coleções importantes: a série completa de vídeos educativos da NASA e o conjunto completo de vídeos avaliados da ACM SIGCHI apresentados em conferências anuais desde 1983 até o presente (retirado de <https://ils.unc.edu/vidarch/>).

Xena

XML Electronic Normalizing for Archives: é uma ferramenta desenvolvida pelos Arquivos Nacionais da Austrália, para ajudar na preservação a longo prazo dos registros digitais, detecta os formatos e os normaliza convertendo os objetos digitais em formatos abertos para preservação. Xena, é um componente da Plataforma de Software de Preservação Digital (DPSP) (retirado de <http://xena.sourceforge.net/>).

2.2. Europa

Na Europa, há muitas conquistas, especialmente no nível das instituições, e abundam projetos em todos os lugares.

Comissão Europeia

As ações empreendidas e realizadas pela Comissão Europeia em matéria de arquivo digital se situam em vários níveis:

Um conjunto de pautas que tenham um impacto indireto significativo no arquivo digital aplicável a todos os países da União. Nesse sentido, podemos citar a Diretiva 1999/93 / CE de 13 de dezembro de 1999, sobre um quadro comunitário para a assinatura eletrônica, ou ainda a Diretiva 2007/2 / CE de 14 de março de 2007, que estabelece uma Infraestrutura de Informação Geográfica na Comunidade Europeia (INSPIRE).

Ação mais proativa e direcionada diretamente para aplicações de arquivo eletrônico, especialmente no campo da gestão documental, com a organização e o financiamento do desenvolvimento do MoReq 2010 (Requisitos padrão para o controle do arquivo eletrônico).

Preparado sob a orientação científica do Fórum DLM e publicado sob os auspícios da Comissão <https://www.moreq.info/>

Um conjunto de ações selecionadas dentro do Programa-Quadro de Pesquisa e Desenvolvimento (PCRD) da União Europeia. Assim, foi destinado um orçamento de várias dezenas de milhões de euros para a análise, pesquisa e experimentação de soluções globais adaptadas a um contexto arquivístico específico, ao desenvolvimento de sinergias entre os atores envolvidos, à educação, ao ensino e à conscientização dentro de comunidades que frequentemente estavam bastante distantes dessas preocupações.

As ações de pesquisa que envolvem um grande número de atores geograficamente dispersos dentro de um mesmo projeto geralmente apresentam uma relação resultado/investimento moderada. Por outro lado, essas ações desempenham um papel real na aproximação de todos os atores da cadeia, bem como dos diversos setores de atividade envolvidos na problemática, na sensibilização, na provisão de informações, resultados de estudos e análises.

Os projetos definidos e financiados no âmbito do PCRD são limitados a alguns anos. A continuidade das atividades realizadas pelos parceiros do projeto após o período de financiamento costuma ser incerta. Além disso, esses projetos são resultado da iniciativa de uma ampla variedade de coletivos e instituições. Disso decorre que é difícil garantir a coerência e a complementaridade dos diferentes projetos. Para suprir essas deficiências, várias organizações importantes no campo do arquivo digital uniram-se em uma entidade chamada “Aliança de Acesso Permanente”. <http://www.alliancepermanentaccess.org/>

Essas organizações também representam o campo do arquivamento patrimonial de documentos e publicações (British Library, bibliotecas nacionais dos Países Baixos, da Alemanha, arquivos nacionais da Suécia) e o campo dos arquivos de dados (Agência Espacial Europeia, Centro Europeu de Pesquisa Nuclear). A orientação geral limita-se ao campo científico no sentido amplo, abrangendo a física, a biologia, a química, assim como as ciências ambientais ou as ciências humanas.

A ambição da Aliança é ser o lugar para o desenvolvimento de colaborações e relações entre arquivos, ser influente e comunicador de suas necessidades perante as autoridades nacionais e europeias. Seu objetivo é garantir que os benefícios da preservação digital se obtenham de maneira efetiva e eficiente. Através de um “Centro Virtual de Excelência” (VCOE), cujo foco é criar uma visão comum e continuar a viagem de integração do panorama europeu da preservação digital

Coalizão de Preservação Digital (Digital Preservation Coalition- DPC)

Existem outro tipo de organizações que atendem o tema da preservação digital, um exemplo é a DPC <https://www.dpconline.org/> uma fundação benéfica impulsionada por seus membros e dedicada ao fomento da preservação digital.

Entre seus objetivos encontra-se4:

- Ajudar membros de todo o mundo a proporcionar acesso adaptável e de longo prazo a conteúdos e serviços digitais, por meio da criação de vínculos com a comunidade, ações de promoção direcionadas, formação

4 Retirado do documento “DPC Prospectus 2020-2021” em espanhol, disponível no site da DPC.

e capacitação, desenvolvimento de infraestrutura e competências, normas e boas práticas, e boa gestão e governança.

- Facilitar um programa projetado para atender às necessidades dos membros e que beneficia a comunidade global dedicada à preservação digital.
- Lutar por um patrimônio digital com o futuro garantido, e oferecer colaboração e assistência a todos os setores, agências e pessoas que compartilham da mesma visão e valores.

2.3. Estados Unidos

Aliança Nacional de Administração Digital (National Digital Stewardship Alliance- NDSA)

Corresponde à outra das iniciativas onde sua participação nacional e internacional baseia-se em uma associação. É uma organização da Federação de Bibliotecas Digitais (DLF) e identifica-se como um consórcio de organizações comprometidas com a preservação a longo prazo da informação digital. <https://ndsa.org/>. Seu trabalho baseia-se na conformação de grupos de trabalho e interesse que focam em amplas áreas da preservação digital.

Entre seus trabalhos e contribuições destacam os Níveis de Preservação Digital (LoP) como um recurso para os profissionais da preservação digital ao construírem ou avaliarem seu programa de preservação digital.

Capítulo 3. Experiências na França

Serão analisadas várias entidades, a primeira diz respeito a uma organização científica, as seguintes são feedbacks de arquivos institucionais nacionais ou locais cujas experiências são de interesse primordial.



NOTA

Esclarece-se que as experiências na França foram apresentadas pelo Portal Internacional Arquivístico Francófono (PIAF) em sua versão original de 2011, razão pela qual não se dispõem de dados recentes que descrevam a situação atual de cada projeto ou entidade.

3.1. A experiência de uma grande organização científica: o Centro Nacional de Estudos Espaciais (CNES)

A retroalimentação do CNES merece ser analisada de várias formas.

Em primeiro lugar, é um testemunho de uma ampla experiência no tema, experiência essa que deu origem a conquistas de valor exemplar e que contribuíram significativamente para o surgimento de uma metodologia hoje amplamente reconhecida.

Em segundo lugar, também revela as lacunas que podem existir entre, por um lado, os engenheiros e pesquisadores da área que estão muito conscientes dos problemas levantados e dos riscos envolvidos, e, por outro lado, os

diretores e presidentes dos departamentos de instituições públicas que têm uma visão mais vaga dessas questões que às vezes consideram um tema essencialmente técnico.

3.1.1 Dados de ciência espacial

As missões espaciais podem ser classificadas em diferentes categorias segundo seus objetivos: algumas têm uma vocação fundamentalmente científica, outras estão voltadas para aplicações de diversas naturezas: telecomunicações, televisão, localização e coleta de dados, meteorologia, observação da Terra, aplicações que podem ser civis ou militares. Algumas missões, em particular todas aquelas que observam e estudam a Terra e seu entorno, têm como objetivo atender simultaneamente a necessidades científicas de longo prazo e a questões econômicas de curto e médio prazo, como o monitoramento da vegetação ou o estudo do ciclo da água.

As missões científicas abrangem um amplo espectro de áreas. O estudo do Universo ocupa aí um lugar importante, destacando-se, em primeiro lugar, as diversas disciplinas da astronomia (astrofísica, astroquímica, astrometria etc.) e a exploração do sistema solar e de seus diversos planetas, ainda muito pouco conhecidos. A observação do Sol e dos fenômenos que ocorrem nele, assim como o estudo dos efeitos desses fenômenos em nosso planeta e em seu entorno ionizado, também fazem parte disso. Outro aspecto muito importante é o do estudo e observação da Terra, sua forma e superfície (geodésia), seus oceanos, e sua evolução em médio e longo prazo em todos os âmbitos. Outras áreas importantes da ciência dão lugar a experimentos em química, biologia, medicina em microgravidade

Todas as observações são realizadas utilizando instrumentos complexos que foram projetados para esse fim e que fornecem medições em forma de dados digitais. Uma característica especial de um grande número de missões é a capacidade desses instrumentos de realizar observações sistemáticas de longa duração e fornecer dados de forma contínua ao longo de vários anos. Isso explica os grandes volumes de dados gerados por essa atividade. Esses dados são transmitidos ao solo por meio de um sinal eletromagnético codificado denominado telemetria. Desde o início da década de 1980, os dados produzidos pelos instrumentos vêm sendo digitalizados a bordo da espaçonave e, em seguida, transmitidos para a Terra.

A necessidade de conservar a maior parte desses dados a longo prazo responde a dois imperativos: um imperativo científico e um imperativo patrimonial.

