

Comparativa entre OpenStreetMap y Cartociudad: caso de estudio de Valencia

N. Fernández Terrones, J.M. De Diego Alarcón y A. Pérez-Navarro

Estudis d'Informàtica, Multimèdia i Telecomunicació, Universitat Oberta de Catalunya,
Rambla Poblenou 156, 08018 Barcelona. [nfernandez0]jde_diegoa|aperezn]@uoc.edu

RESUMEN

OpenStreetMap se inició en 2004 y ha crecido de forma paralela a los proyectos de software libre hasta convertirse en el ejemplo más veterano y de mayor envergadura dentro de lo que se conoce como información geográfica voluntaria (VGI en su acrónimo inglés). El auge del uso de este tipo de datos deja, sin embargo, ciertas preguntas abiertas como por ejemplo: ¿Hasta qué punto son fiables los datos así obtenidos? ¿Cuál es su calidad? En el presente trabajo se ha realizado una comparación de los datos geográficos producidos por voluntarios dentro del proyecto de colaboración OpenStreetMap, con los datos producidos por instituciones y armonizados dentro del proyecto Cartociudad. La intención de la comparación es evaluar la calidad de los primeros respecto de los segundos. Para ello se ha definido el término de calidad cartográfica y se han evaluado los diferentes elementos de calidad cartográfica de OpenStreetMap: precisión espacial y de atributos, compleción, calidad temporal y consistencia lógica. El trabajo se realiza con los datos a dos niveles: municipio y/o provincia de Valencia. Los resultados de este análisis muestran que OpenStreetMap tiene una precisión posicional y temporal más que adecuada para usos geocodificación y cálculo de rutas. Sin embargo la heterogeneidad de la cobertura de datos y ciertas inconsistencias internas pueden comprometer su uso. A pesar de ello, se destaca el potencial del proyecto y de una solución de cálculo de rutas óptimas (OpenRouteService) que utiliza con éxito los datos de OpenStreetMap.

Palabras clave: Información Geográfica Voluntaria, VGI, OpenStreetMap, Cartociudad, OpenRouteService, Calidad Cartográfica, Valencia.

INTRODUCCIÓN

La información geográfica voluntaria (VGI) es un fenómeno que surge dentro del contexto de la Web 2.0, y cuyo auge ha venido impulsado por la cada vez más amplia disponibilidad de receptores GPS, ya sea integrados en móviles o en forma de dispositivos independientes ([1]). Uno de los proyectos más ambiciosos y veteranos dentro de la VGI, considerado el paradigma de este fenómeno, es OpenStreetMap (www.openstreetmap.org). A pesar de que desde diversos sectores se cuestiona la fiabilidad de estos datos creados por voluntarios y se suelen considerar de peor calidad o menos fiables que los productos cartográficos profesionales, no ha sido hasta muy recientemente que colectivos académicos han abordado el tema de la calidad de la VGI: ver por ejemplo [2-4]. Estos estudios se han centrado en comparar

la calidad de OpenStreetMap respecto de cartografías institucionales de diferentes países (Alemania [5], Francia [6] y Reino Unido [2]). En este contexto, el presente trabajo muestra un análisis de la calidad cartográfica de OpenStreetMap comparándolo con la cartografía del proyecto institucional Cartociudad habiéndose escogido la provincia de Valencia para llevar a cabo el estudio.

DATOS COMPARADOS Y CALIDAD CARTOGRÁFICA

Cartociudad es un proyecto de integración y armonización de datos de diferentes instituciones que comenzó en 2006 y su progreso puede verse en la página web del proyecto (<http://www.cartociudad.es>). El propósito de Cartociudad es *“permitir la navegación continua por territorio español, localización directa e inversa de direcciones postales, secciones censales y distritos postales, y la búsqueda de un determinado punto de interés dentro de un área geográfica definida”* [7]. Para ello, se incluye una selección de elementos o fenómenos y se dispone de un modelo conceptual cartográfico. En este modelo se utilizan diferentes primitivas geométricas y se recogen y representan los fenómenos y sus atributos, así como las reglas de consistencia geométrica entre fenómenos y las reglas semánticas en el conjunto de datos.

OpenStreetMap, por el contrario, es un proyecto colaborativo iniciado en 2004 en Inglaterra y que desde sus inicios persigue crear una cartografía digital libre y gratuita para los usuarios. Al ser un proyecto colaborativo cualquier usuario puede contribuir con sus datos, que son subidos al servidor de OpenStreetMap como primitivas geométricas (puntos, líneas, polígonos) a las que se asignan etiquetas con las que se acota el tipo de objeto representado. A diferencia de Cartociudad, donde existe un modelo cartográfico con clases, según [6], OpenStreetMap se basa en un modelo denominado marco de descripción de recursos (RDF del acrónimo inglés de *Resource Description Framework*) que depende completamente del etiquetado de las primitivas.

Diversos autores se han ocupado de definir el término de calidad aplicado al ámbito de los datos cartográficos. Van Oort [8] menciona que los primeros trabajos destacables que hacen referencia a la calidad cartográfica datan de las décadas 60 y 70. Sin embargo, debido al *boom* de los Sistemas de Información Geográfica a partir de los 80, el interés en la calidad de los datos cartográficos aumenta y se manifiesta con la inclusión de capítulos específicos dentro de los textos de Burrough y Longley [9, 10] entre otros. Las razones indicadas en el citado artículo de van Oort para este renovado interés son, entre otras, la accesibilidad a los datos, su creciente uso y el incremento de usuarios no especializados.

