

La Información Geográfica Voluntaria: ¿cómo establecer y evaluar su calidad?

Clara Martínez-de-Ibárreta-Soriano*

Recibido el 14 de marzo de 2016; aceptado el 25 de agosto de 2016

Abstract

A new trend has appeared in the field of cartography production: volunteered geographic information or VGI. Currently, anyone can produce/share/use geographic information with web-based tools such as OpenStreetMap. The professional environment worries about the quality of these data. Analyzing data accuracy is the most important question. VGI considers that the same data collected by many people has a better result than data collected by different individuals. Some former research works determine a comparison between the accuracy of linear and point features collected with VGI and data from official sources. Nevertheless, in many cases there is no possibility to access reliable data sources. How can the quality of voluntary information be quantified? And how to use VGI, thus taking advantage of free data? In this paper of the state of the art about IGV possible answers are presented through the most relevant aspects of this kind of voluntary contributions. Furthermore, methodologies and solutions are shown in several studies to establish a quality assessment procedure. An especially important issue for professionals is to distinguish when and how to use the VGI. Thus, take advantage of free data collection without forgetting the accuracy of the information.

Key words: spatial data quality, crowdsourcing, volunteered geographic information (VGI), OpenStreetMap (OSM).

Resumo

Uma nova tendência tem aparecido no campo da produção cartográfica: a Informação Geográfica Voluntária ou IGV. Atualmente, qualquer indivíduo pode produzir, compartilhar e utilizar a informação geográfica através de ferramentas baseadas na

* ETSI Topografía, Geodesia y Cartografía, Universidad Politécnica de Madrid, Campus Sur, 28031 Madrid, España, correo electrónico: clara.msorianoalumnos.upm.es; <<http://orcid.org>>.

web, tal como o *OpenStreetMap*. A qualidade destes dados é uma preocupação para o ambiente profissional e, analisar sua precisão é de suma importância. O sucesso técnico da abordagem da IGV é baseado no princípio de que o mesmo dado coletado por muitas pessoas gera um resultado melhor do que o dado coletado de forma singular. Pesquisas anteriores procuram determinar a precisão das entidades, lineares e pontuais em sua maioria, coletadas através de IGV, comparando dados de fontes oficiais. No entanto, uma vez que em muitos casos não há nenhuma possibilidade de acesso a fontes confiáveis, como se fazer avaliações quantitativas da qualidade de contribuições voluntárias? Como estabelecer uma avaliação da qualidade para poder se usar a informação geográfica? Neste artigo, do estado da arte do tema IGV e se apresentam possíveis respostas mediante a exposição dos aspectos mais importantes deste tipo de contribuição. Além disso, se mostram metodologias e soluções utilizadas em diversos estudos para se estabelecer um processo de avaliação da qualidade, tema que adquire especial importância para que os profissionais possam distinguir quando e como usar a IGV e, assim, tirar proveito da coleta gratuita dos dados sem se esquecer da precisão das informações.

Palavras chave: *qualidade de dados espaciais, crowdsourcing, Informação Geográfica Voluntária (IGV), OpenStreetMap (OSM)*.

Resumen

Una nueva tendencia ha aparecido en el campo de la producción de cartografía: la Información Geográfica Voluntaria o IGV. Actualmente, cualquier individuo puede producir, compartir y utilizar información geográfica a través de herramientas web, tal como hace OpenStreetMap. La calidad de estos datos es una preocupación para el entorno profesional y analizar su exactitud adquiere gran importancia. El éxito técnico del enfoque de IGV se apoya en considerar que el mismo dato recogido por muchas personas produce un mejor resultado que el de cada contribuyente individual. Investigaciones previas buscan determinar la exactitud de las entidades, lineales y puntuales en su mayoría, recolectadas mediante IGV comparando con datos de fuentes oficiales. Sin embargo, dado que en muchos casos no hay posibilidad de acceder a fuentes fiables, ¿cómo realizar valoraciones cuantitativas de la calidad de las contribuciones voluntarias?, ¿cómo establecer una evaluación de calidad adecuada para poder emplear dicha información geográfica? En este artículo del estado del arte se realiza una revisión del tema IGV y se presentan posibles respuestas mediante la exposición de los aspectos más relevantes de este tipo de contribuciones. Además, se muestran metodologías y soluciones empleadas en diversos estudios para establecer un procedimiento de evaluación de calidad, tema que adquire especial importancia para que los profesionales puedan distinguir cuándo y cómo poder emplear la IGV y, así, aprovechar la recogida gratuita del dato sin olvidar la exactitud de la información.

Palabras clave: *Calidad de los datos espaciales, crowdsourcing, Información Geográfica Voluntaria (IGV), OpenStreetMap (OSM).*

Introducción

El conocimiento del territorio ha sido siempre un objetivo del ser humano ya que disponer de información geográfica (IG) detallada y precisa del espacio le aportaba innumerables ventajas (Ruiz Almar, 2010). Hasta hace poco tiempo la producción de IG ha estado limitada a un grupo pequeño de individuos, generalmente profesionales, debido a que la información ha sido difícil de obtener, de almacenar y de manipular (Ruiz Almar, 2010). Habitualmente, la producción de los datos espaciales ha sido a la vez un deber y un privilegio de Agencias Nacionales de Cartografía o grandes empresas comerciales activas en el dominio geoespacial. Los productos y servicios espaciales así producidos se acompañan con un cierto nivel de garantía debido, entre otros, a la reputación y la credibilidad de la autoridad emisora (Neis *et al.*, 2014). No obstante, este tipo de productos y servicios autorizados vienen con altos costes y términos de licencia restrictivos (Antonioni *et al.*, 2015).

