

Aproximación a la preservación de objetos cartográficos y geográficos digitales

José Luis Colomer Alberich*

Recibido el 24 de abril de 2016; aceptado el 21 de julio de 2016

Abstract

Because of the relatively short lifespan of the digital supports and technologies, there is a need of systematically renewing physical media to prevent its decay and the loss of information. The term *Digital Dark Age* depicts a scenario of almost complete inaccessibility to the digital information of the past due to the obsolescence of the physical media, the formats and, in general, of the systems supporting the digital information. In the geographic domain, it is clear the interest in preserving the information that has been used for decision making. The long term preservation of cartographic and geographic digital objects aims at ensuring its reusability at any time in the future. The paper reviews some of the most relevant topics regarding preservation of digital geographic objects.

Key words: *Digital preservation, digital repositories, digital archives, OAIS.*

Resumo

Os curtos prazos de vigência das tecnologias de informação e de seus suportes digitais exigem sua renovação frequente para prevenir sua obsolescência ou deterioração com a consequente perda da informação armazenada. O termo *Digital Dark Age* anuncia um cenário de impossibilidade prática de se acessar a informação digital devido a obsolescência do suporte físico, dos formatos e, em geral, dos entornos informáticos que a suportam. No âmbito da informação geográfica e cartográfica, parece um dever indesculpável preservar os dados que têm servido de base para a tomada de decisões e conseguir que estes possam ser utilizados nos sistemas de informação do futuro. O artigo apresenta os principais conceitos relacionados com a preservação deste tipo de informação.

* Correos electrónicos: coloner@ati.es; ati_oo_391@gmail.com

Palavras chave: *preservação a longo prazo, repositórios digitais, arquivos digitais, OAIS.*

Resumen

Los cortos plazos de vigencia de las tecnologías de la información y de sus soportes digitales exigen su renovación frecuente para prevenir su obsolescencia o deterioro con la consiguiente pérdida de la información almacenada. El término *Digital Dark Age* anuncia un escenario de práctica imposibilidad de acceder a la información digital debido a la obsolescencia del soporte físico, de los formatos y, en general, de los entornos informáticos que la soportan. En el ámbito de la información geográfica y cartográfica, parece un deber inexcusable preservar los datos que han servido de base para la toma de decisiones y conseguir que éstos puedan ser utilizados en los sistemas de información del futuro. El artículo presenta los principales conceptos relacionados con la preservación de dicho tipo de información.

Palabras clave: *preservación a largo plazo, repositorios digitales, archivos digitales, OAIS.*

Introducción

Es bien conocido el hecho de que la duración de los soportes y formatos de la información digital es más corta que la de los soportes analógicos del pasado, de forma que si no son renovados de forma frecuente, envejecen y la información se pierde. También, la incesante evolución de las tecnologías de la información hace obsoletos los sistemas de proceso con enorme rapidez. El término *digital dark age* expresa un escenario futuro en el que puede ser prácticamente imposible acceder a la información digital del pasado debido a la obsolescencia del formato, del soporte físico o del *software* para procesar los datos.

El impacto que una pérdida de información puede tener sobre cualquier tipo de actividad, bien sea ésta de tipo administrativo, legal, de planificación o científica, es evidente. Por ejemplo, sólo el 20% de los datos de la misión Viking a Marte en 1975 pudieron ser recuperadas en 1999. Con la información adicional que se pudo recuperar dos años después, el profesor Joseph Miller encontró indicios de vida (Spaceflight Now, 2001).

En el ámbito de las aplicaciones y estudios geográficos, es innegable el valor que tiene el poder disponer de la información sobre la que se han basado decisiones pretéritas de relevancia. La preservación a largo plazo de información geográfica pretende conseguir que los datos puedan ser utilizados en los Sistemas de Información Geográficos del futuro.

Para conseguir dicho objetivo, es necesario que las instituciones con responsabilidad en la preservación de datos a largo plazo, garanticen el acceso perenne a la

información que preservan, y ello implica asegurar tanto la fiabilidad técnica como la sostenibilidad económica e institucional. La calificación de “repositorio fiable”, que se otorga a una tal institución una vez certificada su solvencia, va en la línea de garantizar la persistencia de sus servicios en el tiempo.

La preservación a largo plazo se asocia a las responsabilidades de instituciones especializadas —bibliotecas, archivos y cartotecas— que tienen encomendada la función de conservación permanente de la información depositada por las instituciones productoras que la generan. Habitualmente, estas últimas no tienen la preservación a largo plazo como misión principal, lo que implica que sus decisiones sobre la retención y selección de los objetos a preservar son, en principio, diferentes a las de los Archivos. Por otra parte, y dado que una institución productora es donde residen los datos originales, es conveniente organizar y coordinar las responsabilidades de todos los actores que intervienen en el ciclo de vida de la información, desde su captura y producción inicial hasta los procesos necesarios para su preservación a largo plazo.

En definitiva, en la temática de la preservación a largo plazo hay diferentes actores con responsabilidades diversas que conviene definir y coordinar sin ambigüedades. Un país con instituciones dedicadas a la preservación es un indicador de que sus administraciones saben valorar adecuadamente su patrimonio.

El presente artículo pretende aproximar el lector a la temática de la preservación a largo plazo de información geográfica y cartográfica digital. Para ello, el artículo empieza describiendo las tareas que realizan de forma habitual los Centros de Datos, para pasar luego a describir el modelo de referencia OAIS, que sirve de hilo conductor para introducir los principales conceptos de la preservación, y finalizar con la preservación de la información geográfica digital. En el último capítulo se esbozan cuestiones institucionales, de costes y de sostenibilidad.

Las prácticas de preservación en los centros de datos

La gestión de datos que se realiza de forma rutinaria en los centros de cómputo o de datos tiene indudables componentes de preservación. Así, los objetos digitales —archivos con secuencias de bits codificados según un formato— se organizan y gestionan para garantizar su integridad, disponibilidad y acceso, y se borran o se hacen copias perennes en base a su valor, interés o cualquier otro motivo establecido por la institución. Dichas copias se acostumbran guardar en lugares separados en condiciones físicas y ambientales controladas.