Para describir la calidad de los datos espaciales se necesitan diversos elementos de los cuales citamos a continuación aquellos que consideramos relevantes para usos de cálculo de rutas y geocodificación, dos de los usos más extendidos de callejeros en internet. La **precisión posicional**, se refiere a la precisión de los valores de las coordenadas medidos de forma relativa (de unos elementos respecto de otros dentro del mismo *set* de datos) o de forma absoluta (precisión entre los valores del elemento representado y su representación). La **precisión de atributos** no espaciales o temporales se ocupa de cómo son de correctos esos datos respecto de los elementos representados. Un ejemplo de error es una carretera de nivel autonómico cuyos atributos indican que se trata de una carretera de nivel municipal. La **consistencia lógica** es un aspecto que evalúa como de correctos son los datos respecto de las relaciones definidas por el modelo de datos y las relaciones topológicas. Un ejemplo de inconsistencia sería una carretera que se superpone a un edificio, o la falta de conectividad de las líneas en una red de carreteras. La **compleción** es un elemento que mide si los datos son incompletos respecto del número de objetos que se

deberían de representar o si por el contrario hay un exceso de los mismos. La **calidad temporal** evalúa la validez de los datos respecto de cambios ocurridos en los fenómenos representados y por lo tanto mide su actualización.

Algunos de los elementos mencionados forman parte de los metadatos y las instituciones productoras de los datos establecen los requisitos de calidad de los mismos. Con diversos protocolos controlan que esos requisitos son cumplidos por los datos. Por ejemplo, se pueden realizar comprobaciones de la precisión posicional y de atributos de una muestra de los fenómenos representados mediante trabajo de campo y se puede establecer qué porcentaje de los datos comprobados ha de ser correcto. La comprobación en campo de los datos de OpenStreetMap está fuera del alcance de este trabajo y por ello la estrategia adoptada es la siguiente: partiendo de que Cartociudad ha pasado los pertinentes controles de calidad, establecer mediante comparación, si la calidad de OpenStreetMap es suficiente para los mismos usos.

PRECISIÓN POSICIONAL

La calidad posicional se ha evaluado a nivel municipal y/o provincial. Se ha aplicado el análisis a los tres tipos de primitivas presentes en ambos productos, cada una de las cuales representa un tipo de datos diferentes. Así por ejemplo, los puntos representan puntos de interés (POI del acrónimo inglés *Point of Interest*) en OpenStreetMap, y topónimos y números de policía o portales en Cartociudad. Las líneas representan calles y carreteras principalmente, pero también ríos y otros elementos lineales. Finalmente los polígonos representan edificios, masas de agua y otros elementos geográficos.



Figura . Ejemplo de extracción de los vértices del entramado urbano de Cartociudad (puntos amarillos) y detalle del error posicional de OpenStreetMap (puntos verdes).

Puntos. Para estimar la calidad posicional de los puntos, se puede obtener la distancia entre pares de puntos seleccionados ([6]). Debido a la poca coincidencia entre la capa de POIs de OpenStreetMap y los Topónimos de Cartociudad, la estrategia adoptada ha sido obtener un entramado de puntos de los vértices de intersección de la red de vías urbanas dentro del municipio de Valencia mediante la planarización de las líneas (Figura).

Vértices del entramado de vías: Una vez obtenidos los vértices de ambos sets de datos, se ha calculado para cada uno de los puntos, el punto que se encuentra a menor distancia dentro del otro set de datos, (Figura). Se ha utilizado la no coincidencia entre los valores de distancias al punto más próximo para descartar los puntos emparejados incorrectamente. Aun así, todavía quedan puntos que quedan emparejados incorrectamente al no ser puntos equivalentes en ambos sets de datos. Es por ello que dentro de aquellos con distancias coincidentes, se ha necesitado seleccionar manualmente aquellos que se encuentran emparejados correctamente. Una vez realizado este análisis se obtiene un histograma con las distancias, ver Figura

. Se puede apreciar que la distancia entre puntos más frecuente es unos 2-3 metros, obteniéndose una media de 4 metros con una desviación estándar de unos 3 metros. Este es un valor menor al obtenido en otros estudios como por ejemplo [6] que encontraron un error de 6 metros. Según [2], este es un valor en el error de posición esperable a partir de los métodos utilizados para la introducción de datos en OpenStreetMap, ya que se encuentra en el rango de error de los receptores GPS más extendidos y de las imágenes satélite más comúnmente utilizadas para digitalización. Los valores de error máximos obtenidos dentro del municipio de Valencia se sitúan en torno a 25 metros, pero se trata de un número reducido dentro de los puntos muestreados.

Puntos de Interés / Topónimos: Los puntos de interés, denominados topónimos en Cartociudad, demuestran como la falta de criterios o especificaciones a la hora de introducir datos en OpenStreetMap (3.215 puntos en el municipio de Valencia), resultan en una marcada diferencia del número de POIs comparándolos con Cartociudad (49 puntos de interés). Debido a la disparidad entre ambos números, comprobaremos si los 49 puntos de Cartociudad se encuentran representados en OpenStreetMap, y en este caso, cual es la diferencia en posición respecto de Cartociudad. De los 49 topónimos de Cartociudad, únicamente 8 se encuentran representados en OpenStreetMap. El error medio de posición obtenido en este caso es de 17 metros (Figura), con una desviación estándar de 10 metros. La baja representación de los puntos de interés de ambos datos se discutirá en el apartado Atributos de este trabajo.

