

El uso de OSM como herramienta para la identificación de núcleos poblados

Fernando Pino¹, Vladimir Bellini¹, Leandro Stryjek y Martín Moroni¹

¹ Tecnología de la Información, Secretaría de Energía de la Nación. Av. Paseo Colón 171 – 9º Piso, oficina 905, Ciudad Autónoma de Buenos Aires Tel: (011) 43497209 {[mmoroni](mailto:mmoroni@iminplan.gob.ar), [lstryjek](mailto:lstryjek@iminplan.gob.ar), [vbellini](mailto:vbellini@iminplan.gob.ar), fpino@iminplan.gob.ar}

Resumen: A partir de la necesidad de contar con información geográfica a nivel de localidad, se realizó un proyecto que consistió en utilizar calles vectoriales provenientes de Open Street Map (OSM), principalmente en su tipología “residencial”, con una longitud menor a 2km, para identificar envolventes de Núcleos Poblados en forma automática.

Como resultado se obtuvo un conjunto de envolventes al cual se le asignó información del BAHRA, del Censo 2010 y de otras fuentes.

Para esto se utilizó la metodología basada en los trabajos de Cesar Vapñarsky para identificación de aglomerados censales, logrando un conjunto de núcleos poblados útiles para agregar información y planificar políticas públicas.

Este trabajo constituye un avance con respecto a una charla presentada por este equipo en la conferencia State of the Map 2014 realizada en Buenos Aires el día 8/11.

Palabras Claves: Núcleos poblados, aglomerados, BAHRA, políticas públicas, OSM.

1. INTRODUCCIÓN

La Secretaría de Energía de la Nación, en el desarrollo de sus labores diarias, se ve cotidianamente en la necesidad de manejar grandes volúmenes de información geográfica proveniente de múltiples fuentes, la cual muchas veces debe ser georreferenciada por el propio personal de la Secretaría.

De esta manera, se plantea la necesidad de relacionar múltiples datos geográficos, lo cual exige tener una entidad geográfica que permita vincular los distintos datos existentes.

Se estimó que una capa de aglomeraciones de Argentina podría cumplir ese objetivo, pero al indagar sobre el tema se advirtió que no se encuentra disponible hasta el momento una base actualizada de polígonos de localidades de Argentina. Como alternativas existen los puntos publicados por el programa BAHRA y los ejidos del IGN, los cuales no son exhaustivos.

Una de las pocas fuentes de información geografía actualizada y de acceso público en Argentina es Open Street Map (OSM), una plataforma donde distintos usuarios editan y cargan información, principalmente calles y puntos de interés, las cuales están disponibles públicamente y hoy pueden ser consideradas como la base de datos espacial pública más importante de Argentina.

Esta información es validada por una serie de aplicaciones (Keep Right, Osmose, JOSM, Validator, OSM Inspector, Maproulette) creadas para este propósito, además de la colaboración de muchos usuarios que realizan chequeos y correcciones.

A partir de la necesidad de contar con núcleos poblados donde se asienta población y de vincularla con otras fuentes (datos censales, datos energéticos, etc.), se trazó como objetivo el contar con información precisa de los aglomerados de Argentina.

2. MATERIALES Y METODO

A partir de los datos públicos de OSM, se decidió elaborar una base de polígonos de localidades de Argentina utilizando las calles cargadas en OSM.

Se parte de la noción de aglomeración o núcleo poblado como un territorio donde se asienta población para vivir con cierta contigüidad, es decir que se toma la acepción estrictamente física del concepto "localidad".

Dicha acepción define a la localidad como *"la existencia de un conjunto de constructos materiales fijos al territorio, normalmente un entramado de edificios y calles. Según este criterio, una localidad se define como una aglomeración."* (Bertoncello 2004).

En función de este concepto de localidad, se buscó reproducir, con ciertas licencias, los parámetros metodológicos que establece el INDEC para la detección de localidades (Marcos, 2011).

Esta metodología, propuesta originalmente por Cesar Vapñarsky en 1991 (Marcos, 2011), se basa en la identificación de un continuo de edificios y calles que contenga 4 manzanas (con algunas salvedades) o varios mosaicos de 4 manzanas que se encuentren a menos de 1000 metros entre si.

Una diferencia importante entre la presente metodología y la propuesta por Vapñarsky (Marcos, 2011) es que al utilizar OSM, la entidad geográfica tomada para definir una localidad son las "calles".

