

GI+100: Mantenimiento a largo plazo de la información geográfica digital. 16 principios fundamentales aceptados

Joep Cromptoets

Secretario General de EuroSRD

Traducción: Sara León Velasco

Subdirección General de Valoración e Inspección

Dirección General del Catastro

Resumen ejecutivo

Este documento establece 16 principios para la retención y mantenimiento a largo plazo de la información geográfica digital. El documento está dirigido fundamentalmente a los suministradores de información geográfica del sector público europeo (particularmente a aquellos involucrados en cartografía y catastro), con la intención de subrayar el significado de los conceptos fundamentales para el archivo de información geográfica digital. Los suministradores de información geográfica son fundamentalmente las agencias cartográficas como los institutos geográficos, así como los archivos de mantenimiento de información geográfica en relación a un rango más amplio de información digital. Como objetivo adicional, se plantea la posibilidad de que el documento pueda suministrar información

útil a los distribuidores de todo tipo de información geográfica del mundo.

Hay múltiples motivos por los que se desea mantener el acceso a la información, si bien, las principales razones relativas al archivo de información geográfica digital son:

- El cumplimiento de requerimientos legales y
- La explotación a corto y largo plazo de datos archivados para el análisis de cambios sociales, medioambientales (por ejemplo, el cambio climático), económicos, o bien, relativos a la eficiencia en la sustitución de bases de datos.

Este documento establece el camino y describe lo que se debe hacer ahora para probar en el futuro la inversión realizada

por las agencias gubernamentales mundiales en el ámbito de la creación de los Datos Geográficos. Fue aprobado por el Consejo de Administración del European Spatial Data Research Net (EuroSDR)¹, el 30 de mayo de 2013.

Los principios son:

1. El archivo de la información geográfica digital empieza en el momento de la creación del dato, en vez de en el momento de su retirada del sistema activo.
2. Todas las organizaciones deben de tener una política de mantenimiento de sus archivos.
3. Ser selectivo y decidir qué archivar y qué eliminar.
4. Considerar ventanas temporales de conservación de 1, 10 y 100 años.
5. La migración o emulación es inevitable en el medio-largo plazo. Se debe estar preparado para elegir con antelación qué propiedades mantener.
6. El volumen de producción del programa de archivo también deberá ser mantenido a largo plazo para ajustarse a los requerimientos futuros.
7. Archivar no es hacer una copia de seguridad. También deben de hacerse copias de seguridad de los archivos.
8. Los datos geográficos deben ser mantenidos de tal manera que puedan ser usados por no especialistas en la materia.
9. Los objetos de información deberán ser independientes e inteligibles por sí mismos.
10. Conservar la copia original del archivo de 100 años en archivos abiertos no en bases de datos u otros ámbitos complejos.
11. Considerar el mantenimiento de una representación gráfica junto con la representación lógica del dato.
12. Restringir el número de formatos y codificaciones a formatos reconocidos, simples y bien documentados.
13. Preferir los modelos de datos y esquemas sencillos a los complejos.
14. Mantener un sistema de acceso a los datos sencillo. Centrarse en las necesidades básicas de los usuarios comunes –un sistema de archivo visual no necesita ser un SIG en pleno rendimiento.
15. Asegurar la administración efectiva y la calidad de los metadatos asociados al dato.
16. Realizar hipótesis sobre el uso futuro pero sin ser muy restrictivo.

Introducción

Remontándonos a hace 100 años encontramos mucha información geográfica creada en ese momento y todavía accesible y utilizable hoy en formato de mapas en papel. El papel ha resultado ser un buen medio para la conservación a largo plazo si se cuida debidamente. Es sencillo de manejar y la información en formato gráfico es accesible inmediatamente. No obstante, en el mundo digital de hoy en día, la información geográfica se produce de manera predominante en formato digital que depende fuertemente de la tecnología para el almacenamiento y el acceso a los datos. Las estructuras de datos modernas, geometrías y relaciones de datos son indudablemente mucho más potentes y precisas que lo que históricamente podía dibujarse en papel. No obstante, el riesgo de perder la información puede ser mucho más alto, especialmente si no se toman las debidas precauciones. Si miramos 100 años hacia el futuro, es muy difícil decir cómo se asegurará la conservación de los datos digitales y

¹ Organización Europea que se ocupa de la investigación en temas de datos espaciales. [Www.eurosdn.net](http://www.eurosdn.net).

cómo se accederá a la información digital que generemos hoy.

