

MÓDULO 7: Gestión y preservación de documentos digitales

SECCIÓN 12
ESTUDIOS DE CASO



ARCHIVO NACIONAL
COSTA RICA



UNIVERSIDAD DE
COSTA RICA

MÓDULO 7

Gestión y preservación de documentos digitales

SECCIÓN 12

Estudios de caso

Adaptación del Archivo Nacional de Costa Rica
Versión 1, 2024

Este curso fue traducido y adaptado por la Dirección General del Archivo Nacional de Costa Rica en colaboración con la Sección de Archivística de la Universidad de Costa Rica a partir del material original del año 2011 de la Asociación Internacional de Archivos Francófonos disponible en línea en el Portal Internacional Archivístico Francófono. Se aclara que pueden existir variaciones respecto al contenido original. Para acceder al material en francés, visite <https://www.piaf-archives.org/se-former/module-7-gestion-et-archivage-des-documents-numeriques>.



Contenido

Capítulo 1. Objetivo de la sección	4
1.1. Introducción	4
Capítulo 2. Contexto global	4
2.1. Internacional	4
2.2. Europa	9
2.3. Estados Unidos.....	11
Capítulo 3. Experiencias en Francia.....	11
3.1. La experiencia de una gran organización científica: el Centro Nacional de Estudios Espaciales (CNES)	11
3.2. La experiencia de los servicios públicos de archivos franceses	17
Bibliografía	31

Capítulo 1. Objetivo de la sección

1.1. Introducción

En esta sección se presentarán varios ejemplos de implementación de plataformas de archivo electrónico en varios sectores: científico, patrimonial, institucional (archivos, biblioteca, etc.).

En el segundo capítulo, se presentará una breve incursión en proyectos y herramientas globales que evidencian la práctica generalizada por la creación e implementación de archivos electrónicos a nivel internacional. Solo se mencionan algunos de estos logros sin entrar en detalles, pero con la posibilidad de acceder a cada uno de ellos, siguiendo los enlaces. Se pretende ilustrar con esto las actividades en torno al archivo digital y recalcar la necesidad de una cooperación saludable dentro de la globalización de la comunicación, incluso más esencial que en la gestión tradicional de intercambios profesionales.

El tercer capítulo trata de la retroalimentación de dos tipos de experiencias que tuvieron lugar en Francia:

- por un lado, el análisis de las múltiples dificultades encontradas en el pasado en la gestión, sostenibilidad y disponibilidad de la información digital,
- por otro lado, el examen de las primeras lecciones resultantes del establecimiento de organizaciones y plataformas de archivo digital en diversos campos, de la aplicación del modelo de referencia OAIS y otros estándares aplicables al campo del archivo digital.



NOTA

En esta sección no se identifica el texto con caracteres en negrita, no se resaltan términos del glosario que ya han sido asimilados, ni se proponen temas “complementarios”. Por cuanto, se asume un mayor nivel de conocimiento de quienes consultan este módulo.

Si tiene dificultades de comprensión, no dude en retomar el estudio de los apartados anteriores, esta vez acercándose a los “complementos”.

Capítulo 2. Contexto global

A nivel internacional, existen muchas iniciativas, programas de estudio e investigación, logros y proyectos de instituciones públicas y empresas privadas.

2.1. Internacional¹

Se presenta una breve descripción de proyectos, instrumentos, herramientas y aplicaciones que brindan servicios de caracterización de contenido, reconocimiento automático de formatos, comparación automática de objetos digitales,

¹ Este apartado se actualiza al contexto global de 2020-2021, a partir del Trabajo Final de Graduación para optar al Grado de Máster en Administración Universitaria titulado “Modelo de Preservación de Documentos Digitales en la Administración Universitaria Estudio de Caso: Universidad Nacional” por María Gabriela Castillo Solano y Raquel Umaña Alpizar (2018).

identificación y extracción de metadatos, validación de la integridad de los documentos digitales, así como otras funcionalidades de preservación digital que se han desarrollado en el ámbito internacional.

Apache Tika (Content analysis tool kit)

El kit de herramientas de Apache Tika™ detecta el texto y extrae metadatos de objetos digitales, sirve para la indexación de motores de búsqueda, análisis de contenido y traducción (tomado de <https://tika.apache.org/>).

Arca

Arca es un repositorio digital orientado a la preservación de los documentos a través del tiempo, desarrollado por la empresa Business Integrators Systems Limitada. Implementa el modelo OAIS y normas de buenas prácticas archivísticas, que garantiza el valor legal de sus documentos en custodia y, también, que en el futuro estos serán accesibles, representables y legibles (tomado de <http://www.bis.co.cr/Products/Arca>).

Archivemática

Es una aplicación de código abierto diseñada con la finalidad de preservar a largo plazo los contenidos digitales, de tal manera que su autenticidad y fiabilidad perduren durante el paso del tiempo. Archivemática está diseñado de acuerdo con requerimientos del modelo funcional ISO-OAIS (tomado de <https://www.archivematica.org/>).

Catalogue of Criteria for Trusted Digital Repositories

El Nestor Working Group, autores de Catalogue of Criteria for Trusted Digital Repositories, identificaron las condiciones para poder crear depósitos y archivos digitales seguros, auditables y certificables (tomado de http://files.dnb.de/nestor/materialien/nestor_mat_08-eng.pdf).

C3po (Clever, Crafty Content Profiling of Objects)

Es una herramienta de creación de perfiles de contenido para análisis de preservación, utiliza datos meta extraídos de archivos de una colección digital como entrada para generar un perfil del conjunto de contenidos. Está diseñado de manera que se puedan integrar fácilmente diferentes formatos de metadatos procedentes de distintas herramientas. Actualmente, soporta metadatos FITS y metadatos Apache TIKA (tomado de <http://peshkira.github.io/c3po/>).

DAITSS

“Dark Archive in the Sunshine State”: desarrollado por Florida Centre for Library Automation, es un software libre para repositorios digitales, utiliza el esquema de metadatos METS y está basado en OAIS.

DAN

“Documentos de Archivo en la Nube”: es un proyecto desarrollado por la Universidad de la Columbia Británica, Canadá, bajo el contexto del proyecto InterPARES, con el fin de investigar los beneficios y riesgos de guardar documentos de archivo en la *Nube*, identificando y examinando los diversos factores administrativos, operativos, legales y técnicos relativos con el almacenamiento y gestión de documentos de archivo, mediante la tercerización con *Proveedores de Servicios en la Nube* (PSN).

DRAMBORA

El grupo de trabajo RLG-NARA junto con el Center for Research Libraries (CRL), el Digital Curation Centre (DCC) y el Digital PreservationEurope (DPE), desarrollaron el *Digital Repository Audit Method Based on Risk Assessment* (DRAMBORA), una herramienta gratuita que permite evaluar un depósito digital, determinando sus capacidades, debilidades y fortalezas, basada en el análisis de riesgos (tomado de <http://www.repositoryaudit.eu/>).

DROID (Digital Record Object Identification)

Es una herramienta de identificación de formato de fichero, analiza agrupamientos de archivos por formatos y brinda reportes para determinar si se requiere aplicar acciones de migración de formato. Es el primero de una serie planificada de herramientas desarrolladas por los Archivos Nacionales del Reino Unido bajo el paraguas de su servicio de registro técnico de PRONOM (tomado de <http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>).

DSpace

Es un paquete de *software* de repositorio de código abierto utilizado para crear repositorios de acceso abierto para contenido digital publicado. Desarrollado en 2002 por HP y el Instituto Tecnológico de Massachussets (MIT). En 2007 HP y MIT formaron la Fundación DSpace y en 2009 en conjunto con la organización de Fedora Commons se unieron para crear Dura Space (tomado por <http://www.dspace.org/>)².

DuraCloud

Dura Cloud es un servicio de almacenamiento de Dura Space que permite controlar dónde y cómo se conservan los objetos digitales en la *nube*. En 2011, la Universidad de Harvard llevó a cabo un proyecto piloto utilizando esta solución (tomado de <http://duracloud.org/>).

E-journal Archiving

Con fondos de la Fundación Andrew W. Mellon, Harvard llevó a cabo un estudio para la construcción de un repositorio para revistas electrónicas. Como seguimiento LDI financió un proyecto colaborativo con la Biblioteca Nacional de Medicina (NLM) para producir una definición de tipo de documento XML (DTD, por sus siglas en inglés). El DTD está diseñado para aumentar la facilidad de intercambio entre editores y archivos para el contenido de la revista electrónica a nivel de artículo. DTD cuenta con una estructura modular que permite la personalización y que debería ser un objetivo fácil de transformación a partir de contenido XML o SGML codificado

ERA

“Electronic Records Archives”: ERA permite realizar la transferencia y preservación de documentos digitales al NARA, a fin de que esta última garantice la integridad de los documentos digitales y la consulta por parte de los usuarios (tomado de <https://www.archives.gov/era>).

FEDORA

Flexible Extensible Digital Object Repository y Architecture: desarrollado por la Universidad de Cornell en 1997, es un sistema de repositorio de código abierto para la gestión y difusión de contenidos digitales, es administrado por la organización sin fines de lucro Dura Space (tomado de <http://fedorarepository.org/>).

FIDO (Format Identification for Digital Objects):

Es una herramienta de CLI de identificación de formato simple para objetos digitales que utiliza firmas PRONOM convertidas en expresiones regulares. La funcionalidad de FIDO es similar a DROID (tomado de <http://openpreservation.org/technology/products/fido/>).

² DuraSpace es una organización sin fines de lucro que se crea en 2009 cuando se fundan Fedora Commons y la Fundación DSpace. Cabe destacar que este portafolio de tecnologías de código abierto es desarrollado por bibliotecarios, archivistas, tecnólogos e investigadores.

FITS (File Information Tool Set)

Es una iniciativa de la Universidad de Harvard, que desarrollaron conjunto de herramientas de información de objetos digitales que identifica, valida y extrae los metadatos técnicos de diferentes formatos de archivo (<https://projects.iq.harvard.edu/fits/home>).

Functional Requirements for Evidence in Recordkeeping

The Functional Requirements for Evidence in Recordkeeping es un proyecto de la Universidad de Pittsburg, que con el apoyo de la National Historical Publications and Records Commission, desarrollaron un conjunto de requerimientos funcionales de gestión documental para el diseño e implementación de sistemas electrónicos de información (tomado de <http://www.archimuse.com/papers/nhprc>).

GDFR (Global Digital Format Registry)

Es un proyecto que ya no está activo, fue una iniciativa desarrollada por Harvard University Library, cuyo objetivo era construir un registro de formatos de preservación digital.

InSPECT

El proyecto InSPECT (Investigating the Significant Properties of Electronic Content Over Time) se ha dedicado al análisis de los elementos o características esenciales (propiedades significativas) para la conservación de un objeto digital (tomado de <https://searchworks.stanford.edu/view/9520222>).

