

Los usuarios y la calidad de la información geográfica

Irene Lorenzo García*

Recibido el 15 de mayo de 2014; aceptado el 9 de julio de 2014

Abstract

This document presents, based on studies and articles published in the last eight years, the causes which explains the current increase of the inadequate use of spatial data that is given in the industry, and how the position of the users with respect to the information data quality is a key factor in this problem. It is also proposed how the development of Spatial Data Infrastructure (SDI) and the application of related European policies (INSPIRE) promote a substantial improvement in the communication of spatial data quality information (data and metadata) and explains how the implementation of a risk analysis (RA) should improve the process, reducing the problems arising from a wrong interpretation of the information data quality.

Key words: *risk analysis, spatial data quality, SDI, geographic information, producers, users.*

Resumo

O presente documento apresenta, com base em estudos e artigos publicados nos últimos anos, as causas que originaram o atual incremento ao uso inadequado de dados espaciais na indústria, e como a postura dos usuários com respeito a informação de qualidade destes dados é um fator determinante neste problema. Além disso, se mostra como a implantação das Infraestruturas de Dados Espaciais (IDE) acompanhadas de políticas europeias como a diretiva INSPIRE proporcionam uma melhora substancial na transferência de dados e metadados que informam a qualidade da informação geográfica (IG) e discute como a implementação de uma análise de riscos (AR), favoreceria uma melhora no processo de tomada de decisões e reduziria os problemas derivados de uma interpretação errônea da informação de qualidade de dados.

* Universidad Politécnica de Madrid, correo electrónico: irene.lrnz@gmail.com

Palabras clave: *análises de riscos, qualidade de dados espaciais, IDE, informação geográfica, produtores, usuários.*

Resumen

El presente documento presenta, con base en estudios y artículos publicados en los últimos años, las causas que originan el actual incremento del uso inadecuado de datos espaciales que se da en la industria, y cómo la postura de los usuarios con respecto a la información de la calidad de estos datos es un factor determinante en este problema. Así mismo, expone como la implantación de las Infraestructuras de Datos Espaciales (IDE) acompañadas de políticas a nivel europeo, como la normativa INSPIRE proporcionan una mejora sustancial en la transferencia de datos y metadatos que informan de la calidad de la información geográfica (IG) y plantea cómo la implementación de un análisis de riesgos (AR) favorecería una mejora en el proceso de toma de decisiones y reduciría los problemas derivados de una interpretación errónea de la información de calidad de los datos.

Palabras clave: *análisis de riesgos, calidad de datos espaciales, IDE, información geográfica, produtores, usuários.*

Introducción

El uso de datos espaciales ha cambiado significativamente en los últimos años. Los datos espaciales se usan ahora en un creciente número de organizaciones y en numerosas aplicaciones. El acceso gratuito o a costes muy bajos a través de internet ha cambiado el modo de compartir la IG, permitiendo un intercambio cada vez más eficiente y anónimo, dando lugar a largas cadenas desde el productor al usuario final. Al mismo tiempo, la mejora de los sistemas de información geográfica (SIG) hacia sistemas más amigables y accesibles al usuario ha favorecido una mayor penetración de la información geográfica en los procesos de toma de decisiones. Sin embargo, aunque un SIG pueda obtener IG de servidores dispersos, para poder compartirla son necesarios acuerdos entre los propietarios y programas capaces de mostrar dicha información.

El Consejo Superior Geógrafo, órgano director del Sistema Cartográfico Nacional de España, indica que una infraestructura de datos espaciales (IDE) es un sistema informático integrado por un conjunto de recursos (catálogos, servidores, programas, aplicaciones, páginas web, etcétera) que permite el acceso y la gestión de conjuntos de datos y servicios geográficos (descritos a través de sus metadatos), disponibles en internet, cumpliendo una serie de normas, estándares y especificaciones que regulan y garantizan la interoperabilidad de la IG. Así mismo, es necesario establecer un marco legal que asegure que los datos producidos por las instituciones

serán compartidos por toda la administración y que potencie el uso de los mismos por parte de los ciudadanos.

La puesta en práctica de un proyecto IDE se materializa a través de un geoportal que ofrezca como mínimo la visualización de los datos a través de servicios web, la búsqueda de los conjuntos de datos y servicios por medio de sus metadatos y la localización en un mapa mediante un nombre geográfico.

Generalmente, una aplicación IDE implica múltiples productores, usuarios y productos de datos geográficos. Mientras que una IDE nacional conceptualmente es como una institución (con sus normas y procedimientos), su funcionamiento parece similar a una federación, Harvey, F. (2003); (citado en Gebresilassie, M.A.; Ivánová, I. y Morales, J., 2013). Sin embargo, la IDE europea, también conocida como Infrastructure for Spatial Information in Europe (INSPIRE), utiliza un enfoque de IDE más centrada en la creación de una comunidad, actuando como un enorme directorio activo que une a los metadatos, datos y personas, Bernard, L.; Kanellopoulos, I.; Annoni, A. y Smits, P. (2005) (citado en Gebresilassie, M.A.; Ivánová, I. y Morales, J., 2013).

Las implementaciones de infraestructuras de datos espaciales (IDE's) están poniendo en auge la figura de los geoportales frente a los datos. El acceso directo a los datos está perdiendo importancia y la atención se enfoca en los metadatos y servicios de representación de la IG.