La información geográfica que actualmente dibuja el mundo en el que vivimos proporciona un enorme valor al gobierno, empresas, investigadores, así como al público general. Al mismo tiempo, se accede frecuentemente a información histórica que aporta un valor añadido. Entendemos que salvaguardar los datos geográficos para las generaciones futuras, para entender no sólo la historia sino las tendencias históricas, debería ser un objetivo estratégico de las Agencias Cartográficas Nacionales, archivos y demás productores y suministradores de datos.

Si bien, nadie será capaz de predecir de forma precisa cómo será la tecnología dentro de 100 años, se puede hacer mucho para mitigar el riesgo de la pérdida de valor de las bases de datos a lo largo del tiempo. La probabilidad de que las suposiciones o predicciones, técnicas u organizativas, que se hagan en 100 años vista, no se cumplan es muy alta. No obstante, este documento pretende resaltar que es responsable realizar determinadas suposiciones hoy y documentarlas de manera concisa para permitir a las generaciones futuras entender las motivaciones y decisiones actuales. Como consecuencia, hacemos una llamada a los productores y suministradores de datos para empezar a probar sus datos hoy con miras al futuro y así crear recursos sostenibles a largo plazo.

Este documento ha sido creado por el equipo de trabajo del archivo de datos del EuroSDR de 11 Agencias Cartográficas, Archivos Nacionales e Instituciones de Investigación europeas con objeto de buscar la aprobación de una serie de principios por parte del EuroSDR y EuroGraphics así como de la comunidad de Archivos, como por ejemplo, el EBNA (*European Board of National Archivists*), EAG (*European Archives Group*) y EURBICA (*European branch of the International Council of Archives*). Es nuestro objetivo establecer unos principios

y un acercamiento común al archivo de información geográfica. La orientación futura establecida por el equipo de trabajo del archivo de datos del EuroSDR será publicada en <http://tinyurl.com/GI100>.

Glosario

Con frecuencia, existe confusión entre los distintos términos utilizados en las comunidades geográficas y de archivo. Los términos utilizados en este documento tienen lo siguientes significados:

Información geográfica

La información geográfica es²:

- Información sobre lugares de la superficie terrestre.
- Conocimiento de la localización de un elemento determinado.
- Conocimiento sobre lo que hay en una localización específica.

Archivo

El acto de transformar registros administrativos en registros de archivo para establecer las condiciones propicias para la investigación histórica y social. Los registros pueden estar constituidos por cualquier clase de información veraz creada por personas u organizaciones. Archivar incluye la entrada, evaluación, descripción, mantenimiento y provisión de la documentación y contribuye a asegurar las bases legales así como la eficiencia en la administración de datos. Los archivos históricos son responsables de mantener los registros históricos por un tiempo ilimitado. El archivo ocurre normalmente cuando el uso primario de registros ha caído bajo un determinado límite.

² Fuente Goodchild 1997.

Formato de migración

El formato de migración es el proceso por el cual los datos son migrados de un formato a otro para evitar que el primero se vuelva obsoleto.

Medio de migración

El medio de migración es el proceso por el cual los datos son migrados de un medio de archivo a otro (p. ej: de CD a DVD o de CD a cinta).

Principios para el archivo de la información geográfica digital

El equipo de trabajo de archivo de datos de EuroSDR ha identificado y acordado una serie de conceptos prácticos fundamentales y principios relacionados con el archivo de la información geográfica (IG). Éstos están enumerados en este capítulo y pretenden establecer los fundamentos para que los suministradores de información geográfica del sector público y los Archivos europeos conserven la información geográfica catastral para su uso en el futuro. Cabe mencionar que, por mandato legal puede que estos principios no sean seguidos, además, las organizaciones podrán verse obligadas legalmente a separarse de los mismos.

Fuera del dominio de la IG existen ya una serie de conceptos más genéricos y exhaustivos del ciclo vital de los archivos. Estos conceptos han sido fuertemente influenciados por el Sistema de Archivo Abierto de Información (OAIS, ISO 14721). Por ejemplo, el Modelo del ciclo de conservación y mantenimiento de la DDC ha sido diseñado para facilitar una aproximación a la gestión de los materiales digitales de un archivo y, posibilitar de esta manera,

su conservación y mantenimiento, contemplando una selección inicial para la reutilización y el mantenimiento a largo plazo. (Higgins 2008).