InterPARES

El *Proyecto Internacional para la Investigación en Documentos de Archivo Auténticos y Permanentes en Sistemas Electrónicos* (InterPARES), que da inicio en 1999, a cargo de la doctora Luciana Duranti como directora y es uno de los proyectos principales a nivel internacional. Está constituido de 4 etapas, la última es *InterPARES-Trust*, que se enfoca en los factores introducidos por la gestión de documentos de archivo y datos en línea con el proyecto DAN (tomado de <http://www.interpares.org/>).

Jhove

JSTOR/ Harvard Object Validation Environment: es una herramienta que se utiliza para la detección y validación de formatos, e incluso la extracción de metadatos técnicos de manera automática. Ha incorporado a DROID dentro de su estructura para extender los formatos que valida (tomado de <https://jhove.openpreservation.org/>).

LOCKSS

“Lots of Copies Keep Stuff Safe”: desarrollado por la Universidad de Stanford, es un software que comprueba la integridad de los documentos, mantiene múltiples copias en lugares dispares y realiza comparaciones periódicas entre ellos para asegurar la autenticidad de los objetos digitales, utilizado principalmente para revistas electrónicas (<https://www.lockss.org/>).

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

Este proyecto nace en el 2006 con el objetivo de brindar soluciones al tema de la preservación a largo plazo y el aseguramiento del acceso a los documentos de instituciones europeas. PLANETS incluye criterios de TRAC (establece pautas para medir el cumplimiento del modelo OAIS), y NESTOR que se utilizan de manera automática con la herramienta *Plato*.

Plato

Es una herramienta de planificación que sirve de apoyo a las decisiones de preservación de objetos digitales, mediante servicios de caracterización de contenido, acciones de preservación y comparación automática de objetos en una arquitectura orientada a servicios, e integra DROID, JHOVE y FITS. Asimismo, MiniMEE es un servicio de migración de formatos incluido en Plato.

Pronom

Es un registro de formato que se puso en marcha en el 2004 por los Archivos Nacionales de Reino Unido, su capacidad de reconocimiento es de alrededor de 150 formatos; posee una herramienta Open Source, el Digital Record Object Identification (*DROID*) para la identificación y descripción automática del formato de los ficheros, validar el formato de un conjunto de ficheros, además, le permite a Pronom la integración con sistemas informáticos más amplios (tomado de <https://www.nationalarchives.gov.uk/PRONOM/Default.aspx>).

PUIDs (Pronom Unique Identifiers)

Se ha creado bajo el proyecto PRONOM, por las deficiencias de los descriptores MIMEtypes, ya que este distingue entre las diferentes versiones y los subformatos (tomado de http://www.nationalarchives.gov.uk/aboutapps/pronom/pdf/pronom_unique_identifier_scheme.pdf).

RODA “Repository of Authentic Digital Objects”

Es una solución completa de repositorio digital que entrega todas las principales unidades funcionales del modelo de referencia OAIS. RODA es capaz de gestionar y proporcionar acceso a diversos tipos de contenidos digitales producidos por las grandes corporaciones o instituciones públicas. RODA se basa en tecnologías de código abierto y soporta gran parte de los estándares existentes, tales como el Sistema de Información de Archivos Abiertos (OAIS), codificación y transmisión de metadatos estándar (METS), Descripción Archivística Codificada (EAD), Dublin Core (DC) y PREMIS (metadatos para la conservación) (tomado de http://roda-community.org/?locale=es_CL#welcome).

SPIRT

El proyecto SPIRT y su esquema de metadatos RKMS³ (Record keeping Metadata Schema) publicado en 1999, es un proyecto de la Universidad Australiana de Monash. El acrónimo SPIRT proviene de la beca de investigación que dio nombre al proyecto: *Strategic Partnership with Industry- Research & Training*, investigación que fue pensada para construir un marco de trabajo en el que otros estándares de metadatos de otros sectores pudieran desarrollar su propia aplicación (tomado 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)

Son directrices técnicas para digitalizar materiales de archivo que definen los enfoques para la creación de copias digitales para facilitar el acceso y la reproducción. Estas directrices son la base de los procedimientos utilizados por el Laboratorio de imágenes Digital del Laboratorio Preservación de medios especiales de NARA (tomado de <https://www.archives.gov/files/preservation/technical/guidelines.pdf>).

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

Trustworthy Repositories Audit & Certification: Criteria and Checklist

En 2003, Research Library Group RLG y National Archives and Records Administration NARA crearon un grupo de trabajo para analizar la posibilidad de crear repositorios digitales seguros que se pudieran auditar, evaluar y certificar. El reto fue la elaboración de los criterios de certificación y delinear un proceso para certificación aplicable a una amplia gama de repositorios digitales y archivos, desde la preservación institucional y académica de repositorios de grandes archivos de datos y de las bibliotecas nacionales a los servicios de archivo digital a terceros. (RLG-NARA, 2003, p. 1). Es la base de la Norma ISO 16363:2012 Sistemas de transferencia de información y datos espaciales: auditoría y certificación de repositorios digitales confiables.

UDFR (Unified Digital Formats Registry)

La UDFR es una iniciativa de abril de 2009 para crear un único registro de formatos compartidos. El UDFR se desarrolló en el Centro de Curación de la Universidad de California (UC3) con fondos de la Biblioteca del Congreso (tomado de <http://www.udfr.org/>).

VERS

VERS es el proyecto de investigación que promovió el Public Records Office Victoria (PROV) en el año 1996 y que se volvió un estándar cuyo objetivo es garantizar la veracidad y la autenticidad de los documentos digitales; describe una metodología para convertir los registros digitales en formatos a largo plazo conocidos como objetos encapsulados VERS (tomado de <https://prov.vic.gov.au/recordkeeping-government/vers>).

VidArch

El Proyecto VidArch abordó tanto aspectos teóricos como prácticos del contexto de la preservación digital, desarrollando un marco de preservación para el contexto de video digital aplicando a dos importantes colecciones: la serie completa de videos educativos de la NASA y el conjunto completo de videos jurados ACM SIGCHI presentados en conferencias anuales desde 1983 hasta la actualidad (tomado de <https://ils.unc.edu/vidarch/>).

Xena

XML Electronic Normalizing for Archives: es una herramienta desarrollada por los Archivos Nacionales de Australia, para ayudar en la preservación a largo plazo de los registros digitales, detecta los formatos y los normaliza convirtiendo los objetos digitales en formatos abiertos para preservación. Xena, es un componente de la Plataforma de Software de Preservación Digital (DPSP) (tomado de <http://xena.sourceforge.net/>).

2.2. Europa

En Europa, hay muchos logros, especialmente a nivel de las instituciones, y abundan los proyectos en todas partes.

Comisión Europea

Las acciones emprendidas y llevadas a cabo por la Comisión Europea en materia de archivo digital se ubican en varios niveles:

Un conjunto de pautas que tengan un impacto indirecto significativo en el archivo digital aplicable a todos los países de la Unión. En este sentido, podemos citar la Directiva 1999/93 / CE de 13 de diciembre de 1999, sobre un marco comunitario para la firma electrónica o incluso la Directiva 2007/2 / CE de 14 de marzo de 2007 que establece una Infraestructura de información geográfica en la Comunidad Europea (INSPIRE).

Acción más proactiva y dirigida directamente sobre aplicaciones de archivo electrónico, en particular en el campo de la gestión de documentos, con la organización y financiación del desarrollo de MoReq 2010 (Requisitos estándar para el control del archivo electrónico).

Preparado bajo la guía científica del Foro DLM y publicado bajo los auspicios de la Comisión <https://www.moreq.info/>

Un conjunto de acciones seleccionadas dentro del Programa Marco de Investigación y Desarrollo (PCRD) de la Unión Europea. Así, se dedicó un presupuesto de varias decenas de millones de euros al análisis, investigación y experimentación de soluciones globales adaptadas a un contexto de archivo particular, al desarrollo de sinergias entre los actores del problema, la educación, la enseñanza y la sensibilización dentro de comunidades que a menudo estaban muy alejadas de estas preocupaciones.

Las acciones de investigación que involucran a un gran número de actores geográficamente dispersos dentro de un mismo proyecto tienen una relación resultado/ inversión generalmente moderada. Por otro lado, estas acciones juegan un papel real en el acercamiento de todos los actores de la cadena, así como de los diversos sectores de actividad enfrentados a la problemática, en la sensibilización, en la provisión de información, resultados de estudios y análisis.

Los proyectos definidos y financiados en el marco del PCRD están limitados a unos cuantos años. La continuación de la actividad realizada por los socios del proyecto más allá del período de financiación suele ser incierta. Además, estos proyectos son el resultado de la iniciativa de una amplia variedad de colectivos e instituciones. De ello se desprende que es difícil garantizar la coherencia y la complementariedad de los distintos proyectos. Para subsanar estas deficiencias, varias organizaciones importantes en el campo del archivo digital se han unido en una entidad que lleva el nombre de “Alianza de Acceso Permanente” <http://www.alliancepermanentaccess.org/>

Estas organizaciones también representan el campo del archivo patrimonial de documentos y publicaciones (British Library, bibliotecas nacionales de los Países Bajos, Alemania, archivos nacionales de Suecia) y el campo de archivos de datos (Agencia Espacial Europea, Centro Europeo de Investigación Nuclear). La orientación general se limita al campo científico en sentido amplio, abarcando la física, la biología, la química, así como las ciencias ambientales o las ciencias humanas.

La ambición de la Alianza es ser el lugar para el desarrollo de colaboraciones y relaciones entre archivos, ser influyente y comunicador de sus necesidades frente a las autoridades nacionales y europeas. Su objetivo es garantizar que los beneficios de la preservación digital se obtengan de manera efectiva y eficiente. A través de un “Centro Virtual de Excelencia” (VCOE), cuyo enfoque es crear una visión común y continuar el viaje de integración del panorama europeo de la preservación digital

Coalición de Preservación Digital (Digital Preservation Coalition- DPC)

Existen otro tipo de organizaciones que atienden el tema de la preservación digital, un ejemplo es la DPC <https://www.dpconline.org/> una fundación benéfica impulsada por sus miembros y dedicada al fomento de la preservación digital.

Entre sus objetivos se encuentra⁴:

- Ayudar a miembros de todo el mundo a proporcionar acceso adaptable y a largo plazo a contenidos y servicios digitales mediante la creación de vínculos con la comunidad, labores de promoción dirigidas, formación

4 Tomado del documento “DPC Prospectus 2020-2021” en español, disponible en el sitio web de la DPC.

y capacitación, desarrollo de infraestructura y competencias, normas y buenas prácticas, y buena gestión y gobernanza.

- Facilitar un programa diseñado para satisfacer las necesidades de los miembros y que beneficia a la comunidad global dedicada a la preservación digital.
- Luchar por un patrimonio digital con el futuro asegurado, y ofrecer colaboración y asistencia a todos los sectores, agencias y personas que compartan la misma visión y valores.