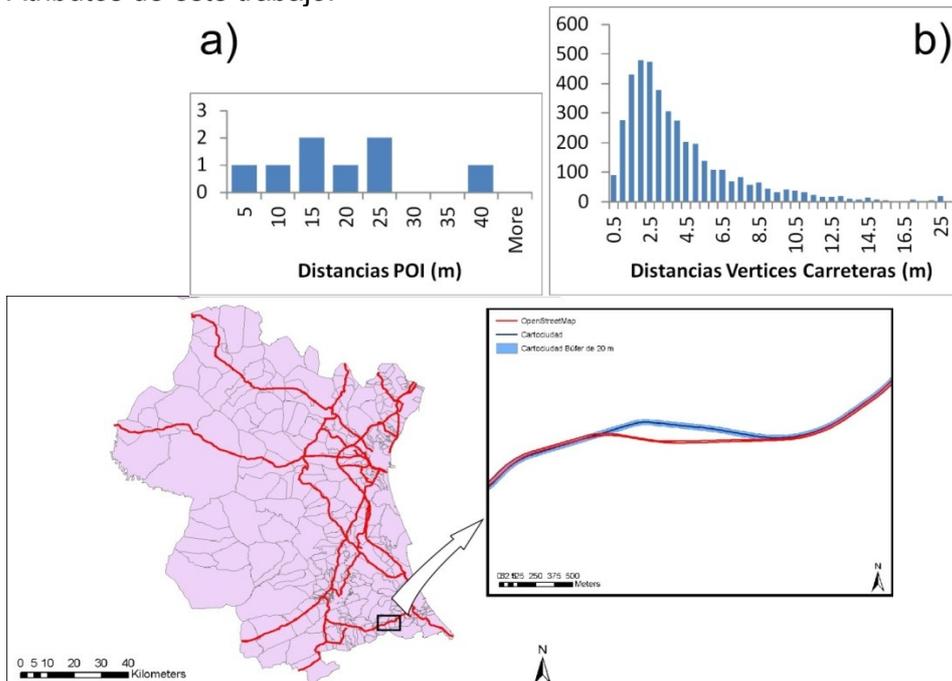


Figura 1. Histograma de las distancias de los vértices de la capa de carreteras de OpenStreetMap ($n=4.083$) y de las distancias de los puntos de interés de OpenStreetMap ($n=8$).

Figura . Red de carreteras (en rojo) utilizadas para el análisis del error posicional. Detalle del búfer de 20 metros (azul) y una zona de error.

Líneas. Se han seleccionado las carreteras principales de la provincia de Valencia, Autopistas y Carreteras Autonómicas de primer nivel (**Figura**). Se aplica el método de Goodchild [11] y utilizado por [2] que consiste en lo siguiente: se crea un búfer de los elementos de la Cartociudad a una distancia que se considera igual al error detectado en la posición de los puntos (4 metros + 3 metros de desviación estándar). Posteriormente, se calcula el porcentaje de longitud de las vías de OpenStreetMap que queda dentro del búfer, y cuyo error posicional es por lo tanto menor que la distancia del búfer. Adicionalmente, se ha creado un búfer con una distancia de 20 metros, que es la distancia utilizado en otros estudios (por ejemplo en [2]) para tener en cuenta que en OpenStreetMap se representan los carriles de ambos sentidos cuando éstos están separados. Del total de 1.344 Km, un porcentaje muy alto de la longitud de las carreteras se encuentra dentro del búfer de 20 metros (1.259 Km, 94 %), y ese porcentaje disminuye, aunque en menos de un 10% cuando consideramos el búfer de 7 metros (1.156 Km, 86 %).

Polígonos. Para medir la precisión posicional de los polígonos se utiliza el método utilizado por [6] basado en Vauglin [12]. Con este método la diferencia entre polígonos se mide como la relación entre el área de intersección de los polígonos respecto del área de unión de los polígonos según la siguiente fórmula:

$$dS = 1 - \frac{S(A \cap B)}{S(A \cup B)}$$

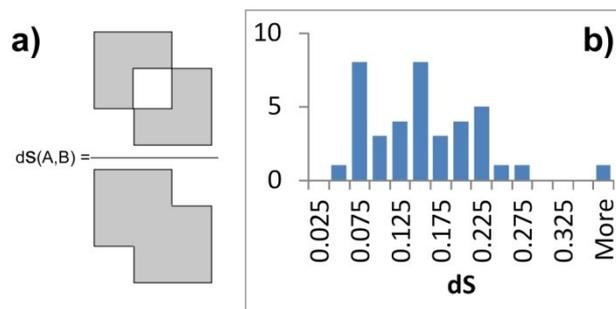


Figura . a) Operaciones geométricas utilizadas para el cálculo del índice dS (modificado de [6]). b) Histograma de los valores dS calculados para los 39 edificios muestreados.

La representación geométrica de la fórmula, con las operaciones implicadas se puede ver en la Figura . Los valores extremos 0 o 1 indican que los polígonos son iguales o no coincidentes en superficie respectivamente. Para aplicar este método se han de escoger y emparejar manualmente pares de elementos que representan las mismas superficies. Sin embargo, se ha encontrado un problema a la hora de emparejar los polígonos de ambos datos. En Cartociudad los elementos urbanos poligonales, corresponden a la digitalización de las manzanas urbanas provenientes de la Dirección General de Catastro, mientras que en OpenStreetMap se han digitalizado edificios, pero no de forma homogénea. No siempre se da una correspondencia de los datos ya que no se está representando el mismo elemento. Un ejemplo puede verse en la Figura . Es por ello que de los 305 polígonos dentro de la capa de edificios de OpenStreetMap se han elegido 39, que se han identificado con su equivalente en Cartociudad. En la mayoría de los casos, se trata de fincas dentro del casco histórico de la ciudad, donde la manzana coincide con el perímetro del edificio. La aplicación de la fórmula para los edificios seleccionados nos da unos resultados que se ven en la Figura , con un valor de dS medio de 0,15 y una desviación estándar de 0,07. De los edificios escogidos, el valor máximo de dS corresponde a 0,35. De acuerdo con la discusión de [6], el valor obtenido indica una diferencia pequeña entre los dos datos. El mencionado autor cita que en un estudio que comparaba datos de