Por otro lado, los perfiles de usuario pueden ser de utilidad para comprender y entender mejor al usuario, según Ivánová, I.; Morales, J.; de By, R.A.; Beshe, T.S. y Gebresilassie, M.A. (2013), con base en sus exigencias en los datos espaciales en cuanto a calidad, los usuarios podrían clasificarse en dos grandes grupos: expertos SIG y usuarios ingenuos.

Los estudios realizados por Van Oort, P.A.J.; Bregt A.K. y Bruin de, S. (2006) han permitido conocer las diferentes reacciones de los usuarios ante la Calidad de los Datos Espaciales, Spatial Data Quality (SQD). Una parte considera que los datos son correctos antes de utilizarlos, otros eligen el mejor entre varios conjuntos de datos y una tercera vertiente usa los datos independientemente de su idoneidad, bien porque se ven obligados a usarlos o porque directamente optan por ignorar la SDQ.

La toma de contacto con los problemas de los usuarios mediante encuestas o la asistencia a conferencias puede ayudar a superar estos problemas; sin embargo, la baja tasa de respuesta e implicación en estos estudios según Boin, A.T.; Hunter, G.J. (2006) y Van Oort, P.A.J.; Bregt A.K. y Bruin de, S. (2006) hace que los resultados obtenidos no puedan considerarse concluyentes o siquiera representativos de los problemas de la comunidad de usuarios.

Parámetros de la Calidad de Datos Espaciales (SDQ)

Tanto en Europa como en España la normalización de la información geográfica digital de las IDE se realiza mediante los organismos de normalización internacio-

nal, Internacional Organization for Standardization (ISO), y europeo, European Comité for Standardization (CEN). Sus comités técnicos son los encargados de generar las familias de normas: el comité internacional ISO/TC211-Geographic Information, cuyos trabajos de normalización dan como resultado la familia de normas ISO 19100, y el comité europeo de normalización EN CEN/TC 287, que adopta la serie ISO 19100 como normativa europea y desarrolla nuevas normas y perfiles en cooperación con ISO/TC211. En el contexto español, la colaboración con los organismos europeos e internacional de normalización se realiza a través del comité técnico AEN/CTN 148 de la Asociación Española de Normalización y Certificación (AENOR). En cuanto a América Latina, el Instituto Geográfico Nacional (IGN) en colaboración con los geoinstitutos iberoamericanos están desarrollando desde el año 2009 una línea de trabajo en cuanto a armonización de la terminología técnica en español en el campo de la información geográfica. Se canaliza a través de una acción de la Red Iberoamericana de Infraestructuras de Información Geográfica (RedI3Geo), que tiene como objetivo armonizar las terminologías utilizadas en las traducciones de las normas ISO 19100 adoptadas como normas nacionales en los países iberoamericanos.

Para conocer la calidad de los datos espaciales, es esencial que los parámetros de calidad se expresen de manera estandarizada y que los términos utilizados abarquen los aspectos que la afectan.

La norma ISO 19157, que sustituye a las normas ISO 19113, 19114 y 19138, proporciona los siguientes principios para describir la SDQ, definiéndose en ella importantes conceptos de calidad (Ariza López, F.J.; García Balboa, J.L. y Amor Pulido R., 2004 e ISO/FDIS 19157, 2013):

Integridad: presencia o ausencia de características, sus atributos y relaciones. Se compone de dos elementos:

- *Comisión*: exceso de datos presentes.
- *Omisión*: datos ausentes.

Consistencia lógica: grado de corrección o conformidad de un conjunto de datos geográficos con respecto a la estructura interna descrita por sus especificaciones. Compuesta de cuatro elementos:

- *Coherencia conceptual*: observación de las reglas del esquema conceptual.
- *Consistencia del dominio*: cumplimiento de los valores de los dominios.
- *Consistencia del formato*: grado en que los datos se almacenan de acuerdo con la estructura física.

- *Consistencia topológica*: corrección de las características topológicas codificadas de manera explícita.

Exactitud posicional: grado de fiabilidad del procedimiento de adquisición de los atributos posicionales del dato geográfico. Parámetro de calidad que describe la exactitud (X, Y, Z) de las posiciones geográficas en un conjunto de datos geográficos. Formada por tres elementos:

- *Precisión absoluta o externa*: cercanía del conjunto de valores de coordenadas a los valores tomados como verdad.
- *Exactitud interna*: cercanía de las características de las posiciones relativas en un conjunto de datos, a sus respectivas posiciones relativas tomadas como verdad.
- *Precisión posicional de los datos reticulados*: cercanía de los valores de posición de los datos reticulares espaciales a los valores tomados como verdad.

Exactitud de atributos o temática: grado de fiabilidad del procedimiento de adquisición de los atributos temáticos del dato geográfico. Parámetro de la calidad que describe la exactitud de los aspectos semánticos de un conjunto de datos geográficos. Se compone de tres elementos:

- *Clasificación por corrección*: comparación de las clases asignadas a las funciones o de sus atributos en un universo de discurso.
- *Corrección atributo no cuantitativo*: medida de si un atributo no cuantitativo es correcta o incorrecta.
- *Exactitud atributo cuantitativo*: cercanía del valor de un atributo cuantitativo a un valor aceptado como conocido de ser cierto.

Calidad temporal: calidad de los atributos temporales y características de las relaciones temporales. Compuesta de tres elementos:

- *Exactitud de una medición de tiempo*: cercanía de las mediciones temporales a los valores aceptados como verdaderos.
- *Consistencia temporal*: corrección de la orden de los eventos.
- *Validez temporal*: validez de los datos con respecto al tiempo.