2.3. Estados Unidos

Alianza Nacional de Administración Digital (National Digital Stewardship Alliance- NDSA)

Corresponde a otra de las iniciativas en donde su participación nacional e internacional se basa en una membresía. Es una organización de la Federación de Bibliotecas Digitales (DLF) y se identifica como un consorcio de organizaciones comprometidas con la preservación a largo plazo de la información digital. <https://ndsa.org/>. Su trabajo se basa en la conformación de grupos de trabajo e interés que se centran en amplias áreas de la preservación digital.

Entre sus trabajos y aportes destacan los Niveles de Preservación Digital (LoP) como un recurso para los profesionales de la preservación digital cuando construyen o evalúan su programa de preservación digital.

Capítulo 3. Experiencias en Francia

Se analizarán varias entidades, la primera concierne a la de una organización científica, las siguientes son retroalimentación de archivos institucionales nacionales o locales cuyas experiencias son de interés primordial.



NOTA

Se aclara que las experiencias en Francia fueron presentadas por el Portal Internacional Archivista Francófono (PIAF) en su versión original de 2011, por lo que no se cuenta con datos recientes que describan la situación actual de cada proyecto o entidad.

3.1. La experiencia de una gran organización científica: el Centro Nacional de Estudios Espaciales (CNES)

La retroalimentación del CNES merece ser analizada de varias formas.

En primer lugar, es testimonio de una amplia experiencia en el tema, experiencia que ha dado lugar a logros que tienen un valor ejemplar y que han contribuido en gran medida al surgimiento de una metodología hoy ampliamente reconocida.

En segundo lugar, también revela las brechas que pueden existir entre, por un lado, los ingenieros e investigadores de la materia que son muy conscientes de los problemas planteados y los riesgos incurridos, y, por otro lado, los

directores y presidentes de los departamentos de instituciones públicas que tienen una visión más vaga de estas cuestiones que a veces consideran un tema esencialmente técnico.

3.1.1 Datos de ciencia espacial

Las misiones espaciales se pueden clasificar en diferentes categorías según sus objetivos: algunas tienen una vocación fundamentalmente científica, otras están orientadas a aplicaciones de diversa índole: telecomunicaciones, televisión, localización y recolección de datos, meteorología, observación de la Tierra, aplicaciones que pueden ser civiles o militares. Algunas misiones, en particular todas aquellas que observan y estudian la Tierra y su entorno, tienen como objetivo satisfacer al mismo tiempo necesidades científicas a largo plazo y cuestiones económicas a corto y medio plazo como el seguimiento de la vegetación o el estudio del ciclo del agua.

Las misiones científicas cubren un amplio espectro de campos. El estudio del Universo ocupa allí un lugar importante, siendo en primer lugar las diversas disciplinas de la astronomía (astrofísica, astroquímica, astrometría, etc) y la exploración del sistema solar y sus diversos planetas aún muy mal conocidas. La observación del sol y los fenómenos que allí ocurren, el estudio de los efectos de estos fenómenos en nuestro planeta y su entorno ionizado también forman parte de ella. Otro aspecto muy importante es el del estudio y observación de la Tierra, su forma y superficie (geodesia), sus océanos, su evolución a medio y largo plazo en todos los ámbitos. Otras áreas importantes de la ciencia dan lugar a experimentos en química, biología, medicina en gravedad cero.

Todas las observaciones se llevan a cabo utilizando instrumentos complejos que han sido diseñados para este propósito y que proporcionan mediciones en forma de datos digitales. Una característica especial de un gran número de misiones es la capacidad de estos instrumentos para realizar observaciones sistemáticas de larga duración y proporcionar datos de forma continua durante varios años. Esto explica los grandes volúmenes de datos generados por esta actividad. Estos datos se transmiten al suelo mediante una señal electromagnética codificada denominada telemetría. Desde principios de la década de 1980, los datos producidos por los instrumentos se han digitalizado a bordo de la nave espacial y luego se han transmitido a tierra.

La necesidad de conservar la mayor parte de estos datos a largo plazo responde a dos imperativos: un imperativo científico y un imperativo patrimonial.

3.1.2. Cuestión de formatos de almacenamiento y grabación

A principios de la década de 1970, los datos se registraron por primera vez en cintas magnéticas de 7 pistas (6 pistas para datos y una llamada pista de comparación para la verificación de errores de bits) con densidades de grabación de 200, 556 y luego 800 bpi (bits por pulgada). Estas cintas desaparecieron rápidamente y fueron reemplazadas por cintas de 9 pistas (8 para datos más una para la comprobación) en relación con el desarrollo de conjuntos de caracteres de ocho bits (por ejemplo, EBCDIC de IBM). Las densidades de las cintas de 9 pistas evolucionaron de 800 bpi a 1600 bpi y luego a 6250 bpi hasta finales de la década de 1980. Más allá de este período, esta tecnología dejó de evolucionar y prácticamente desapareció a finales de la década de 1990. Las capacidades de grabación eran bajas en comparación con nuestros medios actuales.

En 1990, todos los datos de las misiones científicas del CNES se almacenaron en decenas de miles de cintas almacenadas en las instalaciones del centro informático del CNES en Toulouse. Casi todas estas cintas y los archivos de datos que contenían tenían la estructura patentada de las poderosas (¡por el momento!) computadoras del centro de datos de

control. Cada colección de cintas magnéticas fue gestionada, controlada y mantenida por los equipos del proyecto que eran propietarios de los datos contenidos en estas colecciones.

Fue en este momento que la situación y las perspectivas a mediano plazo cambiaron significativamente debido a tres factores: el primero y más restrictivo fue la desaparición anunciada de las tecnologías de almacenamiento de cinta magnética de 6250 bpi. En segundo lugar, la evolución general de la informática en un entorno científico, el CNES ha previsto el cierre de las máquinas de Control de Datos del centro informático, basadas en el sistema operativo NOS / VE y su sustitución por máquinas basadas en el sistema UNIX. Por último, existía el deseo de hacer que los datos científicos fueran accesibles y utilizables por la comunidad en general.

3.1.3. Implementación de STAF y migración de datos

Ante esta situación, el CNES tomó la decisión, en 1992, de poner en marcha un servicio de almacenamiento central que ofreciera una garantía real de conservación a largo plazo de los bits, sea cual sea la tecnología de almacenamiento utilizada. Este servicio especializado encargado de perpetuar los archivos es el STAF (File Transfer and Archiving Service). Opera desde 1994 y se presenta como una entidad independiente de proyectos o servicios de archivo. Estos últimos son clientes de STAF y se dirigen a él mediante un conjunto de comandos básicos que permiten, en particular, solicitar el almacenamiento o restitución de un archivo o un conjunto de archivos. Estas comunicaciones pasan por la red interna del CNES. Por lo tanto, STAF tiene una misión muy simple:

- recibir archivos sin tener que conocer su formato o su contenido de información,
- garantizar la conservación a largo plazo de estos archivos,
- garantizar su integridad,
- garantizar su confidencialidad,
- devolver los archivos a solicitud.

Por lo tanto, se decidió y se emprendió la migración inmediata y esencial de datos en cintas al STAF en 1994. Duró cinco años, se centró en más de 60.000 cintas magnéticas, aproximadamente 500.000 archivos, y supuso una movilización momentánea de una serie de actores que realmente tomaron conciencia de la gran vulnerabilidad de los datos y de las múltiples causas que pueden conducir a su pérdida. El número de archivos perdidos en esta ocasión por motivos de almacenamiento fue muy bajo pero no nulo debido a la existencia de algunas cintas con una densidad obsoleta (800 bpi) para las que el CNES ya no contaba con ningún equipo de lectura, y cintas que no se pueden leer debido a la degradación física.

La migración reveló de inmediato que la mayoría de los datos presentaban estructuras lógicas y codificaciones exclusivas de los sistemas operativos que se habían utilizado para crear esos datos, los sistemas operativos en proceso de salida. Como resultado, los archivos no eran portátiles y, no se podían leer en otro sistema. Por tanto, una primera tarea consistió en eliminar de los archivos toda la información adicional específica del sistema operativo. Esto se puede hacer utilizando un software de utilidad disponible en el propio sistema operativo.

Hubo que emprender una segunda operación mucho más delicada. En la mayoría de los casos, los archivos contenían resultados de procesamiento científico como secuencias de números enteros y reales codificados en binario. La representación binaria de muy alta precisión de números reales, de 128 bits de tamaño, era una representación patentada. CNES eligió utilizar una representación estándar de números (en particular IEEE para números reales), lo

que requería una transformación de los datos. Se trata, pues, aquí de transformaciones o migraciones de formato. Estas transformaciones presentaron varias dificultades:

- no se podían ejecutar automáticamente y suponían un desarrollo de software específico para cada colección de archivos,
- no fueron reversibles debido a los inevitables errores de redondeo relacionados con las últimas cifras significativas.

Por lo tanto, requirieron un esfuerzo de validación considerable. Para mayor seguridad, los archivos originales se conservaron durante algunos años y luego se destruyeron.

Además, los elementos descriptivos de estos datos, que permitían conocer su significado, eran en ocasiones inexactos o incompletos, o incluso no siempre estaban disponibles. Por lo tanto, durante la misma operación calificada como “rehabilitación de datos”, fue necesario reformatear los datos dotarles de estructuras independientes de los sistemas operativos, reconstituir los metadatos en la medida de lo posible y migrar estos datos al STAF. Varios ingenieros dedicaron varios años a esta operación, tuvieron que recurrir a expertos científicos aún disponibles para reconstituir los metadatos, confiaron en empresas de servicios para el desarrollo de software de transformación de formatos, consumieron recursos, máquinas considerables. La mayor parte de los datos se guardaron, pero esta operación permitió medir el costo de los arreglos tardíos. Dado el ritmo actual de desarrollo de las tecnologías digitales, tal rescate probablemente ya no sea posible en la actualidad.

Otros signos de aceleración de la obsolescencia tecnológica también marcaron este período. El ejemplo de los documentos textuales del campo de la ofimática es elocuente. El CNES utilizó un primer sistema ofimático disponible en el mercado en la segunda parte de la década de 1980. Era un sistema propietario que constituía el avatar electrónico de la máquina de escribir tradicional y permitía la entrada, maquetación e impresión de documentos de texto aprovechando las posibilidades de la Informática. Se incautaron grandes cantidades de documentos utilizando este sistema. A principios de la década de 1990, con el desarrollo de la microcomputación y los inicios del monopolio de Microsoft sobre la automatización de oficinas, la mayoría de los otros sistemas propietarios existentes desaparecieron del mercado, pero los documentos permanecieron. Sin la más mínima posibilidad técnica de migrar documentos al paquete de software Word para DOS, que fue la primera versión de Word utilizada por CNES, los documentos fueron incautados nuevamente. Seis años más tarde, CNES descubrió que los documentos guardados en Word para DOS solo eran parcialmente compatibles con Word 97 para Windows. Para todos los documentos que debían conservarse, se pudo recuperar el texto, pero se asumió por completo el diseño de miles de tablas complejas.