mapas topográficos y de catastro de Francia, el 70% de los polígonos analizados se sitúan en torno a 0,15 y 0,45. Otro parámetro que se puede medir para comparar ambos *set* de datos es el de compacidad según un índice de Miller dado por la siguiente fórmula:

$$C = \frac{2\pi Area}{Perímetro^2}$$

El índice de compacidad, nos permite evaluar la diferencia de formas entre los edificios de Cartociudad y de OpenStreetMap. Los resultados se pueden ver en la Tabla , donde se puede observar que las diferencias en las formas de edificios seleccionados son de hecho muy pequeñas.

	OpenStreetMap	Cartociudad
Máximo	0,426684	0,428946
Mínimo	0,153408	0,152767
Media	0,345672	0,339181
Desviación Estándar	0,053799	0,057333

Tabla . Valores estadísticos de la compacidad de los edificios analizados.

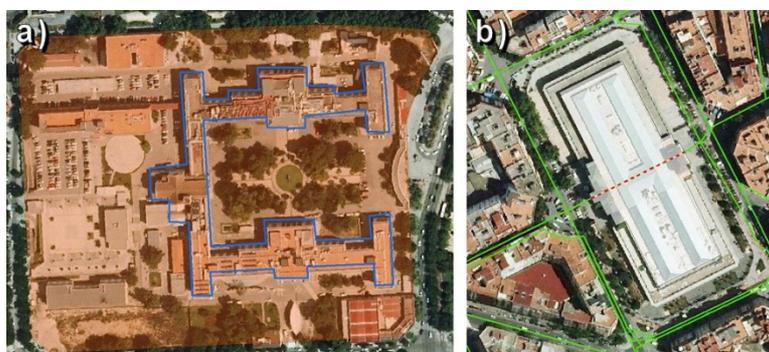


Figura . a) Discrepancia entre manzana de Cartociudad (rojo) y perímetro del edificio en OpenStreetMap (azul). b) Ejemplo de solapamiento entre polígono y línea de OpenStreetMap. Imagen de fondo: PNOA.

CONSISTENCIA LÓGICA

Se ha realizado un sencillo test de consistencia interna a nivel de municipio para detectar posibles errores del modelo de datos de OpenStreetMap. Debido a la carencia de un modelo explícito de datos como tal, el análisis propuesto es muy básico y en todo caso permite detectar si existe solapamiento entre fenómenos que en principio no deberían de solaparse. Este es el caso de por ejemplo vías (urbanas o interurbanas) y los polígonos que representan edificios. Se han detectado 11 tramos de vías que se solapan a diferentes polígonos de edificios y que representan un total de 308 m. Un ejemplo puede verse en la Figura . Es importante mencionar que un solapamiento podría deberse a pasos subterráneos, o a edificios puente. Sin embargo ninguno de los atributos de los elementos detectados contiene una descripción que indique tal característica y por lo tanto se consideran todas inconsistencias, ya sea de tipo geométrico o por omisión de atributos. El total de longitud solapada (308 m) es un porcentaje marginal si se tiene en cuenta que la suma de la longitud de las líneas dentro del término municipal representadas en OpenStreetMap es de 1.437 Km. Sin embargo, hay que tener en cuenta, que los polígonos de edificios, representan un área también marginal dentro del área de estudio (termino municipal de Valencia). En total hay 305 elementos poligonales que representan edificios, y que suponen un área de 98 Has., mientras que dentro de Cartociudad, hay 4.878 polígonos que representan manzanas y que suponen un área de 2.669 Has. Por lo tanto el área cubierta por polígonos de OpenStreetMap, representa únicamente un 4% del área cubierta por polígonos de Cartociudad. Es de esperar pues, que a medida que la representación en forma de primitiva de polígono aumente, también aumente el número de solapamientos. Dentro de la consistencia lógica de los datos también se ha evaluado la conectividad de las líneas. Una conectividad adecuada es necesaria para poder

crear una red que sirva para realizar los cálculos de rutas. Para evaluar la conectividad se han obtenido los vértices de los extremos de las líneas y se ha calculado el número de líneas que intersectan los puntos. Posteriormente, se ha creado un búfer de 5 metros sobre los puntos y sobre ese búfer se han calculado el número de líneas que intersectan. Asumiendo que la distancia usada para el búfer es el rango de error posicional, las diferencias en el número de líneas que intersectan, se asume que es debido a la no conectividad de las líneas. De 6.364 vértices que se intersectan con una única línea, 734, intersectan con más de una cuando se realiza un búfer. Si esto se interpreta como debido a error de conectividad, podemos concluir que un 12% de los puntos analizados tienen problemas de conectividad.