Linaje: registro de las características, de las fuentes de datos, de las operaciones, que han creado el producto final, con el fin de poder reconstruir la historia de un producto o servicio.

Usabilidad: se basa en los requisitos del usuario. Todos los elementos de calidad pueden ser usados para evaluar la facilidad de uso. La evaluación de usabilidad se puede basar en las necesidades específicas de los usuarios que no se pueden describir utilizando los elementos de calidad descritos anteriormente. En este caso, el elemento de usabilidad se utiliza para describir la información específica de calidad sobre la idoneidad de un conjunto de datos para una aplicación en particular o la conformidad con una serie de requisitos.

Disponibilidad, intercambio y uso de la IG

Cualquier usuario dentro de una organización tiene acceso a un servidor interno donde se alojan los datos espaciales. El problema radica en que no se tiene noción alguna de quién está realmente utilizando un conjunto de datos determinados. Desde varios puntos de vista, esta práctica puede no ser la más adecuada; por un lado, los productores no están bien informados acerca de las necesidades de los usuarios, por otro lado, los usuarios desconocen la calidad de los datos y la experiencia de otros usuarios.

Sirva como muestra de los problemas que pueden surgir, el siguiente ejemplo de cadena de geodatos propuesto por Masser, Omran, Eten y Van (2007), citado en Van Oort, P.A.J. (2008).

Uno o más topógrafos u operadores fotogramétricos producen un conjunto de datos espaciales. Estas personas proporcionan el conjunto de datos a su distribuidor. Posiblemente, pero no necesariamente, el distribuidor de datos pide a los productores acceder a los datos de registro. El receptor no usa necesariamente el conjunto de datos él mismo, sino que puede tratarse de un gestor que pone el conjunto de datos a disposición de los usuarios de su organización mediante un servidor interno. El propio servidor contiene decenas o cientos de bases de datos; del mismo modo, a él pueden acceder decenas o cientos de personas pertenecientes a la organización. Ante esto, es imposible conocer qué persona está accediendo a un conjunto de datos en particular.

Una organización de gran tamaño puede a su vez estructurarse en departamentos subsidiarios con su gestor de datos interno. Dentro de este departamento un subconjunto de datos se utiliza en un proyecto por un equipo de expertos en SIG, transmitiéndose entre los miembros del mismo y llevando a cabo cada uno sus propias operaciones SIG en particular sobre el conjunto de datos.

Posiblemente en esta etapa diferentes conjuntos de datos espaciales se combinan fusionando un conjunto de cadenas de datos en una nueva cadena única. Basado en las salidas que el GIS va a proporcionar, el jefe de equipo escribe el informe de recomendaciones técnicas de los datos; sin embargo, no ha manejado realmente el conjunto de datos, por lo que redacta las recomendaciones de los datos sin conocer el origen y calidad de los mismos. Como consecuencia, el responsable final, que se basa en el informe anterior para la toma de decisiones, probablemente no sea

consciente de las características técnicas de los datos originales utilizados para el análisis.

Cadenas como la de este ejemplo, forman en su conjunto una red de nodos y enlaces. Los nodos pueden ser personas, organizaciones de usuarios y sitios web. Administradores de datos y geoportales son centros de la red, con muchos enlaces salientes. Cada nodo es parte de por lo menos una red de datos geográficos (enlace: productor -> usuario) y también puede ser parte de la red de personas (usuario -> productor y del usuario -> usuario).

La red de datos geográficos se utiliza para transferir datos desde el productor a los usuarios, estando los enlaces dirigidos sólo en esa dirección. Como consecuencia de esta configuración, es casi imposible encontrar a los usuarios, mientras que es muy fácil encontrar los centros y el productor de datos. Las características de estas redes difieren en función de si los datos son compartidos de modo completamente anónimo a través de Internet entre usuarios de todo el mundo, o bien de forma menos anónima, como es el caso de los paquetes de datos comerciales, con la comunidad de usuarios circunscrita a los límites de un país.

Potencialmente, este tipo de red tiene serias desventajas (Van Oort, P.A.J., 2008):

- Para los productores es imposible conocer y responder a las necesidades de los usuarios.
- Los gestores de datos desconocen si debe ordenarse una nueva actualización de un conjunto de datos, pues les es imposible saber el número de usuarios dentro de la organización que están utilizando los datos.
- Cuando un gestor de datos no puede saber si un conjunto de datos está realmente en uso, la producción de nueva información puede estar en riesgo.
- Los usuarios finales no pueden compartir entre ellos ideas interesantes, innovaciones y preocupaciones.

Van Oort, P.A.J. (2008) presenta como una alternativa para mitigar las desventajas anteriores los enlaces entre usuarios pertenecientes a una red de personas que trabajan con información geográfica geo-people network. La gente colabora a la vez que trabaja con su conjunto de datos, y si es necesario pueden contactar con el productor para solicitar información.

De este modo, los usuarios y productores tienen conocimiento de la calidad de sus conjuntos de datos. Al trabajar con los datos, los usuarios pueden identificar errores o elementos que no están incluidos dentro de los metadatos proporcionados por los productores, facilitando de este modo la aparición de innovaciones para hacer frente a estos errores. La comunicación al productor y otros usuarios de los errores localizados y actuaciones planteadas para su resolución beneficia a todas las partes.