3.1.2. Questão de formatos de armazenamento e gravação

No início da década de 1970, os dados foram registrados pela primeira vez em fitas magnéticas de 7 trilhas (6 trilhas para dados e uma chamada trilha de comparação para verificação de erros de bits), com densidades de gravação de 200, 556 e depois 800 bpi (bits por polegada). Essas fitas desapareceram rapidamente e foram substituídas por fitas de 9 trilhas (8 para dados mais uma para verificação), em função do desenvolvimento de conjuntos de caracteres de oito bits (por exemplo, o EBCDIC da IBM). As densidades das fitas de 9 trilhas evoluíram de 800 bpi para 1600 bpi e depois para 6250 bpi até o final da década de 1980. Além desse período, essa tecnologia deixou de evoluir e praticamente desapareceu no final da década de 1990. As capacidades de gravação eram baixas em comparação com nossos meios atuais.

Em 1990, todos os dados das missões científicas do CNES foram armazenados em dezenas de milhares de fitas guardadas nas instalações do centro de informática do CNES em Toulouse. Quase todas essas fitas e os arquivos de dados que continham tinham a estrutura proprietária dos poderosos (por enquanto!) computadores do centro de dados de

controle. Cada coleção de fitas magnéticas foi gerenciada, controlada e mantida pelas equipes do projeto que eram proprietárias dos dados contidos nessas coleções.

Foi nesse momento que a situação e as perspectivas de médio prazo mudaram significativamente devido a três fatores: o primeiro e mais restritivo foi o desaparecimento anunciado das tecnologias de armazenamento em fita magnética de 6250 bpi. Em segundo lugar, a evolução geral da informática em um ambiente científico: o CNES previu o encerramento das máquinas de Controle de Dados do centro de informática, baseadas no sistema operacional NOS / VE e sua substituição por máquinas baseadas no sistema UNIX. Por fim, havia o desejo de tornar os dados científicos acessíveis e utilizáveis pela comunidade em geral.

3.1.3. Implementação de STAF y migração de dados

Diante dessa situação, o CNES tomou a decisão, em 1992, de implementar um serviço de armazenamento central que oferecesse uma garantia real de conservação a longo prazo dos bits, independentemente da tecnologia de armazenamento utilizada. Esse serviço especializado encarregado de perpetuar os arquivos é o STAF (File Transfer and Archiving Service). Opera desde 1994 e apresenta-se como uma entidade independente de projetos ou serviços de arquivo. Estes últimos são clientes do STAF e se comunicam com ele por meio de um conjunto de comandos básicos que permitem, em particular, solicitar o armazenamento ou a restituição de um arquivo ou de um conjunto de arquivos. Essas comunicações passam pela rede interna do CNES. Portanto, STAF tem uma missão muito simples:

- receber arquivos sem ter que conhecer seu formato ou seu conteúdo de informação,
- garantir a conservação a longo prazo desses arquivos,
- garantir sua integridade,
- garantir sua confidencialidade,
- devolver os arquivos mediante solicitação.

Portanto, decidiu-se e iniciou-se a migração imediata e essencial dos dados em fitas para o STAF em 1994. Durou cinco anos, concentrou-se em mais de 60.000 fitas magnéticas, aproximadamente 500.000 arquivos, e envolveu uma mobilização momentânea de diversos atores que realmente tomaram consciência da grande vulnerabilidade dos dados e das múltiplas causas que podem levar à sua perda. O número de arquivos perdidos nessa ocasião por motivos de armazenamento foi muito baixo, mas não nulo, devido à existência de algumas fitas com densidade obsoleta (800 bpi) para as quais o CNES já não possuía nenhum equipamento de leitura, e fitas que não podem ser lidas devido à degradação física.

A migração revelou imediatamente que a maioria dos dados apresentava estruturas lógicas e codificações exclusivas dos sistemas operacionais que haviam sido usados para criar esses dados, sistemas operacionais que estavam em processo de descontinuação. Como resultado, os arquivos não eram portáteis e, não podiam ser lidos em outro sistema. Portanto, uma primeira tarefa consistiu em eliminar dos arquivos todas as informações adicionais específicas do sistema operacional. Isso pode ser feito utilizando um software utilitário disponível no próprio sistema operacional.

Foi necessário iniciar uma segunda operação muito mais delicada. Na maioria dos casos, os arquivos continham resultados de processamento científico, como sequências de números inteiros e reais codificados em binário. A representação binária de altíssima precisão de números reais, com tamanho de 128 bits, era uma representação proprietária. CNES optou por utilizar uma representação padrão de números (em particular a IEEE para números reais), o

o que exigia uma transformação dos dados. Trata-se, portanto, aqui, de transformações ou migrações de formato. Essas transformações apresentaram várias dificuldades:

- não podiam ser executadas automaticamente e exigiam o desenvolvimento de software específico para cada coleção de arquivos,
- não foram reversíveis devido aos inevitáveis erros de arredondamento relacionados às últimas casas decimais significativas.

Portanto, exigiram um esforço considerável de validação. Para maior segurança, os arquivos originais foram preservados durante alguns anos e depois destruídos.

Além disso, os elementos descritivos desses dados, que permitiam entender seu significado, eram às vezes imprecisos ou incompletos, e nem sempre estavam disponíveis. Portanto, durante a mesma operação, qualificada como “reabilitação de dados”, foi necessário reformatar os dados, dotá-los de estruturas independentes dos sistemas operacionais, reconstituir os metadados na medida do possível e migrar esses dados para o STAF. Vários engenheiros dedicaram vários anos a essa operação, tiveram que recorrer a especialistas científicos ainda disponíveis para reconstituir os metadados, confiaram em empresas de serviços para o desenvolvimento de software de transformação de formatos e consumiram recursos e máquinas consideráveis. A maior parte dos dados foi preservada, mas essa operação permitiu medir o custo dos ajustes tardios. Dado o ritmo atual de desenvolvimento das tecnologias digitais, um resgate como esse provavelmente já não seria possível hoje em dia.

Outros sinais de aceleração da obsolescência tecnológica também marcaram esse período. O exemplo dos documentos textuais do campo da informática administrativa é eloquente. O CNES utilizou um dos primeiros sistemas de automação de escritório disponíveis no mercado na segunda metade da década de 1980. Era um sistema proprietário que representava o avatar eletrônico da máquina de escrever tradicional e permitia a digitação, formatação e impressão de documentos de texto, aproveitando os recursos da informática. Grandes quantidades de documentos foram produzidas utilizando esse sistema. No início da década de 1990, com o desenvolvimento da microinformática e o início do monopólio da Microsoft sobre a automação de escritórios, a maioria dos outros sistemas proprietários existentes desapareceu do mercado, mas os documentos permaneceram. Sem a menor possibilidade técnica de migrar os documentos para o pacote de software Word para DOS que foi a primeira versão do Word utilizada pelo CNES, os documentos foram novamente produzidos nesse novo sistema. Seis anos depois, o CNES descobriu que os documentos salvos no Word para DOS eram apenas parcialmente compatíveis com o Word 97 para Windows. Para todos os documentos que precisavam ser preservados, foi possível recuperar o texto, mas o formato de milhares de tabelas complexas foi totalmente perdido.

O STAF realizou as inevitáveis migrações de suporte, que não são visíveis para os clientes do serviço, e demonstrou o interesse e a eficácia dos princípios sobre os quais foi construído, ou seja, a total independência da função de armazenamento em relação às demais entidades funcionais de um serviço de arquivamento a longo prazo, colocando em prática o que hoje se chama virtualização do armazenamento.

3.1.4. Sistemas genéricos de arquivamento para reduzir custos

É possível, até certo ponto, definir formatos de dados e metadados que sejam completamente independentes dos sistemas operacionais e tecnologias. Essa escolha limita a vulnerabilidade desses dados e metadados às mudanças nessas tecnologias. Por outro lado, não é possível construir um sistema de arquivo digital, composto por hardware e software que sejam independentes da tecnologia. Além disso, sabendo que o sistema de arquivamento digital é o meio pelo

qual poderemos receber os dados para arquivar, armazenar, gerenciar e tornar acessíveis, esse sistema deve ser sustentável. Sujeito a todos os caprichos da obsolescência tecnológica, será necessário assegurar a manutenção desse sistema para que ele permaneça em funcionamento contínuo.

Periodicamente, devido ao desaparecimento de uma determinada tecnologia utilizada, não são mais realizados trabalhos de manutenção, sendo necessária a reconstrução parcial do sistema e, portanto, seu financiamento.

A questão de limitar e, se possível, reduzir os custos de manutenção do sistema de arquivamento em funcionamento contínuo é um ponto crítico ao qual o CNES tentou responder com duas opções.