COMPLECIÓN

La comparación de compleción se realiza a nivel de provincia y se utilizan los municipios como elementos comparativos en vez de una red regular tal y como se hace en diversos trabajos, [2, 5, 6]. Esta división municipal se utiliza con el objetivo de poder comparar los resultados obtenidos con datos socioeconómicos disponibles en el Instituto de Estadística de la Generalitat Valenciana (<http://www.ive.es/>) y analizar si hay alguna correlación espacial entre ambos. Para evaluar la compleción se comparan elementos lineales y puntuales. Tal y como se puede apreciar en los mapas de la Figura , hay diversos municipios en los que la suma de la longitud de los elementos lineales de OpenStreetMap, supera a la de Cartociudad. Hay que tener en cuenta que el catalogo de fenómenos representados en Cartociudad está reducido y limitado a aquellos fenómenos que se consideran de utilidad para su propósito (navegación) y que no se incluyen algunos tipos de elementos como pueden ser pistas no transitables por vehículos, senderos, caminos peatonales en parques etc. Por lo tanto, si OpenStreetMap contuviera toda la información representada en Cartociudad, además de los mencionados elementos “extra”, la suma de longitudes debería de ser siempre superior en el caso de OpenStreetMap. Sin embargo, apreciamos en la Figura , que esto no es así, y que únicamente en algunos municipios es tal el caso. Una observación similar se podría hacer con los puntos de interés, ya que como hemos visto para el municipio de Valencia, la diferencia es enorme (3.215 frente a 49), pero no sucede lo mismo en todos los municipios. Estas diferencias, en cuanto a municipios con más elementos cartografiados se pueden comparar con el mapa de densidades, Figura . A grandes rasgos, se aprecia una correlación entre municipios de mayor densidad y aquellos que contienen más elementos cartografiados en OpenStreetMap. En este sentido cabe mencionar algo que ya se sugiere en Hacklay [2], y que el autor denomina la división social y digital en la VGI. Áreas rurales, más deprimidas, o aquellas que no despiertan el interés de los voluntarios se quedan con una cobertura cartográfica menor que las áreas urbanas o aquellas con mayor presencia de potenciales cartógrafos voluntarios.

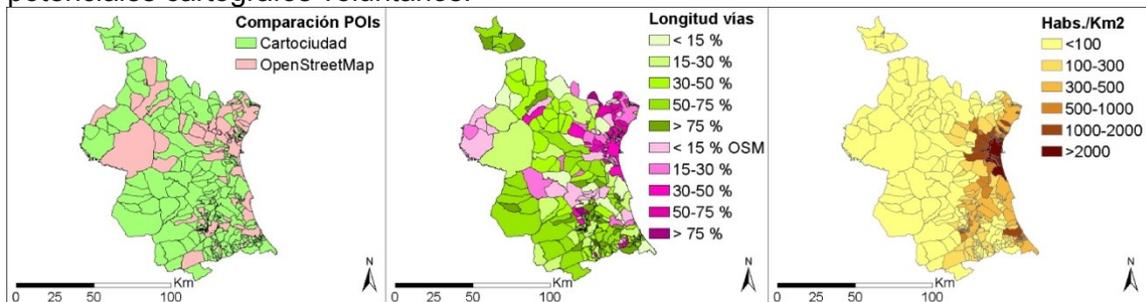


Figura . Mapas comparando la diferencia en longitud de las vías y la cantidad de puntos de interés (Verde=Cartociudad > OpenStreetMap). Mapa de densidad de los municipios de la provincia.

ATRIBUTOS

Debido a la diferencia en el modelo de datos, no se puede realizar una evaluación de los atributos de OpenStreetMap con Cartociudad de una manera directa. Es por ello que la evaluación de la calidad de los atributos de OpenStreetMap ha consistido en determinar cómo están de completos esos atributos, sin valorar si son correctos o no. Por lo tanto se ha analizado el porcentaje de elementos no clasificados en OpenStreetMap.

Puntos de Interés/Topónimos. Los POIs de OpenStreetMap tienen un campo de clasificación de Categoría, que en el caso del municipio corresponde a 10 categorías diferentes, ver Figura . Además hay un campo cuya etiqueta aporta mayor detalle en la categoría y el nombre. Un ejemplo es un punto que corresponde a la categoría “Automotive” y cuyo nombre es “Fuel: La Dehesa”, de esta manera dentro de la categoría de automoción, forma parte de una sub-categoría “Fuel”. Ambos atributos se encuentran todos rellenos, por lo tanto todos los puntos quedan identificados y simbolizados. Por el contrario, los topónimos disponibles en Cartociudad, disponen únicamente de un campo de texto, con el nombre asignado, y no se asignan a diferentes categorías. Se ha considerado de interés establecer una clasificación de los 49 puntos de Cartociudad, en 8 categorías diferentes, para representar al igual que con los datos de OpenStreetMap, el porcentaje de cada una de las categorías. Esta clasificación no es directamente comparable con los datos de OpenStreetMap, pero da una idea de la distribución en categorías de los datos de Cartociudad. El resultado puede verse en la Figura . Si analizamos la correspondencia entre OpenStreetMap y Cartociudad teniendo en cuenta la clasificación en categorías que hemos realizado (2), veremos que mientras que los puntos de Cartociudad dentro de Museos, Teatros y Cultura, así como los de Transporte y Automoción tienen una representación alta en OpenStreetMap, los puntos dentro del resto de categorías no están representados dentro de OpenStreetMap (al menos como primitivas de puntos). Sin embargo, aunque no sea el caso, podría encontrarse representados como polígonos de edificios. Estos datos nos indican la disparidad en cuanto a lo que se consideran puntos de interés en Cartociudad y se representa en el callejero, y lo que por el contrario los voluntarios geográficos encuentran interesante y por lo tanto cartografían para que quede incluido en OpenStreetMap.

Categoría	Cartociudad	Representación en OpenStreetMap	
	Cantidad	Cantidad	Porcentaje
Deportes	5	0	0%
Educación	12	1	8%
Espacios Verdes	5	0	0%
Gobierno y Inst. Públicas	10	0	0%
Museos, Teatros, Cultura	4	3	75%
Salud	3	1	33%
Templo Religioso	2	0	0%
Transporte y Automoción	8	5	62%

Tabla . Tabla de clasificación de los topónimos de Cartociudad y su representación en OpenStreetMap.