El STAF ha realizado las inevitables migraciones de soporte y no son visibles para los clientes del servicio, ha demostrado el interés y la eficacia de los principios sobre los que se construyó, a saber, la total independencia de la función de almacenamiento, a partir de las demás entidades funcionales de un servicio de archivo a largo plazo, pone en práctica lo que ahora se denomina virtualización del almacenamiento.

3.1.4. Sistemas de archivo genéricos para reducir costos

Es posible, hasta cierto punto, definir formatos de datos y metadatos que son completamente independientes de los sistemas operativos y tecnologías. Esta elección limita la vulnerabilidad de estos datos y metadatos a los cambios en estas tecnologías. Por otro lado, no es posible construir un sistema de archivo digital, compuesto por hardware y software que sean independientes de la tecnología. Además, sabiendo que el sistema de archivo digital es el medio por

el cual podremos recibir los datos a archivar, almacenar, gestionar, hacer accesibles, este sistema debe ser sostenible. Sujeto a todos los caprichos de la obsolescencia tecnológica, será necesario asegurar el mantenimiento de este sistema para que permanezca en funcionamiento permanente.

Periódicamente, debido a la desaparición de una determinada tecnología utilizada, no se realizan más trabajos de mantenimiento, por lo que se requerirá la reconstrucción parcial del sistema y por tanto financiar.

La cuestión de limitar y, si es posible, reducir los costes de mantenimiento del sistema de archivo en funcionamiento permanente es un punto crítico al que el CNES ha tratado de responder con dos opciones.

- Una elección de arquitectura consistente en estructurar los sistemas en bloques funcionales independientes entre sí para que cualquier cambio tecnológico importante que provoque modificaciones profundas en un bloque no repercuta en los demás. En el pasado, el CNES se había encontrado con el caso de sistemas monolíticos en los que un cambio limitado en un dominio inducía una propagación en cadena de modificaciones en todo el sistema, con consecuencias significativas en términos de costos de modificación y validación.
- Una opción genérica dirigida a construir sistemas que puedan ser reutilizados por varias aplicaciones dentro del establecimiento y por varias organizaciones que llevan a cabo actividades de archivo de datos científicos. Esta elección tiene como objetivo compartir los costos de desarrollo, luego el mantenimiento y la actualización por parte de los distintos sitios de usuarios del sistema. Hasta 1995, cada misión espacial científica condujo al desarrollo de un sistema dedicado que permitía la recepción, procesamiento, difusión y archivo de datos de esta misión. CNES se dio cuenta rápidamente de que sería imposible en el futuro mantener tantos sistemas en funcionamiento como misiones espaciales pasadas. Por tanto, era imperativo construir un sistema de acceso y gestión de datos capaz de ofrecer funciones de acceso a los datos para todas las misiones de una misma disciplina científica. o incluso varias disciplinas científicas distintas.

Se desarrolló un primer sistema genérico de este tipo, el SIPAD (Information, Preservation and Access to Data System), para asegurar la gestión y provisión de datos del Plasma Physics Data Center (<http://cdpp.cesr.fr>). Se puso en servicio por primera vez en 1999. Este sistema se basó principalmente en un producto comercial de gestión de datos técnicos llamado Metaphase.

Después de solo unos años de operación, parecía necesario resolver un conjunto de problemas técnicos relacionados con SIPAD: controlar el rendimiento del acceso a la base de datos, rendimiento que se deteriora al mismo tiempo que aumenta el número de conjuntos de datos, eliminar la dependencia de SIPAD sobre el producto Metaphase cuya sostenibilidad no estaba garantizada, introducir nuevas funciones con miras a la interrogación automatizada, tener posibilidades de especialización avanzada de la interfaz hombre-máquina. La escala de las necesidades de cambios ha llevado al desarrollo de un sistema completamente nuevo, el SIPAD-NG. Por lo tanto, la puesta en servicio de este nuevo sistema en 2006 requirió una revisión previa del diseño de la base de datos. Los metadatos se extrajeron globalmente del SIPAD, luego se transformaron y enriquecieron de acuerdo con las nuevas especificaciones de metadatos, luego se validaron e ingresaron en el SIPAD-NG. El sistema SIPAD-NG ahora tiene una base sólida: es utilizado dentro del CNES por varias entidades separadas a cargo del archivo de datos.

3.1.5. Una metodología que se está consolidando

El trabajo de rescate y “rehabilitación de datos” estuvo acompañado de un análisis metodológico de qué hacer y qué no hacer en el futuro. En junio de 1993 se redactó y distribuyó un primer documento de especificaciones para el archivo a largo plazo de datos espaciales. Subraya la importancia del patrimonio de datos científicos y tecnológicos preservado y mantenido desde principios de la década de 1970 y observa tres desarrollos esenciales:

- el aumento constante de los volúmenes de datos producidos,
- el aumento significativo de los períodos mínimos de retención requeridos (varias décadas),
- la necesidad de una accesibilidad cada vez más amplia a estos datos por parte de la comunidad científica.

Esta primera especificación ya contiene todos los requisitos esenciales que se aplican a los datos científicos:

- identificación de toda la información que debe asociarse con los datos de destino para su archivo: descripción sintáctica y semántica de archivos, parámetros de calibración, (Información de representación), metadatos en DIF (Formato de intercambio de directorios) (Información de descripción), base documental que describe la misión, experiencia, instrumento (Información de procedencia y contexto),
- el requisito de mantener la integridad física de toda la información digital,
- el requisito de métodos de codificación estandarizados y estructuras de archivos independientes de los sistemas operativos,
- accesibilidad de los datos a los usuarios autorizados.

Es a partir de sus experiencias prácticas y sus primeras reflexiones metodológicas que el CNES pudo participar fructíferamente en la redacción del modelo OAIS y que luego asumió la responsabilidad directa de redactar el estándar PAIMAS (Norma ISO 20652:2006 Sistemas de transferencia de información y datos espaciales- Interfaz productor-archivo- Estándar abstracto de metodología) y el estándar PAIS (Norma ISO 20104:2015 Sistemas de transferencia de información y datos espaciales- Especificación de interfaz productor-archivo). El modelo OAIS permite analizar los sistemas desarrollados en el CNES con un nuevo punto de vista y un nuevo vocabulario. Son estas interacciones entre el enfoque pragmático nacido de la experiencia de campo y la reflexión metodológica las que gradualmente dan la solidez y confiabilidad necesarias a los sistemas de archivo digital.

Es también sobre esta base que se abrió una nueva rama denominada “Ingeniería de datos” dentro del Repositorio Normativo del CNES, que define el organigrama de las normas aplicables a sus proyectos y estructuras. Además, de un resumen de las necesidades en materia de sostenibilidad y acceso a los datos, esta rama del Repositorio incluye una serie de normas y recomendaciones aplicables a los proyectos de producción y datos, así como a los servicios encargados de archivar estos datos.

3.1.6. Posible distorsión entre necesidades y decisiones

En 2000 se elaboró un plan estratégico del CNES para el período 2001-2005. Este plan tuvo en cuenta, en términos generales, el problema de la gestión, difusión y explotación de los datos de los experimentos espaciales. Este plan dio lugar a trabajos para traducir sus principales orientaciones en acciones concretas.

La primera acción propuesta en este contexto fue más que simbólica: “Preparar una decisión para la firma del Presidente del CNES para hacer valer las responsabilidades y objetivos del CNES para la valoración, archivo y provisión

de productos (de datos)". Esta situación ilustra, si es necesario, la importancia de convencer a los líderes y tomadores de decisiones de la urgencia del problema.

3.1.7. Conclusiones sobre la experiencia del CNES

En cuanto a la implementación del archivo digital de datos científicos espaciales, podemos ofrecer las siguientes conclusiones provisionales:

- en términos de almacenamiento, consideramos que hasta el nivel de petabytes, las necesidades de preservación física de archivos se resuelven con un nivel satisfactorio de confiabilidad,
- la descripción semántica de datos y el desarrollo de metadatos descriptivos son los elementos clave de la reutilización de datos en el futuro. Estos metadatos aún dependen de ontologías y terminologías en rápida evolución. Esta es una dificultad real que no debe ignorarse,
- la sostenibilidad de los sistemas informáticos desarrollados para la transferencia de datos, su gestión y su difusión plantea problemas de distinta índole. El desafío aquí es controlar y minimizar los costos de desarrollo, mantenimiento y actualización de estos sistemas. Vimos cómo se podía imaginar esto, con respecto a los archivos de datos,
- se puede observar que en disciplinas científicas han surgido archivos de datos estándar, se han desarrollado rápidamente herramientas gratuitas para el procesamiento, análisis, visualización y servicios de valor agregado. Por el contrario, las disciplinas para las que realmente no ha surgido ningún formato están muy penalizadas. De ahí la necesidad de que estas disciplinas emprendan o aceleren el trabajo en esta dirección.

A nivel político y más específicamente en las decisiones de implementación, la situación sigue siendo incierta y el archivo de datos solo se cubre parcialmente. Las directrices de la CNES para el archivo a largo plazo de todos los datos espaciales para los que existe esta necesidad, quedan por aclarar y formalizar.

El CNES no tiene un repositorio central para todos sus datos y, por lo tanto, no sabe cómo dar una cuenta completa de sus activos. Además, no hay nada que sugiera que todos los datos de los proyectos del CNES estén realmente archivados. Ciertos datos no están bajo la responsabilidad de ningún centro de archivo identificado, lo que obviamente no augura ninguna garantía de sostenibilidad. Otros están definitivamente perdidos, como fue el caso de algunas misiones en el pasado. Incluso si comienza a surgir un proyecto para constituir un repositorio global, serán necesarios varios años antes de llegar a una situación gestionada y controlada.

En la práctica, muchas misiones espaciales se organizan en el marco de la cooperación internacional. Por tanto, una agencia espacial no puede imponer sus normas a todas las demás. Esto refuerza, si es necesario, la necesidad de una cooperación lo más estrecha posible entre las agencias sobre el archivo de datos, especialmente porque la comunidad de usuarios también es internacional.

3.2. La experiencia de los servicios públicos de archivos franceses

Los Archivos Nacionales, ubicados en Fontainebleau, han estado recibiendo, controlando, conservando y facilitando el acceso a archivos digitales desde principios de la década de 1970 como parte del programa de Constanza. Han seguido una metodología probada, especialmente apta para la recepción de archivos estadísticos estructurados. Con el desarrollo de la administración electrónica y la producción de documentos digitales con firmas electrónicas, se ha

vuelto crucial desarrollar nuevas herramientas que integren sistemas más automáticos. El uso de lenguajes XML se ha vuelto común para garantizar la integridad y durabilidad de los archivos, incluida la identificación, control y conversión de formatos si es necesario, así como la duplicación de archivos en sitios remotos para garantizar la trazabilidad.

3.2.1. El contexto en el que se llevaron a cabo los experimentos

El marco legal abarca tanto los archivos en papel como los digitales desde la promulgación de la Ley de Archivos del 9 de enero de 1979. Durante mucho tiempo, los archivistas apenas prestaron atención a la tecnología digital, salvo en los Archivos Nacionales, que archiva datos científicos, especialmente bases de datos estadísticas de las principales organizaciones científicas, desde hace unos treinta años.