Además, se refresca el soporte físico cada pocos años para prevenir los efectos de su envejecimiento, y se migra la información a soportes de mayor capacidad, velocidad y fiabilidad, y de coste menor (*United States Geological Survey* —USGS, 2012). Si bien dichas acciones aseguran que los datos se almacenen en un soporte no obsoleto, en ningún caso garantizan que las aplicaciones puedan continuar sien-

do operativas cuando se cambia la arquitectura del sistema: sistema operativo, lenguajes de programación, sistema de archivos y bases de datos, formatos, modelo de proceso centralizado o distribuido, etc. Un cambio de ese tipo desencadena largos y costosos procesos de reingeniería y de migración de datos que pueden durar años.

Más allá de la preservación física de los datos en un entorno tecnológico cambiante, preservar a largo plazo significa también garantizar su usabilidad actual y su disponibilidad y posibilidad de reutilización en el futuro. El término “curación” (traducción literal del término inglés *curation*) engloba la totalidad de las acciones necesarias para asegurar la reusabilidad de los datos a largo plazo. Por simplicidad, el presente artículo usará el término preservación como sinónimo de curación, y de “Archivo”, escrito con mayúscula, para designar una institución dedicada a la preservación a largo plazo.

El largo plazo

En preservación, el largo plazo comprende un período de tiempo suficientemente dilatado como para llegar a comprometer la integridad de la información por deterioro del soporte, cambio de tecnología, cambio de la tipología de los usuarios que la utilizan, etc. Una definición más concreta es la del grupo de trabajo de preservación de la organización de Agencias Cartográficas Europeas EuroSDR (EuroSDR, 2016), que propone períodos de retención de un año para los archivos en línea en los servidores corporativos, 10 años en archivos de transición operando en modo casi tiempo real, y 100 o más años para los archivos de larga duración fuera de línea. La Figura 1 corresponde a una arquitectura operacional que coincide esencialmente con la clasificación citada.

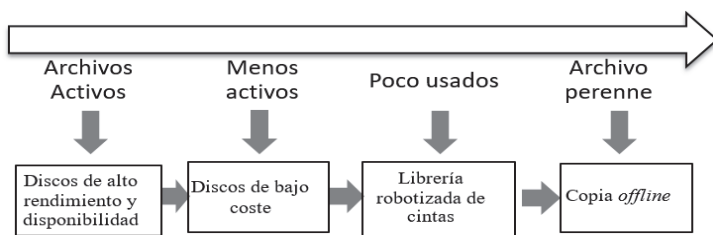


Figura 1. Arquitectura de almacenamiento y estrategias típicas de migración: refresco del soporte cada 5-6 años y cambio formato cada 10-30 años.

Es interesante estudiar la propuesta de la *National Digital Stewardship Alliance* (NDSA), que sintetiza en una sencilla tabla (NDSA, 2012) los requerimientos y acciones básicas a implantar en un sistema de preservación a largo plazo. El objetivo de sus autores fue el resumir la temática de la preservación de forma sencilla, pero representativa, comprensible y libre de demasiado argot técnico, para así conseguir una

mayor difusión y aceptación de los conceptos de preservación. La propuesta ha sido adoptada por el USGS (2014), que es la institución que, entre otras, tiene a su cargo el archivo a largo plazo de las imágenes captadas por los satélites Landsat desde 1972.

La propuesta de la NDSA divide las tareas de preservación en cinco grupos: almacenamiento y ubicación, integridad de los datos, seguridad, metadatos y formatos, y para cada uno de ellos, especifica una serie de acciones a emprender a cuatro niveles de complejidad progresiva: protección, conocimiento, monitoreo, y reparación del material preservado. La propuesta se puede aplicar a cualquier tipo de datos digitales y en ella se reconocen alguna de las acciones que realiza de forma rutinaria un centro de datos, lo que ofrece una vía de implantación progresiva de la preservación en una institución.

El ciclo de vida de los datos

El ciclo de vida de un objeto digital empieza con su creación por parte de un productor, y sigue con su ingreso en el repositorio digital una vez comprobada la integridad de los datos, las acciones periódicas de cambio de soporte y formato, y los procesos necesarios para permitir a los usuarios acceder a los datos. Aunque hay principios comunes, cada institución define su ciclo de vida particular tal y como puede verse en la compilación de la referencia (Committee on Earth Observation Satellites —CEOS, 2012).

El ciclo de vida contempla también la baja de objetos preservados. Un mecanismo interesante para prolongar la vida del objeto es hacer pública una situación de inminente retirada de una colección de un Archivo con la idea de que aparezcan otras instituciones (o empresas) interesadas en la colección a eliminar (American Society of Photogrammetry and Remote Sensing —ASPRS, 2010).

Los formatos

Un formato codifica secuencias de bits de acuerdo con una especificación determinada. Por tanto, es necesario conservar la documentación de dicha especificación y conocer los paquetes software que lo saben manejar. Una forma eficiente de hacerlo es estableciendo un registro (Geremew, 2006), que no es más que un recurso centralizado capaz de guardar información diversa como, por ejemplo, diccionarios de datos, esquemas de base de datos, plantillas, y vocabularios. En particular, un registro puede guardar información de formatos: por ejemplo, la base de datos del sistema creado por los Archivos Nacionales del Reino Unido (PRONOM, 2016), que contiene información de más de 800 formatos digitales diferentes. Otro ejemplo es la Librería del Congreso de los Estados Unidos (NDSA, 2016) que, además, define factores de sostenibilidad para ayudar a la elección de los formatos más adecuados para la preservación:

- Divulgación: existencia de especificaciones completas
- Aprobación: grado de utilización del formato
- Transparencia: capacidad de poder inspeccionarlos con un editor de texto
- Dependencias externas: de sistemas físicos, sistema operativo o programas
- Patentes: grado en que éstas inhiben la capacidad de preservar del Archivo
- Mecanismo de protección (cifrado): que inhiba preservar el contenido

Evidentemente, dichos factores compiten entre sí: por ejemplo, el formato PDF para preservar la apariencia original de un documento *versus* la sencillez de un archivo XML. Si bien el buen sentido aconseja decidir sobre formatos mientras aún están vigentes, a menudo dicha buena práctica se olvida por el hecho de que los fabricantes de sistemas acostumbran a asegurar la compatibilidad entre formatos y *software* “mientras los usuarios lo pidan”, una expresión que se antoja demasiado vaga como para no preocuparse del tema.