El orden de los principios sigue el ciclo vital de los datos, desde el diseño y creación hasta el mantenimiento y conservación (incorporación, anotación, migración), así como el acceso a los datos archivados. Los puntos de acción sugeridos y las preguntas en las que se pueda requerir una mayor investigación se indican con el siguiente símbolo: *

Principio 1: El archivo de la información geográfica digital empieza en el momento de la creación del dato, en vez de en el momento de su retirada del sistema activo

El archivo de una base de datos con frecuencia se contempla como un tema secundario, sin embargo, el valor a largo plazo de una base de datos debería evaluarse desde un principio. Si esta evaluación ocurre, los requisitos del archivo están claros desde el inicio para poder actuar en consecuencia.

- *– Consultar al principio con los responsables y expertos en la conservación de datos digitales a largo plazo. Por ejemplo: archivos, librerías o centros de datos, dependiendo de la legislación nacional.
- *– Definir si la conservación a largo plazo es deseable o necesaria.
- *– Determinar y documentar el periodo de retención. Éste puede ser cambiado con posterioridad si cambian los requerimientos pero se clarifican las necesidades de archivo desde el momento inicial. Esta evaluación debería realizarse para todas las bases de datos existentes.

Principio 2: Todas las organizaciones deben tener una política de mantenimiento de archivos

- *– El eje de cualquier caso de archivo es el establecimiento y acuerdo de un plan de conservación común así como de unos objetivos comunes entre los generadores de datos y los que archivan.
- *– Al elaborar un archivo se requiere tener una visión global. Se recomienda la consulta de un alto número de expertos para formular un plan de conservación eficiente. Usar un vocabulario común y un modelo de referencia (como el modelo OAIS) mejorará la claridad y la comprensión. Uno de los objetivos clave de la estrategia de archivo/conservación es la mitigación de riesgos contra la pérdida y la corrupción.
- *– Los objetivos de conservación de un archivo deben quedar definidos y archivados en su correspondiente política de archivo. La política debe contemplar los requerimientos tanto de los suministradores de datos como de los futuros usuarios.
- *– Se debe establecer un buen régimen de dirección de los productores de datos y de los que archivan para asegurar que la política esté correctamente implementada en el futuro.
- *– Los archivos deben ser capaces de evolucionar –el archivero no debe ser rígido en la forma de pensar o de proceder.

Principio 3: Ser selectivo y decidir qué archivar y qué eliminar

Archivar es tanto un asunto económico como un reto técnico. Los beneficios a largo plazo pueden ser intangibles, así que se recomienda focalizarse en beneficios a corto y medio plazo. El archivo a largo plazo será

menos desafiante si se consideran, preparan y ejecutan bien acciones a medio plazo. El ratio de supervivencia de los datos será mejor si se selecciona menos material pero se archiva correctamente que si se archiva de manera deficiente una mayor cantidad de material.

- *– El archivo debe definir el periodo de retención previsto para cada base de datos individual, producto o grupo característico. Así mismo, el archivo deberá conservar la documentación que haya sido seleccionada para archivar junto con el motivo del archivo. Esto significa que deberá existir una explicación de por qué determinados aspectos de una base de datos son más importantes en el corto y/o largo plazo.
- *– Cuando los datos temáticos o de referencia provienen de diferentes bases de datos/organizaciones es mejor que sean archivados por separado. Se requiere más investigación en esta área con objeto de identificar mejores métodos para combinar y archivar de forma que estén sincronizados.
- *– Para datos que se encuentran en actualización continua se requieren consideraciones sobre la mejor manera de capturar los cambios. Son posibles cualquiera de los dos acercamientos siguientes si bien, se requiere más trabajo para entender las ventajas y desventajas para una combinación óptima:
 - Archivo de instantáneas de la base de datos. Este método es más fácil de implementar, pero puede producir exceso de información y, por lo tanto, requerir más almacenamiento. Puede también hacer más difícil de sincronizar diferentes bases de datos en una fecha posterior.