Usuarios desconocedores de la calidad de la IG

Tipos de usuario

No existe un único perfil de usuario, sino que existen diferentes roles en cuanto a los consumidores que trabajan con información geográfica. En ocasiones, es posible que una misma persona desempeñe diferentes roles. Van Oort, P.A.J. (2008) define cuatro tipologías:

- *Intermediario*: gestor de datos, comprador/vendedor.
- *GIS/usuario gráfico*: trabaja directamente con el conjunto de datos espaciales.
- *Usuario de alto nivel*: no trabaja directamente con los datos espaciales, pero los conoce ya que se utilizan en procesos donde está implicado.
- *Ex usuario*: trabajó en algún momento con datos espaciales, pero no en la actualidad.

Así mismo, en los últimos tiempos algunas organizaciones del sector han creado un problema adicional al introducir en su estructura usuarios sin la calificación adecuada para el trabajo con IG, pero que se ven obligados a trabajar con ella. Esto implica dos consecuencias: primera, que parte del personal necesite ayuda para el acceso a datos y metadatos, y segunda, que la pérdida de personal calificado pueda provocar la pérdida de datos y metadatos al menos temporalmente, volviendo todo ello más vulnerable a la organización.

Usuarios y calidad de la IG

En las dos últimas décadas, un medio ampliamente utilizado por los productores de datos para que profesionales y usuarios conozcan y valoren la SDQ es la publicación de metadatos (Devillers *et al.*, 2010 y Devillers, R.; Stein, A.; Bédard, Y.; Chrisman, N.; Fisher, P. y Shi, W., 2010). Sin embargo, la generación de metadatos es una tarea ardua que consume tiempo y no mejora en numerosas ocasiones la capacidad del usuario para comprender los posibles usos de los datos.

Goodchild, M.F. (2008) define metadatos 2.0 como un planteamiento que recoge las experiencias de los usuarios que han intentado acceder y explotar un conjunto de datos en una serie de posibles aplicaciones. Los metadatos permiten conocer la SDQ, bien sea mediante normas ISO que definen parámetros de calidad o por medio de informes de evaluación.

La información de la calidad puede almacenarse con los datos que describe o por separado en diferente estructura. La aplicación de informes de evaluación de la calidad se utilizan para gestionar, organizar y estructurar la SDQ. La estructura y la gestión de información de calidad están directamente relacionadas con los datos espaciales que describe y su contenido. La aplicación de estos informes, bien sea en

forma agregada o en diferentes niveles de detalles, es importante para facilitar el acceso y conocer las características exactas de calidad de los datos espaciales, de una manera tal que los usuarios pueden consultarla. En cuanto a los informes, hay varios tipos (Gebresilassie, M.A.; Ivánová, I. y Morales, J., 2013):

- *Catálogo de metadatos*: es la descripción de la información de datos espaciales incluyendo la calidad de los mismos. El catálogo de registros representa un conjunto de datos con una estructura específica (McHugh, J.; Abiteboul, S.; Goldman, R.; Quass, D. y Widom, J., 1997), la información de calidad de datos espaciales es parte de los metadatos. Este tipo de estructura carece de una fuerte conexión con el propio conjunto de datos.
- *Sistema de gestión de base de datos*: almacena los datos de información de calidad con el mismo o similar esquema que los datos espaciales descritos. Es útil para conocer, recuperar y actualizar las bases de datos espaciales (BBDDEE) y su información de calidad. Devillers, R. y R. Jeansoulin (2006) asocian explícitamente información sobre la SDQ a una BBDDEE, ya que a un nivel adecuado puede mejorar el acceso eficiente a los datos. Sin embargo, para decidir sobre la idoneidad del conjunto de datos, las herramientas para descubrir y recuperar procesos destinados específicamente a la calidad no son las mismas que las de los datos espaciales.
- *Informes basados en texto*: son informes de texto separados de los datos que describen. Este tipo de descripción de la calidad de los datos espaciales representada y organizada en un archivo puede ayudar a los usuarios a obtener conocimientos resumidos sobre las características generales de datos espaciales. Sin embargo, ayuda poco en el suministro de descripciones detalladas de los conjuntos de datos espaciales y es difícil utilizarla con procesos automatizados.

En general, la vinculación de información sobre SDQ tiene enormes ventajas para los usuarios. Por ejemplo, el problema de la descripción de la calidad resumida se puede reducir. También ayuda a determinar el comportamiento de un conjunto de datos en una aplicación espacial, tanto en la fácil recuperación de la calidad como en los datos espaciales. Por lo tanto, estos diversos modelos espaciales de calidad de datos, tienen sus propias fortalezas y debilidades asociadas a la descripción de la calidad suficiente, de una manera fácil para los usuarios de diferentes orígenes y distintos niveles de entendimiento.

Expresando la calidad

Pese a que los componentes utilizados para definir la calidad detallada con anterioridad pueden ser apropiados para registrar datos como un activo, la pregunta surge cuando se plantea si son adecuados para comunicar la calidad a los consumidores

de datos espaciales. La mayoría de estos componentes se centran en las características relacionadas con la producción de los datos (calidad interna), pero habría que plantearse qué información adicional de calidad sería recomendable incluir para lograr que estos datos sean “aptos para el uso” (Boin, A.T.; Hunter, G.J., 2006 y Gebresilassie, M.A.; Ivánová, I. y Morales, J., 2013).