Por tanto, y dado que no actuar a tiempo suele ser el preludio de la pérdida de información y de unos costes elevados para recuperarla, es aconsejable monitorear frecuentemente la evolución de los sistemas comerciales. Una alternativa es desarrollar un formato propio como, por ejemplo, el *Software Independent Archiving of Relational Databases* (SIARD, 2016) creado por los Archivos Federales Suizos, y de uso libre.

La selección de la información a preservar

Obviamente, no es posible ni sostenible preservarlo todo. Por tanto, deben establecerse criterios de selección (*appraisal*) razonados y documentados exhaustivamente basados en la importancia de la información, uso previsto, valor, y otros similares. Por motivos fáciles de entender, la selección es el proceso más crítico de la preservación.

Forma parte de la selección el definir el nivel de detalle al que se quiere llegar en la preservación: por ejemplo, para datos científicos de misiones espaciales, la Agencia Europea del Espacio preserva los datos originales, los productos derivados, la información auxiliar, la calibración y la validación, los indicadores de calidad, el *software* de generación de productos, de visualización y de control de calidad, y la documentación relativa a la misión (ESA, 2012). La razón de tanta exhaustividad es la consideración de que nuevos avances científicos aplicados a los datos antiguos pueden dar lugar a resultados más precisos o nuevos.

También es común establecer acuerdos entre los productores y las instituciones de preservación para depositar el producto final en un Archivo y conservar los datos originales en la institución productora. Por tanto, el productor asume también responsabilidades de preservación en el ciclo de vida del producto.

Acceso a los datos preservados

Un Archivo debe considerarse como un recurso más de información y no únicamente como un lugar donde guardarla a largo plazo. En consecuencia, un Archivo debe implantar servicios de localización y descarga de los objetos preservados y en los formatos que se determinen.

Los canales de distribución de datos digitales tradicionales —las copias en soporte magnético y la descarga de ficheros por Internet— han evolucionado hacia los servicios en línea. Por ejemplo, los servicios web de las Infraestructuras de Datos Espaciales (IDE) son hoy en día la modalidad preferida para acceder a la información geográfica. En los dos primeros casos, se distribuyen “paquetes” conteniendo los datos y sus metadatos; en el último, el proveedor asume la mayor servidumbre que representa dar un servicio en línea en cuanto a asegurar su velocidad y disponibilidad, lo que redundará en mayores costes. Aunque actualmente los Archivos no suelen ofrecer servicios de datos en línea, un argumento a favor de hacerlo sería dar mayor visibilidad al Archivo en la medida que le permitiría integrarse fácilmente en iniciativas de la sociedad de la información.

Enlaces externos

La volatilidad de los enlaces web es una preocupación constante a pesar de las buenas prácticas de diseño de identificadores uniformes de recursos (Uniform Resource Identifier —URI) y de las normas nacionales promulgadas para hacerlos lo más persistentes posible (ISA, 2012). No extraña, pues, que algunos Archivos los utilicen solamente para enlazar con entornos controlados. Por ejemplo, a publicaciones almacenadas en el mismo Archivo (Hodge, 2000). Una solución más robusta es establecer mecanismos entre Archivos para informar de cualquier modificación en los enlaces. Ese tipo de mecanismo tiene precedentes en el proyecto SCIDIP-SE de la Agencia Espacial Europea (ESA, 2016).

El modelo conceptual OAIS

Hasta el momento se han descrito aspectos de preservación a largo plazo que, en mayor o menor medida, corresponden a buenas prácticas comunes en el manejo de datos digitales. No obstante, y muy a menudo, las soluciones que se implementan son específicas para problemas de preservación concretos. Por ejemplo, las derivadas de las recomendaciones del grupo de preservación EuroSDR para la información geográfica y cartográfica digital (Röndsdorf *et al.*, 2011).

Por tanto, parece conveniente establecer unos conceptos y una terminología común para poder evaluar y comparar diferentes estrategias y soluciones de los sistemas de preservación, sus modelos de datos y procesos, y los estándares existentes a utilizar o los nuevos a desarrollar. Con ese objetivo, el *Consultative Committee*

for Space Data Systems (CCSD) desarrolló el *Open Archives Information System* (OAIS - ISO14721:2012) como un modelo conceptual para formalizar sistemas de preservación de datos a largo plazo. OAIS es

...un Archivo, que es una organización y que puede pertenecer a una más grande, con sistemas y personas que han aceptado la responsabilidad de preservar información y ponerla a disposición de una cierta comunidad designada.

El modelo OAIS contempla todas las funciones de un sistema de preservación: la entrada al sistema, el almacenamiento y manejo de datos, y el acceso y la diseminación de la información. Asimismo, contempla la migración a nuevos soportes y formatos, los modelos de datos, el intercambio de información digital entre Archivos y sus interfaces, y otros servicios de alto nivel. Igualmente define el conjunto mínimo de características que debe poseer un Archivo para ser considerado “Archivo OAIS”. Finalmente, define y especifica conceptos y terminología.

OAIS introduce dos conceptos importantes: la comunidad de usuarios (la “comunidad designada”) para quien se preserva la información y que posee una serie de conocimientos (la “base de conocimientos”) sobre la información preservada, y la “información de representación”, indispensable para interpretar el significado de los datos preservados. La relación se puede visualizar como sigue (Figura 2):

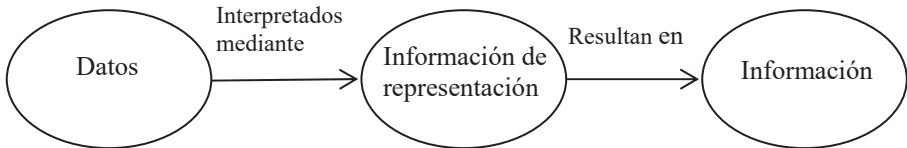


Figura 2. Información de representación y su papel en la interpretación de los datos.

Por ejemplo, tratar una imagen digital como una pura secuencia de bits no produce una imagen si no se conoce el número de bits consecutivos que forman un píxel y que el valor numérico del píxel indica el nivel de gris. La información de representación debe proporcionar dicho conocimiento. Si la comunidad designada ya lo posee o dispone de un programa de visualización de imágenes, entonces no es necesario almacenar ninguna información de representación sobre lo que es una imagen digital; en caso contrario, ésta deberá describirse con todo detalle.