Durante la última década, la aparición de Información Geográfica Voluntaria, IGV, (término acuñado por Goodchild en 2007) proporciona una alternativa a la disponibilidad de datos espaciales (Goodchild, 2007a). Esto es debido, en gran medida, a las mejoras en la tecnología de las comunicaciones y la disponibilidad de la información que están impactando en el campo de la geografía así como el acceso a los sistemas globales de navegación por satélite, GNSS (Esmaili *et al.*, 2013; Haklay, 2010; Rana *et al.*, 2009; Tulloch *et al.*, 2003; Coleman *et al.*, 2009). Los sistemas GNSS y los receptores permiten a los usuarios no entrenados adquirir su posición con facilidad y exactitud, lo que supone simplificar el proceso de recopilación de IG (Haklay, 2010). Además, la aparición de Internet permite una gran demanda y difusión de datos geoespaciales (Rana *et al.*, 2009; Tulloch *et al.*, 2003). A todo ello, se suma la llegada de los servicios basados en conceptos de Web 2.0 (Goodchild, 2007b) donde todo individuo, desde cualquier lugar del mundo, puede acceder a una ingente cantidad de datos territoriales y puede disponer de herramientas que le permiten tratar y editar esa información eficientemente y, además, dotarla de semántica (Ruiz Almar, 2010). Por tanto, IGV combina elementos de la Web 2.0, inteligencia colectiva y neogeografía (Castelein *et al.*, 2010), definiendo como neogeografía a la Geografía cuando incorpora aportaciones de usuarios no especialistas.

En este contexto ha aparecido el concepto de *crowdsourcing*, que se podría traducir al castellano como *colaboración abierta distribuida* (Ruiz Almar, 2010). El mismo consiste en derivar tareas dejándolas a cargo de un grupo numeroso de personas o una comunidad que, de forma voluntaria, decide producir información de un lugar de su interés. El concepto no es específico a la IG. El ejemplo más famoso

de sitios web de *crowdsourcing* es *Wikipedia* donde usuarios voluntarios pueden crear y editar artículos de la enciclopedia digital, que se administra por una organización sin ánimo de lucro y que fue creada en lengua inglesa en 2001 (Esmaili *et al.*, 2013).

Algunos autores han denominado a este movimiento: explosión geográfica, ya que las Tecnologías de la Información Geográfica (TIG) han propiciado la revalorización de lo espacial en el entorno cotidiano de la sociedad. Los tres principales factores que han provocado el nacimiento de la *explosión geográfica* son: su papel como elemento reestructurador de la geografía en todas sus escalas, la aparición de nuevos códigos y prácticas disciplinarias y, por último, la constitución de nuevas representaciones del espacio y la sociedad global (Ruiz Almar, 2010).

Las ventajas de recoger información mediante los voluntarios son la rápida adquisición y, en muchas ocasiones, el acceso gratuito de la misma. Incluso, es posible actualizar a través de los ciudadanos los acontecimientos espaciales locales en un corto periodo de tiempo, datos que, de otra manera, tardarían mucho tiempo en ser recolectados por fuentes oficiales (Esmaili *et al.*, 2013). Sin embargo, determinar la exactitud y la calidad del proceso de producción de estos datos aún son un reto para los investigadores (Esmaili *et al.*, 2013), alcanzando incluso a la utilidad de los datos y la consistencia de los mismos (Haklay, 2010). Hay importantes limitaciones a la IGV relacionadas con la calidad y el nivel de confianza que se puede tener sobre ellos (Jackson *et al.*, 2013). En este momento es prioritario encontrar los mecanismos necesarios para garantizar la calidad, para detectar y eliminar errores, y para construir un nivel comparable de confianza y seguridad al que las agencias nacionales de cartografía han disfrutado tradicionalmente (Goodchild, 2007b).

Las instituciones deberían asumir que los datos voluntarios están en continuo cambio y que sería necesario introducir nuevas reglas que establecieran y equilibrasen las pautas de recolección voluntaria de datos como con los productores tradicionales. La definición de criterios para decidir si un trabajo voluntario puede o no ser utilizado sería de máxima importancia. El autor (Ruiz Almar, 2010) reseña que “los voluntarios, por su parte, si desean que la información que generan tenga el valor que se le pretende atribuir, deberían someter su trabajo a una serie de estándares que midieran la equidad, la seguridad, la calidad y la autoridad de los mismos”.

El conocimiento de la calidad de los datos espaciales es importante por dos razones: importancia económica y jurídica en los procesos de toma de decisiones y la posibilidad de combinarse con otros para diferentes propósitos (Esmaili *et al.*, 2013). Por ejemplo, los mapas de OpenStreetMap (OSM) se utilizan en diferentes proyectos comerciales como mapas de fondo por lo que el conocimiento de su calidad es fundamental (Exel *et al.*, 2010).