- Uma escolha arquitetônica que consiste em estruturar os sistemas em blocos funcionais independentes entre si, para que qualquer mudança tecnológica importante que provoque modificações profundas em um bloco não repercuta nos demais. No passado, o CNES já havia enfrentado casos de sistemas monolíticos em que uma mudança limitada em um domínio causava uma propagação em cadeia de modificações em todo o sistema, com consequências significativas em termos de custos de modificação e validação.
- Uma opção genérica voltada para construir sistemas que possam ser reutilizados por várias aplicações dentro da instituição e por várias organizações que realizam atividades de arquivamento de dados científicos. Essa escolha tem como objetivo compartilhar os custos de desenvolvimento, depois a manutenção e a atualização por parte dos diferentes locais usuários do sistema. Até 1995, cada missão espacial científica conduziu ao desenvolvimento de um sistema dedicado que permitia a recepção, processamento, difusão e arquivamento de dados desta missão. CNES percebeu rapidamente que seria impossível, no futuro, manter em funcionamento tantos sistemas quantas missões espaciais passadas. Portanto, era imperativo construir um sistema de acesso e gestão de dados capaz de oferecer funções de acesso aos dados para todas as missões de uma mesma disciplina científica ou inclusive várias disciplinas científicas diferentes.

Foi desenvolvido um primeiro sistema genérico desse tipo, o SIPAD (Information, Preservation and Access to Data System), para assegurar a gestão e provisão de dados do Plasma Physics Data Center (<http://cdpp.cesr.fr>). Entrou em operação pela primeira vez em 1999. Este sistema baseou-se principalmente em um produto comercial de gestão de dados técnicos chamado Metaphase.

Após apenas alguns anos de operação, parecia necessário resolver um conjunto de problemas técnicos relacionados ao SIPAD: controlar o desempenho do acesso ao banco de dados, desempenho que se deteriora à medida que aumenta o número de conjuntos de dados; eliminar a dependência do SIPAD em relação ao produto Metaphase, cuja sustentabilidade não estava garantida; introduzir novas funções visando a consulta automatizada; e ter possibilidades de especialização avançada da interface homem-máquina. A escala das necessidades de mudanças levou ao desenvolvimento de um sistema completamente novo, o SIPAD-NG. Portanto, a entrada em operação desse novo sistema em 2006 exigiu uma revisão prévia do design do banco de dados. Os metadados foram extraídos globalmente do SIPAD, depois transformados e enriquecidos de acordo com as novas especificações de metadados, posteriormente validados e inseridos no SIPAD-NG. O sistema SIPAD-NG agora tem uma base sólida: é utilizado dentro do CNES por várias entidades separadas sob responsabilidade do arquivo de dados.

3.1.5. Uma metodologia que está se consolidando

O trabalho de resgate e “reabilitação de dados” foi acompanhado de uma análise metodológica sobre o que fazer e o que não fazer no futuro. Em junho de 1993 foi redigido e distribuído um primeiro documento de especificações para o arquivamento a longo prazo de dados espaciais. Ressalta a importância do patrimônio de dados científicos e tecnológicos preservado e mantido desde o início da década de 1970 e observa três desenvolvimentos essenciais:

- o aumento constante dos volumes de dados produzidos,
- o aumento significativo dos períodos mínimos de retenção requeridos (várias décadas),
- a necessidade de uma acessibilidade cada vez mais ampla a esses dados por parte da comunidade científica.

Esta primeira especificação já contém todos os requisitos essenciais que se aplicam aos dados científicos:

- identificação de todas as informações que devem ser associadas aos dados a serem arquivados: descrição sintática e semântica dos arquivos, parâmetros de calibração (informações de representação), metadados em DIF (Formato de intercâmbio de diretórios) (informações descritivas), base documental que descreve a missão, experiência, instrumento (informações de procedência e contexto).
- o requisito de manter a integridade física de toda a informação digital,
- o requisito de métodos de codificação padronizados e estruturas de arquivos independentes dos sistemas operacionais,
- acessibilidade dos dados aos usuários autorizados.

Foi a partir de suas experiências práticas e primeiras reflexões metodológicas que o CNES pôde participar frutiferamente na redação do modelo OAIS e que depois assumiu a responsabilidade direta de redigir o padrão PAIMAS (Norma ISO 20652:2006 Sistemas de transferência de informação e dados espaciais - Interface produtor-arquivo - Padrão abstrato de metodologia) e o padrão PAIS (Norma ISO 20104:2015 Sistemas de transferência de informação e dados espaciais - Especificação de interface produtor-arquivo). O modelo OAIS permite analisar os sistemas desenvolvidos no CNES com um novo ponto de vista e um novo vocabulário. São essas interações entre a abordagem pragmática nascida da experiência prática e a reflexão metodológica que gradualmente conferem a solidez e a confiabilidade necessárias aos sistemas de arquivamento digital.

É também com base nisso que foi aberta uma nova área chamada “Engenharia de Dados” dentro do Repositório Normativo do CNES, que define a estrutura organizacional das normas aplicáveis aos seus projetos e estruturas. Além de um resumo das necessidades em termos de sustentabilidade e acesso aos dados, essa área do Repositório inclui uma série de normas e recomendações aplicáveis aos projetos de produção e dados, bem como aos serviços responsáveis pelo arquivamento desses dados.

3.1.6. Possível distorção entre necessidades e decisões

Em 2000 foi elaborado um plano estratégico do CNES para o período 2001-2005. Este plano levou em consideração, de forma geral, o problema da gestão, difusão e utilização dos dados dos experimentos espaciais. Esse plano deu origem a trabalhos para traduzir suas principais diretrizes em ações concretas.

A primeira ação proposta nesse contexto foi mais que simbólica: Preparar uma decisão para a assinatura do Presidente do CNES a fim de afirmar as responsabilidades e os objetivos do CNES quanto à valorização, arquivamento e provisão.

de produtos (de dados)”. Essa situação ilustra, se necessário, a importância de convencer os líderes e tomadores de decisão sobre a urgência do problema.

3.1.7. Conclusões sobre a experiência do CNES

Quanto à implementação do arquivamento digital de dados científicos espaciais, podemos apresentar as seguintes conclusões provisórias:

- em termos de armazenamento, consideramos que até o nível de petabytes, as necessidades de preservação física de arquivos são resolvidas com um nível satisfatório de confiabilidade,
- a descrição semântica de dados e o desenvolvimento de metadados descritivos são os elementos-chave da reutilização de dados no futuro. Esses metadados ainda dependem de ontologias e terminologias em rápida evolução. Essa é uma dificuldade real que não deve ser ignorada,
- a sustentabilidade dos sistemas informáticos desenvolvidos para a transferência de dados, sua gestão e sua difusão apresentam problemas de natureza distinta. O desafio aqui é controlar e minimizar os custos de desenvolvimento, manutenção e atualização desses sistemas. Vimos como isso podia ser imaginado, no que diz respeito aos arquivos de dados,
- Pode-se observar que, em disciplinas científicas, surgiram arquivos de dados padronizados, e ferramentas gratuitas para processamento, análise, visualização e serviços de valor agregado foram rapidamente desenvolvidas. Pelo contrário, as disciplinas para as quais realmente não surgiu nenhum formato estão bastante prejudicadas. Daí a necessidade de que essas disciplinas iniciem ou acelerem o trabalho nessa direção.

No nível político e, mais especificamente, nas decisões de implementação, a situação continua incerta e o arquivamento de dados é coberto apenas parcialmente. As diretrizes do CNES para o arquivamento a longo prazo de todos os dados espaciais para os quais existe essa necessidade ainda precisam ser esclarecidas e formalizadas.

O CNES não tem um repositório central para todos seus dados e, portanto, não sabe como prestar uma conta completa de seus ativos. Além disso, não há nada que sugira que todos os dados dos projetos do CNES estejam realmente arquivados. Certos dados não estão sob a responsabilidade de nenhum centro de arquivo identificado, o que obviamente não promete nenhuma garantia de sustentabilidade. Outros estão definitivamente perdidos, como foi o caso de algumas missões no passado. Mesmo que comece a surgir um projeto para constituir um repositório global, serão necessários vários anos antes de se chegar a uma situação gerenciada e controlada.

Na prática, muitas missões espaciais são organizadas no âmbito da cooperação internacional. Portanto, uma agência espacial não pode impor suas normas a todas as demais. Isto reforça, se necessário, a necessidade de uma cooperação o mais estreita possível entre as agências sobre o arquivamento de dados, especialmente porque a comunidade de usuários também é internacional.

3.2. A experiência dos serviços públicos de arquivos franceses

Os Arquivos Nacionais, localizados em Fontainebleau tem recebido, controlado, conservado e facilitado o acesso a arquivos digitais desde o início da década de 1970 como parte do programa Constanza. Eles seguiram uma metodologia comprovada, especialmente adequada para o recebimento de arquivos estatísticos estruturados. Com o desenvolvimento da administração eletrônica e a produção de documentos digitais com assinaturas eletrônicas,

tornou-se crucial desenvolver novas ferramentas que integrem sistemas mais automáticos. O uso de linguagens XML tornou-se comum para garantir a integridade e durabilidade dos arquivos, incluindo a identificação, controle e conversão de formatos, se necessário, assim como a duplicação dos arquivos em sites remotos para assegurar a rastreabilidade.