Evidentemente, los conocimientos de la comunidad designada tienen implicaciones sobre el tipo, cantidad y detalle de la documentación que hay que guardar junto con los datos para garantizar que éstos sean interpretables y, por tanto, reutilizables en el futuro. Es fácil de entender que cuando la base de conocimientos de la

comunidad designada cambia, también hay que cambiar la información de representación.

Arquitectura y paquetes de información

Esencialmente, OAIS se apoya en un modelo de información propio para almacenar datos sobre la información preservada. En lo que sigue, el artículo resume sucintamente los principales elementos de OAIS. La referencia (CCSDS, 2012) describe completamente el modelo.

En un Archivo OAIS participan tres actores: el Productor (*Producer*), el Gestor (*Manager*), y el Consumidor (*Consumer*). Los actores pueden ser una institución o un ordenador interactuando en un entorno como el de la Figura 3.

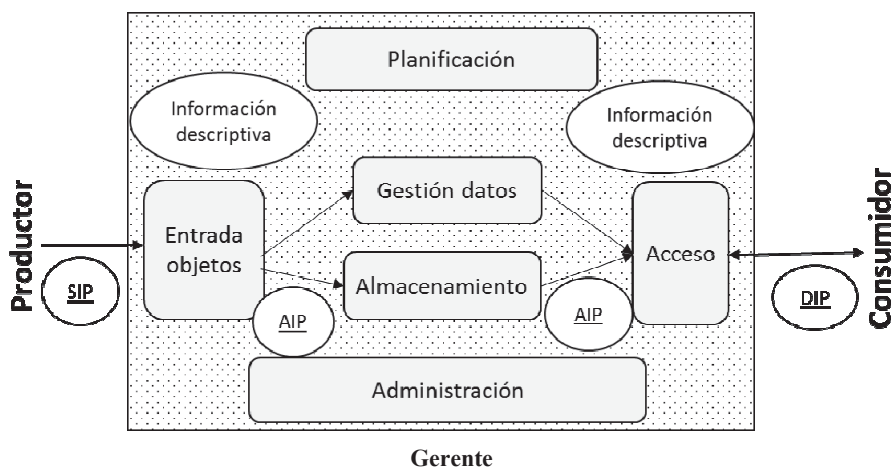


Figura 3. Arquitectura funcional de un Archivo OAIS (de CCSDS, 2012).

El Productor son las personas o sistemas que entregan la información a preservar al Archivo. Lo hacen con “paquetes de información de entrega” (*Submission Information Packages* —SIP) que contiene dos grupos de datos: por una parte, los datos a preservar (la secuencia de bits) con su información de representación y, por otra, la información descriptiva de la preservación (*Preservation Description Information* —PDI). Forma parte del PDI el linaje, el contexto para relacionar los datos con otros, identificadores, la información para la comprobación de la integridad de los datos, y los derechos de propiedad y acceso.

Tanto la información de representación como la información descriptiva del paquete entregado pueden ser parciales y no cumplir con los requerimientos establecidos por el Archivo. Por tanto, ambas se amplían y completan antes de conver-

irse en un “paquete de archivo” (*Archival Information Package* —AIP), que es el objeto que se maneja físicamente y que se almacena junto con sus propios metadatos. Parte de la información de dichos metadatos se cargan en el sistema de información (*Archival Information System*) que gestiona el sistema.

El Consumidor interactúa con el sistema para obtener la información deseada, que se distribuye mediante el “paquete de distribución” (*Dissemination Information Package* —DIP). Por razones prácticas, es conveniente que los DIP sean capaces de entregar los datos con los formatos vigentes en cada momento. Finalmente, el Manager realiza funciones de alto nivel tales como determinar los objetivos y alcance del Archivo, supervisar su funcionamiento, establecer políticas de precios del servicio, realizar tareas de promoción institucional, etc., mientras que el Administrador es la figura responsable de las operaciones del Archivo.

Un Archivo OAIS debe cumplir un mínimo de requerimientos. Por ejemplo, debe poseer suficientes derechos de copia como para poder distribuir la información que está preservando, así como poder determinar la comunidad designada, disponer de procedimientos de contingencia (incluyendo la eventualidad de la desaparición del propio Archivo), y poner los datos a disposición de la comunidad designada. Cualquier modificación implica modificar la información asociada a los datos, y estos cambios deben estar perfectamente documentados y registrados. Todos estos requerimientos deben estar certificados y auditados.

Metadatos de preservación

OAIS define conjuntos de metadatos necesarios para implementar las funcionalidades de preservación y es información indispensable para la preservación a largo plazo. El documento *Preservation Metadata: Implementation Strategies* (PREMIS, 2016), describe dichos metadatos en base a los conceptos y requerimientos de OAIS. PREMIS define unidades semánticas (piezas de información) sobre los objetos, eventos, agentes y derechos y que pueden relacionarse entre sí.

Al igual que OAIS, PREMIS es neutral respecto a la implementación, pero existe un conjunto de esquemas XML disponibles para facilitar su utilización en sistemas de preservación. *El Diccionario PREMIS*, el esquema, y otros recursos están mantenidos por la Biblioteca del Congreso de los Estados Unidos; la última versión 3.0 (noviembre de 2015) es el estándar de facto empleado en los Archivos digitales.

Los paquetes de información y su transferencia a Archivos OAIS

Una propuesta de paquetes XML para el transporte de información en base a OAIS es el Metadata Encoding and Transmission Standard (METS, 2016). Otros no inspirados en OAIS son los esquemas BagIt y Fedora Object XML (FOXML).

METS consta de una cabecera, metadatos descriptivos del objeto, metadatos administrativos, la lista de archivos que constituyen el objeto digital, un mapa de los componentes del objeto, una lista de enlaces entre los componentes del mapa, y una lista de comportamientos asociados al objeto. Los metadatos pueden empaquetarse en el documento METS o pueden estar enlazados a una fuente externa como, por ejemplo, un registro. METS y el paquete de información OAIS son similares en la medida que ambos contienen información y sus metadatos en un paquete.