Es importante también estudiar la motivación de los voluntarios para contribuir en proyectos de IGV (Goodchild *et al.*, 2012), la cual puede ser muy diversa en

función de las regiones geográficas del mundo y el análisis de los factores socio-económicos y culturales donde se desarrollan las actividades (Neis *et al.*, 2014).

Este documento está organizado de la siguiente manera. En la Sección 2, se presentan las plataformas IGV más utilizadas haciendo una breve descripción de ellas. Además, se describen brevemente los conceptos que conforman la calidad. En la Sección 3 se detallan los métodos y conclusiones extraídas de diversos estudios que han intentado establecer procesos para conseguir cuantificar la calidad en la información voluntaria. En la Sección 4 se habla de las posibles vías futuras de actuación. Por último, en la Sección 5 se exponen las conclusiones.

Plataformas IGV

Los proyectos OSM, *Wikimapia*, *Google Earth Hacks*, *Wikiloc* y *Geo-Wiki* son ejemplos de plataformas IGV (Esmaili *et al.*, 2013). OSM es una iniciativa, creada en 2006, que tiene por objeto crear y proporcionar datos geográficos libres, tales como callejeros y mapas de carreteras, así como de proporcionar los medios para liberar dichos datos bajo licencia ODbL para que cualquiera que lo desee pueda usarlos, editarlos y compartirlos.

La comunidad de colaboradores y desarrolladores de OSM ha generado una base de datos de IG muy completa así como las herramientas para editarla, visualizarla y analizarla de forma fácil y amigable (Borges *et al.*, 2013). Una de las actividades que desarrolla es la organización de talleres locales, denominados *mapping party*, cuyo objetivo es crear y anotar el contenido de áreas geográficas localizadas según la voluntad de los usuarios (Haklay, 2010). Uno de los tipos de datos más comunes que se recolectan con estas actividades es de carácter puntual, al que OSM denomina Punto de Interés (POI), indicando sus atributos y descripciones (Jackson *et al.*, 2013).

Wikimapia es una aplicación web que integra mapas de *Google Maps* con un glosario de descripciones muy amplias de diversos lugares, mientras que OSM es un callejero digital de las ciudades más importantes, siendo ambos de dominio público (Goodchild, 2007b). Es una iniciativa donde todos los contribuyentes editan o corrigen de manera anónima y no existe un mecanismo supervisor o disciplinario aplicable a usuarios problemáticos.

Las mencionadas no son las únicas iniciativas exitosas. Entre otras, destacan *Google Earth Hacks* que proporciona información sobre las nuevas aplicaciones, *Wikiloc* (Castelein *et al.*, 2010) que comparte datos GNSS de excursionistas o la *Wikipedia* donde se definen contextos y personajes históricos aunque no tiene la componente espacial de la IGV (Esmaili *et al.*, 2013; Goodchild, 2007b). Cabe mencionar además proyectos como *Geo-Wiki* que es una red global de voluntarios cuyo objetivo es ayudar a mejorar la calidad de la información sobre la cobertura global de la Tierra mediante la adición de información sobre la distribución de

especies, hábitats y ecosistemas apoyado por el conocimiento local de los contribuyentes (Foody *et al.*, 2013).

OSM puede ofrecer un terreno muy fructífero para la investigación de plataformas IGV, con la ventaja de contar cada vez con más seguidores que la nutren al incluir información de varios contribuyentes de manera simultánea. El resultado de dicha investigación puede llegar a desafiar algunas suposiciones bien establecidas dentro de la Geomática ya que esta nueva vía de producción de datos modifica las estructuras existentes (Haklay, 2010). La manera de afrontar estos cambios por parte de las agencias oficiales de producción de IG debe evolucionar modificando sus estructuras de recolección de datos para afrontar la competencia que supone la IGV (Haklay, 2010). En consecuencia, en el caso de poder determinar la calidad de los datos IGV, algunos autores creen que quizás fuera posible remodelar los métodos de producción cartográfica actuales y estandarizar el uso de este tipo de información (Ruiz Almar, 2010; Haklay, 2010).

Procesos no tradicionales de control

En el campo de la IG, los principios y directrices de la ISO 19157 son los criterios seguidos para la evaluación de la calidad. La norma define como elementos de calidad de datos: exhaustividad, integridad, consistencia lógica, exactitud posicional, linaje, usabilidad y exactitud temática (Antoniou *et al.*, 2015; Coleman *et al.*, 2009).

La preocupación por la calidad es tal vez mayor ahora que en el pasado con el uso generalizado de las nuevas tecnologías para la cartografía a través de Internet (Haklay, 2010), que presentan tanto retos como oportunidades para la difusión de la IG (Feick *et al.*, 2010). La calidad de la documentación es un factor importante cuando se trabaja con datos espaciales, especialmente con IGV ya que no existen especificaciones para la creación de datos (Antoniou *et al.*, 2015). Es necesario, por tanto, caracterizar individualmente la exactitud de cada voluntario o fuente de datos para luego adjudicársela a los productos obtenidos.