3.2.1. O contexto em que os experimentos foram realizados

O marco legal abrange tanto os arquivos em papel quanto os digitais desde a promulgação da Lei de Arquivos de 9 de janeiro de 1979. Durante muito tempo, os arquivistas prestaram pouca atenção à tecnologia digital, exceto nos Arquivos Nacionais, que arquivam dados científicos, especialmente bases de dados estatísticas das principais organizações científicas, há cerca de trinta anos.

De fato, a produção massiva de documentos públicos permaneceu em papel por diversos motivos, principalmente pelas exigências legais que requeriam que os atos tivessem validade probatória por meio de um documento escrito com uma marca de validação, como um selo ou uma assinatura. A falta de interoperabilidade entre os sistemas de informação também contribuiu para essa situação, obrigando à materialização da informação ao sair de seu ambiente de produção.

Por fim, uma grande quantidade de informações externas (cartas de outras administrações, cidadãos) que chegavam aos serviços em formato papel continuava sendo o “formato” comum mais adequado e o mais fácil de generalizar.

Apesar disso, desenvolveram-se bases de dados nas administrações, substituindo gradualmente os registros em papel por dados mais complexos e ricos, que permitiam um acompanhamento preciso de casos e situações, assim como a edição de cartas e formulários.

No entanto, a consciência sobre a necessidade de arquivar os dados dessas aplicações surgiu tardiamente entre os arquivistas, já que as operações de transferência se concentravam em documentos antigos produzidos exclusivamente em papel. Além disso, a gestão dessas aplicações empresariais está nas mãos dos departamentos de informática, que não eram os interlocutores dos arquivistas, os quais até então haviam trabalhado exclusivamente com os produtores dos documentos.

Surgiu uma nova consciência com o desenvolvimento da administração eletrônica, que permitia a produção de originais digitais, desde que o processo de negócio fosse desmaterializado. Ao mesmo tempo, graças à atuação da Direção Geral de Modernização do Estado, avançava a interoperabilidade entre os sistemas de informação, que por si só podiam viabilizar essa virtualização.

Ao mesmo tempo, a digitalização massiva de arquivos em papel trouxe desafios quanto à conservação de arquivos em formatos de preservação, sem compressão. Inicialmente, foram feitas cópias em CD-R duplicados, que foram armazenadas em depósitos tradicionais, utilizados para arquivos em papel. No entanto, reconheceu-se tardiamente a necessidade de monitorar e testar esses meios, o que levou vários departamentos a optarem por manter os arquivos em servidores de armazenamento, como discos ou fitas LTO3 ou 4, fornecidos pelos departamentos de TI. Essa decisão foi tomada especialmente quando os volumes de arquivos aumentaram consideravelmente.

3.2.2. Projeto PILAE da Direção de Arquivos da França (DAF)

O Projeto PIL@E, realizado pela Direção de Arquivos da França (DAF), foi lançado no final de 2006, sob a supervisão do Departamento de Inovação Tecnológica e Padronização da DAF, recebeu uma valiosa assistência na gestão de projetos da Direção Geral de Modernização do Estado. (DGME).

3.2.2.1. Contexto

Este projeto, apoiado pelo diretor dos Arquivos da França e pelo diretor dos Arquivos Nacionais, foi considerado estratégico devido às lições essenciais que proporcionaria à comunidade arquivística como um todo. A transição para a nova produção digital nativa requeria uma adaptação nos métodos de trabalho, habilidades e formação, o que representava um desafio inovador para o cliente, o diretor do projeto e os usuários finais da ferramenta. A gestão da mudança se apresentava como um aspecto especialmente complicado.

Uma característica destacada foi a falta de medições e conhecimento preciso sobre os volumes futuros, já que a digitalização completa dos processos de negócio ainda estava em andamento na maioria dos casos.

Decidiu-se implementar o plano piloto no serviço de arquivos eletrônicos do Arquivo Nacional em Fontainebleau, com o objetivo de receber, preservar e comunicar arquivos nativamente digitais produzidos pelos serviços centrais do Estado durante um período transitório de 2009 a 2013. Os Arquivos Nacionais, por meio do projeto de um novo centro de arquivos em Pierrefitte-sur-Seine, empreenderam um esforço significativo para renovar e revisar seu sistema de informação, abrangendo a gestão, comunicação, descrição e distribuição de arquivos em papel dos serviços do Governo Central.

Portanto, o PIL@E deveria ser capaz de receber, armazenar e permitir a busca e consulta de uma variedade de objetos nativamente digitais, incluindo dados de bases de dados comerciais, documentos eletrônicos, intranets colaborativas, mensagens eletrônicas e documentos de uma cadeia de virtualização completa. Além disso, o PIL@E se basearia no modelo OAIS e nas recomendações do repositório de interoperabilidade geral, focando em formatos de documentos compatíveis, políticas de arquivo e formatos padrão de metadados.

3.2.2.2. Características do cofre eletrônico

O diretor do projeto propôs utilizar o pacote de software que comercializava, conhecido como o cofre eletrônico comunicante (CFEC), que oferecia uma série de funções necessárias para o armazenamento seguro e a rastreabilidade:

- funções de verificação de impressões digitais, assinaturas de arquivos, carimbos de tempo baseados em uma fonte externa de tempo “segura” e funções que permitem verificações periódicas desses parâmetros ao longo do tempo
- funções que garantem a rastreabilidade do sistema: edição de logs de eventos que permitem acompanhar as diversas operações de conexão à ferramenta, transferências de objetos para arquivar, consultas, exclusões; os registros são selados regularmente para que a informação desses registros não possa ser alterada,
- funções que permitem controlar as operações de replicação síncrona de dados, cada vez que uma transferência é aceita e validada pelo arquivista: gravação em dois locais remotos (replicação entre sites), junto com replicação dentro do site; no total, uma replicação em quatro pontos de armazenamento em quatro servidores distribuídos em dois locais remotos.

Além disso, a segurança dos dados é garantida por sistemas de redundância tanto dos dados quanto dos bancos de dados, acessos e instalações elétricas, junto com um sistema de backup tradicional integrado à política geral de garantia do departamento de sistemas de informação do ministério.

O cofre eletrônico é baseado em blocos de código aberto e garante a reversibilidade de todos os arquivos, diários e outras evidências geradas, permitindo sua restituição e recuperação em outro sistema, se necessário. Por exemplo, PIL@E poderia funcionar eventualmente com outro cofre eletrônico que não seja a da Cecurity.com.

3.2.2.3. Funções desenvolvidas além do cofre eletrônico

Além dessas funções básicas, era necessário desenvolver outras funções específicas para o arquivamento seguro de dados e documentos dos serviços dos produtores (administrações dos serviços do Governo Central).

Decidiu-se excluir o acesso do público em geral que desejasse consultar arquivos digitais comunicáveis por duas razões principais: a falta de uma reflexão aprofundada sobre o nível de segurança necessário para abrir uma aplicação na Internet que gerenciasse e permitisse o acesso a dados e documentos confidenciais, e a relevância questionável dessa abertura devido a uma cobertura inicialmente desigual entre as diversas áreas administrativas suscetíveis de serem abrangidas, agravada pela natureza não transferível de alguns desses documentos e dados

O núcleo dos desenvolvimentos se concentra na implementação do padrão para troca de dados para arquivo (SEDA), focado nos processos de “transferência” e “investigação e comunicação”.

Foram criadas interfaces e mensagens específicas para os diversos processos entre os ministérios e os Arquivos Nacionais: o envio por parte dos arquivistas depositantes, o recebimento por parte dos arquivistas (processo de controle/validação/rejeição) e o aviso de recebimento. O processo de busca e consulta foi simplificado, eliminando a necessidade da intervenção dos arquivistas do Arquivo Nacional (busca baseada nos direitos de acesso aos arquivos a partir dos metadados e download).

Quanto ao formato de metadados definido pelo padrão de troca de dados para arquivo, é crucial para automatizar a descrição das principais categorias de arquivos enviados periodicamente. Foi proposto um método principal que implica a transferência 'manual' de pagamentos já formatados no formato padrão de intercâmbio. Em seguida, o sistema verifica a transferência recebida (conformidade com o esquema padrão de intercâmbio, verificação de impressões digitais, identificação de formatos) e envia uma mensagem de aceitação ou erro em resposta, conforme os problemas encontrados. Uma vez aceita, a transferência é verificada por um arquivista dos Arquivos Nacionais, que pode modificar ou enriquecer os metadados e decidir se valida ou não o envio.

Este método de funcionamento (importação de arquivos no formato padrão de intercâmbio) é o mais relevante para evitar interrupções no carregamento e o reingresso de metadados, especialmente com grandes volumes de arquivos. É especialmente útil para envios periódicos da mesma categoria de arquivos, onde o trabalho de especificação já foi realizado previamente para formatar o padrão de intercâmbio.

Outro método consiste na inserção manual de metadados por meio de telas de entrada que exibem todos os campos do comprovante de envio definidos pelo padrão de intercâmbio. O arquivista deve determinar a classificação dos arquivos em papel e, em seguida, inserir os metadados associados a cada nível.