Finalmente, pueden mencionarse otras iniciativas para estandarizar la interfaz entre los productores y los Archivos OAIS. Por ejemplo, el *Methodology Abstract Standard* (PAIMAS) ISO 20652:2006 aporta una descripción estandarizada de las interacciones entre productores y un Archivo OAIS, o el *Producer-Archive Interface Standard* (PAIS) para definir los objetos de información digitales a transferir de un productor a un Archivo y para empaquetar estos objetos en formato de SIP. En definitiva, PAIMAS y PAIS son iniciativas para formalizar y hacer más robusto y comprensible el paso de información desde el Productor a un Archivo OAIS.

La certificación TRAC

La neutralidad del modelo OAIS en lo que se refiere a la implementación crea indefiniciones cuando debe precisarse lo que es “ser un Archivo OAIS”. El método de auditoría Trusted Repositories Audit & Certification: Criteria and Checklist (TRAC, 2007) es una interpretación particular, pero detallada, de los requisitos de conformidad OAIS para certificar tanto la suficiencia técnica como la sostenibilidad económica e institucional de un Archivo. TRAC puede emplearse también como guía para establecer los aspectos organizativos e institucionales de un Archivo. TRAC es la base del documento Trusted Digital Repository ISO16363:2012. TRAC contiene una lista de verificación para comprobar:

- Aspectos institucionales: viabilidad, gobierno, estructura organizativa y de personal, marco normativo, sostenibilidad financiera, contratos, licencias y obligaciones.
- Gestión de los objetos digitales: procesos de adquisición de objetos, estrategias de preservación y procesos para actualizarlos en un entorno de cambio tecnológico, metadatos para localizar y gestionar los objetos en el sistema, y capacidad para producir y difundir versiones correctas y auténticas de los objetos.
- Infraestructura: gestión del riesgo y la seguridad.

TRAC verifica también si existen mecanismos para identificar la identidad del productor del material digital, la existencia de una nomenclatura que genere identificadores persistentes, únicos y visibles para los objetos archivados, la existencia de mecanismos para asegurar que las copias de los objetos estén sincronizadas, y si se

realiza un análisis sistemático de los factores de riesgos asociados a la seguridad de los datos, del sistema, del personal y de la infraestructura física.

La información geográfica y su preservación

Los datos geográficos (o datos espaciales) tienen una serie de características que los hacen diferentes a otro tipo de información. Por ejemplo:

- Tienen coordenadas para determinar su posición en un sistema de referencia espacial bien determinado. El registro EPSG proporciona los parámetros geodésicos de cada sistema y las transformaciones entre ellos.
- Maneja tipos de datos *raster* y también vector, que son representaciones de los objetos del mundo real. No obstante, sus significados no son únicos y la semántica depende de las aplicaciones. Tampoco existe una forma de especificar la semántica de forma procesable directamente por un ordenador.
- La generación de información geográfica la efectúan instituciones diversas y no necesariamente de forma simultánea y coordinada. No obstante, los análisis espaciales requieran combinar dicha información producida en esas condiciones.
- No existe un esquema para generar identificadores únicos universales, y no todas las clasificaciones temáticas son únicas. Una sugerencia útil es usar la clasificación temática incluida en el estándar de metadatos geográficos ISO19115 (FGDC, 2006).
- Los formatos acostumbra a ser propietarios y ligados a productos comerciales. Tampoco existe un formato estandarizado para empaquetar datos y metadatos, sino sólo soluciones tales como usar empaquetamientos ZIP o TAR.
- A partir de los datos espaciales pueden generarse mapas y visualizaciones diferentes con sólo cambiar la simbología. De ahí la recomendación de preservar un mapa junto con los datos.
- Las visualizaciones en pantalla desaparecen al terminar la sesión. Afortunadamente, el estándar WMS del *Open Geospatial Consortium* (OGC) permite conservar tanto la relación de los distintos servicios geográficos web invocados para la visualización, y la apariencia (estilo) de cada uno de ellos.
- El derecho de uso de los datos geográficos de terceros da lugar a casuística diversa. Por ejemplo, el caso de información temática libre compilada sobre una base cartográfica con restricciones de uso y que se quisiera preservar (y distribuir) de forma conjunta.
- El elevado volumen de datos, que puede llegar a exceder la capacidad máxima de un archivo individual, obliga a o bien preservar bases de datos (propietarias), o bien a hacerlo en base a trozos discretos (teselas) con los consiguientes problemas de geometría y topología de los objetos en los bordes de los mismos.

- Es estándar de metadatos ISO 19115 no fue diseñado para la preservación. Un problema añadido es que éstos pueden generarse asincrónicamente con los datos y que no hay ningún formato de datos que permita empaquetarlos conjuntamente con ellos. El resultado son dos conjuntos de datos prácticamente independientes, lo induce a frecuentes incoherencias de actualización entre ellos.

Como puede observarse, la casuística es variada y la lista de temas a resolver no es completa, con lo que las soluciones a adoptar requieren de una evaluación detallada e individualizada (DPC, 2009). En los apartados que siguen se amplían algunos de los casos citados en la lista anterior.

El estándar OAIS para la información geográfica

Como se ha dicho, OAIS permite diseñar los procesos de preservación digital abstractando los detalles de implementación. En otras palabras, OAIS es neutral en cuanto a la implementación y tampoco ha definido estándares sobre algunos de los conceptos que utiliza. Por ejemplo, no existe un estándar para el SIP ni tampoco su equivalente para información geográfica.

Por tanto, OAIS puede utilizarse para definir los objetivos de un proyecto de preservación manteniendo la libertad en la implementación. Por ejemplo, una implementación específica para la información geográfica inspirada en OAIS es el proyecto *National Geospatial Digital Archive* (NGDA) financiado por la Librería del Congreso de Estados Unidos (Janée, 2008).

Actualmente está en fase de elaboración el estándar ISO19165 que aplica el modelo OAIS a la temática de la preservación información geográfica.

Formatos para la información geográfica

En principio, hay libertad de elección de formatos, pero las buenas prácticas aconsejan escoger formatos sencillos y abiertos. Suiza propone los formatos TIFF y GeoTIFF, Interlis (para base de datos), Shape, GML y GeoPDF (Ellipse, 2013), y Estados Unidos GeoTIFF, GML, KML y TIGER/Line como formatos preferidos, y Shape y otros, como aceptables (NARA, 2014).