Conocer cómo se recoge la información, y los procesos de calidad que pasan, ayuda a los usuarios de la IGV a saber cuándo es adecuado utilizar dichos datos (Foody *et al.*, 2013).

Para ello se disponen de algunos métodos y mecanismos que se pueden aplicar a) la revisión por pares (también con procesos de *crowdsourcing*), donde la calidad de datos puede ser asegurada tras considerar las opiniones dadas por múltiples contribuyentes supuestamente independientes, b) medidas sociales, que se centran en la autoevaluación de los contribuyentes como medida aproximada de la calidad de sus propias contribuciones, y c) la coherencia geográfica, a través de un análisis de la consistencia mutua de las entidades (Goodchild *et al.*, 2012).

A pesar de las dudas sobre los contenidos generados, recogidos por personas no expertas y sin seguir estándares de calidad, el gran volumen de información puede

equilibrar la balanza de la falta de control de calidad (Baeza-Yates, 2009). Sin embargo, el usuario no puede evaluar la calidad de los datos de forma aislada pues necesita conocer el propósito y la necesidad de la información que está recogiendo. Ahí es donde se requiere a la comunidad IGV para que, de común acuerdo, se tracen unas pequeñas directrices que servirán para dar coherencia a los datos procedentes de diversos colaboradores (Esmaili *et al.*, 2013). Por otra parte, los voluntarios no están obligados a crear metadatos y, además, no todos los aspectos de la calidad pueden evaluarse cuantitativamente (Esmaili *et al.*, 2013).

Por todo lo expuesto anteriormente, la gestión de la calidad de las contribuciones voluntarias requiere el desarrollo de herramientas que asesoren a los contribuyentes durante el proceso de creación de la entidad y que ayude identificar y corregir rápidamente los datos ya existentes de calidad cuestionable (Ahmed Loai Ali, 2014). Por lo tanto, se hace esencial localizar una manera de medir o cuantificar la exactitud de los datos IGV (Esmaili *et al.*, 2013).

Se entiende por credibilidad como la verosimilitud de una fuente o un mensaje. Las personas que contribuyen no son independientes y es necesario analizar cómo la nueva contribución influye en la credibilidad. Se hace ineludible encontrar una manera de fomentar la comunicación igualitaria (Spielman, 2014).

La credibilidad de la IGV se basa en que una muestra representativa de personas haga su aportación personal con honradez, compromiso y exactitud (Coleman *et al.*, 2009). Las evaluaciones de la calidad de la información construida colectivamente o a través de los esfuerzos de la comunidad podrían convertirse en nuevos árbitros de la credibilidad. Como por ejemplo, en el estudio de Haklay (2010) donde se demuestra que la exactitud posicional mejora a medida que el número de editores aumenta, pero siempre que el número de editores no supere el trece (Coleman *et al.*, 2009). De hecho, el 84% de las modificaciones sobre un objeto geográfico se hicieron por un 12% de los contribuyentes (Coleman *et al.*, 2009).

Procesos tradicionales de control

Para medir la exactitud posicional de la IGV es muy común comparar los conjuntos de datos IGV con un conjunto de datos acreditado con una exactitud, al menos, tres veces mejor que la analizada, de la misma zona, que actúa como datos de referencia sobre el terreno (Esmaili *et al.*, 2013; Haklay, 2010; Jackson *et al.*, 2013; Foody *et al.*, 2013).

En consecuencia, el principal problema es la elección de dicha base. No obstante, emplear este método para la IGV es, en ocasiones, imposible debido a que no hay acceso a conjuntos de datos de campo adecuados para dicha comparación (Esmaili *et al.*, 2013). Otras veces, la IGV es la única fuente de IG disponible (Neis *et al.*, 2014).

En las siguientes secciones se expondrán los casos prácticos que los investigadores han ido poniendo en marcha para la definición de mecanismos que ayuden a evaluar la calidad de la IGV.

Métodos de validación de la IGV

Otro aspecto a tener en cuenta es que la IGV encaja bien en el modelo de una Infraestructura de Datos Espaciales (IDE), facilitando el intercambio de IG entre los individuos de una comunidad. Los metadatos, las normas, la interoperabilidad, la política y la organización de las IDE serán útiles y válidos incluso en un contexto IGV (Castelein *et al.*, 2010). Para poder aprovechar las nuevas tecnologías, las IDE deben ser abiertas a nuevos actores para permitir la combinación con los entornos IGV. En este sentido, ha habido poca investigación empírica, siendo la convergencia entre IGV y las IDE una posible vía de actuación futura (Castelein *et al.*, 2010).

Casi todas las investigaciones existentes se basan en la comparación de los datos IGV con los datos oficiales precisos (Esmaili *et al.*, 2013; Haklay, 2010; Jackson *et al.*, 2013; Foody *et al.*, 2013). A continuación, se presentan métodos ensayados para los diversos escenarios que se pueden encontrar.

Entidades puntuales

La evaluación de la exactitud posicional en las entidades puntuales ha sido el primer caso en el control de la calidad de plataformas de recogida de información voluntaria.