Carreteras, Espacios Verdes y Edificios. Se han analizado los datos de carreteras y vías a nivel de provincia y se ha determinado cómo están de completos dos atributos: el tipo de vía y el nombre de la misma. Los resultados pueden verse en la Tabla . Mientras que Cartociudad dispone de prácticamente el 100% de sus datos etiquetados, no pasa lo mismo con OpenStreetMap, donde casi un 10% de los tramos

lineales se encuentra sin clasificar, y casi un 60% no tiene un nombre atribuido. En la misma tabla se pueden ver los resultados de los atributos de espacios verdes y de edificios de OpenStreetMap a nivel de municipio. Se aprecia la desigualdad en cuanto a etiquetado. Mientras que los espacios verdes están etiquetados al 100% en su atributo tipo, no pasa lo mismo con los edificios o con el campo nombre de ambos datos.

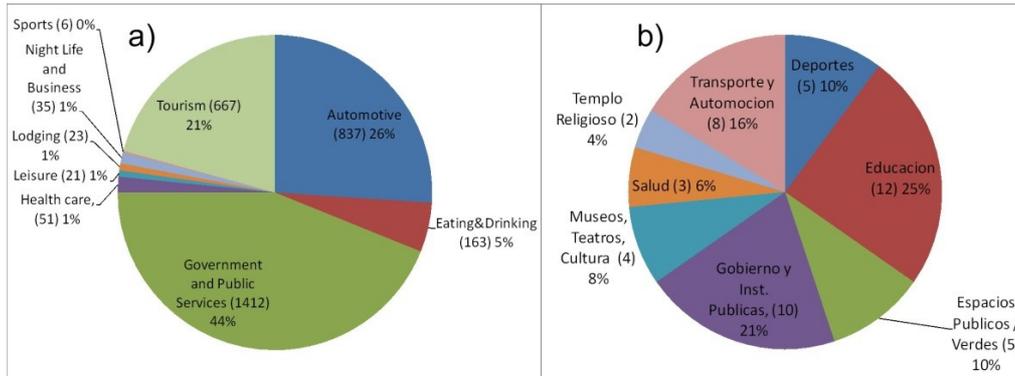


Figura . a) Clasificación temática de los 3.215 POIs de OpenStreetMap del municipio de Valencia. b) Clasificación temática de los 49 Topónimos de Cartociudad del municipio de Valencia.

	Longitud o Total	Tipo (vacío)	Nombre (vacío)
OpenStreetMap	12.434 Km	1.087 Km	7.285 Km
Cartociudad	13.112 Km	466 Km	0 Km
Espacios Verdes	386	0	299
Edificios	305	282	219

Tabla . Longitud en kilómetros y % de tramos o carreteras sin clasificar o sin nombre y número y % de elementos sin etiquetas del tipo Espacios Verdes y Edificios de OpenStreetMap.

CALIDAD TEMPORAL

Una de las ventajas que se suele atribuir a la VGI, son las posibilidades de actualización de sus datos, ya que son los mismos usuarios los que detectan y corrigen los errores y carencias sin necesidad de pasar por un largo proceso de revisión. El proyecto Cartociudad, introdujo los datos relativos a la provincia de Valencia en el año 2008, por lo tanto es evidente que no se incluirán actualizaciones posteriores a ese año. Sin embargo la red de carreteras se basa principalmente en la cartografía 1:25.000 del Instituto Geográfico Nacional, cuyos años de edición para las hojas que cubren el municipio de Valencia van desde el año 1984 al 2005. Es por lo tanto destacable, que en algunos casos, se trata de información muy anterior a 2008. Dentro del municipio de Valencia, se han detectado al menos ocho tramos de carreteras construidas anteriormente a 2008 que suman un total de 9 Km y que no se encuentran incorporadas en Cartociudad, pero si en OpenStreetMap. Dos ejemplos se muestran en la Figura .

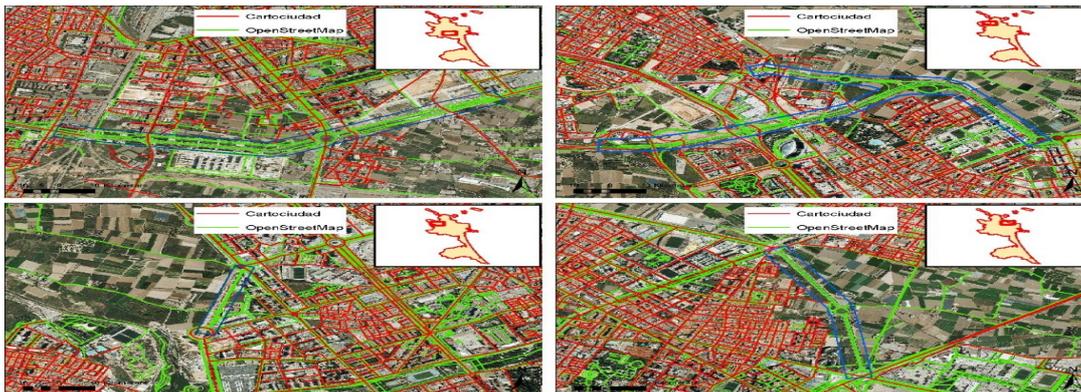


Figura . Ejemplos de carreteras construidas antes de 2008 y no contempladas en Cartociudad, pero que sí se incluyen en OpenStreetMap. El tramo en cuestión se señala con un polígono azul.