É evidente que esse trabalho é tedioso e deverá ser repetido a cada novo envio. A descrição estará relacionada a todo esse conjunto e será um resumo. Uma vez completada a entrada e anexados os arquivos de dados correspondentes, o sistema formatará automaticamente o padrão de intercâmbio. O processo é então idêntico.

3.2.2.4. Questão de conversões de formato

Com base em um acordo estabelecido para a gestão de uma categoria de arquivos, os arquivistas depositantes realizarão sua transferência. Este acordo especifica se o departamento de produção autoriza ou não as conversões de formatos ao ingressar no PIL@E, caso os formatos iniciais não estejam em conformidade com o sistema geral de referência de interoperabilidade para sua conservação a longo prazo. O PIL@E conta com ferramentas detalhadas de identificação de formatos e, de acordo com as regras estabelecidas, pode realizar conversões de formatos (por exemplo, de arquivos da suíte Microsoft Office para o formato PDF/A). Esses processamentos são realizados de forma interativa ou em lotes, especialmente quando se trata de grandes volumes.

Os arquivistas responsáveis pela verificação serão notificados caso sejam detectadas anomalias nas conversões, mas mesmo em caso de falha (ou seja, se nenhuma conversão puder ser realizada), ainda poderão aceitar a transferência. Os formatos originais são preservados no Pacote de Informação de Arquivo (AIP), e os usuários que desejarem solicitá-los posteriormente podem optar pelo formato original ou pelo formato resultante da conversão.

No entanto, são necessários testes exaustivos fora do PIL@E para avaliar com maior precisão os “riscos” associados a determinadas conversões e definir estratégias de acordo com esses riscos, especialmente conforme os tipos de formatos. Foram observados obstáculos em documentos gráficos, por exemplo, durante o projeto, em que os formatos de destino recomendados geralmente não permitiam conservar algumas funcionalidades (especialmente links) dos formatos originais, que infelizmente eram proprietários.

3.2.2.5. Conclusões sobre PILAE

A implementação completa do PIL@E e sua entrada em produção estão previstas para o início de 2010. Os anos de transição que se seguirão permitirão que o produto seja testado e devidamente preparado para a transição para um sistema de maior escala.

Para antecipar essas evoluções, várias questões devem ser estudadas com maior precisão. É crucial concretizar os aspectos relacionados aos formatos de codificação dos documentos e assegurar o uso de identificadores únicos e permanentes para SIP/AIP/DIP. Além disso, surgirão problemas relacionados à segurança dos dados, como a necessidade de um cofre para todos os dados e documentos, a criptografia de dados sensíveis e a arquitetura necessária para abrir o sistema na Internet para o público em geral.

A infraestrutura de armazenamento também deverá ser refinada e ampliada para incluir outras funcionalidades à medida que os volumes aumentem, como a monitoração automática dos meios e a reflexão sobre os tipos de meios utilizados.

Provavelmente seja apropriado considerar o uso de um formato de empacotamento padronizado como METS para os respectivos formatos de SIP, AIP e DIP.

É importante ter em mente que, uma vez criados os AIP, não é possível modificá-los, a menos que se escolham soluções alternativas, como a adição de metadados ou arquivos adicionais.

É normal que todas essas questões surjam durante a implementação de um plano piloto, especialmente porque o campo é inovador e os diferentes atores do mercado envolvidos ainda operam em ambientes compartimentados. Isso inclui especialistas em segurança e assinatura eletrônica, fornecedores de soluções de armazenamento, desenvolvedores de software de gestão de documentos eletrônicos e sistemas de gestão de registros, assim como editores de software de arquivos, presentes no âmbito patrimonial.

Finalmente, uma grande incógnita permanece sobre como os arquivistas dos ministérios e dos arquivos nacionais adotarão o PIL@E. Este é um novo domínio para eles, pois embora os princípios de arquivamento permaneçam os mesmos, a administração eletrônica representa uma mudança significativa nos métodos de trabalho e no apoio à gestão de projetos por parte dos produtores e especialistas em TI, bem como na adoção do conhecimento em termos de formatos XML e linguagens.

3.2.3. A plataforma de arquivo eletrônico do Conselho Geral de Yvelines

O Conselho Geral de Yvelines é a primeira autoridade local na França que desenvolveu uma plataforma piloto. O projeto foi iniciado em 2005 como uma implementação experimental nesse departamento, focando na transferência de documentos sujeitos ao controle de legalidade. A plataforma de teletransmissão assume essa função de transmissão e, uma vez concluída a transação e esgotados os recursos legais, encarrega-se de transferir os documentos para arquivamento.

Este projeto foi realizado internamente no Conselho Geral, com uma estreita colaboração entre os serviços de TI e os Arquivos Departamentais. As funcionalidades principais da plataforma estão muito próximas às do projeto PILAE. O projeto Yvelines se inspira no modelo OAIS, assim como no estudo de custos de plataformas realizado pelo departamento dos Arquivos da França e, finalmente, nas especificações funcionais do PILAE. No entanto, as funcionalidades se concentram na recepção, controle e validação de arquivos que já estão no formato padrão de intercâmbio de dados para arquivo, sem abordar as funções de identificação, controle e possível conversão de formatos de documentos.

Este projeto é interessante por várias razões, especialmente porque teve o efeito de reposicionar os Arquivos Departamentais muito antes no ciclo de vida do documento. Neste caso, os documentos são recebidos rapidamente após sua produção, assim que a transação é validada, sem necessidade de esperar o vencimento do período de vigência administrativa. Além disso, espera-se que seja fornecida uma forma de colaboração para os municípios que, por não possuírem plataformas de arquivo eletrônico, desejem depositar as atas que produzem.

3.2.4. A plataforma de arquivo eletrônico do Conselho Geral de Aube

A plataforma de arquivo eletrônico do Conselho Geral de Aube também apresenta uma abordagem interessante, pois adotou uma estratégia de colaboração ao recuperar a aplicação desenvolvida pelo Conselho Geral de Yvelines, realizando algumas melhorias e atualizações funcionais.

3.2.4.1. Contexto e desafios

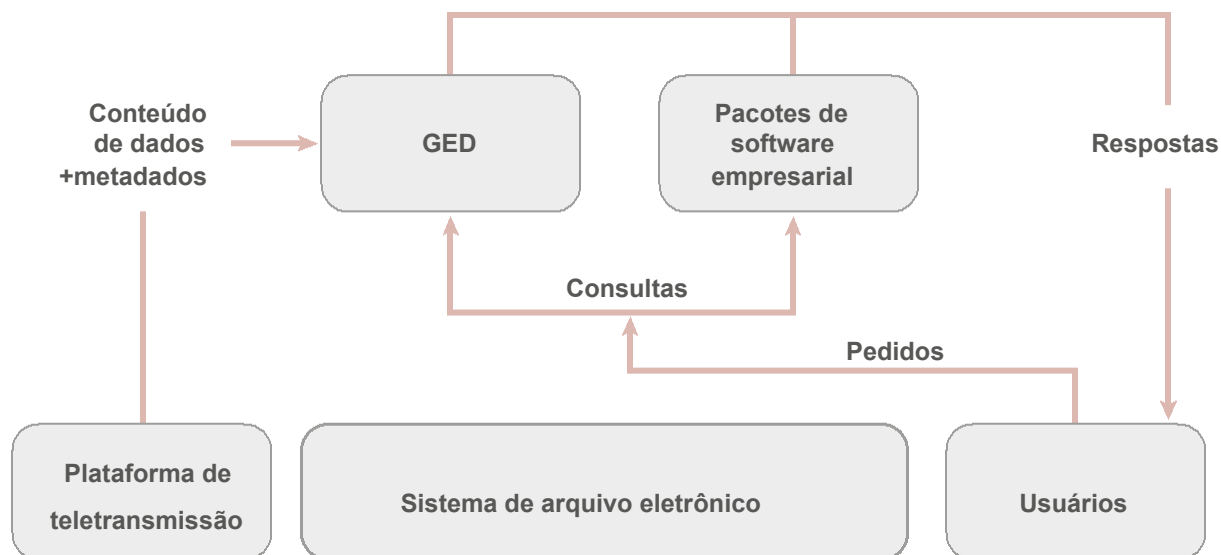
O contexto e os desafios são similares aos de Yvelines, com a implementação de projetos de administração eletrônica e a adoção de ferramentas de assinatura eletrônica por parte do Conselho Geral, o que gerou documentos digitais que precisavam ser preservados. Isso impulsionou a necessidade de desenvolver ferramentas de arquivo eletrônico.

A estreita colaboração entre os serviços de informática e os arquivos departamentais, por meio de um comitê de arquivo eletrônico, perante o qual todos os projetos de transformação, incluindo o arquivo, devem ser considerados desde o início do projeto.

Foi realizada uma profunda reflexão organizativa, jurídica e técnica.