Un primer intento ha sido estudiar las intersecciones de carreteras comparando OSM con datos comerciales (Ahmed Loai Ali *et al.*, 2014). Sin embargo, en este apartado se tratará la ubicación de puntos de interés (POI) de OSM (Jackson *et al.*, 2013).

Jackson *et al.* (2013) comparan la posición de los POI obtenidos de tres conjuntos de datos: uno de referencia (*Oak Ridge National Laboratory*, ORNL), un proyecto colaborativo (OSM) y un híbrido (OSMCP) enlazado con el (United States Geological Survey, USGS) que le ha realizado su propio control de calidad (Jackson *et al.*, 2013). El producto resultante es consecuencia de la interpretación que realiza el colaborador voluntario de su entorno (Coleman *et al.*, 2009). Por ejemplo, si se registran colegios como entidades puntuales, un colaborador ubicará el colegio en función de su punto de acceso y otro interpretará que debe indicar el centro del edificio para ubicarlo en el mapa (Jackson *et al.*, 2013). La capacidad del voluntario de interpretar correctamente donde debe registrar una entidad está ligada a la educación, habilidades y experiencia que tenga en el ámbito cartográfico (Foody *et al.*, 2013). Además, se concluye que la gran cantidad de contribuciones sobre un mismo objeto permitirá ubicarlo en la mejor posición posible.

Entidades lineales

Las métricas que se utilizan para elementos lineales son varias. Una posible es comparando el porcentaje de la línea de un conjunto de datos que se encuentra a cierta distancia de la misma característica en otro conjunto de datos de exactitud más alta (Haklay, 2010). El uso de un *buffer* permite determinar una zona donde dos conjuntos de datos de dos fuentes distintas pueden tener discrepancia pero tratarse como si no las hubiera.

Haklay (2010) emplea como referencia datos extraídos de *Meridian 2*, que es una plataforma comercial basado en un conjunto de datos vectorial recogido por la agencia cartográfica británica *Ordnance Survey*, centrado en el análisis medioambiental (Haklay, 2010; Ahmed Loai Ali *et al.*, 2014). La evaluación de la exactitud posicional de OSM para entidades lineales puede llevarse a cabo contra *Meridian 2* cuyos nodos son de gran exactitud (Helbich *et al.*, 2010). A pesar de que *Meridian 2* sólo contempla las carreteras generales, obviando las secundarias, se puede emplear para diversas aplicaciones SIG y comparaciones siendo, no obstante, OSM más completa respecto a la información lineal difundida (Haklay, 2010). De hecho, si en OSM se encuentran datos erróneos u omitidos, con un corto proceso se actualizará o completará la información mientras que errores u omisiones en *Meridian 2* requieren contactar con el productor lo que, en consecuencia, requiere más tiempo (Haklay, 2010).

Meridian 2 realiza un proceso de generalización del eje de las carreteras que provoca la disminución de nodos en estas entidades y se obtiene una exactitud de cinco metros dentro de una región o *buffer* de veinte metros de enlace entre los nodos (Haklay, 2010). Respecto a OSM, Haklay (2010) declara que un 70% de sus datos aseguran la exactitud posicional. En términos de calidad, se concluye que los resultados obtenidos con OSM no son diferentes a los conjuntos de datos comerciales provenientes de *Meridian 2* (Haklay, 2010).

Procedimiento crowdsourcing para auditar una clasificación

En ciertos controles de calidad se emplea la validación cruzada, donde los propios voluntarios intercambian sus opiniones con otros voluntarios sobre el trabajo desarrollado por cada uno (Foody *et al.*, 2013).

Proyectos como *Geo-Wiki* promovieron el trabajo de voluntarios para que buscaran en una serie de imágenes satelitales y asignaran una etiqueta con el tipo de cobertura de la Tierra a partir de una lista definida de diez clases, sin establecer ninguna restricción por edad o experiencia cartográfica previa. Debido a ello, el conjunto de etiquetas eran muy variadas y la exactitud desconocida e incierta (Foody *et al.*, 2013). La experiencia de Foody *et al.* (2013) proporciona una forma de evaluar la exactitud de las fuentes de datos de voluntarios sin utilizar datos de

referencia para comparar, centrándose en el problema concreto de la clasificación por etiquetas de la cobertura terrestre.

La clasificación que realizó el estudio involucra al coeficiente *Kappa* de Cohen e intentaba medir el grado de acuerdo de las clasificaciones entre el conjunto de etiquetas indicadas por un voluntario con las etiquetas de cada uno de los otros voluntarios. Dicho coeficiente se calculó para cada voluntario con el fin de proporcionar un indicador de acuerdo a la referencia (Foody *et al.*, 2013). La exactitud de los datos que aportó cada voluntario se comunicará según un modelo de etiquetas elaborado para este fin (Foody *et al.*, 2013). Comparando las regiones de OSM definidas en el proyecto, el coeficiente *Kappa* se encuentra entre 0.227 y 0.663, lo que indica el nivel bajo de acuerdo entre los diferentes conjuntos de etiquetas de cobertura (Foody *et al.*, 2013).

Foody *et al.* (2013) concluyen que cada voluntario aislado proporciona un conjunto de etiquetas con escasa relación con los datos de referencia (*kappa* <0.67) mientras que los datos de múltiples voluntarios podrían combinarse de manera constructiva para el análisis y la producción de IG de utilidad por métodos IGV.