Preservación de bases de datos SIG

Actualmente, los datos geográficos se suelen almacenar en bases de datos relacionales y a manejarlos mediante software de Sistemas de Información Geográfica (SIG). Por tanto, además del formato, el sistema gestor de bases de datos y el paquete de programas SIG son dos nuevos elementos a controlar en lo que se refiere a su obsolescencia tecnológica.

La prudencia aconseja no preservar la información en bases de datos sino hacerlo en archivos sencillos, por capas temáticas, y con un modelo de datos plano, aunque sea a costa de perder alguna de las facilidades que ofrece una base de datos para la implementación de modelos de datos complejos (GeoMAPP, 2011a). Un ejemplo de esto último, y que sirve además para esbozar el problema de la preservación en un entorno de actualizaciones continuas, es el problema de preservar la información en un sistema que almacene la historia de todas las acciones —altas, bajas y modificaciones— realizadas sobre los objetos mediante atributos de temporalidad y versionado de los objetos. Claramente, el entorno ideal para la implementación de dicho sistema y de las aplicaciones para manejarlo es una base de datos.

No obstante, a la hora de preservar la información con los criterios de sencillez apuntados antes, tomar instantáneas (*snapshots*) de la base de datos en archivos planos no es suficiente porque se pierde la capacidad de reconstruir el pasado entre ellos. Una solución a analizar con detalle es complementar dichas instantáneas guardando también las transacciones de actualización entre ellas.

Metadatos de preservación para la información geográfica

Como se ha dicho, los metadatos ISO 19115 para la información geográfica no contemplan elementos de preservación OAIS, especialmente los de descripción y representación. A. Shaon y A. Woolf (Shaon, 2011) han propuesto un perfil del estándar ISO19115 que incorpora metadatos de representación, procedencia, identificadores, acciones de preservación realizadas, información para la autenticación, anotaciones, etc. En el mismo sentido, en ISO 19165, la información descriptiva de la preservación OAIS se modela como una especialización del estándar de metadatos ISO 19115-1.

Otro ejemplo de metadatos geoespaciales basadas en OAIS son los del proyecto GeoMAPP. La publicación (GeoMAPP, 2011b) describe los elementos geoespaciales añadidos a los otros metadatos de preservación del modelo OAIS, mientras que la referencia (GeoMAPP, 2011c) muestra los elementos del estándar de metadatos geográficas del *Federal Geographic Data Committee* (FGDC) y del ISO 19115 que pueden aportar información útil para la preservación a largo plazo. Hay que resaltar que, al no existir estándares de implementación, todos los ejemplos acabados de mencionar pueden decir que cumplen OAIS cuando, evidentemente, las implementaciones tecnológicas son diferentes.

Paquetes de información geográfica

En ausencia de una norma de implementación concreta, deben definirse soluciones particulares para la composición de los paquetes de preservación. Por ejemplo, los paquetes geo-SIP y geo-DIP del proyecto suizo Ellipse contienen los datos, sus

metadatos y las del sistema de archivado, el modelo UML y otra documentación (por ejemplo, una visualización en forma de mapa). Ellipse es una implementación OAIS.

Ante la falta de un estándar para empaquetar físicamente la información geográfica y sus metadatos, es corriente hacerlo en archivos con formatos ZIP o TAR. Otros formatos geográficos como KMZ, MEF (relacionado con el producto GeoNetwork), y LPK y MPK (relacionados ambos con los productos ESRI), son formatos más avanzados y ofrecen posibilidades interesantes para la preservación (GeoMAPP, 2011d) basadas en su estructura interna de directorios. ISO19165 propone el formato *Open Packaging Convention* (ISO 29500-2) como estándar para el empaquetamiento de la información geográfica y de los metadatos y elementos descriptivos.

La comunidad designada y la selección de la información a preservar

A la vista de la complejidad y características de la información geográfica, parece adecuado que la “comunidad designada” sea la directamente involucrada en su uso profesional, entendiéndose que será dicha comunidad la que, en su momento, decidirá sobre la oportunidad de generar productos para otros sectores de mercado como, por ejemplo, las aplicaciones de consumo.

Con respecto a la información a preservar, parece obvio que ésta debe incluir la mínima indispensable para la gestión de un país o región. La selección debe incluir imágenes aéreas por cuanto constituyen la base para la interpretación del territorio. Dichas imágenes pueden preservarse con los estándares de Observación de la Tierra del anexo B del documento (ESA, 2012), que detalla la información y metadatos a preservar de imágenes ópticas aéreas y de sensores hiperespectrales, lidar y radar.

Aspectos institucionales, organizativos y de sostenibilidad

Es frecuente citar el hecho de que el 80% de los datos generados por las administraciones públicas tienen una componente espacial. Ello supone la existencia de innumerables instituciones en todos los ámbitos de las administraciones cuya producción de información geográfica está, de alguna forma, contemplada en la legislación. No obstante, dicha legislación acostumbra a tratar la preservación de forma genérica, sin precisar el alcance, las responsabilidades y las competencias de las instituciones respecto la preservación a largo plazo de la información que producen. Por ejemplo, para establecer de forma rutinaria la aportación de información geográfica a las instituciones de Archivo.

Por otra parte, las Infraestructuras de Datos Espaciales son una realidad tangible en muchos países que culminan iniciativas de ordenación de la información geográfica. Parecería, pues, ventajoso, aprovechar su existencia para auspiciar iniciativas

que fomentaran la colaboración entre las instituciones geográficas y las especializadas en preservación. La realización de estudios y proyectos piloto conjuntos se antoja como una buena forma de iniciar dicha colaboración.

Desde el punto de vista tecnológico, OAIS establece que los Archivos pueden interaccionar de formas diferentes: independientes unos de otros, de forma cooperativa adoptando estándares comunes (por ejemplo, definiendo un mismo SIP), o bien de forma federada compartiendo soluciones y herramientas. En el modelo federado, cada Archivo continúa definiendo y sirviendo a su propia comunidad designada, pero se introduce el concepto de comunidad designada ampliada para tener en cuenta una comunidad de usuarios más extensa.