- Archivo de los cambios sólo en actualizaciones/registros. Este método minimiza el exceso de información, significando esto una disminución del espacio de almacenamiento requerido, pero no se ha probado que sea inalterable ante los cambios de software y de sistemas.

Principio 4: Considerar ventanas temporales de conservación de 1, 10 y 100 años

Considerar ventanas temporales ayudará a planificar el sistema de archivo mediante la identificación las consideraciones clave en las fases del ciclo vital de la IG. Como establece el principio1, esta planificación deberá empezar en una fase temprana.

- *– Las ventanas temporales de 1, 10 y 100 años son sólo propuestas y podrán adaptarse atendiendo a las necesidades.
 - 1 año: salvaguarda operacional, centrado en las necesidades a corto plazo, pueden ser apropiados los formatos de propiedad y soluciones especializadas.
 - 10 años: estratégico, archivo de negocio, centrado en la reutilización y el acceso a los datos. Nos referimos a éste como un Archivo Transitorio, puesto que establece la conexión entre las necesidades de los distribuidores de datos a corto plazo y las de los archivos.
 - 100 años: Archivo a largo plazo cuyo objetivo es la conservación. Se centra en la robustez contra la pérdida de datos y la corrupción, así como por su capacidad en relación al mantenimiento de datos y la migración. Los datos se guardan en archivos planos de formato abierto.
- *– Se necesitará mover datos entre archivos, con diferentes soluciones

técnicas y por lo tanto, se requerirá planificación.

El acceso a los datos del archivo de 100 años puede plantearse a través de una réplica del archivo decenal.

- *– El archivo de 100 años estará bajo la custodia de un Archivo Nacional. Para el archivo de transición se deberá decidir si la responsabilidad le corresponde a los Archivos Nacionales o bien a las Agencias Cartográficas Nacionales. La progresión de los datos en el ciclo vital de archivo queda representada en el siguiente diagrama:

Principio 5: La migración o emulación es inevitable en el medio-largo plazo. Se debe estar preparado para elegir con antelación qué propiedades mantener

La codificación y formato en la que se entrega una IG al archivo no tiene por qué ser necesariamente en la que se mantenga. Es altamente probable que los formatos originales no puedan ser mantenidos y leídos en el futuro. Por lo tanto, la capacidad de leer la información y archivar el contenido requiere una reelaboración o migración a un formato soportado. La emulación está actualmente en una fase inicial y por lo tanto, es preferible la migración como estrategia para mantener la IG accesible en un periodo más amplio de tiempo.

- *– Definir las propiedades de la información geográfica (propiedades significativas) que deben ser retenidas en el momento de la migración o la emulación. (Nestor, 2008 and 2011).
- *– No asumir que los datos archivados no se degradan –se deben considerar mecanismos para comprobar y reportar la degradación de datos.

- *– Elegir la versión apropiada para la conservación de datos (original/ digital/ análogo...).
- Documentar las elecciones y razonamientos, este tipo de información ayudará a futuros archiveros y usuarios a entender y determinar futuras versiones actualizadas de los datos.
- *– Habilitar, si es posible, anotaciones para los datos archivados y conservar dichas anotaciones junto con otra información contextual que pueda añadir valor a los datos. Las anotaciones no deberán cambiar el contenido original de los datos.
- *– Asegurar que se controlan las correcciones de datos archivados. Las correcciones no deberán contaminar la autenticidad del dato original. Pueden ser necesarios un conjunto separado de metadatos que describan los cambios y el control.

La necesidad de un control de los cambios es un ejemplo de por qué se recomienda considerar la capacidad de evolución del archivo, el archivero no debe ser rígido en su forma de pensar o de proceder. La automatización de la colaboración facilitará los esfuerzos de conservación (ver, por ejemplo, <http://www.openplanetsfoundation.org/>)

Principio 6: El volumen de producción del programa de archivo también deberá ser mantenido a largo plazo para ajustarse a los requerimientos futuros

Los documentos descriptivos de la política y plan de archivo deberán estar asociados con los datos geográficos relevantes para establecer el contexto en el que se tomaron las diferentes decisiones en el momento de archivo o previo al mismo.