Otros aspectos relevantes para estimar la calidad

El autocontrol que ejercen los mismos usuarios voluntarios, basado en el conocimiento directo que tienen del territorio, es el principal mecanismo para verificar los datos que recogen. Sin embargo, dichos mecanismos no dejan de ser parciales, poco rigurosos y están sujetos al albedrío de usuarios con unos conocimientos limitados (Ruiz Almar, 2010).

Esmaili *et al.* (2013) aseguran que es posible indicar la calidad de los datos empleando los metadatos, pues puede ser determinada por la experiencia del colaborador en proyectos de SIG (Sistemas de Información Geográfica). Sin embargo, el metadato es el encargado de comunicar información sobre el dato y no el que determina la calidad del mismo.

Sin perjuicio de lo anterior, la inexactitud de los datos *crowdsourced* podría ser estimada por dos tipos de metadatos: los metadatos convencionales y aquellos metadatos donde es el sistema el que automáticamente guarda ciertas características de los datos, por ejemplo, la resolución de la imagen utilizada (Esmaili *et al.*, 2013). Las contribuciones voluntarias han provocado la aparición de los denominados metadatos 2.0, donde el usuario aporta gran cantidad de contenido de forma similar a las *wikis*. Sin embargo, proyectos como OSM aún no los añaden aunque los datos podrían ser criticados o controlados por otros voluntarios (Goodchild, 2008b). Además, los metadatos de la IGV deberían indicar, mediante un identificador, qué voluntario ha recogido dicha IG.

Otra novedad que introduce la IGV respecto a los procedimientos tradicionales es el uso de niveles de confianza de la IG recolectada.

Algunos esfuerzos se han dirigido a categorizar los tipos de contribuyentes que existen en las plataformas colaborativas estableciendo indirectamente, de esta manera, niveles de confianza (Arsanjani *et al.*, 2013). Normalmente, la IG producida por agencias oficiales (como los institutos geográficos) siempre ha tenido el sello de garantía de calidad, pues los que se dedican a ello son profesionales. Si la IGV obtiene la misma confianza que el consumidor deposita en los productos de las agencias oficiales, los datos que se recopilan y editan voluntariamente podrán ser utilizados sin reserva en muchas aplicaciones donde, en la actualidad, sólo se emplea la IG oficial. Hay otros estudios que analizan qué tipo de factores pueden contribuir a una evaluación de la fiabilidad de los datos y la reputación de los contribuyentes en un sistema de IGV (Forati *et al.*, 2016). Forati *et al.* (2016) proponen un método que crea automáticamente las puntuaciones de reputación y confiabilidad de cada voluntario, lo cual permitiría ayudar a catalogar y categorizar los tipos de colaboradores (Arsanjani *et al.*, 2013).

En el trabajo de Arsanjani *et al.* (2013), los usuarios se clasifican en cinco grandes grupos en función de la cantidad de contribuciones y de cómo son dichos contenidos: neófitos o principiantes, aficionados interesados (regular), aficionados expertos (nivel intermedio), expertos y cartógrafos profesionales tal como muestra la Figura 1 (Coleman *et al.*, 2009). Para realizar esta categorización se tienen en cuenta tres indicadores de calidad: exactitud posicional, exhaustividad y exactitud semántica que consigue cada voluntario (Arsanjani *et al.*, 2013).

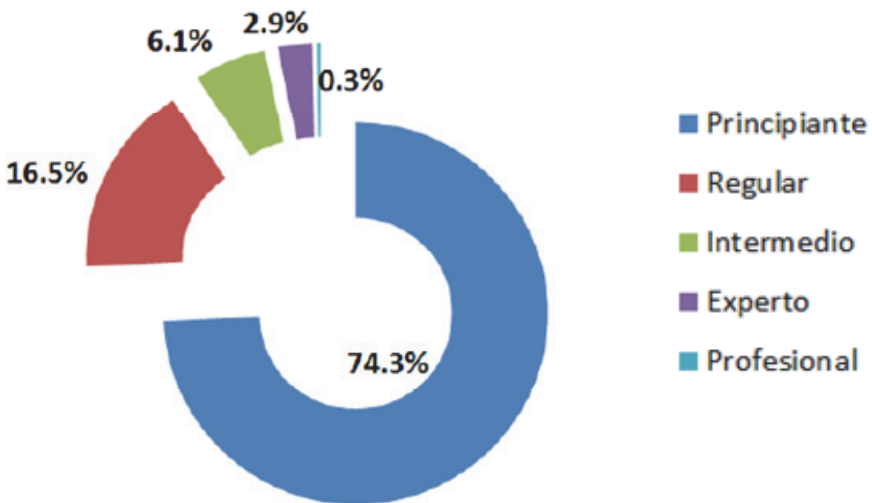


Figura 1. Nivel de experiencia de los colaboradores de OSM.
Fuente: Arsanjani *et al.*, 2013.

Por otro lado, actualmente, el campo de la calidad de la IGV sigue generando debate haciendo surgir nuevos estudios e investigaciones relativas a los tipos de indicadores de calidad que pueden ser medidos.