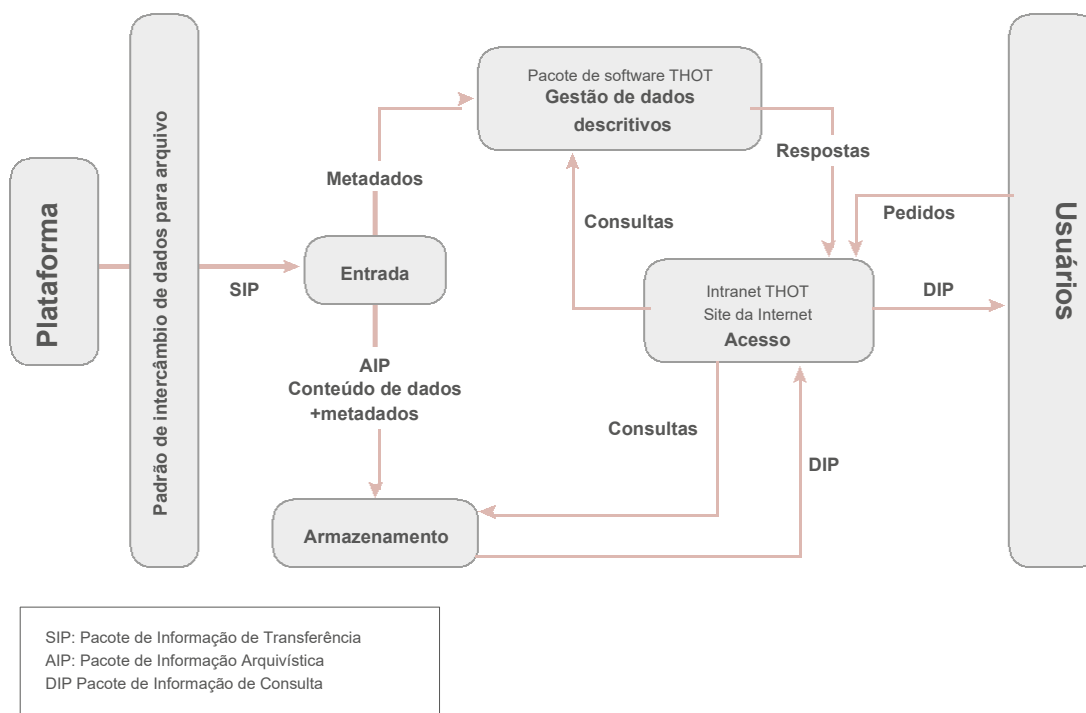
- Decisão para qualquer aplicação comercial de depositar os dados/documentos/fluxos na plataforma de arquivo assim que os trâmites forem concluídos, a fim de assegurar a informação, mantendo também os dados nas aplicações comerciais durante sua vigência. Trata-se de uma mudança importante na organização vigente até então, pois apenas os arquivos denominados definitivos eram sustentados pelos arquivos departamentais, uma vez vencido o período de vigência administrativa. Para os Arquivos, trata-se de intervir muito cedo no processo do ciclo de vida da informação e assim ter maior legibilidade e responsabilidade.
- Estabelecimento de uma contratualização sistemática por meio de assinatura para cada nova categoria documental suportada, por um lado um contrato de serviços entre o serviço produtor, o serviço de arquivos e o departamento de informática, e por outro lado, um protocolo de pagamento, desta vez com o pagamento pelo serviço.
- Acesso aos arquivos depositados (nota: o mesmo acesso está configurado para arquivos independentemente do seu suporte):
 - para os departamentos de produção, por meio de suas aplicações comerciais ou de uma gestão documental eletrônica simples, mas adaptada às suas necessidades.
 - e para outros usuários (leitores de arquivos), através do software de gestão de arquivos anteriormente utilizado para pesquisa e acesso a arquivos em papel (software Thot).

GESTÃO DE SOLICITAÇÕES DE CONSULTA DE UNIDADES PRODUTORAS



Gestão de solicitações de consulta para departamentos produtores (por meio do seu GED / Gestão eletrônica de documentos comerciais) e para pesquisadores e cidadãos, através do software de gestão documental utilizado pelos arquivos departamentais de Aube para arquivos tanto em papel quanto digitais

GESTÃO DE SOLICITAÇÕES DE CONSULTA DA COMUNIDADE DE PESQUISADORES E CIDADÃOS



Gestão de solicitações de consulta para departamentos produtores (por meio do seu GED / Gestão eletrônica de documentos comerciais) e para pesquisadores e cidadãos, através do software de gestão documental utilizado pelos arquivos departamentais de Aube, para arquivos tanto em papel quanto digitais.

3.2.4.2. A plataforma de confiança

Outra inovação: a implementação da plataforma de arquivo confiável direcionada a:

- verificar a assinatura dos documentos transferidos que estão assinados,
- para dar fé desta verificação,
- editar um relatório de verificação que por sua vez esteja assinado,
- assinar o comprovante de envio, uma vez emitido o relatório.

A plataforma também implementará a identificação e conversão, se for necessário, dos formatos dos documentos transferidos.

3.2.5. Comentários de uma importante instituição patrimonial: a Biblioteca Nacional da França (BnF)

A preocupação pela sustentabilidade da informação digital nos últimos anos se estendeu a outros atores importantes do patrimônio, como o Instituto Nacional do Audiovisual (INA) e a Biblioteca Nacional da França (BnF). Um exemplo destacado desta preocupação apresenta-se a seguir.

3.2.5.1. Contexto digital na BnF

A distribuição de materiais digitais foi uma aventura de longa data para a BnF, com mais de 15 anos de programas de digitalização e uma atividade que vem experimentando um forte crescimento com a digitalização massiva, a consulta de documentos audiovisuais em formato digital e o acesso a acervos online de lançamentos, entre outros.

Esses projetos deram lugar à provisão de um número crescente de serviços, como estações de consulta interna, a biblioteca digital Gallica desde 1997, a venda de reproduções digitais sob demanda e a possibilidade de coletar metadados através do protocolo OAI (Iniciativa de Arquivo Aberto), que permite a construção de ferramentas de pesquisa eficientes e a identificação duradoura de recursos digitais. Esses esforços mostram o compromisso da BnF em ser um ator relevante na sociedade da informação, oferecendo ferramentas e serviços para que seus usuários possam acessar, compreender, manipular e colaborar com os recursos digitais.

Os primeiros experimentos de digitalização de imagens fixas e impressos surgiram nos anos 1990-1991. O meio de gravação que foi preservado foi a fita magnética DAT, fornecida em duplicado pelos prestadores de serviços de digitalização, mas ela tinha uma vida útil de 2 a 3 anos, o que gerava desafios em termos de migração de dados.

Naquele momento, a BnF não estava plenamente consciente da magnitude do problema e via o arquivo apenas como um centro de custos em lugar de um seguro para os dados. A equipe buscava uma solução que exigisse a menor quantidade de habilidades e a menor intervenção possível, mas que garantisse a facilidade de leitura a longo prazo. Optou-se por soluções óticas e magneto-óticas, considerando que ofereciam uma oportunidade satisfatória em comparação com os discos magnéticos e as soluções de fita, que eram menos confiáveis ou requeriam ciclos frequentes de migração e atualização. A desvantagem dos suportes CD-R (disco compacto gravável) é sua instabilidade no tempo. Entretanto, uma solução baseada em tecnologia francesa permitiu resolver este problema.

3.2.5.2. Projeto Century Disk e primeiras migrações

O Century Disk, desenvolvido pela empresa francesa Digipress em Caen, oferecia uma técnica inovadora. Consistia em gravar os masters, que em seguida serviam de matriz para prensar os CD-ROM de policarbonato. Este método tinha duas vantagens principais: em primeiro lugar, os dados eram gravados verdadeiramente como um CD-ROM, não somente eram expostos como no caso do CD-R; em segundo lugar, o backup utilizado era vidro temperado no lugar de policarbonato, o que proporcionava estabilidade e durabilidade excepcionais para acessar os dados durante um século ou mais. O processo de produção assemelhava-se ao utilizado para gravar circuitos impressos muito pequenos, como processadores, e requeria um ambiente livre de pó. Graças à inércia físico-química do vidro, obtinha-se uma gravação praticamente eterna. Embora a capa refletora poderia apresentar problemas a longo prazo, teoricamente era possível “voltar a metalizar” os suportes.

Entretanto, essa tecnologia tinha um custo elevado, mais de 207 € por GB, o que à primeira vista parecia astronômico, mas considerando que esse custo poderia ser amortizado ao longo de pelo menos 20 anos, o preço caía para 11 € por GB ao ano, o que o tornava mais aceitável, mesmo pelos padrões atuais. Esse preço foi comparado ao custo de investimento em termos de recursos humanos e materiais necessários para o controle e a regeneração de mídias CD-R a cada 3 anos, durante um período de 20 anos, e os custos eram semelhantes. Diante do risco de perder uma fonte potencial de economia em um contexto de orçamento cada vez mais apertado, essa opção parecia a melhor: pagar mais agora para não se preocupar durante 20 anos ou mais.

O projeto foi lançado em 1997, quando as primeiras fitas estavam em risco, e foram armazenadas em duplicata nos depósitos do novo edifício da BnF, em condições controladas de umidade e temperatura. No entanto, as dúvidas sobre qual estratégia adotar duraram mais de dois anos, enquanto as fitas continuavam se degradando. Ao mesmo tempo, em 1997, a BnF lançou seu projeto de biblioteca digital: Gallica. A necessidade de migrar os dados dos DAT tornou-se imperativa para manter a Gallica online.