Principio 7: Archivar no es hacer una copia de seguridad. También deben de hacerse copias de seguridad de los archivos

- *– Las copias de seguridad constituyen un plan de mitigación contra cualquier fallo de sistema para posibilitar la recuperación de los datos y se considera una labor estándar de informática de cualquier sistema operativo que contenga datos críticos. En cambio, el archivo tiene como objetivo el mantenimiento y acceso a información a largo plazo de una manera más eficaz.
- *– Es necesario al menos hacer dos copias de seguridad del archivo en lugares diferentes. Una de las copias de seguridad debe estar en un ámbito seguro.

Principio 8: Los datos geográficos deben ser mantenidos de tal manera que puedan ser usados por no especialistas en la materia

La posibilidad de acceso y supervivencia del dato será mayor si está estructurado de una forma conocida por los archiveros y usuarios. El formato de malla en fotografías aéreas podría ser transformado a un formato, como por ejemplo el TIFF, que es más comúnmente utilizado por archiveros y usuarios.

- *– Documentar las migraciones, formatos y estructura de modo que puedan ser comprendidas por archiveros, comisarios y usuarios.
- *– Documentar las razones por las cuales se aplican determinadas acciones de mantenimiento a los datos (por ejemplo, la migración). Este tipo de información constituye el historial de mantenimiento de la base de da-

tos y ayudará a futuros archiveros a entender y determinar los requisitos de mantenimiento de la actualización de la base de datos.

- *– Archivar cronológicamente de forma independiente versiones de geodatos temáticos (datos climáticos o medio-ambientales) y datos de georreferenciación (ortofotos).
- *– Archivar especificaciones, definiciones de sistemas coordinados y materiales anecdóticos que puedan ayudar a interpretar y comprender los datos en un futuro.

Principio 9: Los objetos de información deberán ser independientes e inteligibles por sí mismos

Cuanto más descriptivo sea el objeto de información, más fácil será de interpretar a largo plazo.

Principio 10: Conservar la copia original del archivo de 100 años en archivos abiertos no en bases de datos u otros ámbitos complejos

Un repositorio abierto posibilita una utilización más eficiente de los datos (especialmente, si se usan archivos de lectura automática), y apoya las futuras migraciones. Este acercamiento también facilita un acceso a los datos mejor, o al menos más simple, si bien hay que ser conscientes de que puede implicar cierta pérdida de información.

Principio 11: Considerar el mantenimiento de una representación gráfica junto con la representación lógica del dato

Cuando la representación de los datos no pueda ser usada fácilmente por personas

que no sean especialistas, se recomienda mantener una imagen ráster o un conjunto de vectores autodiseñados de toda la extensión geográfica de la base de datos.

Si esto no es posible, conviene mantener un ejemplo de representación del área tipo para conservar los datos como fueron trasladados.

Principio 12: Restringir el número de formatos y codificaciones a formatos reconocidos, simples y bien documentados

En el futuro, se requerirá menos esfuerzo para migrar o emular datos, puesto que deberán soportarse un menor número de formatos. Se espera que la industria llegue a un consenso sobre los formatos más comunes en el campo del mantenimiento de la IG.

- *– Establecer estándares y buenas prácticas.
- *– En el archivo de 100 años, no utilizar codificaciones binarias con la posible excepción de los datos ráster.

Cualquier formato comprimido deberá ser abierto, documentado y usado de forma generalizada.

Principio 13: Preferir los modelos de datos y esquemas sencillos a los complejos

Un número alto de los modelos de IG de hoy en día son bastante complejos puesto que han sido diseñados para soportar los requisitos modernos como, por ejemplo, expresar atribuciones y relaciones entre entidades.

- *– Una base de datos altamente normalizada, en la que todas las propiedades del objeto estén distribuidas

y requieran una conexión para su uso, puede beneficiar a corto plazo los objetivos de acceso y uso de datos pero, sin embargo, no resulta útil para el mantenimiento a largo plazo, dado que hay un alto riesgo de que las conexiones entre las propiedades puedan perderse a lo largo del tiempo.

- *- Si bien puede ser interesante mantener las propiedades, el equilibrio entre la conservación de la riqueza y la pérdida de datos por contar con modelos excesivamente difíciles de interpretar, deberá ser una elección individualizada.
- *- Considerar, siempre que sea viable, la migración de datos complejos a un modelo de datos más sencillo en el momento de la incorporación en el archivo de los 100 años. En determinadas circunstancias (como por ejemplo formatos obsoletos) esto puede iniciarse como parte del proceso de archivo decenal.