Senaratne *et al.* (2016) exponen una revisión literaria exhaustiva sobre los indicadores de calidad que han sido estudiados hasta la fecha y los artículos donde se mencionan: exactitud posicional, exactitud temática, consistencia topológica, completitud, exactitud temporal, exactitud geométrica, exactitud semántica, linaje, usabilidad, credibilidad, fiabilidad, reputación, experiencia, conocimiento de la zona de estudio y reconocimiento. Además, se presenta un cuadro resumen de estos indicadores de calidad respecto del método empleado para evaluar cada uno de dichos tipos, realizando un agrupamiento por motivación social, geográfico, de colaboración y de técnicas (Goodchild *et al.*, 2012; Senaratne *et al.* 2016).

Si bien se han utilizado la mayoría de los métodos para evaluar la exactitud posicional, la exactitud temática y la consistencia topológica existen otros métodos que han abordado el resto de las medidas de calidad e indicadores (Coleman *et al.*, 2009; Senaratne *et al.* 2016). Una posible vía exploratoria es centrarse en otros enfoques posibles para manejar las medidas de calidad y los indicadores asociados a ella (Senaratne *et al.* 2016).

Discusión y conclusiones

En este artículo se realiza una introducción a la IGV de tal manera que sirva como documento preliminar para otros proyectos de investigación de la calidad de la IGV. A continuación, se establecen una serie de conclusiones obtenidas durante el proceso de escritura de este artículo.

La IGV ha llegado para quedarse, por lo que conocer y estimar de forma práctica sus características más importantes es de especial interés. Los esfuerzos futuros deben aproximarse a una combinación entre el conocimiento adquirido por los profesionales a lo largo de años de formación, y la experiencia con las aportaciones rápidas y masivas de los voluntarios, obteniendo productos nuevos o actualizando los existentes más rápidamente, siendo una alternativa razonable y beneficiosa para ambas partes.

Se plantea la cuestión de cuánta diferencia hay entre la norma de calidad y la calidad obtenida al comparar la información recogida por voluntarios y la obtenida por fuentes oficiales, si bien es verdad que el resultado siempre está afectado de error experimental. La norma permite garantizar que los atributos de los objetos están dentro de un rango especificado. El usuario deposita una confianza en el proveedor porque este le garantiza que la IG tiene calidad y sabe que el producto final pasó por un procedimiento formal de control de calidad.

Ignorar la IGV sería un grave error. En la medida de lo posible, la IGV no debe ser desestimada sino que se deben crear los mecanismos de acreditación necesarios

que permitan establecer su credibilidad para que pueda ser incorporada y aceptada para aplicaciones cada vez más críticas.

Otro factor que debe tenerse en cuenta es la manera en la que se recogen los datos. Por una parte, los organismos gubernamentales y oficiales están comprometidos a proporcionar una cobertura total del país, independientemente de la lejanía de la ubicación o el estado socioeconómico de sus habitantes. Por otro lado, IGV se basa en las decisiones de los contribuyentes sobre las áreas en las que les interesa recopilar información y, por lo tanto, no hay garantías de obtener una distribución homogénea de las actividades voluntarias por el territorio. En este sentido, sería de especial relevancia crear mecanismos de estímulo que motiven la colaboración voluntaria no solamente en las zonas de interés sino la recolección de IG a lo largo y ancho de todo el territorio.

Las entidades oficiales en la práctica no pueden oponerse a esta iniciativa, y muy por el contrario, deberían contribuir ofreciendo sus conocimientos y sus habilidades a los nuevos actores geográficos, especialmente a los voluntarios, de manera que el conocimiento acumulado a lo largo de los años no se diluya y llegue a ser visible en los contextos colaborativos. Conseguir un mecanismo de control de calidad integrado que permita a los participantes calificar la calidad de la contribución de los otros participantes es una posibilidad práctica y razonable para atribuirle características de confiabilidad a la IGV. Además, sería recomendable ahondar en otros aspectos de la IGV como la motivación de los contribuyentes, las tendencias futuras y las aplicaciones dentro de la cartografía.

Respecto a los criterios que deberían aplicarse a la IGV destacan los siguientes:

- Trazar una serie de directrices, acordadas por la comunidad IGV que cuente con el respaldo de los estándares y de las entidades oficiales, de tal manera que se de consistencia y coherencia a la información registrada por los voluntarios.
- Crear un sistema de reparto de zonas para evitar exceso de información en ciertas áreas e inexistencia en otras. La colaboración no sería obligatoria pero si recomendable ya que es un reparto de tareas.
- Motivar un control de calidad realizado sobre los datos que recoge el propio voluntario. Además, establecer una validación cruzada entre los colaboradores. El objetivo sería que los propios voluntarios se auto corrigieran apoyándose en ciertas normas establecidas por la comunidad IGV.
- Implantar una jerarquía de contribuciones respecto a voluntarios más cualificados y menos cualificados en todas las plataformas IGV, motivando la participación ya que cuantas más personas participen mayor será la calidad de la información.

Siguiendo estas pautas, algunas de las cuales ya se están intentando implementar, se puede conseguir una mejor evaluación de la calidad de los datos procedentes de IGV y, en consecuencia, obtener en un futuro cercano productos cartográficos creados a partir no sólo de entidades oficiales sino de plataformas voluntarias.