GEOCODIFICACIÓN Y RUTAS ÓPTIMAS

Uno de los usos atribuidos a Cartociudad es el de la geocodificación directa e inversa de direcciones. Para ello Cartociudad dispone de los números de policía o portales proporcionados por Catastro. Esto permite la geocodificación directa, introduciendo el nombre de una vía y el número de portal o punto kilométrico en el caso de carreteras, obtener las coordenadas del punto correspondiente. Esta es una de las acciones fundamentales de cualquier navegador también para el cálculo de rutas óptimas. Por el momento, los números de portales no están muy extensamente representados en OpenStreetMap. No existe un elemento específico para la creación de número de portales, y su introducción se realiza como en el resto de elementos mediante la asignación de la etiqueta apropiada. Para crear números de portal se utiliza la etiqueta *addr:housenumber*. La escasez de números de portales se puede comprobar analizando la cantidad de etiquetas *addr:housenumber* existentes dentro del estado español (<http://tagwatch.stoecker.eu/>), ver Tabla .

	España		Alemania	
	Julio 2011	Enero 2012	Julio 2011	Enero 2012
Valores Diferentes	3.149	-	15.576	-
Total Usados	123.555	127.317	1.700.078	2.101.185
Usado en Nodos	114.035(92%)	116.918 (92%)	801.476(47%)	894.576 (43%)
Usado en Rutas	5.103(4%)	5.903 (5%)	897.843(53%)	1.205.601 (57%)
Usado en Relaciones	4.417(4%)	4.496 (4%)	759 (0%)	1.008 (0%)
	Diferencia:	3.762(%)	Diferencia:	401.107 (24%)

Tabla . Estadísticas de la utilización de la etiqueta *addr:housenumber* dentro de OpenStreetMap para España y Alemania (fuente: <http://tagwatch.stoecker.eu> , consultado el 9/07/2011 y el 30/01/2012).

Hay un método cuyo uso se ha extendido para la generación y etiquetado de números de portal y que se denomina *método Karlsruhe* ya que es en esa ciudad alemana donde se comenzó a etiquetar de esta manera. Este método utiliza no solo elementos puntuales para asignar números a las casas, sino también elementos lineales sobre los que se introducen números, para interpolarlos linealmente sobre el segmento de la línea. Esta forma de añadir los números de portales se puede observar en diversos municipios pero sobre todo en el mapa OpenStreetMap de Karlsruhe donde se pueden apreciar los números de portales introducidos mediante las diferentes formas. Por un lado, hay números sueltos, y por otro se pueden ver algunos números en los extremos de líneas discontinuas. Se observa como a diferencia de España, el número de etiquetas *addr:housenumber* es mayor en Alemania (Tabla) y al menos la mitad de ellas se encuentran sobre rutas (líneas), tal y como se introducen según el *método Karlsruhe*.

Aunque la ausencia de etiquetas de portales o números de policía podría poner en cuestión el potencial de OpenStreetMap para la geocodificación y cálculo de rutas óptimas, existen soluciones. Uno de los ejemplos que resuelve este tema en la actualidad es OpenRouteService <http://www.openrouteservice.org/> que permite no solo geocodificar direcciones usando los datos de OpenStreetMap, sino que también permite el cálculo de rutas óptimas para coches, peatones y ciclistas. La geocodificación utilizada por OpenRouteService se basa en el método descrito en [13-15]. El autor utiliza un enfoque basado en la probabilidad para hacer una estimación efectiva de los números de policía. La aproximación se basa en tres parámetros diferentes: distancia media entre dos números de portales, el *offset* entre la primera casa y el inicio de la calle y la dirección de la calle, utilizando como restricción la numeración secuencial alternante extendida en la mayor parte de Europa. Gracias a estos parámetros, se consigue una estimación de la posición de un número de portal que no está presente en los datos. El estudio concluye que es posible aproximar de manera efectiva la posición de números de portal mediante dos enfoques de probabilidad diferentes, a pesar de que los resultados dependen completamente de la zona y de los datos disponibles para la zona. El error posicional si se compara con los datos cuando están disponibles, puede llegar a ser de 142 metros cuando no se dispone de suficientes datos en la zona, y de 31 metros cuando se dispone de datos sobre los que interpolar, sin embargo donde se dispone de suficientes datos, el error puede ser inferior al error obtenido con la herramienta de geocodificación proporcionada por Google [15].

La ausencia total de números de portal en el área de Valencia de OpenStreetMap compromete la efectividad de los datos para geocodificación en el área de estudio, y de hecho según pruebas realizadas en la zona, al introducir un número de portal en la búsqueda de <http://www.openrouteservice.org/> se obtienen resultados muy diferentes de los de Cartociudad. Sin embargo, es esperable, que a medida que los datos se vayan nutriendo de los números de portales necesarios, se puedan realizar geocodificaciones adecuadamente. En la Tabla se puede ver la evolución de la etiqueta *addr:housenumber* desde la realización del análisis. Mientras que en España apenas se da un crecimiento del 3%, en Alemania es del 24%. Además, vemos que en España, el uso de la etiqueta *addr:housenumber* sobre rutas (tal y como se hace con el método Karlsruhe) sigue siendo marginal (ha pasado de 4% a 5%) comparado con Alemania, donde ha aumentado de un 53% a un 57%.