Se han propuesto elementos adicionales para representar la calidad, como las limitaciones del producto, ejemplos de uso de los datos e interrelaciones con otros conjuntos de datos (Boin, A.T. y Hunter, G.J., 2006). Sin embargo, existe la sensación que si bien estos componentes podrían añadirse, su inclusión incrementaría la carga de trabajo de los productores, ya de por sí elevada con los componentes actuales fijados en las normas.

Es necesario considerar como factor destacado que si se desea mejorar la comunicación de la calidad a los consumidores, se hace preciso tener en cuenta el contexto en el cual ésta se convierte en un problema. Habitualmente los usuarios comienzan a preocuparse por la calidad de los datos después de comparar dos conjuntos de datos independientes del mismo tema al solaparlos conjuntamente, pero esto plantea problemas en un entorno del mundo real, donde las restricciones técnicas pueden limitar la capacidad de superponer datos directamente.

El control de calidad: el caso concreto del proyecto IDE Cartociudad (España)

Cartociudad es la base de datos de red viaria, cartografía urbana, divisiones censal y postal de ámbito nacional, cuyos datos son oficiales y provienen de la Dirección General del Catastro (DGC), el Instituto Nacional de Estadística (INE), la Sociedad Estatal de Correos y Telégrafos S.A. (Correos), el IGN y otros organismos autonómicos. Producida por el IGN en colaboración con distintas comunidades autónomas, Cartociudad se está convirtiendo en uno de los referentes en callejeros oficiales a nivel nacional.

La explotación de los datos de Cartociudad se realiza a través de Internet empleando servicios web accesibles desde el geoportal de Cartociudad e implementados siguiendo los estándares del Open Geospatial Consortium (OGC). Estos servicios permiten llevar a cabo operaciones de visualización de información, localización de direcciones, unidades administrativas, códigos postales y secciones censales y cálculo de áreas de influencia y de enrutamiento entre direcciones de cualquier lugar de España.

Las herramientas de control de calidad de Cartociudad han sido diseñadas satisfaciendo los requerimientos y los flujos de trabajo específicos del proyecto. Así, los controles de calidad que se realizan a los nuevos datos de Cartociudad para validarlos se agrupan en tres tipos, que se aplican en tres fases diferentes (González Jiménez, A.; Rubio Iglesias, J.M.; Velasco Tirado, A.; González García, J.; Verdejo Herreras, P. y Andrés Yusá, G., 2012):

- Controles básicos de aceptación, que consisten en verificar que se cumplen las condiciones mínimas para poder incorporar los datos: que las tablas cumplan el modelo de datos, que existan datos de los municipios objeto de la revisión, que no existan duplicados, etcétera.
- Contraste de los nuevos datos con los ya existentes en la base de datos: que no se inserten fenómenos que ya existían comparándolos por su identificador, ni se traten de eliminar fenómenos que no existiesen.
- Garantizar que la base de datos consolidada (donde se hayan integrado las actualizaciones con los datos ya existentes) cumple todos los requisitos del modelo de Cartociudad en cuanto a consistencia conceptual (integridad referencial, listas controladas, etcétera) y consistencia topológica (tramos conectados, polígonos postales cerrados, etcétera).

El punto de vista de los consumidores

Boin, A.T. y Hunter, G.J. (2006) al amparo de las encuestas realizadas para su estudio, consideran que los intereses de los consumidores en la calidad tienden a estar fuertemente relacionados con temas prácticos:

- Prevalencia de consultas sobre la completitud y edad de los datos.
- Se alude a la consistencia lógica cuando se refiere a la superposición de diferentes conjuntos de datos.
- La exactitud posicional es motivo de preocupación, especialmente con relación a las dimensiones reales de parcelas.
- Exactitud de atributos.
- No se presta atención al linaje o los métodos de producción.

Sorprendentemente, el vocabulario dominante de los consumidores está más relacionado con las expectativas y fiabilidad de lo que los datos incluyen que con los elementos de calidad de los mismos. No mencionan los términos “metadatos”, “consistencia lógica” ni “linaje” e incluso la palabra “exactitud” es utilizada escasamente.

La percepción de los metadatos publicados en Internet les parece “incomprensible”; ya de por sí, el término “metadato” es un término confuso incluso para profesionales de la información geográfica, y cuando se introduce en presentaciones o textos es invariablemente seguido de la frase “esto es, datos acerca de los datos”. De modo que tal vez habría que plantearse la cuestión de por qué se continúa utilizando con consumidores no expertos, si es necesario incluir una frase adicional para explicar su significado a personas dentro de la industria.

Con base en esto, se puede argumentar que mientras que los metadatos pueden ser muy útiles para un proveedor de datos o una organización de proveedores de servicios, no hay signos que indiquen que este formato de implementación de metadatos sea útil para comunicar la calidad. Un gran número de usuarios primerizos de datos espaciales no comprenden gran parte de los términos de los estándares actuales. Más aún, parecen estar más preocupados con los problemas de calidad externos tales como: ¿qué datos espaciales están disponibles para una región en particular? y ¿qué tan completos son? También son conscientes de que los datos no son perfectos, sino lo mejor disponible. En este sentido, la calidad parece ser un concepto con un gran significado para los consumidores únicamente cuando se comparan diferentes conjuntos de datos.