En OSM se considera como “calle” aquella entidad que el usuario de OSM decidió definir y conceptualizar como “calle”, lo cual marca una ventaja: la carga masiva de datos; y una desventaja: la fiabilidad de la información, dada por la falta de un criterio común a la hora de definir una entidad geográfica, más allá de lo que existe en la documentación de OSM.

En este punto radica una diferencia importante entre ambas metodologías: mientras en el caso de Vapñarsky (Marcos, 2011), la entidad geográfica que sirve para la definición de la localidad es construida por el operador, en el segundo caso es heredada de lo que cargan los usuarios de OSM.

En función de estas diferencias, se tomó la decisión de utilizar los segmentos de calles de OSM sólo en sus tipologías urbanas (“residential”, “pedestrian” y living_street”), con menos de 2 km de longitud (para evitar incluir rutas o caminos cargados con una tipología incorrecta) y en un conjunto de al menos 10 segmentos que no disten entre sí más de 1000 m.

De esta forma se intenta reproducir, parcialmente, la metodología de Vapñarsky, ya que un mosaico de 4 manzanas puede contener de 6 a 8 calles según su disposición. Debido a que la existencia de calles no garantiza manzanas edificadas, se decidió considerar 10 segmentos de calle como el valor mínimo para identificar una aglomeración, la cual, con la nueva metodología llamaremos Núcleo Poblado.

A continuación se ilustra lo comentado en los párrafos precedentes.



Figura 1: Puntos del BAHRA y ejido urbano del IGN sobre Google Maps. Saladillo, Prov. De Buenos Aires

La figura 1 permite apreciar la ausencia de cartografía en uno de los servidores de mapas más difundidos (Google Maps), más los puntos del BAHRA y el ejido urbano del IGN disponible para la Ciudad de Saladillo (Buenos Aires)



Figura 2: Puntos del BAHRA y ejido urbano del IGN sobre imagen satelital. Saladillo, Prov. De Buenos Aires

La figura 2 la imagen satelital de Google Earth (2015), frente a las mismas capas de la figura 1. Como se aprecia, el ejido urbano actual excede al delimitado por la capa de ejidos urbanos del IGN.



Figura 3: Puntos del BAHRA, ejido urbano del IGN y calles de OSM sobre imagen satelital. Saladillo, Prov. De Buenos Aires



Figura 4: Puntos del BAHRA, ejido urbano del IGN y calles de OSM (con calles filtradas por tipología y longitud de 2 km) sobre imagen satelital. Saladillo, Prov. De Buenos Aires

Las figuras 3 y 4 incluyen las calles de OSM. Puntualmente la figura 4 muestra las calles que, filtradas por tipo y longitud, permitieron a partir de la utilización de nuestra metodología, identificar núcleos urbanos en forma más precisa. Además se observa el nivel de actualización de OSM frente a una imagen satelital actual.

Al conjunto de calles definido bajo los criterios precedentes se les aplicó un proceso (Concave_Hull) para generar una envolvente que encierre todas las geometrías -calles filtradas y seleccionadas- dentro de una nueva geometría.

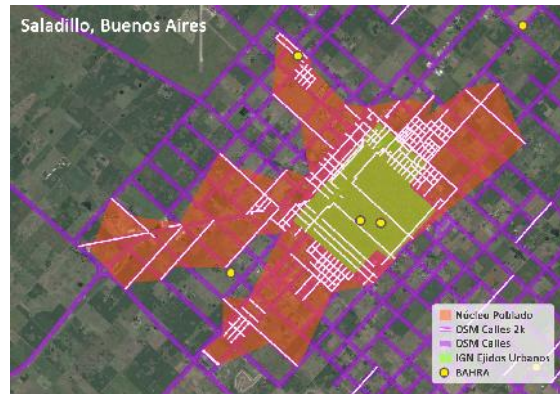


Figura 5: Puntos del BAHRA, ejido urbano del IGN, calles de OSM (con calles filtradas por tipología y longitud de 2 km) y envolvente obtenido sobre imagen satelital. Saladillo, Prov. De Buenos Aires

La figura 5 muestra un nuevo polígono: el Núcleo Poblado resultante, el cual evidencia ser una entidad más actualizada que los ejidos del IGN pero a su vez presenta algunas extensiones de la mancha que pueden considerarse exageradas. Esto evidencia las ventajas y desventajas de este proceso.