Bibliografía

- Ahmed Loai Ali, A. y Schmid, F. (2014). "Data Quality Assurance for Volunteered Geographic Information", *Lecture Notes in Computer Science*, 8728:126-141.
- Antoniou, V. y Skopeliti, A. (2015). "Measures and Indicators of VGI Quality: an Overview", *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II-3/W5, pp. 345-351.
- Arsanjani, J.; J. Barron; C. Bakillah, M. y Helbich, M. (2013). "Assessing the Quality of OpenStreetMap Contributors together with their Contributions", *16th AGILE Conference*, Leuven, pp. 14-17.
- Baeza-Yates, R. (2009). "User generated content: how good is it?", *3rd workshop on Information credibility on the web*, pp. 1-2.
- Borges, C.; E. Pijoan; A. Sorrosal; G. Oribe-Garcia; I. González, M. y Esteban, O. K. (2013). "Uso de fuentes de información geográfica voluntarias en proyectos de ingeniería", *Actas de las VII Jornadas de SIG Libre*, Girona 6-8 marzo, España, pp. 1-22.
- Castelein, W. Grus, L. Crompvoe, J. y Bregt, A. (2010). "A characterization of volunteered geographic information", *13th AGILE International Conference on Geographic Information Science*, pp. 1-10.
- Coleman, D. Georgiadou, Y. y Labonte, J. (2009). "Volunteered Geographic Information: The Nature and Motivation of Producers", *International Journal of Spatial Data Infrastructures Research*, 4:332-358.
- Esmaili, R.; Naseri, F. y Esmaili, A. (2013). "Quality Assessment of Volunteered Geographic Information", *American Journal of Geographic Information System*, 2(2):19-26.
- Exel, M. Van; Dias, E. y Fruijtjer, S. (2010). "The impact of crowdsourcing on spatial data quality indicators", *GIScience*, Zurich, Switzerland, pp. 1-4.
- Feick, R. y Roche, S. (2010). "Valuing Volunteered Geographic Information (IGV): Opportunities and Challenges Arising from a New Mode of GI Use and Production", *2nd GEOValue Workshop*, HafenCity University, Hamburg, Germany, pp. 75-79.
- Foody, G.; M. See; L. Fritz; S. Van der Velde; M. Perger, C. Schill, C. y Boyd, D.S. (2013). "Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project", *Transactions in GIS*, 17(6):847-860.

- Forati, A.M. y Karimipour, F. (2016). "A VGI Quality Assessment Method for VGI based on Trustworthiness Related Work", *GI_Forum*, 1:3-11.
- Goodchild, M.F. (2007a). "Citizens as sensors: The world of volunteered geography", *GeoJournal*, 69:211-221.
- (2007b). "Citizens as sensors: web 2.0 and the volunteering of geographic information", *Geofocus (Editorial)*, 7:8-10.
- (2008b). "Spatial Accuracy 2.0", *8th international symposium on spatial accuracy assessment in natural resources and environmental sciences*, World Academic Union, Shanghai, pp. 1-7.
- Goodchild, M.F. y Li, L. (2012). "Assuring the quality of volunteered geographic information", *Spatial Statistics*, 1:110-120.
- Haklay, M. (2010). "How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets", *Environment and Planning B: Planning and Design*, 37(4):682-703.
- Helbich, M. Amelunxen; C. Neis, P. y Zipf, A. (2010). "Investigations on Location Accuracy of Volunteered Geographic Information Using OpenStreetMap Data", *GIScience 2010 Workshop*, Zurich, Switzerland.
- Jackson, S.; Mullen, W.; Agouris, P.; Crooks, A.; Croitoru, A. y Stefanidis, A. (2013). "Assessing Completeness and Spatial Error of Features in Volunteered Geographic Information", *ISPRS International Journal of Geo-Information*, 2(2):507-530.
- Neis, P. y Zielstra, D. (2014). "Recent Developments and Future Trends in Volunteered Geographic Information Research: The Case of OpenStreetMap", *Future Internet*, 6(1):76-106.
- Rana, S. y Joliveau, T. (2009). "NeoGeography: an extension of mainstream geography for everyone made by everyone?", *Journal of Location Based Services*, 3(2):75-81.
- Ruiz Almar, E. (2010). "Consideraciones acerca de la explosión geográfica: Geografía colaborativa e información geográfica voluntaria acreditada", *GeoFocus (Artículos)*, 10:280-298.
- Senaratne, H.; Mobasheri, A.; Ali, A.L.; Capineri, C. y Haklay, M.M. (2016). "A Review of Volunteered Geographic Information Quality Assessment Methods", *International Journal of Geographical Information Science*, pp. 1-27.
- Spielman, S.E. (2014). "Spatial collective intelligence? Credibility, accuracy, and volunteered geographic information", *Cartography and Geographic Information Science*, 41(2):115-124.
- Tulloch, D.L. y Shapiro, T. (2003). "The Intersection of Data Access And Public Participation : Impacting GIS Users' Success ?", *Urisa Journal*, 15(APA II), pp. 55-60.