En efecto, la producción masiva de documentos públicos permaneció en papel debido a varios motivos, principalmente las exigencias legales que requerían que los actos tuvieran validez probatoria mediante un documento escrito con una marca de validación como sello o firma. La falta de interoperabilidad entre sistemas de información también contribuyó a esta situación, obligando a la materialización de la información al abandonar su entorno de producción.

Finalmente, una gran cantidad de información del exterior (cartas de otras administraciones, ciudadanos) que llegaba a los servicios en formato papel, seguía siendo el “formato” común más adecuado y el más fácil de generalizar.

A pesar de ello, se desarrollaron bases de datos en las administraciones, reemplazando gradualmente los registros en papel por datos más complejos y ricos que permitían un seguimiento preciso de casos y situaciones, así como la edición de cartas y formularios.

Sin embargo, la conciencia sobre la necesidad de archivar los datos de estas aplicaciones surgió tardíamente entre los archivistas, ya que las operaciones de transferencia se centraban en documentos antiguos producidos exclusivamente en papel. Además, la gestión de estas aplicaciones empresariales está en manos de los departamentos de informática, que no eran los interlocutores de los archivistas que hasta entonces habían trabajado exclusivamente con los productores de los documentos.

Apareció una nueva conciencia con el desarrollo de la administración electrónica, que permitía la producción de originales digitales siempre que se desmaterializara el proceso de negocio. Al mismo tiempo, gracias a la actuación de la Dirección General de Modernización del Estado, avanzaba la interoperabilidad entre sistemas de información, que por sí solos podían posibilitar esta virtualización.

Al mismo tiempo, La digitalización masiva de archivos en papel planteó desafíos en cuanto a la conservación de archivos en formatos de preservación, sin comprimir. Inicialmente, se realizaron copias en CD-R duplicados que se almacenaron en almacenes tradicionales, utilizados para los archivos en papel. Sin embargo, se reconoció tarde la necesidad de monitorear y probar estos medios, lo que llevó a varios departamentos a optar por mantener los archivos en servidores de almacenamiento, como discos o cintas LTO3 o 4, proporcionados por los departamentos de TI. Esta decisión se tomó especialmente cuando los volúmenes de archivos aumentaron considerablemente.

3.2.2. Proyecto PILAE de la Dirección de Archivos de Francia (DAF)

El Proyecto PIL@E, llevado a cabo por la Dirección de Archivos de Francia (DAF), fue lanzado a finales de 2006 bajo la supervisión del Departamento de Innovación Tecnológica y Estandarización de la DAF, y recibió una valiosa asistencia en la gestión de proyectos de la Dirección General de Modernización del Estado (DGME).

3.2.2.1. Contexto

Este proyecto, respaldado por el director de los Archivos de Francia y el director de los Archivos Nacionales, se consideró estratégico debido a las lecciones esenciales que proporcionaría a la comunidad archivística en su conjunto. La transición hacia la nueva producción digital nativa requería una adaptación en los métodos de trabajo, habilidades y formación, lo que suponía un desafío innovador para el cliente, el director del proyecto y los usuarios finales de la herramienta. La gestión del cambio se presentaba como un aspecto especialmente complicado.

Una característica destacada fue la falta de mediciones y conocimiento preciso sobre los volúmenes futuros, ya que la digitalización completa de los procesos de negocio aún estaba en curso en la mayoría de los casos.

Se decidió implementar el plan piloto en el servicio de archivos electrónicos del Archivo Nacional en Fontainebleau, con el objetivo de recibir, preservar y comunicar archivos nativamente digitales producidos por los servicios centrales del Estado durante un período transicional de 2009 a 2013. Los Archivos Nacionales, a través del proyecto de un nuevo centro de archivos en Pierrefitte-sur-Seine, emprendieron un esfuerzo significativo para renovar y revisar su sistema de información, abarcando la gestión, comunicación, descripción y distribución de archivos en papel de los servicios del Gobierno Central.

Por lo tanto, PIL@E debía ser capaz de recibir, almacenar, y permitir la búsqueda y consulta de una variedad de objetos nativamente digitales, incluidos datos de bases de datos comerciales, documentos electrónicos, intranets colaborativas, mensajes electrónicos y documentos de una cadena de virtualización completa. Además, PIL@E se basaría en el modelo OAIS y en las recomendaciones del repositorio de interoperabilidad general, centrándose en formatos de documentos compatibles, políticas de archivo y formatos de metadatos estándar.

3.2.2.2. Características de la caja fuerte electrónica

El director del proyecto propuso utilizar el paquete de software que comercializaba, conocido como la caja fuerte electrónica comunicante (CFEC), que ofrecía una serie de funciones necesarias para el almacenamiento seguro y la trazabilidad:

- funciones de verificación de huellas dactilares, firmas de archivos, marcas de tiempo basadas en una fuente de tiempo externo “seguro” y funciones que permiten verificaciones periódicas de estos parámetros a lo largo del tiempo,
- funciones que aseguran la trazabilidad del sistema: edición de logs de eventos que permiten seguir las distintas operaciones de conexión a la herramienta, transferencias de objetos a archivar, consulta, borrados; los registros se sellan regularmente para que la información de estos registros no se pueda cambiar,
- funciones que permiten controlar las operaciones de replicación de datos síncronos, cada vez que una transferencia es aceptada y validada por el archivista: escritura en dos sitios remotos (replicación entre sitios) junto con replicación dentro del sitio; en total, una replicación en cuatro puntos de almacenamiento en cuatro servidores distribuidos en dos sitios remotos.

Además, la seguridad de los datos se garantiza mediante sistemas de redundancia tanto de los datos como de las bases de datos, accesos e instalaciones eléctricas, junto con un sistema de respaldo tradicional integrado en la política general de garantía del departamento de sistemas de información del ministerio.

La caja fuerte se basa en ladrillos de código abierto y garantiza la reversibilidad de todos los archivos, diarios y otras evidencias generadas, lo que permite su restitución y recuperación en otro sistema si es necesario. Por ejemplo, PIL@E podría funcionar eventualmente con otra caja fuerte que no sea la de Cecurity.com.

3.2.2.3. Funciones desarrolladas más allá de la caja fuerte electrónica

Además de estas funciones básicas, era necesario desarrollar otras funciones específicas para el archivo seguro de datos y documentos de los servicios de productores (administraciones de servicios del Gobierno Central).

Se decidió excluir el acceso al público en general que deseara consultar archivos digitales comunicables por dos razones principales: la falta de una reflexión exhaustiva sobre el nivel de seguridad necesario para abrir una aplicación en Internet que gestionara y permitiera el acceso a datos y documentos confidenciales, y la relevancia cuestionable de esta apertura debido a una cobertura inicialmente desigual entre las diversas áreas administrativas susceptibles de ser cubiertas, agravada por la naturaleza no transferible de algunos de estos documentos y datos.

El núcleo de los desarrollos se centra en la implementación del estándar para el intercambio de datos para archivo (SEDA), enfocado en los procesos de “transferencia” e “investigación y comunicación”.

Se han creado interfaces y mensajes específicos para los diversos procesos entre los ministerios y los Archivos Nacionales: el envío por parte de los archivistas depositantes, la recepción por parte de los archivistas (proceso de control/validación/rechazo) y el acuse de recibo. Se ha simplificado el proceso de búsqueda y consulta, eliminando la necesidad de intervención de los archivistas del Archivo Nacional (búsqueda basada en derechos de acceso a los archivos a partir de metadatos y descarga).

Respecto al formato de metadatos definido por el estándar de intercambio de datos para archivo, es crucial para automatizar la descripción de las principales categorías de archivos enviados periódicamente. Se ha propuesto un método principal que implica la transferencia “manual” de pagos ya formateados en el formato estándar de intercambio. Luego, el sistema verifica la transferencia recibida (cumplimiento del esquema estándar de intercambio, verificación de huellas dactilares, identificación de formatos) y envía un mensaje de aceptación o error en respuesta, según los problemas encontrados. Una vez aceptada, la transferencia es verificada por un archivista de los Archivos Nacionales, quien puede modificar o enriquecer los metadatos y decidir si valida o no el pago.

Este método de funcionamiento (importación de archivos en formato estándar de intercambio) es el más relevante para evitar interrupciones de carga y el reingreso de metadatos, especialmente con grandes volúmenes de archivos. Es especialmente útil para pagos periódicos de la misma categoría de archivos, donde el trabajo de especificación ya se ha realizado previamente para formatear el estándar de intercambio.

Otro método consiste en el ingreso manual de metadatos a través de pantallas de entrada que muestran todos los campos del comprobante de pago definidos por el estándar de intercambio. El archivista debe determinar la clasificación de los archivos en papel y luego ingresar los metadatos asociados a cada nivel.

Es evidente que este trabajo es tedioso y deberá repetirse con cada nuevo pago. La descripción se relacionará con todo este conjunto y será un resumen. Una vez completada la entrada y adjuntados los archivos de datos correspondientes, el sistema formateará automáticamente el estándar de intercambio. El proceso es entonces idéntico.

3.2.2.4. Cuestión de conversiones de formato

Es sobre la base de un acuerdo establecido para la gestión de una categoría de archivos, los archivistas depositantes llevarán a cabo su transferencia. Este acuerdo especifica si el departamento de producción autoriza o no las conversiones de formatos al ingresar a PIL@E, en caso de que los formatos iniciales no cumplan con el sistema general de referencia de interoperabilidad para su conservación a largo plazo. PIL@E cuenta con herramientas detalladas de identificación de formatos y, de acuerdo con las reglas establecidas, puede realizar conversiones de formatos (por ejemplo, de archivos de la suite de Microsoft Office a formato PDF/A). Estos procesamientos se llevan a cabo de manera interactiva o en lotes, especialmente cuando se trata de grandes volúmenes.

Los archivistas encargados de la verificación serán notificados si se detectan anomalías en las conversiones, pero incluso en caso de fallo (es decir, si no se puede realizar ninguna conversión), aún pueden aceptar la transferencia. Los formatos originales se conservan en el Paquete de Información de Archivo (AIP), y los usuarios que deseen solicitarlos posteriormente pueden optar por el formato original o el formato posterior a la conversión.

No obstante, se requieren pruebas exhaustivas fuera de PIL@E para evaluar con mayor precisión los “riesgos” asociados con determinadas conversiones y definir estrategias de acuerdo con estos riesgos, especialmente según los tipos de formatos. Se han observado obstáculos en documentos gráficos, por ejemplo, durante el proyecto, donde los formatos de destino recomendados generalmente no permitían conservar algunas funcionalidades (especialmente enlaces) de los formatos originales, que lamentablemente eran propietarios.

3.2.2.5. Conclusiones sobre PILAE

La implementación completa de PIL@E y su puesta en producción están programadas para principios de 2010. Los años de transición que seguirán permitirán que el producto sea probado y preparado adecuadamente para la transición a un sistema a mayor escala.