Costes

El análisis costo/beneficio del proyecto GeoMapp (GeoMAPP, 2011e) indica que los costes operacionales de preservar 40TB de información durante 10 años se acerca a los 1,800 dólares por Terabyte cada año. En el estudio citado se tienen cuenta los costes recurrentes y no recurrentes, el personal, los sistemas, el manejo de datos, etc., representando el coste de personal casi el 90% del coste total.

Por otra parte, a partir del estudio durante cinco años de los costes de tres proyectos de preservación no relacionados con datos geográficos, M. Termens (Termens, 2012) concluye que el coste del soporte físico es poco relevante comparado con los costes de personal. Dicha constatación revela la necesidad de minimizar las operaciones manuales optimizando, por ejemplo, las manipulaciones y controles en la ingesta de los paquetes de preservación. El interés y la necesidad por estandarizar los procesos OAIS va en la línea de no tener que desarrollar procesos a medida para cada tipo de información a preservar.

Sostenibilidad

Es indudable que el tema de la preservación a largo plazo cae dentro de la categoría de acciones con beneficios intangibles, lo que la hace vulnerable a avatares económicos. Una estrategia posible para hacerla más robusta es considerar que el producto y todo su ciclo de vida forman un todo indisoluble para, así, conseguir que la preservación sea incluida automáticamente en las dotaciones presupuestarias de los planes de producción de información.

No obstante, una actitud más activa para asegurar la sostenibilidad es la propuesta por la *Blue Ribbon Task Force on Sustainable Digital Preservation and Access* (BRTF, 2010) que arguye que la sostenibilidad de la preservación no es solamente buscar más subvenciones, sino que también es construir una actividad económica basada en proposiciones de valor, incentivación a actuar, y responsabilidades claras en los roles de la preservación. El estudio concluye diciendo que, sin

estos componentes, los esfuerzos en la preservación y el material que preservan tienen pocas perspectivas de persistir en el tiempo.

Una consecuencia inmediata de dicha idea es que la implantación de la preservación a largo plazo debería contemplar un plan de negocio y una estimación coste/beneficio, tema éste de no poca dificultad en el ámbito de la información geográfica, y más cuando deben hacerse proyecciones al futuro.

Finalmente, y en misma línea de construir una actividad económica, es indispensable emprender acciones de visibilidad y notoriedad en forma de seminarios, congresos, reuniones, anuncios, y proyectos piloto. Por ejemplo, una sencilla aplicación web de comparación de cartografía o de imágenes antiguas con las modernas se convierte en un éxito instantáneo.

Conclusiones

En la medida que la preservación a largo plazo prolonga el valor de la inversión realizada en producir información, hay que fomentarla e implantarla entre las instituciones con responsabilidades sobre ella. Añadir conceptos de preservación a largo plazo en los procesos de gestión de datos de un centro de cómputo es un buen punto de partida que permite aproximarse al tema de forma progresiva. Por otra parte, la neutralidad del modelo OAIS con respecto a la implementación es una oportunidad para poner en marcha soluciones particulares sencillas que, no por sencillas, dejan por ello de ser OAIS, es decir, siguiendo un modelo de referencia estándar, aceptado y certificable.

En el caso particular de la preservación de información geográfica, los diferentes contextos, aplicaciones, productos y la complejidad de modelos de datos y de formatos, aconsejan un análisis riguroso de las experiencias, problemas, soluciones y barreras para su adopción (Locher, 2016). Adoptar estándares y buenas prácticas de preservación en los sistemas de información geográficos que manejan dicha información parece el camino correcto a seguir para fomentar la preservación a largo plazo.

Considerando los costes de producción de la información cartográfica y geográfica, la preservación debería estar contemplada explícitamente en las tareas de las instituciones productoras como una fase más en el ciclo de vida del producto, de forma que el coste asociado a un primer nivel de preservación estuviera contemplado en los costes de producción de la información.

Finalmente, archivos nacionales, cartotecas e instituciones productoras de información cartográfica y geográfica deberían coordinarse aportando cada uno su experiencia, acordando responsabilidades, e impulsando planes de negocio que aseguren la sostenibilidad presente y futura de la preservación. El estándar en desarrollo ISO19165 establecerá un marco acordado para la preservación a largo plazo de información cartográfica y geográfica digital.

Bibliografía

- ASPRS (2010). “Data Preservation and Archiving Committee Purge Alert”, American Society of Photogrammetry and Remote Sensing, <<http://www.asprs.org/Committee-General/Data-Preservation-and-Archiving-Committee-Purge-Alert.html>>
- BRTF (2010). “Sustainable Economics for a Digital Planet: Ensuring Long Term Access to Digital Information”, Blue Ribbon Task Force on Sustainable Digital Preservation and Access, <http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf>
- CCSDS (2012). “Reference Model for an Open Archival Information System (OAIS)”, The Consultative Committee for Space Data Systems, <<http://public.ccsds.org/publications/archive/650x0m2.pdf>>
- CEOS (2012). “Data Life Cycle Models and Concepts. CEOS Version 1.2”, Committee on Earth Observation Satellites, Working Group on Information Systems and Services (WGISS), <http://wgiss.ceos.org/dsig/whitepapers/Data_Lifecycle_Models_and_Concepts_v12.docx>
- DPC (2009). “Preserving Geospatial Data”, Digital Preservation Coalition Technology Watch Series Report - Report 09-01, <http://www.dpconline.org/component/docman/doc_download/363-preserving-geospatial-data-by-guy-mcgarvasteve-morris-and-gred-greg-janee>
- Ellipse (2013). “Project Ellipse - Concept for the archiving of official geodata under federal legislation-Concept Report”, Swiss Federal Archives and SwissTopo, <<http://www.swisstopo.admin.ch/internet/swisstopo/en/home/topics/geodata/geodata.archive.parsysrelated1.59693.downloadList.56737.DownloadFile.tmp/konzeptberichtellipsev1.3publikationen.pdf>>
- ESA (2012). “Long Term Data Preservation-Earth Observation Preserved Data Set Content. LTDP/PDSC”, Long Term Data Preservation Working Group, <https://earth.esa.int/documents/1656065/1681917/LTDP_PDSC_4.0.pdf>
- (2016). “Preservation Overview”, Science Data Infrastructure for Preservation-Earth Science, <<http://www.scidip-es.eu/preservation/preservation-overview/>>
- EuroSDR (2016). EuroSDR Spatial Data Research Archiving Working Group, <<http://www.eurohdr.net/archiving>>
- FGDC (2006). “Metadata Quick Guide”, Federal Geographic Data Committee, <<http://www.spatial.redlands.edu/redlandsinstitute/external/isotopiccategoriesfrommetadataquickguide.pdf>>
- GeoMAP (2011a). “Archival Challenges Associated with the Esri Personal Geodatabase and File Geodatabase Formats”, Geospatial Multistate Archive and Preservation Partnership (GeoMAPP), <http://www.geomapp.net/docs/Geodatabase_Report_v1.0_final_20111206.pdf>