A los polígonos obtenidos se les asignaron los nombres de múltiples fuentes. En orden, las fuentes de las cuales surgieron los nombres fueron la base de asentamientos del BAHRA, de Centros Poblados del IGN, de Ejidos Urbanos del IGN, de Countries de Wikimapia y de "Places" de OSM.

El sentido de heredar los nombres de estas diferentes fuentes y en el orden mencionado, obedece a que el proceso por el cual se asignan los nombres de las localidades es una consulta espacial. Por lo tanto, cuando hay corrimientos o falta de puntos de una base inmediatamente se busca información en la siguiente. De esta manera se lograron asignar nombres a prácticamente el total de las localidades identificadas. Esta asignación de nombres permite que el producto generado sea interoperable entre dichas fuentes.

Utilizando esta metodología se logra obtener el área de 2327 núcleos poblados en todo el territorio Argentino. Esta cantidad es inferior a la que nos presenta el BAHRA y similar a la de ejidos urbanos de IGN.

Sin embargo, si se filtran en BAHRA todas las zonas rurales, los sitios edificados y las entidades municipales, el resultado de dicha consulta es similar a la cantidad de núcleos poblados obtenidos por nuestra metodología. Este hecho valida nuestros resultados.

Por otro lado, al comparar ambas bases de datos se advierte que en el BAHRA, existen múltiples tipologías de asentamientos humanos, incluyendo a

asentamientos en desuso. Dicha diferencia puede observarse en el caso del AMBA, donde el BAHRA presenta tanto entidades administrativas (“Gerli”), de gobierno local (“Ciudad Autónoma de Buenos Aires”) y sitios edificados (“Carupa”). Esto permite dar cuenta de la complementariedad de ambas herramientas donde la base de Núcleos Poblados solo refleja las localidades en su sentido físico, como lo muestra la siguiente figura:



Figura 6: AMBA en base a OSM, asentamientos del BAHRA e imagen satelital

Posteriormente los Núcleos Poblados fueron vinculados con ciertos datos censales y energéticos de interés particular para la Secretaría de Energía. La información del Censo 2010 fue vinculada mediante consultas espaciales con los radios censales provistos por el INDEC, asignando información solo de aquellos radios que superen en 50% la intersección de superficies con los polígonos generados.

Al mismo tiempo se les asignó a los polígonos información energética proveniente de otras fuentes. El total de la información asignada a los núcleos se resume en el siguiente cuadro:

Tabla 1: Información vinculada a los núcleos poblados

| Fuente | Información |
|-----------------------|---------------------------------------|
| Censo 2010 | Habitantes |
| Censo 2010 | Viviendas |
| Censo 2010 | Hogares |
| Censo 2010 | Hogares con NBI (%) |
| Censo 2010 | Hogares con red pública de agua |
| Censo 2010 | Hogares que utilizan garrafa |
| Censo 2010 | Hogares con desagüe público |
| Secretaría de Energía | Cooperativa Eléctrica Local |
| Secretaría de Energía | Empresa distribuidora de electricidad |

Esta capa de información se encuentra publicada en el visor SIG web de la Secretaría de Energía, para ser consultada públicamente por cualquier usuario en: <http://sig.se.gob.ar/>



Figura 7: Núcleos Poblados visualizados en el visor SIG web de la Secretaría de Energía

3. EQUIPAMIENTO Y TECNOLOGÍA UTILIZADA

En cuanto a la Geodatabase, se decidió alojar los datos espaciales del proyecto bajo el motor PostgreSQL con su extensión PostGIS, utilizando distintos esquemas, agrupando la información por orígenes de datos.

Incorporamos información externa a través de importación de shapefiles (fuentes IGN / BAHRA / OSM Geofabrik).

Para generar los Núcleos Poblados utilizamos scripts SQL, y una serie de funciones propias de PostgreSQL/PostGIS: `array_to_string()`, `array_agg()`, `sum()`, `count()`, `round()`, `ARRAY[]`, `st_length()`, `st_geography()`, `st_concavehull()`, `st_union()`, `st_dump()`, `st_buffer()`, `st_contains()`, `st_intersects()`, `st_intersection()`.

Para cada paso del proceso se generó su tabla temporal, su vista y sus índices necesarios (espacial/no-espacial).

Para comprobar visualmente los cambios en las geometrías utilizamos tanto QuantumGIS v2.6 con su módulo de conexión a DB-PostGIS, como la extensión de PostGIS Layers para el Pgadmin III.