Para anticipar estas evoluciones, se deben estudiar con mayor precisión varias cuestiones. Es crucial concretar los aspectos relacionados con los formatos de codificación de los documentos y asegurar el uso de identificadores únicos y permanentes para SIP/AIP/DIP. Además, surgirán problemas relacionados con la seguridad de los datos, como la necesidad de una caja fuerte para todos los datos y documentos, el cifrado de datos sensibles y la arquitectura necesaria para abrir el sistema en Internet para el público en general.

La infraestructura de almacenamiento también deberá ser refinada y ampliada para incluir otras funcionalidades a medida que aumenten los volúmenes, como la monitorización automática de los medios y la reflexión sobre los tipos de medios utilizados.

Probablemente sea apropiado considerar el uso de un formato de empaquetado estandarizado como METS para los respectivos formatos de SIP, AIP y DIP.

Es importante tener en cuenta que una vez que se han creado los AIP, no es posible modificarlos, a menos que se elijan soluciones alternativas, como la adición de metadatos o archivos adicionales.

Es normal que todas estas preguntas surjan durante la implementación de un plan piloto, especialmente porque el campo es innovador y los diferentes actores del mercado involucrados aún operan en entornos compartimentados. Esto incluye a especialistas en seguridad y firma electrónica, proveedores de soluciones de almacenamiento, desarrolladores de software de gestión de documentos electrónicos y sistemas de gestión de registros, así como a los editores de software de archivos, presentes en el ámbito patrimonial.

Finalmente, una gran incógnita sigue siendo cómo los archivistas de los ministerios y los archivos nacionales adoptarán PIL@E. Este es un nuevo dominio para ellos, ya que aunque los principios de archivo siguen siendo los mismos, la administración electrónica representa un cambio significativo en los métodos de trabajo y la asistencia en la gestión de proyectos por parte de productores y especialistas en TI, así como la adopción del conocimiento en términos de formatos XML y lenguajes.

3.2.3. La plataforma de archivo electrónico del Consejo General de Yvelines

El Consejo General de Yvelines es la primera autoridad local en Francia que ha desarrollado una plataforma piloto. El proyecto se inició en 2005 como una implementación experimental en este departamento, centrándose en la transferencia de documentos sujetos a control de legalidad. La plataforma de teletransmisión asume esta función de transmisión y, una vez finalizada la transacción y superados los recursos legales, se encarga de transferir los documentos para su archivo.

Este proyecto fue llevado a cabo internamente en el Consejo General, con una estrecha colaboración entre los servicios de TI y los Archivos Departamentales. Las funcionalidades principales de la plataforma están muy cercanas a las del proyecto PILAE. El proyecto Yvelines se inspira en el modelo OAIS, así como en el estudio de costos de plataformas realizado por el departamento de los Archivos de Francia y, finalmente, en las especificaciones funcionales de PILAE. Sin embargo, las funcionalidades se centran en la recepción, control y validación de archivos que ya están en el formato estándar de intercambio de datos para el archivo, sin abordar las funciones de identificación, control y posible conversión de formatos de documentos.

Este proyecto es interesante por varias razones, especialmente porque ha tenido el efecto de reubicar los Archivos Departamentales mucho antes en el ciclo de vida del documento. En este caso, los documentos se reciben rápidamente después de su producción, tan pronto como se valida la transacción, sin necesidad de esperar el vencimiento del período de vigencia administrativa. Además, se espera que se proporcione una forma de colaboración para los municipios que, al no tener plataformas de archivo electrónico, deseen depositar las actas que producen.

3.2.4. La plataforma de archivo electrónico del Consejo General de Aube

La plataforma de archivo electrónico del Consejo General de Aube también presenta un enfoque interesante, ya que adoptó una estrategia de colaboración al recuperar la aplicación desarrollada por el Consejo General de Yvelines, realizando algunas mejoras y actualizaciones funcionales.

3.2.4.1. Contexto y desafíos

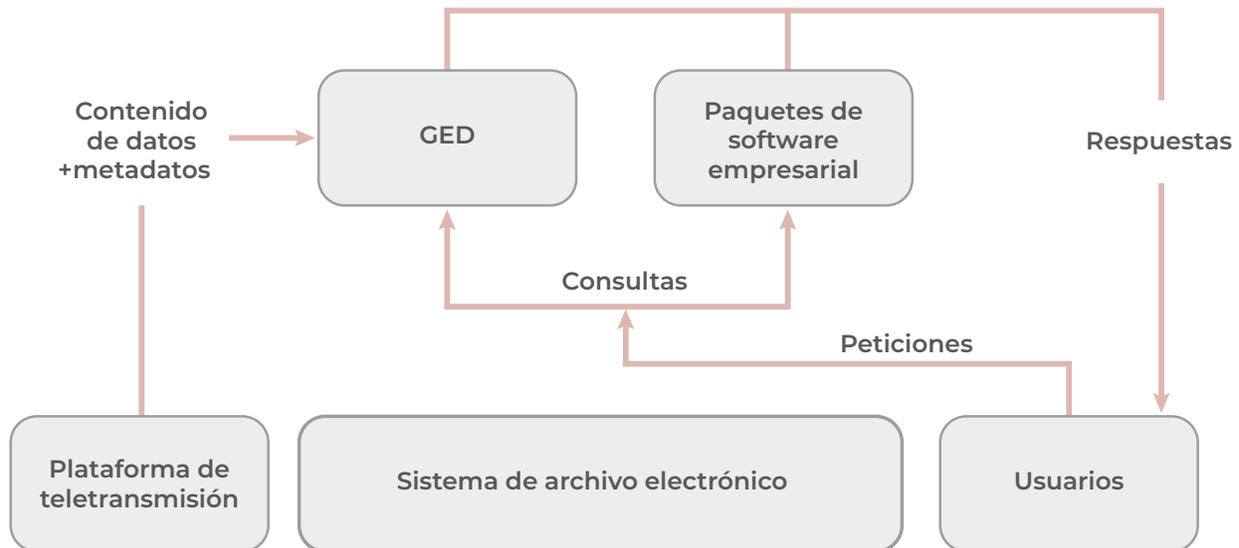
El contexto y los desafíos son similares a los de Yvelines, con la implementación de proyectos de administración electrónica y la adopción de herramientas de firma electrónica por parte del Consejo General, lo que generó documentos digitales que necesitaban ser conservados. Esto impulsó la necesidad de desarrollar herramientas de archivo electrónico.

La estrecha colaboración entre los servicios informáticos y los archivos departamentales, a través de un comité de archivo electrónico, ante el cual todos los proyectos de transformación, incluido el archivo, deben ser tenidos en cuenta desde el inicio del proyecto.

Se realizó una profunda reflexión organizativa, jurídica y técnica.

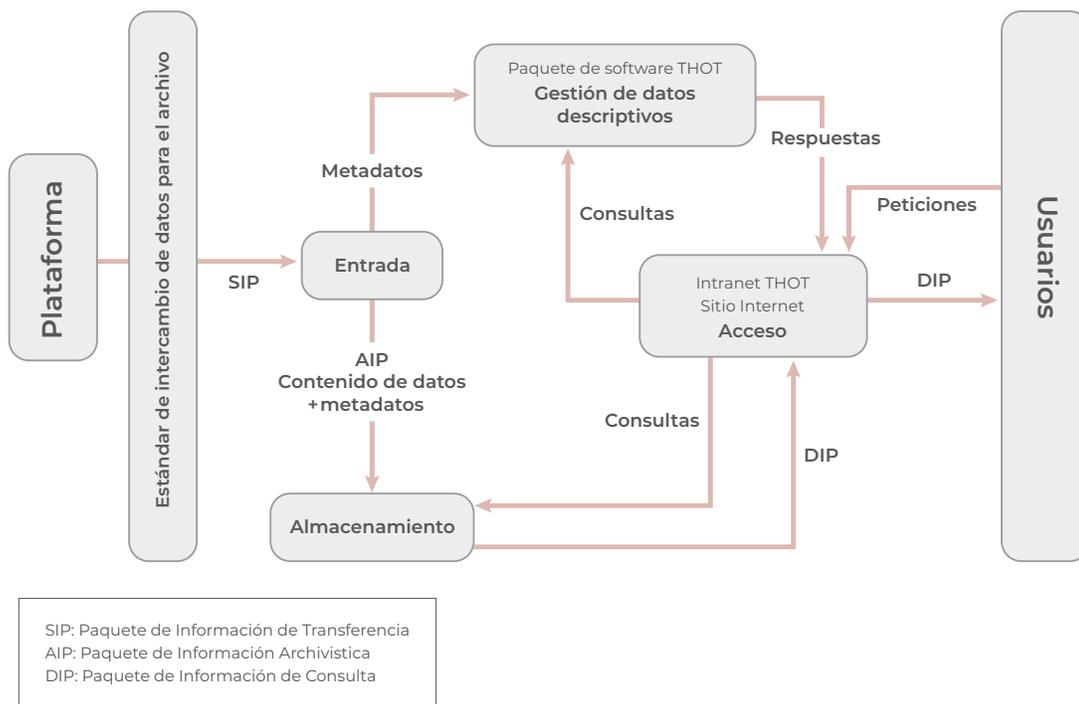
- Decisión para cualquier aplicación comercial, de depositar los datos / documentos / flujos en la plataforma de archivo, tan pronto como se completen los trámites, con el fin de asegurar la información, manteniéndose también los datos en las aplicaciones comerciales durante su vigencia. Se trata de un cambio importante en la organización hasta entonces vigente, ya que solo los archivos denominados definitivos fueron sustentados por los archivos departamentales, una vez vencido el período de vigencia administrativa. Para Archivos, se trata de intervenir muy temprano en el proceso del ciclo de vida de la información y así tener una mayor legibilidad y responsabilidad.
- Establecimiento de una contractualización sistemática mediante la firma para cada nueva categoría documental soportada, por un lado un contrato de servicios entre el servicio productor, el servicio de archivos y el departamento de informática y, por otro lado, un protocolo de pago esta vez con el pago de servicio.
- Acceso a archivos depositados (nota: el mismo acceso está configurado para archivos independientemente de su apoyo):
 - para los departamentos de producción, a través de sus aplicaciones comerciales o de una gestión documental electrónica sencilla pero adaptada a sus necesidades
 - y para otros usuarios (lectores de archivos), a través del software de gestión de archivos utilizado anteriormente para la investigación y el acceso a archivos en papel (software Thot).

GESTIÓN DE SOLICITUDES DE CONSULTA DE UNIDADES PRODUCTORAS



Gestión de solicitudes de consulta para departamentos productores (a través de su EDM / Gestión electrónica de documentos comerciales) y para investigadores y ciudadanos, a través del software de gestión documental utilizado por los archivos departamentales de Aube para archivos tanto en papel como digitales

GESTIÓN DE SOLICITUDES DE CONSULTA DE LA COMUNIDAD DE INVESTIGADORES Y CIUDADANOS



Gestión de solicitudes de consulta para departamentos productores (a través de su EDM / Gestión electrónica de documentos comerciales) y para investigadores y ciudadanos, a través del software de gestión documental utilizado por los archivos departamentales de Aube para archivos tanto en papel como digitales

3.2.4.2. La plataforma de confianza

Otra innovación: la implementación de la plataforma de archivo confiable dirigida a:

- verificar la firma de los documentos transferidos que están firmados,
- para dar fe de esta verificación,
- editar un informe de verificación que a su vez esté firmado,
- firmar el comprobante de envío, una vez emitido el informe.