SIG y SQD

Las dos causas de riesgo expuestas previamente se relacionan directamente con el *software* utilizado para el trabajo y el análisis de la información geográfica. Como ya se menciona en la introducción, el *software* SIG apenas dispone de herramientas para trabajar con calidad de datos, y la edición de sus metadatos en numerosos paquetes comerciales aún deja mucho que desear. Por otro lado, el hecho de poder trabajar con datos de diferentes formatos sin que la aplicación obligue a comprobar su calidad, facilita la interoperabilidad y la producción, pero por otro lado, en manos de un usuario inexperto puede suponer un problema.

La calidad de los datos espaciales está a menudo limitada por la calidad de sus fuentes y las operaciones espaciales realizadas sobre las bases de los datos de un SIG, como selección del sistema de referencia y producto cartesiano, no siempre trabajan correctamente debido a su exactitud y completitud, dependiente de la SDQ.

Estas preocupaciones han disminuido recientemente con la fusión de la tecnología SIG y las herramientas espaciales que permiten, extraer, transformar y cargar —Extract, Transform and Load (ETL)—, mediante la cual se mueven datos desde diferentes fuentes, se reformatean, limpian, se cargan en una BBDD y finalmente se analiza la información. Por tanto, estas herramientas hacen posible la conversión de datos, mejorando así la interoperabilidad de los mismos con distintos formatos, así como entre las distintas aplicaciones SIG, y proporcionando además la posibilidad de automatizar flujos de procesado de datos. Gracias a la incorporación de esta tecnología se abre un nuevo camino para las IDE, ya que facilita en gran medida la incorporación de multitud de datos heterogéneos en este tipo de plataformas.

Un caso de uso muy interesante, es el del Alto Comisionado de la Organización de las Naciones Unidas para Refugiados (ACNUR) (Arévalo, J.; Martínez Izquierdo, C. y Simonazzi Domínguez, W., 2009). Esta organización durante operaciones humanitarias recoge una gran cantidad de datos en muy diversos formatos, la integración de éstos se realizaba de forma manual y en muchas ocasiones no se sacaba

el máximo partido a la información. Desde el año 2008, el ACNUR ha mejorado notablemente su organización gracias al uso de la herramienta espacial ETL, pudiendo compilar varios datos en distintos formatos Excel, shapes, etcétera, que necesitan ser actualizados regularmente, almacenándolos en una única base de datos, para posteriormente publicar la información por medio de un visor cartográfico. Una de las principales características presentes en los ETL, de especial interés para el mundo SIG, es la capacidad de computación en paralelo que ofrecen, la cual se basa en la premisa de que “grandes problemas se pueden dividir en problemas más pequeños” para después ser resueltos de forma concurrente, salvando de esta forma la limitación en *hardware*, que se hacen evidentes a la hora de trabajar con grandes volúmenes de datos. Este paradigma, sin embargo, añade un mayor nivel de complejidad en el desarrollo de un ETL geoespacial, ya que existen ciertas limitaciones impuestas por los datos espaciales y los procesos SIG que se aplican a ellos que hacen complicada su ejecución en paralelo.

Productores y comunicación de información a los usuarios

Levesque, M.A.; Bédard, Y.; Gervais, M. y Devillers, R. (2007) plantean cómo desde algunos ámbitos de la industria, se expresa la necesidad de cambiar la aptitud de los productores de IG hacia los usuarios finales, en el contexto del consumo masivo de IG.

Generalmente, los productores deben mostrar cuidado y diligencia. Más específicamente, una de las labores más importantes que un productor de IG debería satisfacer es informar adecuadamente a los usuarios, si bien es verdad que la importancia de este deber varía en función del contexto, la naturaleza del producto y el conocimiento del consumidor (usuario).

Debido a la complejidad, la naturaleza técnica y el peligro potencial de la IG combinada con la posible falta de experiencia del usuario, la importancia del deber del productor de informar adecuadamente a los usuarios en el contexto actual del consumo masivo de IG sería particularmente alta.

Este deber de informar implica tanto el deber de aconsejar como de advertir. En el contexto de la IG, informar implica comunicar la información incluida en los metadatos o cualquier otra descripción de los datos que pudiese ser necesaria para comprender las características de un conjunto de datos dado, pero sin orientar la decisión del usuario.

El deber de aconsejar se caracteriza por la necesidad de juzgar el contenido de la información transmitida, con atención al tipo de información (por ejemplo, IG digital vectorial), su importancia y el uso a que está destinada.

El deber de advertir debería comunicarse para prevenir usos erróneos y permitir a los usuarios apreciar adecuadamente el valor de los datos. Las advertencias deben ser claras, completas y actualizadas.

Se plantea también que ambos casos impliquen el deber de identificar y revelar los riesgos potenciales relacionados con el uso de unos datos determinados.

Todo ello requiere por parte del productor entender las necesidades, objetivos e incluso el uso al que están destinados los datos.

Herramientas para prevenir el uso erróneo de la IG

Es necesario plantear una aproximación genérica para ayudar a los productores y consumidores a identificar los riesgos potenciales del mal uso de la información geográfica y las diferentes alternativas para hacerle frente, como la comunicación de advertencias sobre la sensibilidad de esta información a los usuarios finales. Pero previamente tal vez sea necesario aclarar qué se entiende por riesgo en el ámbito de la información geográfica.

La noción de riesgo se relaciona con la “combinación de la probabilidad de ocurrencia de un daño y el grado de gravedad del mismo”, ISO/IEC Guide 51:1999 (1999), considerando el término “daño” en el campo que nos ocupa como el uso inadecuado de los datos o la interpretación errónea de los mismos.