A migração para as bibliotecas de discos ópticos digitais (DON) durou dois anos, de 1998 a 1999, e foi observada uma perda de 2,5% das fitas. Essa observação acelerou o projeto de transferência para o Century Disk. As fitas ilegíveis foram copiadas da segunda cópia, utilizando o CD-R como mídia intermediária. A maioria dos CDs do Century foi gravada durante o ano 2000, mas 6% das fitas se perderam durante a operação.

O aumento constante do acervo e o rápido crescimento na quantidade de visitantes ao site da Gallica tornaram necessária uma nova mudança na arquitetura de armazenamento. As bibliotecas de discos DON não foram projetadas para suportar uma demanda tão elevada. Além disso, a falta de um algoritmo de cache de disco no nível dos servidores de consulta dificultava a gestão dos pedidos devido à grande disparidade na demanda de acesso. Em 2001, foi necessária uma nova migração. Desta vez, os dados foram copiados para um sistema de armazenamento totalmente baseado em disco, sem perda de informação durante essa operação.

3.2.5.3. Projeto SPAR

Uma necessidade

Com o surgimento de novos projetos de digitalização, como o da imprensa do século XIX, surgiu uma reflexão sobre os formatos adequados. Dado o estado frágil das publicações de imprensa, com papel de baixa qualidade, a Biblioteca Nacional da França (BnF) optou por mudar de uma digitalização voltada para a divulgação, que priorizava a distribuição fácil, mas comprometia a conservação, para uma digitalização focada na preservação. Essa última buscava manter o documento em seu formato original pelo maior tempo possível, sem alterações. Portanto, abandonou-se o formato JPEG, considerado muito arriscado devido à sua compressão destrutiva, em favor do formato TIFF sem compressão, o que aumentou em quatro vezes o tamanho dos arquivos gerados.

Ao mesmo tempo, os meios para a produção de suportes cinematográficos, especialmente em cores, estavam desaparecendo rapidamente, o que levou à digitalização progressiva de documentos coloridos.

Além desses desenvolvimentos, surgiram novas missões, como a coleta da web francesa. As primeiras coleções web surgiram em 2002 com o projeto de coleta de sites das eleições presidenciais, o que resultou no acúmulo de vários terabytes de dados. Essas coleções continuaram se expandindo, alcançando mais de 150 TB em 2008. A partir de 2005, a BnF também começou a explorar o depósito legal eletrônico como alternativa ao depósito em papel.

Essas mudanças tiveram repercussões significativas nas capacidades de armazenamento, o que obrigou o departamento de Tecnologias da Informação (TI) a aumentá-las regularmente. Com o tempo, o projeto se tornou mais heterogêneo e difícil de manter. Em 2004, diante da crescente complexidade, o departamento de sistemas de informação decidiu lançar um contrato para adquirir uma infraestrutura de acervos digitais escalável em termos de capacidade de armazenamento e que cumprisse os critérios de sustentabilidade em termos de abertura, distribuição e independência tecnológica. Assim nasceu o que se conhece hoje como a infraestrutura SPAR.

A direção da BnF então sentiu a necessidade de ir além nesse processo e iniciar o projeto de um sistema real de preservação: SPAR (Preservar o patrimônio digital da BnF, arquivar todos os seus dados e distribuir o acesso a eles).

Princípios adotados

Para os administradores de coleções digitais, principalmente bibliotecários, existe o risco de perder o contato com os materiais que devem ser preservados. Portanto, era essencial estabelecer um ambiente de confiança e responsabilidade. Por isso, os grupos de trabalho formados no âmbito do projeto concordaram em implementar:

- **Contratos de Serviço:** Esses contratos definem os compromissos tanto do Arquivo com seus usuários quanto dos usuários com o Arquivo. Qualquer alteração no número de cópias de um tipo de documento ou na velocidade de distribuição deve levar em consideração esses contratos, o que implica orçamentos e decisões administrativas bem fundamentadas. O objetivo é evitar assumir compromissos que não possam ser cumpridos.
- **Auditabilidade do Sistema:** Busca-se que o sistema e os procedimentos implementados possam ser auditados periodicamente por pessoas ou organismos externos ao Arquivo, demonstrando assim sua conformidade com os padrões, principalmente o padrão OAIS, e os compromissos estabelecidos nos contratos de serviço. Essa transparência é um compromisso tanto dos usuários quanto dos responsáveis pelo Arquivo para garantir o cumprimento dos compromissos

Os grupos de trabalho definiram quem são os produtores, independentemente da estrutura organizacional atual da BnF. Foram selecionados oito tipos de produtores com base em suas características técnicas e legais específicas.

Além disso, foram estabelecidos modelos de dados e de preservação:

- o modelo METS para o formato de empacotamento,
- o modelo PREMIS para a conservação de metadados,
- e o formato ODRL para formalizar as licenças de uso de documentos:

Os documentos são distribuídos aos usuários por meio de sistemas dedicados, como o Gallica, o banco de imagens e as estações de consulta interna. O SPAR garante que o solicitante esteja autorizado a fazer a solicitação e que o tipo de documento e o formato correspondente sejam acessíveis a ele. Além disso, assegura que sejam respeitadas as restrições de distribuição estabelecidas nas licenças de uso. Por exemplo, se um documento não deve ser impresso, o SPAR informa o solicitante e ajusta o sistema de distribuição de acordo. As licenças de uso são um meio de formalizar a transferência de responsabilidades, garantindo assim a integridade e o cumprimento das políticas de uso estabelecidas.

Os grupos de trabalho estabeleceram o número de versões que seriam preservadas de um documento. Assim, será mantido um máximo de três versões. A versão inicial que nunca será modificada novamente. A versão atual e a versão imediatamente anterior. Essa precaução permitirá que a BnF volte para a versão anterior caso seja observado um erro após uma transformação. Apenas os documentos que possuem contratos de serviço específicos, como é o caso dos registros administrativos que têm períodos de retenção definidos, podem ser destruídos.

A equipe responsável pela gestão de riscos elaborou uma lista preliminar de riscos e avaliou cada um deles. Embora tenham tentado definir um escopo específico para os riscos, perceberam que essa tarefa era extensa demais. Em vez disso, estabeleceram procedimentos para monitorar e lidar com os riscos, o que permitirá que o repositório de riscos evolua ao longo do tempo

Estudo de viabilidade

Era fundamental garantir que a implementação de um sistema OAIS fosse viável para a BnF. A instituição não tinha a intenção de desenvolver um sistema do zero, mas sim construir a partir de uma solução existente que pudesse ser compartilhada com outros. Nesse sentido, foi realizado um estudo de viabilidade baseado no modelo OAIS, em paralelo com outras atividades e sem pré-determinar as opções escolhidas. Esse exercício foi delicado porque a oferta ainda não estava madura no mercado. Inicialmente, foi necessário explorar os produtos existentes que poderiam atender às necessidades.

Vinte produtos foram avaliados em três áreas: qualidade funcional, qualidade técnica e durabilidade comprovada. Esses produtos pertenciam a três categorias diferentes:

- software de gestão de conteúdo que poderia ser adaptado às necessidades, mas nenhum foi selecionado nessa categoria,
- software de “repositório institucional”, como DSPACE ou FEDORA, que geralmente são softwares dedicados ao acesso, mas que incluem mecanismos interessantes de transferência e gestão de dados,
- Software especificamente projetado para a gestão de arquivos.