La plataforma también implementará la identificación y conversión, si es necesario, de los formatos de los documentos transferidos.

3.2.5. Comentarios de una importante institución patrimonial: la Biblioteca Nacional de Francia (BnF)

La preocupación por la sostenibilidad de la información digital en los últimos años se ha extendido a otros actores importantes del patrimonio, como el Instituto Nacional del Audiovisual (INA) y la Biblioteca Nacional de Francia (BnF). Un ejemplo destacado de esta preocupación se presenta a continuación.

3.2.5.1. Contexto digital en la BnF

La distribución de materiales digitales ha sido una aventura de larga data para la BnF, con más de 15 años de programas de digitalización y una actividad que está experimentando un fuerte crecimiento con la digitalización masiva, la consulta de documentos audiovisuales en formato digital y el acceso a archivos web de estrenos, entre otros.

Estos proyectos han dado lugar a la provisión de un número creciente de servicios, como estaciones de consulta interna, la biblioteca digital Gallica desde 1997, la venta de reproducciones digitales bajo demanda y la posibilidad de recolectar metadatos a través del protocolo OAI (Iniciativa de Archivo Abierto), que permite la construcción de herramientas de investigación eficientes y la identificación duradera de recursos digitales. Estos esfuerzos muestran el compromiso de la BnF en ser un actor relevante en la sociedad de la información, ofreciendo herramientas y servicios para que sus usuarios puedan acceder, comprender, manipular y colaborar con los recursos digitales.

Los primeros experimentos de digitalización de imágenes fijas e impresos surgieron en los años 1990-1991. El medio de grabación que se conservó fue la cinta magnética DAT suministrada por duplicado por los proveedores de servicios de digitalización, pero tenían una vida útil de 2 a 3 años, lo que planteaba desafíos en términos de migración de datos.

En aquel entonces, la BnF no estaba plenamente consciente de la magnitud del problema y veía el archivo solo como un centro de costos en lugar de un seguro para los datos. El equipo buscaba una solución que requiriera la menor cantidad de habilidades y la menor intervención posible, pero que garantizara la facilidad de lectura a largo plazo. Se optó por soluciones ópticas y magnetoópticas, considerando que ofrecían una oportunidad satisfactoria en comparación con los discos magnéticos y las soluciones de banda, que eran menos confiables o requerían ciclos frecuentes de migración y actualización. La desventaja de los soportes CD-R (disco compacto grabable) es su inestabilidad en el tiempo. Sin embargo, una solución basada en tecnología francesa permitió resolver este problema.

3.2.5.2. Proyecto Century Disk y primeras migraciones

El Century Disk, desarrollado por la empresa francesa Digipress en Caen, ofrecía una técnica innovadora. Consistía en quemar los masters, que luego servían de matriz para prensar los CD-ROM de policarbonato. Este método tenía dos ventajas clave: en primer lugar, los datos se grababan verdaderamente como un CD-ROM, no solo se exponían como en el caso del CD-R; en segundo lugar, el respaldo utilizado era vidrio templado en lugar de policarbonato, lo que proporcionaba estabilidad y durabilidad excepcionales para acceder a los datos durante un siglo o más. El proceso de producción se asemejaba al utilizado para grabar circuitos impresos muy pequeños, como procesadores, y requería un entorno libre de polvo. Gracias a la inercia físico-química del vidrio, se obtenía un grabado prácticamente eterno. Aunque la capa reflectante podría plantear problemas a largo plazo, teóricamente era posible “volver a metalizar” los soportes.

Sin embargo, esta tecnología era costosa, más de 207 € por GB, lo que inicialmente parecía astronómico. Pero considerando que este costo se podía amortizar durante al menos 20 años, el precio disminuía a 11 € por GB al año, lo que resultaba más aceptable incluso para los estándares actuales. Se comparó este precio con el costo de inversión en términos de recursos humanos y materiales necesarios para el control y la regeneración de soportes CD-R cada 3 años durante un período de 20 años, y los costos eran similares. Ante el riesgo de perder una fuente potencial de ahorro en un contexto de presupuesto cada vez más ajustado, esta opción parecía la mejor: pagar más ahora para no preocuparse durante 20 años o más.

El proyecto se lanzó en 1997 cuando las primeras bandas estaban en riesgo y se almacenaron por duplicado en los almacenes del nuevo edificio de la BnF, en condiciones controladas de humedad y temperatura. Sin embargo, las dudas sobre qué estrategia adoptar duraron más de dos años mientras las bandas se iban degradando. Al mismo tiempo, en 1997, la BnF lanzó su proyecto de biblioteca digital: Gallica. La necesidad de migrar los datos de los DAT se volvió imperativa para mantener Gallica en línea.

La migración a las bibliotecas de discos ópticos digitales (DON) duró dos años, de 1998 a 1999, y se observó una pérdida del 2,5% de las bandas. Esta observación aceleró el proyecto de transferencia al Century Disk. Las cintas ilegibles se copiaron de la segunda copia, utilizando el CD-R como soporte intermedio. La mayoría de los CD de Century se grabaron durante el año 2000, pero el 6% de las cintas se perdieron durante la operación.

El incremento constante del acervo y el rápido aumento en la cantidad de visitantes al sitio de Gallica hicieron necesario cambiar nuevamente la arquitectura de almacenamiento. Las bibliotecas de discos DON no estaban diseñadas para soportar una demanda tan elevada. Además, la falta de un algoritmo de caché de disco a nivel de los servidores de consulta dificultaba la gestión de solicitudes debido a la gran disparidad en la demanda de consulta. En 2001, se requirió una nueva migración. En esta ocasión, los datos se copiaron en un sistema de almacenamiento completamente basado en disco, sin que se perdiera información durante esta operación.

3.2.5.3. Proyecto SPAR

Una necesidad

Con el surgimiento de nuevos proyectos de digitalización, como el de la prensa del siglo XIX, se planteó una reflexión sobre los formatos adecuados. Dado el estado frágil de las publicaciones de prensa, con papel de baja calidad, la Biblioteca Nacional de Francia (BnF) optó por cambiar de una digitalización de difusión, que priorizaba la distribución fácil pero comprometía la conservación, a una digitalización para la preservación. Esta última buscaba mantener el documento en su formato original durante el mayor tiempo posible, sin alteraciones. Por lo tanto, se abandonó el formato JPEG, considerado demasiado riesgoso debido a su compresión destructiva, en favor del formato TIFF sin comprimir, lo que aumentó cuatro veces el tamaño de los archivos generados.

Al mismo tiempo, los medios para la producción de soportes cinematográficos, especialmente en color, estaban desapareciendo rápidamente, lo que llevó a la digitalización progresiva de documentos a color.

Además de estos desarrollos, surgieron nuevas misiones, como la recopilación de la web francesa. Las primeras colecciones web surgieron en 2002 con el proyecto de recopilación de sitios web de elecciones presidenciales, lo que resultó en la acumulación de varios terabytes de datos. Estas colecciones continuaron expandiéndose, alcanzando más de 150 TB en 2008. A partir de 2005, la BnF también comenzó a explorar el depósito legal electrónico como alternativa al depósito en papel.

Estos cambios tuvieron repercusiones significativas en las capacidades de almacenamiento, lo que obligó al departamento de Tecnologías de la Información (TI) a aumentarlas regularmente. Con el tiempo, el proyecto se volvió más heterogéneo y difícil de mantener. En 2004, ante la creciente complejidad, el departamento de sistemas de información decidió lanzar un contrato para adquirir una infraestructura de fondos digitales escalable en términos de capacidad de almacenamiento (“escalable”) y que cumpliera con los criterios de sostenibilidad en términos de apertura, distribución e independencia tecnológica. Así nació lo que se conoce como la infraestructura SPAR.

La dirección de la BnF sintió entonces la necesidad de ir más allá en este proceso y poner en marcha el proyecto de un sistema de preservación real: SPAR (Preservar el patrimonio digital de la BnF, archivar todos sus datos y distribuir el acceso a estos).

Principios adoptados

Para los administradores de colecciones digitales, principalmente bibliotecarios, existe el riesgo de perder el contacto con los materiales que deben preservarse. Por lo tanto, era esencial establecer un ambiente de confianza y responsabilidad. Es por ello que los grupos de trabajo formados en el marco del proyecto acordaron implementar:

- **Contratos de Servicio:** Estos contratos definen los compromisos tanto del Archivo con sus usuarios como de los usuarios con el Archivo. Cualquier cambio en el número de copias de un tipo de documento o en la velocidad de distribución debe tener en cuenta estos contratos, lo que implica presupuestos y decisiones administrativas bien fundamentadas. El objetivo es evitar asumir compromisos que no se puedan cumplir.
- **Auditabilidad del Sistema:** Se busca que el sistema y los procedimientos implementados puedan ser auditados periódicamente por personas u organismos externos al Archivo, demostrando así su conformidad con los estándares, principalmente el estándar OAIS, y los compromisos establecidos en los contratos de servicio. Esta transparencia es un compromiso tanto de los usuarios como de los responsables del Archivo para garantizar el cumplimiento de los compromisos.

Los grupos de trabajo han definido quiénes son los productores, independientemente de la estructura organizativa actual de la BnF. Se seleccionaron ocho tipos de productores en función de sus características técnicas y legales específicas.

Además, se establecieron modelos de datos y conservación:

- el modelo METS para el formato de envasado,
- el modelo PREMIS para la conservación de metadatos,
- y el formato ODRL para formalizar las licencias de uso de documentos:

Los documentos se distribuyen a los usuarios a través de sistemas dedicados como Gallica, banco de imágenes y estaciones de consulta interna. SPAR garantiza que el solicitante esté autorizado para realizar la solicitud y que el tipo de documento y formato correspondiente sean accesibles para él. Además, se asegura de que se respeten las restricciones de distribución establecidas en las licencias de uso. Por ejemplo, si un documento no debe ser impreso, SPAR informa al solicitante y ajusta el sistema de distribución en consecuencia. Las licencias de uso son un medio para contractualizar la transferencia de responsabilidades, garantizando así la integridad y el cumplimiento de las políticas de uso establecidas.

Los grupos de trabajo establecieron el número de versiones que se conservarían de un documento. Así, se mantendrá un máximo de tres versiones. La versión inicial que nunca se volverá a tocar. La versión actual y la versión inmediatamente anterior. Esta precaución permitirá que BnF vuelva a la versión anterior en caso de que se observe un error después de una transformación. Solo los documentos que tienen contratos de servicio específicos, como es el caso de los registros administrativos que tienen períodos de retención definidos, pueden destruirse.