Partiendo de las limitaciones ya planteadas que tienen los metadatos para la transmisión de la calidad existen dos aproximaciones (*a priori* y *a posteriori*) para ayudar a los expertos y usuarios finales a evaluar la conveniencia de uso de un conjunto de datos y reducir así el riesgo de un uso incorrecto (Levesque, M.A.; Bédard, Y.; Gervais, M. y Devillers, R., 2007).

Aproximación a priori

La aproximación *a priori*, implica actuaciones antes de que los datos estén en posesión de los usuarios finales:

- Restringir el uso de conjuntos de datos a aquellos usos para los que sean apropiados, identificándolos previamente.
- Encapsular los datos con un *software* para asegurar la adecuación entre los datos y las operaciones disponibles en el producto.
- Ayudar a los usuarios a identificar el ajuste de los datos al uso destinado mejorando las herramientas existentes de selección de datos.
- Requerimiento de la opinión profesional de un experto en geomática.
- Desarrollar herramientas que puedan ayudar al experto a hacer frente a tareas complejas. Estas herramientas darían al experto la capacidad de integrar, gestionar y visualizar la calidad de los datos a diferentes niveles de granularidad.

Aproximación a posteriori

Como su propio nombre indica actúa después de que los usuarios estén en posesión de los datos:

- Mejora de las capacidades del *software* GIS para una mejor gestión y uso de la información de calidad de los datos espaciales.
- Comunicar la calidad de los datos espaciales mediante técnicas de visualización, como alternativa a los metadatos tradicionales.
- Proporcionar advertencias a los usuarios finales cuando se realicen operaciones ilógicas.
- Diseño de *software* GIS capaz de ofrecer funcionalidades básicas para manejar los errores, error-aware GIS.
- Diseño de *software* GIS con capacidades adicionales de gestionar, actualizar, explorar, evaluar y comunicar información de calidad, quality-aware GIS, tratando también cuestiones relacionadas con el contexto de los usuarios de GIS y los patrones de uso.

El análisis de riesgos

Dentro de los procesos de toma de decisiones donde es necesario recurrir a análisis espaciales, un uso inadecuado de la información geográfica, como por ejemplo datos con mala calidad, tendría consecuencias nefastas.

Van Oort, P.A.J.; Bregt A.K. y Bruin de, S. (2006) estiman que la mera decisión por parte de los responsables de llevar a cabo un análisis de riesgos (AR) que cuantifique los riesgos derivados de un mal uso de la información de calidad y analice las probabilidades de error, implicaría una clara mejora del proceso de toma de decisiones. El problema radica en que el AR raramente es adoptado o cuantificado debido a las siguientes causas (Van Oort, P.A.J.; Bregt A.K. y Bruin de, S., 2006):

- La calidad de los datos espaciales es frecuentemente ignorada por los usuarios que son reticentes, por desconocimiento, a dedicar recursos al análisis de los riesgos que un uso inadecuado de los datos implicaría.
- Falta de ejemplos prácticos que muestren necesidades y beneficios del análisis de riesgos.
- Falta de conocimientos de los usuarios para analizar los resultados del análisis de riesgos.
- Carencia de metodologías para analizar riesgos debidos a la calidad de los datos espaciales.
- Falta de herramientas en el *software* GIS actual.
- La documentación sobre calidad de la información geográfica es escasa.
- Las investigaciones iniciales estaban enfocadas a la propagación de errores en información geográfica en lugar de al análisis de riesgos.

Usuarios, percepción positiva del análisis de riesgos

Superados estos obstáculos, las encuestas realizadas (Van Oort, P.A.J.; Bregt A.K. y Bruin de, S., 2006) permiten describir tres escenarios donde el análisis de riesgos es percibido de manera positiva por todos los actores del proceso de toma de decisiones (responsables e interesados), aceptando el uso de recursos en análisis de riesgos:

- En el caso de que información acerca de la calidad de los datos fuese recibida al comienzo del proceso de toma de decisiones, fase de documentación e información (DO), las partes interesadas podrían y deberían detectar e informar a los responsables del proceso sobre los errores existentes en la información geográfica disponible, para así reducir los riesgos a un nivel dentro de unos límites aceptables.
- Cuando el objetivo de la decisión puede expresarse numéricamente y los resultados de la calidad son recibidos, ya sea en la fase de toma de decisiones (DT) de manos de corresponsables de la decisión, pues esto genera mayor confianza en la información, o bien cuando los resultados de calidad son recibidos al final del proceso de toma de decisiones en la evaluación de la decisión (DE), y así el responsable puede valorar los riesgos debidos a la calidad de los datos utilizados.
- Cuando en el proceso de toma de decisiones confluyen diferentes intereses, que pueden derivar en consecuencias externa al proceso, los usuarios también son favorables al análisis de riesgos.

Conclusiones

Es necesario sensibilizar a las partes implicadas, así como mejorar las capacidades y conocimientos de los usuarios, conjugándolo con el desarrollo de herramientas y documentación que ayuden a mejorar el trabajo en el ámbito de la calidad y a comprender y valorar las opiniones de los usuarios.