[1]

[2] CONCLUSIONES

Los resultados indican que OpenStreetMap tiene una precisión posicional más que aceptable para usos de geocodificación y cálculo de rutas óptimas y se ha detectado que existe una mayor actualización de los datos de OpenStreetMap que de los de CartoCiudad. Sin embargo se han encontrado algunos aspectos que pueden poner en compromiso la idoneidad de OpenStreetMap para los mencionados usos. Por un lado se ha detectado una carencia importante en la compleción de los etiquetados de ciertos elementos y teniendo en cuenta que OpenStreetMap no tiene un modelo cartográfico conceptual con clases, el etiquetado de las primitivas es necesario. También se han detectado ciertas inconsistencias, como solapamientos entre elementos geométricos o falta de conectividad en la red de calles. Se ha observado también que la cobertura de los datos de OpenStreetMap es muy heterogénea y tiene una compleción muy desigual por zonas dándose mayor diferencia positiva a favor de OpenStreetMap en las zonas más densamente pobladas.

Se puede decir que la falta de un modelo cartográfico conceptual, la heterogeneidad en cuanto a personas que contribuyen al proyecto y la diversidad de fuentes de datos que se utilizan en OpenStreetMap pueden ser los principales factores

que resultan en estos mencionados *handicaps*. Estas observaciones podrían poner en compromiso su uso para labores de geocodificación, cálculo de rutas óptimas etcétera que requieren una consistencia geométrica y de atributos, y no solo una buena precisión posicional y temporal. Sin embargo, el proyecto OpenStreetMap es relativamente joven, y la cantidad de información contribuida desde 2004 es enorme, igual que lo es su potencial. Algunos autores como [6], reflexionan, que es quizás hora de encontrar un equilibrio entre la libertad de acción de los que contribuyen al proyecto y unas especificaciones que ayuden a solucionar los problemas mencionados, si se quiere que OpenStreetMap sea una base de datos fiables para ciertos usos. A pesar de todo, se ha visto un ejemplo de cómo con creatividad se pueden solventar ciertas carencias, como es el caso de la disponibilidad de número de portales, y se pueden utilizar los datos de OpenStreetMap como cartografía de base para una plataforma de geocodificación y cálculo de rutas óptimas (www.openrouteservice.org).

El análisis presentado se ha realizado con datos de Junio de 2011, pero desde entonces OpenStreetMap ha seguido creciendo. Es necesario destacar la evolución de OpenStreetMap en este lapso de tiempo y así poder contextualizar las conclusiones que se derivan de este estudio. La Tabla recoge una comparación entre dos fechas: Junio de 2011 y Enero de 2012. Se aprecia un aumento en mayor porcentaje dentro de la provincia que dentro del municipio. Teniendo en cuenta que las zonas urbanas son las que se han cartografiado más intensamente desde el inicio y que aún quedan grandes áreas con peor cobertura fuera del municipio, esta diferencia de ritmo es esperable. Sin embargo, el número de edificios muestra una tendencia inversa y un aumento más que notable. El cómo han evolucionado los diferentes elementos de calidad cartográfica estudiados en este lapso de tiempo no se aborda en el presente trabajo, pero es probable que la inclusión de fenómenos representados mediante primitiva de polígonos suponga un impacto negativo en la consistencia lógica de los datos.

	Municipio			Provincia		
	Junio 2011	Enero 2012	Diferencia	Junio 2011	Enero 2012	Diferencia
Número POIs	3.215	3.336	121(4%)	5.405	6.500	1.095 (20%)
Número Edificios	331	612	281 (85%)	705	1.059	354 (50%)
Kms Carreteras	1.437	1.544	107(7%)	12.434	13.690	1.256(10%)

Tabla . Evolución de los datos de OpenStreetMap entre Junio de 2011 y Enero de 2012

REFERENCIAS

[3]

[4] [1] Goodchild, M.F., Citizens as sensors: the world of volunteered geography. *GeoJournal*, 2007. 69(4): p. 11.

[5] [2] Haklay, How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning B: Planning and design*, 2010. 37: p. 23.

[6] [3] Goodchild, M.F., Beyond metadata: towards a user-centric description of data quality, 5th international symposium on spatial data quality. 2007: Eindhoven, Netherlands. p. 11.

[7] [4] Goodchild, M.F., Assertion and authority: user-generated geographic content, GI-Forum 2008. Salzburg.

[8] [5] Zielstra, D.a.Z., A comparative study of proprietary geodata and volunteered geographic information for germany, Agile 2010.

[9] [6] Girres, J.-F. & Touya, G., Quality assessment of the french OpenStreetMap dataset. *Transactions in GIS*, 2010. 14(4): p. 435-459.

[10] [7] González, R., Velasco, González, Ruiz, Especificaciones del producto Cartociudad v 8.0. 2010.

[11][8] van Oort, P., Spatial data quality: from description to application, Centre for geo-information. 2006, PhD, Salzburg.

[12] [9] Longley, G., Maguire, R., *Geographic information systems and science*. 2001: John Wiley & sons 472.

[13] [10] Burrough, P.A., *Principles of geographical information systems for land resources assessment*. 1986, Oxford: Oxford University press. 129.

[14] [11] Goodchild, M.F.& Hunter, G.J., A simple positional accuracy measure for linear features. *International journal of geographical information science*, 1997. 11(3): p. 8.

[12] Vauglin, Modèles statistiques des imprécisions géométriques des objets géographiques linéaires. 1997, Marne-la-Vallée University.

[15] [13] Amelunxen, C., An approach to geocoding based on volunteered spatial data, in Department of geography, chair of Geoinformatics,. 2009, University of Heidelberg: Heidelberg. p. 124.

[16] [14] Amelunxen, C., An approach to geocoding based on volunteered spatial data, *Geoinformatik 2010*. 2010: Kiel.

[15] Amelunxen, C., On the suitability of volunteered geographic information for the purpose of geocoding, GI-Forum 2010. 2010: Salzburg.