- (2011b). “Archival Metadata Elements for the Preservation of Geospatial Datasets”, Geospatial Multistate Archive and Preservation Partnership, <http://www.geomapp.net/docs/GIS_OAIS_Archival_Metadata_v1.0_final_20110921.pdf>
- (2011c). “Utilizing Geospatial Metadata to Support Data Preservation Practices”, Geospatial Multistate Archive and Preservation Partnership, <http://www.geomapp.net/docs/GeoMetadata_Items_for_Preservation_2011_0110.pdf>
- (2011d). “Emerging Trends in Content Packaging for Geospatial Data”, Geospatial Multistate Archive and Preservation Partnership, <http://www.geomapp.net/docs/ContentPackaging_v1.0_final_20111202.pdf>
- (2011e). “Geoarchiving Comprehensive Cost-Benefit Analysis Guidance”, Geospatial Multistate Archive and Preservation Partnership, <http://www.geomapp.net/docs/03_Geoarchiving_Cost-Benefit_Analysis_Guidance_20111231.pdf>
- Geremew, M. Song and S. Jaja, J. (2006). “Using Scalable and Secure Web Technologies to Design a Global Digital Format Registry Prototype: Architecture, Implementation, and Testing”, Archiving Conference, <<http://www.umiacs.umd.edu/~joseph/focus-archiving06.pdf>>
- Hodge, G. (2000). “Best practices for digital archiving”, D-Lib Magazine January, vol. 6, no. 1, <<http://www.dlib.org/dlib/january00/01hodge.html>>
- ISA (2012). “Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC”, Interoperability Solutions for European Public Administrations, <https://joinup.ec.europa.eu/sites/default/files/D7.1.3%20-%20Study%20on%20persistent%20URIs_0.pdf>
- Janée, G.; Mathena, J. and Frew, J. (2008). “A Data Model and Architecture for Long Term Preservation”, National Geospatial Digital Archive, <<http://www.ngda.org/research/Tech%20Arch/jcdl-paper.pdf>>
- Locher, A. (2016). “Starting Points for Lowering the Barrier to Spatial Data Preservation”, *Journal of Map & Geography Libraries*, vol. 12, núm. 1, pp. 28-51, <<http://dx.doi.org/>>
- METS (2016). “Metadata Encoding and Transmission Standard”, The Library of the Congress, <<http://www.loc.gov/standards/mets/mets-home.html>>
- NARA (2014). “Revised Format Guidance for the Transfer of Permanent Electronic Records-Tables of File Formats”, National Archives, <<http://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html>>
- NDSA (2012). “NDSA Levels of Digital Preservation: Release Candidate One”, The Library of the Congress, National Digital Stewardship Alliance, <<https://blogs.loc.gov/digitalpreservation/2012/11/ndsas-levels-of-digital-preservation-release-candidate-one/>>

- (2013). “Issues in the Appraisal and Selection of Geospatial Data”, Geospatial Content Team of the National Digital Stewardship Alliance, <http://www.digitalpreservation.gov/documents/NDSA_AppraisalSelection_report_final102413.pdf>
- (2016). “Sustainability of Digital Formats - Planning for Library of Congress Collections”, National Digital Stewardship Alliance, <http://www.digitalpreservation.gov/formats/intro/format_eval_rel.shtml>
- PREMIS (2016). “Preservation Metadata Implementation Strategies Working Group”, Library of the Congress-Standards, <<https://www.loc.gov/standards/premis/pig.html>>
- PRONOM (2016). “The Technical Registry PRONOM”, The National Archives, <<http://www.nationalarchives.gov.uk/aboutapps/PRONOM/default.htm>>
- Röndsdorf, C. *et al.* (2016). “GI+100: Long term preservation of digital Geographic Information-16 fundamental principles agreed by National Mapping Agencies and State Archives”, EuroSDR Archiving Working Group, <http://eurosdrireland.net/archiving/GI+100%20-%2016%20EuroSDR%20Archiving%20Principles_V3%201.pdf>
- Shaon, A. and Wolf, A. (2011). “Long-term Preservation for Spatial Data Infrastructures: a Metadata Framework and Geo-portal Implementation”, *D-Lib Magazine*, vol. 17, nos. 9-10, <<http://www.dlib.org/dlib/september11/shaon/09shaon.html>>
- SIARD (2016). “Software Independent Archiving of Relational Databases (SIARD)”, Swiss Federal Archives, <<https://www.bar.admin.ch/bar/en/home/archiving/tools/siard-suite.html>>
- Spaceflight Now (2001). “Signature of life on Mars found in decades-old data”, <<http://spaceflightnow.com/news/n0107/29marslife/>>
- Térmens, M. (2012). “La sostenibilitat econòmica i tècnica dels repositoris de preservació digital”, Facultat de Biblioteconomia i Documentació, <<http://bd.ub.edu/pub/termens/docs/liligall-31.pdf>>
- TRAC (2007). “Trustworthy Repositories Audit & Certification: Criteria and Checklist”, Center for Research Libraries, <http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf>
- USGS (2012). “Archive and Records Management Fiscal Year 2012 Archive Media Trade Study”, United States Geological Survey, <<http://eros.usgs.gov/government/records/media/2012-Archive-Media-Trade-Study.pdf>>
- (2014). “USGS Guidelines for the Preservation of Digital Scientific Data”, <http://www.digitalpreservation.gov/ndsa/working_groups/documents/USGS_Guidelines_for_the_Preservation_of_Digital_Scientific_Data_Final.pdf>