La implantación cada vez mayor de infraestructuras de datos espaciales (IDE's), junto a políticas como la normativa europea INSPIRE, va enfocada a incrementar la eficiencia y efectividad de la comunicación entre productores y usuarios. Las tecnologías de IDE's, normas y políticas están mejorando la eficiencia de la transferencia de datos y metadatos, de manera que los usuarios puedan encontrarlos y acceder a ellos de manera más sencilla, aunque la comunicación efectiva entre proveedores y usuarios sigue siendo un problema.

Es preciso mejorar en el conocimiento de los tipos de información de calidad que demandan los usuarios, cómo obtienen actualmente esta información, así como las fortalezas y debilidades de la forma en que la información de calidad es presentada y comunicada vía web para poder plantear soluciones más efectivas de comunicación.

Una parte de los usuarios desconoce o no sabe interpretar la información de calidad contenida en los metadatos. A fin de lograr que los usuarios tengan en cuenta la información de calidad de los datos a la hora de trabajar con ellos, es necesario plantear la utilización de la terminología del consumidor de los datos en lugar de recurrir a los términos demasiado técnicos de la industria, así como tratar de centrarse en la manera de describir la idoneidad y confiabilidad del producto en lugar del método de producción.

El trabajo en el ámbito de la calidad de datos espaciales implica un esfuerzo para hacer percibir a las partes implicadas la necesidad y los beneficios que el análisis de riesgos reportaría a la calidad de los datos espaciales y los procesos de toma de decisiones.

Bibliografía

- Arévalo, J.; Martínez Izquierdo, C. y W. Simonazzi Domínguez (2009). "BeETLe: Herramienta ETL geo-espacial libre", IV Jornadas de SIG Libre, Servei de Sistemes d'Informació Geogràfica y Teledetecció (Sigte), Universitat de Girona, European Topic Centre on Land Use and Spatial Information, European Environment Agency, Universidad Autónoma de Barcelona, Campus Bellaterra, Barcelona, España, p. 11.
- Ariza López, F.J.; García Balboa, J.L y R. Amor Pulido (2004). "Casos prácticos de calidad en la producción cartográfica", Servicio de Publicaciones de la Universidad de Jaén.
- Bernard, L.; Kanellopoulos, I.; Annoni, A. y P. Smits (2005). "The European Geportal - One Step Towards the Establishment of a European Spatial Data Infrastructure", *Computers, Environment and Urban Systems*, pp. 15-31.
- Boin, A.T. y G.J. Hunter (2006). "Do Spatial Data Consumers Really Understand Data Quality Information?", 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Cooperative Research Centre for Spatial Information, Department of Geomatics, University of Melbourne, edited by M. Caetano and M. Painho, pp. 215-224.
- Devillers, R. y R. Jeansoulin (2006). "Fundamentals of Spatial Data Quality (Geographical Information Systems series)", *ISTE London*, p. 310.
- Devillers, R.; Stein, A.; Bédard, Y.; Chrisman, N.; Fisher, P. y W. Shi (2010). "Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities", pp. 387-400.
- Gebresilassie, M.A.; Ivánová, I. y J. Morales (2013). "User Profiling Methods for Spatial Data Quality Models", Global Geospatial Conference 2013, Addis Ababa, Ethiopia, p. 14.
- González Jiménez, A.; Rubio Iglesias, J.M.; Velasco Tirado, A.; González García, J.; Verdejo Herreras, P. y G. Andrés Yusá (2012). "Nuevos retos del Proyecto

- Cartociudad”, X Congreso Topcart 2012 – I Congreso Iberoamericano de Geomática y C.C. de la Tierra, Centro Nacional de Información Geográfica, p. 9.
- Goodchild, M.F. (2008). “Spatial Accuracy 2.0”, 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, University of California, Santa Barbara, p. 7.
- Harvey, F. (2003). “Developing Geographic Information Infrastructures for Local Government: The Role of Trust”, *Canadian Geographer / Le Géographe canadien*, pp. 28-36.
- ISO/FDIS 19157 “Geographics information – Data Quality, as sent to the ISO, Central Secretariat for Issuing as FDIS (2013)”, ISO International Organization for Standardization, p. 170.
- ISO/IEC Guide 51:1999 (1999). “Safety aspects – Guidelines for their inclusion in standards”, ISO International Organization for Standardization.
- Ivánová, I.; Morales, J.; de By, R.A.; Beshe, T.S. y M.A. Gebresilassie (2013). “Searching for Spatial Data Resources by Fitness for Use”, *Journal of Spatial Science*, pp. 15-28.
- McHugh, J.; Abiteboul, S.; Goldman, R.; Quass, D. y J. Widom (1997). “Lore: A Database Management System for Semi Structured Data”, *ACM Sigmod Record*, pp. 54-66.
- Levesque, M.A.; Bédard, Y.; Gervais, M. y R. Devillers (2007). “Towards Managing the Risks of Data Misuse for Spatial Datacubes”, *Canada NSERC Industrial Research Chair in Geospatial Databases for Decision Support*, Dép. des Sciences Géomatiques, Université Laval, Department of Geography, Memorial University of Newfoundland, p. 8.
- Van Oort, P.A.J.; Bregt A.K. y S. Bruin de (2006). “Do Users Ignore Spatial Data Quality?”, 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Centre for Geo-Information, Wageningen University, edited by M. Caetano y M. Painho, pp. 236-246.
- Van Oort, P.A.J. (2008). “Geo-Data User Chains”, Proceedings of the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Centre for Geo-Information, Wageningen University, edited by M. Caetano y M. Painho, pp. 398-405.