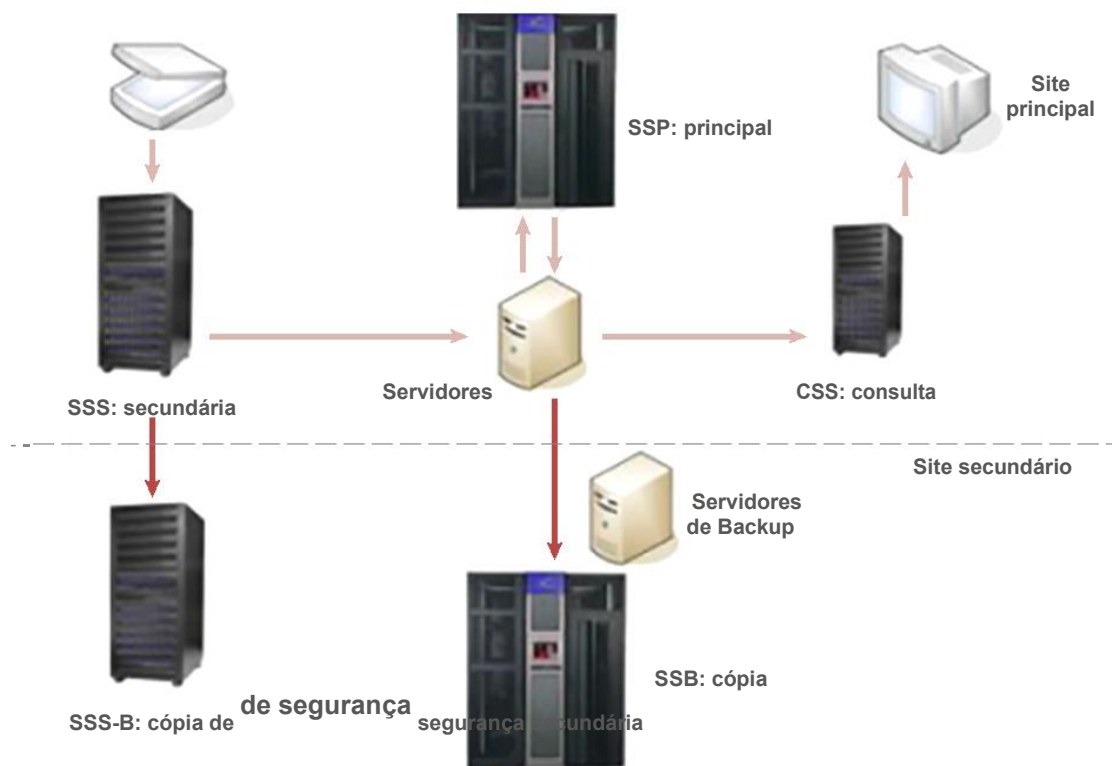
Após essa exploração inicial, seis programas de software foram selecionados para uma avaliação mais detalhada, baseada em um questionário com 132 critérios divididos em 9 categorias que abordavam as funcionalidades definidas pelo padrão OAIS, bem como aspectos técnicos e organizacionais. Este estudo demonstrou que as soluções identificadas podiam atender às necessidades da BnF.

Preservação da informação digital na BnF

A BnF tem a responsabilidade de conservar o patrimônio cultural que abriga, o que inclui também a preservação de suas coleções digitais com o mesmo compromisso que suas coleções físicas. Para a instituição, essa preservação deve contemplar uma perspectiva multissecular.

Portanto, SPAR seria mais que um simples repositório de dados seguro. O sistema garantiria a integridade dos dados por meio da criação de múltiplas cópias, monitoramento contínuo do estado dos meios de armazenamento e a realização das transformações necessárias em caso de obsolescência tecnológica para garantir a acessibilidade contínua. Então, por exemplo, quando o formato de imagem JPEG se tornar obsoleto, o SPAR poderá transformar as imagens afetadas para o formato JPEG do futuro. Além disso, o SPAR permitiria recuperar os dados a qualquer momento e conservar até nove cópias de um objeto digital.

A INFRAESTRUTURA SPAR



Infraestrutura de armazenamento de SPAR

3.2.6. A experiência de uma grande organização de educação superior: PAC, a plataforma CINES

O mundo da educação superior e da pesquisa tem se preocupado há anos com a difusão, publicação e acesso à informação digital. No entanto, a preocupação com a preservação digital é mais recente, como demonstra um dos enfoques das equipes muito grandes (TGE ADONIS) nas ciências humanas, centrado na preservação a longo prazo.

3.2.6.1. Centro Nacional de TI para a Educação Superior (CINES)

O Centro Informático para a Educação Superior (CINES), localizado no sul da França, em Montpellier (Hérault), foi estabelecido em 1999, sucedendo o CNUSC (Centre National Universitaire Sud de Calcul), criado em 1980. Supervisionado pela DGRI (Direção Geral de Pesquisa e Inovação) e pela DGES (Direção Geral de Educação Superior) do Ministério da Educação Superior e Pesquisa, possui duas missões principais: informática digital intensiva e preservação a longo prazo de documentos eletrônicos. Desde 2004, CINES tem trabalhado no desenvolvimento de um serviço de arquivamento a longo prazo do patrimônio científico.

A Informação Científica e Técnica (IST) abrange toda a informação produzida ou recebida pelos setores de pesquisa e educação. Inclui uma variedade de conteúdos e formatos, desde revistas científicas, teses, relatórios, anais, trabalhos especializados, manuais, bibliografias, resumos, pré-publicações, patentes, mapas, até dados estatísticos e bancos de imagens e vídeos.

O sucesso de uma estratégia nacional para o arquivamento a longo prazo de documentos eletrônicos produzidos pela comunidade científica e técnica requer enfrentar vários desafios, que vão desde o desenvolvimento de novas competências comerciais (com o estabelecimento de um serviço dedicado ao arquivamento, definição de novos processos e métodos a serem implementados e dominados para gerir um sistema de arquivamento), até a conscientização da comunidade sobre a importância da preservação a longo prazo (com o fortalecimento das colaborações entre profissionais de informática, arquivistas e bibliotecários, a organização de jornadas de conscientização e informação sobre o arquivamento a longo prazo, etc.).

Essa implementação também deve ser realizada em conjunto com os demais serviços de arquivos envolvidos nessa cadeia, de acordo com a origem e a natureza dos documentos (Arquivos Departamentais, Arquivos Nacionais, BnF), respeitando o contexto legislativo.

3.2.6.2. Implementação da plataforma PAC (plataforma de arquivo CINES)

Para cumprir sua missão de arquivamento, o CINES lançou o projeto PAC, com o objetivo de estabelecer um serviço de arquivamento digital sustentável. A equipe é composta por um diretor de projeto, quatro engenheiros e um arquivista.

Na fase 1, foi desenvolvida internamente uma primeira plataforma para validar os serviços esperados no projeto de arquivamento eletrônico de teses, com uma capacidade de armazenamento inicial de 300 GB. A plataforma baseou-se em padrões da indústria, como o modelo OAIS e o formato de metadados Dublin Core, e foi deliberadamente limitada a uma lista de formatos de arquivo padrão e amplamente utilizados, como HTML, PDF, TXT, XML, GIF, JPEG, TIFF e PNG, assim como WAV. A arquitetura foi baseada em software de código aberto (Java, PostgreSQL, Jhove, ImageMagick), e os primeiros testes de arquivamento de teses foram realizados em março de 2007, com o início da produção no final daquele mesmo ano.

Na fase 2, no final de 2007, foi lançada uma licitação para adquirir uma plataforma de armazenamento capaz de lidar com grandes volumes (20 TB expansíveis para 40 TB). Essa plataforma também se baseou em padrões e normas do campo, utilizando hardware SUN, software Arcsys e software de código aberto como Java e MySQL, junto com Jhove e ImageMagick para a identificação, validação e caracterização de formatos. Os primeiros testes de transferência e migração de documentos começaram em março de 2008, com o início das operações no final do segundo trimestre daquele ano.

A plataforma oferece diversas funcionalidades, como o recebimento e controle dos SIPs (cumprimento de metadados, controle e validação do formato do arquivo), a criação do AIP, o armazenamento (cópia múltipla do AIP em diferentes suportes, verificação periódica da integridade do AIP, migração tecnológica, geração de relatórios e estatísticas) e o acesso (autenticação do solicitante, consulta e comunicação).

3.2.6.3. Arquivamento de teses e revistas eletrônicas de humanidades e ciências sociais

Quanto às teses, o processo foi iniciado após a promulgação do decreto de 7 de agosto de 2006, que estabelece as modalidades de depósito, notificação, reprodução, distribuição e conservação das teses ou obras apresentadas em defesa para obtenção do doutorado.

Os estudantes de doutorado apresentam suas teses em formato eletrônico na biblioteca universitária onde defendem seu trabalho. Depois, as bibliotecas enviam as teses eletrônicas para a Agência Bibliográfica para a Educação Superior (ABES) por meio da ferramenta STAR. Após três etapas de validação, as teses elegíveis para arquivamento são transferidas para a plataforma PAC, e uma cópia pode estar disponível online para a comunidade da Internet através de um site de divulgação.

O projeto de arquivamento de revistas foi iniciado em 2006 como resposta a um projeto de digitalização em massa e preservação de coleções retrospectivas de revistas de humanidades e ciências sociais, liderado pela equipe de Persée (Université Lumière - Lyon 2). A cadeia de digitalização garante a digitalização em massa, a centralização e automação do processamento, bem como o arquivamento a longo prazo dos dados.

Esta cadena de documentación incluye herramientas para la descripción de la colección, herramientas de archivo y seguimiento, pasos de control de calidad y validación, y, finalmente, herramientas para la difusión de los datos generados.

Bibliografia

BANAT-BERGER F., HUC C., DUPLOUY L., *L'Archivage numérique à long terme, les débuts de la maturité?* (Primera obra de síntesis sobre el archivo digital en lengua francesa) Paris, La Documentation française, 2009

BANAT-BERGER F., HUC C., Module 7 - Gestion et archivage des documents numériques. Portail International Archivistique Francophone. 2011. <https://www.piaf-archives.org/se-former/module-7-gestion-et-archivage-des-documents-numeriques> (Se identifica en el texto como PIAF)

Castillo Solano, M.G y Umaña Alpízar, R. (2018) *Modelo de Preservación de Documentos Digitales en la Administración Universitaria Estudio de Caso: Universidad Nacional*.



ARCHIVO NACIONAL
COSTA RICA



UNIVERSIDAD DE
COSTA RICA