Principio 14: Mantener un sistema de acceso a los datos sencillo. Centrarse en las necesidades básicas de los usuarios comunes - un sistema de archivo visual no necesita ser un SIG en pleno funcionamiento

Un sistema que es sencillo de operar, mantener, migrar y emular será más beneficioso a largo plazo que un sistema con muchas funcionalidades.

- *- El archivo de 100 años sólo debe incluir un acceso básico mientras que, los entornos basados en requerimientos a corto plazo, pueden contar con más funcionalidades.
- *- Decidir el tipo de nivel de servicio de acceso, si es necesario, basado en el contexto (estático, estable, etc.).

Principio 15: Asegurar la administración efectiva y la calidad de los metadatos asociados al dato

- *- Los metadatos almacenados en el archivo deben ser semántica y sintácticamente válidos. Por ejemplo, los registros de metadatos basados en XML pueden validarse contra su esquema XML correspondiente para asegurar la validez estructural. La validación semántica es más compleja y puede implicar el uso de un vocabulario específico definido por el archivo, preferiblemente a través de la colaboración con el usuario.
- *- Definir los tipos de metadatos necesarios para posibilitar un descubrimiento eficiente, una traslación precisa, una comprensión y reutilización continuada, así como el mantenimiento efectivo de los datos a largo plazo.
- *- Utilizar estándares y formatos de metadatos apropiados y ampliamente utilizados.
- *- Aplicar mecanismos eficientes y apropiados para manejar los cambios hechos a lo largo del tiempo a los metadatos.
- *- Considerar la habilitación de usuarios para anotar los metadatos en el archivo con objeto de facilitar la mejora de los mismos.
- *- Definir una serie de principios que constituyan la guía operativa para el mantenimiento y conservación de los metadatos. Las políticas de mantenimiento y conservación de los metadatos normalmente será políticas subsidiarias de las de mantenimiento del archivo de datos. Debe de haber claras referencias a las reglas de mantenimiento y archivo relativas a aspectos legales y a otros temas que afecten el mantenimiento y conservación de los metadatos. Todos estos

elementos se regirán por una política de archivo de datos central.

Principio 16: Realizar hipótesis sobre el uso futuro pero sin ser muy restrictivo

Es mejor hacer suposiciones sobre el uso futuro y documentarlas, que intentar concebir una previsión detallada. La expedición de directrices ayudará a focalizarse y a limitar los esfuerzos de conservación. El cómo y de qué manera estas suposiciones se realizan debe ser una valoración individual.

Entender y documentar, si es posible, el nivel de conocimiento, experiencia técnica y otras prácticas de los usuarios que podrían impactar en la forma en que los datos mantenidos son encontrados, accedidos y utilizados. El archivo requiere monitorizar los cambios de los usuarios en el tiempo. Cualquier cambio sustancial deberá ser analizado y gestionado por las medidas de mantenimiento aplicadas a los datos archivados.

- *– Identificar los consumidores potenciales (humanos, software...) a los que el dato mantenido les será beneficioso en términos de interpretación y utilización, y documentar dichas suposiciones.

Conclusiones

Estos 16 principios son el resultado de las discusiones entre las agencias cartográficas nacionales y archiveros de diferentes contextos legislativos e instituciones del ámbito europeo. La facilidad con la que el grupo discutió y acordó estos principios (en cinco reuniones a lo largo de dos años) apoya la idea de que hay numerosos aspectos comunes en los diversos acercamientos nacionales al conjunto de retos que se les presentan. Quizá, el mayor reto es el conseguir un criterio común de los expertos interdisciplinarios de los diferentes países con objeto de facilitar el perfeccionamiento del consenso de buenas prácticas, la comprensión de las diferentes perspectivas y la capacidad de aprender de contextos más amplios. Definitivamente, el grupo defiende un acercamiento común frente a la publicación unilateral de las políticas de archivo. De hecho, dado que uno de los retos clave es el acceso a la información protegida en ficheros de formato digital a largo plazo, la creación de los archivos geográficos digitales basada en un conjunto de principios comunes, es probablemente el mejor medio para garantizar que nuestros datos más preciados se mantendrán accesibles en los años que quedan por venir.