4. RESULTADOS

A modo de síntesis, pueden exponerse los siguientes resultados cuantitativos sobre la nueva base de Núcleos Poblados y su comparativa con 2014:

Tabla 2: Comparativa de resultados entre los Núcleos Poblados creados en 2014 y 2015

| Fecha de creación de la base de Núcleos Poblados | 11/2014 | 04/2015 | Diferencias |
|--|------------|------------|-------------------|
| Total Núcleos Poblados | 2.247 | 2.327 | +80 |
| Total calles OSM dentro de Núcleos | 530.694 | 578.346 | +47.652 |
| Superficie total Núcleos (km2) | 18.242 | 19.620 | +1.378 |
| Superficie del país en Núcleos (%) | 0,66 | 0,7 | +0.4 |
| Total población dentro de Núcleos | 37.448.152 | 38.783.088 | +1.334.936 |
| Población dentro de Núcleos (%) | 93,34 | 96,68 | +3.34 |

Tabla 3: Resultados del análisis demográfico de los Núcleos Poblados 2015

| | |
|--|---------------|
| Superficie Argentina: | 2.870.000 km2 |
| Superficie Núcleos Poblacionales OSM (Aglomerados): | 19.620 km2 |
| Superficie Aglomerados imagen nocturna: | 11.218 km2 |
| Población Argentina: | 40.117.000 |
| Población Núcleos Poblacionales OSM (Aglomerados): | 38.783.088 |
| Cantidad Aglomerados OSM: | 2.327 |
| Cantidad Aglomerados imagen nocturna: | 1201 |
| % Superficie Núcleos Poblacionales (Aglomerados): | 0,7 |
| % Población Argentina en Aglomerados OSM: | 96,68 |

5. CONCLUSIONES

El presente trabajo permitió volver a aplicar, parcialmente y con licencias, la metodología propuesta por Vapñarsky para la identificación de localidades, en el sentido físico del concepto, tal cual se propuso en 1991.

A partir de dicha metodología y con la importante contribución de trabajos previos como el BAHRA y la base de datos geográfica del IGN y la carga de datos masiva, anónima y desinteresada que realizan miles de usuarios en OSM, se ha logrado avanzar hacia la identificación automatizada de Núcleos Poblados en Argentina, basada en información dinámica y viva. Dicho insumo es imprescindible para el planeamiento y la gestión de políticas públicas en general.

Se considera que esta experiencia ha permitido abrir nuevos horizontes de trabajo en la identificación de aglomeraciones, e invita a continuar experimentando y mejorando las herramientas que permitan lograr este tipo de objetivos.

6. AGRADECIMIENTOS

Se agradece el apoyo brindado por todo el equipo de Tecnología de la Información de Secretaría de Energía de la Nación.

7. REFERENCIAS

Ares, Sofía Estela; Mikkelsen, Claudia Andrea; Sabuda, Fernando Gabriel (2011). Identificación de localidades en el partido de General Pueyrredón a partir de la Implementación de Tecnologías de Información Geográfica (TIGS), *Geograficando*, 7(7), 33-50. Recuperado de http://www.memoria.fahce.unlp.edu.ar/art_revistas/pr.5091/pr.5091.pdf

Bertoncello, Rodolfo (2004). Buenos Aires ¿Quién es la “Reina del Plata”? *Población en Buenos Aires*, 1(0), 15-25. Recuperado de: http://www.buenosaires.gob.ar/areas/hacienda/sis_estadistico/nuevoinforme/poblacion%20de%20obsas%20Completa.pdf

Marcos, Mariana (2011). Base cartográfica para el estudio de diferencias interurbanas en la Aglomeración Gran Buenos Aires: Procedimientos técnicos para su realización. *Geografía y Sistemas de Información Geográfica (GEOSIG)*, 3(3), 1-21. Recuperado de <http://www.gesig-proeg.com.ar/documentos/revista-geosig/2011/II- Metodologia/01-II-CARTO-AGBA.pdf>

Vapñarsky, Cesar (2004). Cuando el caos caracteriza la división oficial del territorio del Estado. A propósito de los municipios argentinos, *Población en Buenos Aires*, 1(1),9-35. Recuperado de http://www.buenosaires.gob.ar/areas/hacienda/sis_estadistico/poblacion01.pdf

8. LICENCIA

Este Resumen Extendido se realiza bajo la licencia Creative Commons Atribución 3.0. Las características de esta licencia pueden consultarse en: <http://creativecommons.org/licenses/by/3.0/legalcode>