El equipo encargado de la gestión de riesgos ha desarrollado una lista preliminar de riesgos y ha evaluado cada uno de ellos. Aunque intentaron definir un ámbito específico para los riesgos, se dieron cuenta de que esta tarea era demasiado extensa. En su lugar, establecieron procedimientos para monitorear y abordar los riesgos, lo que permitirá que el repositorio de riesgos evolucione con el tiempo.

Estudio de factibilidad

Era fundamental asegurar que la implementación de un sistema OAIS fuera factible para la BnF. La institución no tenía la intención de desarrollar un sistema desde cero, sino más bien construir sobre una solución existente que pudiera compartirse con otros. En este sentido, se llevó a cabo un estudio de viabilidad basado en el modelo OAIS, en paralelo con otras actividades y sin prever las opciones elegidas. Este ejercicio fue delicado porque la oferta no estaba madura en el mercado. Inicialmente, fue necesario explorar los productos existentes que podrían satisfacer las necesidades.

Se evaluaron veinte productos en tres áreas: calidad funcional, calidad técnica y durabilidad demostrada. Estos productos pertenecían a tres categorías distintas:

- software de gestión de contenidos que podría adaptarse a las necesidades, pero ninguno fue seleccionado en esta categoría,
- software de “repositorio institucional”, como DSPACE o FEDORA, que generalmente son software dedicados al acceso pero que incluyen interesantes mecanismos de transferencia y gestión de datos,
- Software específicamente diseñado para la gestión de archivos.

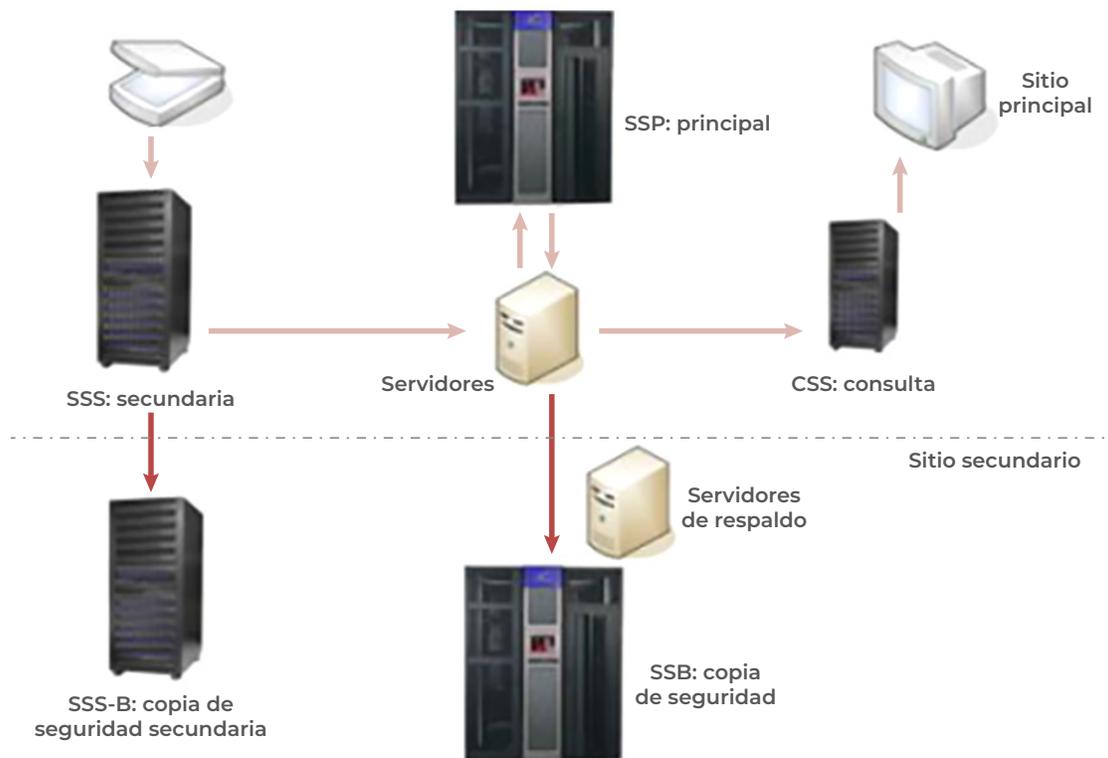
Tras esta exploración inicial, se seleccionaron seis programas de software para una evaluación más detallada, basada en un cuestionario de 132 criterios divididos en 9 categorías que abordaban las funcionalidades definidas por el estándar OAIS, así como aspectos técnicos y organizativos. Este estudio demostró que las soluciones identificadas podían satisfacer las necesidades de la BnF.

Preservación de la información digital en la BnF

La BnF tiene la responsabilidad de conservar el patrimonio cultural que alberga, lo que incluye también la preservación de sus colecciones digitales con el mismo compromiso que sus colecciones físicas. Para la institución, esta preservación debe contemplar una perspectiva multiseccular.

Por lo tanto, SPAR sería más que un simple repositorio de datos seguro. El sistema garantizaría la integridad de los datos mediante la creación de múltiples copias, monitoreo continuo del estado de los medios de almacenamiento y la realización de transformaciones necesarias en caso de obsolescencia tecnológica para garantizar la accesibilidad continua. Entonces, por ejemplo, cuando el formato de imagen JPEG se vuelva obsoleto, SPAR podrá transformar las imágenes afectadas al formato de imagen JPEG del mañana. Además, SPAR permitiría recuperar los datos en cualquier momento y conservar hasta nueve copias de un objeto digital.

LA INFRAESTRUCTURA SPAR



Infraestructura de almacenamiento de SPAR

3.2.6. La experiencia de una gran organización de educación superior: PAC, la plataforma CINES

El mundo de la educación superior y la investigación ha estado preocupado durante años por la difusión, publicación y acceso a la información digital. Sin embargo, la preocupación por la preservación digital es más reciente, como lo demuestra uno de los enfoques de los equipos muy grandes (TGE ADONIS) en las ciencias humanas, centrado en la preservación a largo plazo.

3.2.6.1. Centro Nacional de TI para la Educación Superior (CINES)

El Centro Informático para la Educación Superior (CINES), ubicado en el sur de Francia, en Montpellier (Hérault), fue establecido en 1999, sucediendo al CNUSC (Centre National Universitaire Sud de Calcul), creado en 1980. Supervisado por la DGRI (Dirección General de Investigación e Innovación) y la DGES (Dirección General de Educación Superior) del Ministerio de Educación Superior e Investigación, tiene dos misiones principales: la informática digital intensiva y la preservación a largo plazo de documentos electrónicos. Desde 2004, CINES ha estado trabajando en el desarrollo de un servicio de archivo a largo plazo del patrimonio científico.

La Información Científica y Técnica (IST) abarca toda la información producida o recibida por los sectores de investigación y educación. Incluye una variedad de contenidos y formatos, desde revistas científicas y tesis informes, actas, trabajos especializados, manuales, bibliografías, resúmenes, prepublicaciones, patentes, mapas, hasta datos estadísticos y bancos de imágenes y videos.

El éxito de una estrategia nacional para el archivo a largo plazo de documentos electrónicos producidos por la comunidad científica y técnica requiere enfrentar varios desafíos, que van desde el desarrollo de nuevas habilidades comerciales (con el establecimiento de un servicio dedicado al archivo, definición de nuevos procesos y métodos a implementar y dominar para gestionar un sistema de archivo), hasta la sensibilización de la comunidad sobre la importancia de la preservación a largo plazo (con el fortalecimiento de colaboraciones entre informáticos, archivistas y bibliotecarios, la organización de jornadas de sensibilización e información sobre el archivo a largo plazo, etc.).

Esta implementación también debe realizarse en conjunto con los demás servicios de archivos involucrados en esta cadena en función del origen y naturaleza de los documentos (Archivos Departamentales, Archivos Nacionales, BnF) respetando el contexto legislativo.

3.2.6.2. Implementación de la plataforma PAC (plataforma de archivo CINES)

Para cumplir con su misión de archivo, CINES ha lanzado el proyecto PAC, con el objetivo de establecer un servicio de archivo digital sostenible. El equipo está compuesto por un director de proyecto, cuatro ingenieros y un archivista.

En la fase 1, se desarrolló internamente una primera plataforma para validar los servicios esperados en el proyecto de archivo electrónico de tesis, con una capacidad de almacenamiento inicial de 300 GB. La plataforma se basó en estándares de la industria, como el modelo OAIS y el formato de metadatos Dublin Core, y se limitó deliberadamente a una lista de formatos de archivo estándar y ampliamente utilizados, como HTML, PDF, TXT, XML, GIF, JPEG, TIFF y PNG, así como WAV. La arquitectura se fundamentó en software de código abierto (Java, PostgreSQL, Jhove, ImageMagick), y las primeras pruebas de archivo de tesis se realizaron en marzo de 2007, con el inicio de la producción a finales de ese mismo año.

En la fase 2, a finales de 2007 se lanzó una licitación para adquirir una plataforma de almacenamiento capaz de manejar grandes volúmenes (20 TB ampliables a 40 TB). Esta plataforma también se basó en estándares y normas del campo, utilizando hardware SUN, software Arcsys y software de código abierto como Java y MySQL, junto con Jhove e ImageMagick para la identificación, validación y caracterización de formatos. Las primeras pruebas de transferencia y migración de documentos comenzaron en marzo de 2008, con el inicio de operaciones a fines del segundo trimestre de ese año.

La plataforma ofrece diversas funcionalidades, como la recepción y el control de SIPs (cumplimiento de metadatos, control y validación del formato de archivo), la creación del AIP, el almacenamiento (copia múltiple del AIP en diferentes soportes, verificación periódica de la integridad del AIP, migración tecnológica, generación de informes y estadísticas) y el acceso (autenticación del solicitante, consulta y comunicación).

3.2.6.3. Archivo de tesis y revistas electrónicas de humanidades y ciencias sociales

En cuanto a las tesis, el proceso se inició tras la promulgación del decreto del 7 de agosto de 2006, que establece las modalidades de depósito, notificación, reproducción, distribución y conservación de las tesis u obras presentadas en defensa para obtener el doctorado.

Los estudiantes de doctorado presentan sus tesis en formato electrónico en la biblioteca universitaria donde defienden su trabajo. Luego, las bibliotecas envían las tesis electrónicas a la Agencia Bibliográfica para la Educación Superior (ABES) mediante la herramienta STAR. Después de tres etapas de validación, las tesis elegibles para el archivo se transfieren a la plataforma PAC, y una copia puede estar disponible en línea para la comunidad de Internet a través de un sitio de difusión.

El proyecto de archivo de revistas se inició en 2006 como respuesta a un proyecto de digitalización masiva y preservación de colecciones retrospectivas de revistas de humanidades y ciencias sociales, liderado por el equipo de Persée (Université Lumière - Lyon 2). La cadena de digitalización garantiza la digitalización masiva, la centralización y automatización del procesamiento, así como el archivo a largo plazo de los datos.

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.

Bibliografía

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