

Sistema de Información Geográfica como apoyo a las actividades del Censo Nacional de Población, Hogares y Viviendas 2010 de la provincia de Buenos Aires

Fernando Aliaga¹, Flavia Bonomo², Diego Delle Donne², Guillermo Durán², Javier Marengo², Rodrigo Sotelo³

Correos Electrónicos: fernando.aliaga@improntait.com (Aliaga), flavia@bonomo.com.ar (Bonomo), ddelledo@ungs.edu.ar (Delle Donne), gduran@dm.uba.ar (Durán), jmarengo@ungs.edu.ar (Marengo), rsotelo@estadistica.ec.gba.gov.ar (Sotelo)

Resumen. Los Sistemas de Información Geográfica (GIS)⁴ constituyen una plataforma tecnológica de gran utilidad para las agencias de estadística, brindando software apropiado para la administración de una base de datos geográfica y la generación de cartografía digital o impresa. Con el objeto de apoyar las actividades para el Censo Nacional de Población Viviendas y Hogares 2010, la Dirección Provincial de Estadística de la provincia de Buenos Aires (DPE) encomendó a Impronta IT S.A. (IM) la actualización de la plataforma GIS y el desarrollo de aplicaciones y algoritmos para la segmentación automática de viviendas. A través de un contrato de vinculación tecnológica, se incorporaron al equipo de trabajo expertos en optimización del Departamento de Matemática y Computación de la Facultad de Ciencias Exactas y Naturales de la UBA. El trabajo presenta el caso de vinculación Universidad – Industria – Gobierno que permitió incorporar novedosos algoritmos de optimización al proyecto GIS para el Censo Nacional 2010 en la provincia de Buenos Aires.

Palabras claves: vinculación, censo, segmentación, programación entera, optimización, sistemas de información geográfica, GIS.

¹ Impronta IT S.A., Ciudad Autónoma de Buenos Aires, Argentina.

² Grupo de Investigación en Teoría de Grafos y Optimización del Departamento de Matemática y Computación, Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires.

³ Dirección Provincial de Estadística, Provincia de Buenos Aires, Argentina.

⁴ Sistemas de Información Geográfica o GIS por sus siglas en inglés (Geographic Information Systems), ver explicación en <http://www.esri.com/what-is-gis/index.html>

1 Introducción

La Dirección Provincial de Estadística de la provincia de Buenos Aires (DPE) tiene sus orígenes en el Registro Estadístico de la Provincia de Buenos Aires, creado el 13 de diciembre de 1821. Entre sus tareas principales se encuentra la de coordinar las tareas relativas a los operativos censales que se desarrollen en el territorio de la provincia de Buenos Aires

La provincia de Buenos Aires es la más grande y poblada de la República Argentina: con 304.907 km² de superficie [2], está dividida administrativamente en 134 partidos con 5.581.761 de viviendas y 15.594.428 habitantes [3]. Sobre un total de 40.091.359 de habitantes para todo el país, representa un 39 % del total nacional.

El objetivo básico del Censo Nacional de Población, Hogares y Viviendas es “dar cuenta en forma precisa, de la cantidad de personas, hogares y viviendas que existen en un momento determinado en todo el territorio nacional” [4]. Para asegurar que todas las viviendas y personas sean censadas, es necesario determinar y delimitar con anterioridad el área que deberán recorrer los censistas y la carga de trabajo que se le asignará a cada uno de ellos. A este proceso se lo denomina “segmentación” de viviendas.

Con el objeto de apoyar las actividades para el Censo Nacional de Población Viviendas y Hogares 2010 en la provincia de Buenos Aires, la DPE encomendó a Impronta IT la actualización de la plataforma GIS, el desarrollo e implementación de aplicaciones y el desarrollo de un algoritmo para la segmentación automática de viviendas, todo lo cual se denominó “Proyecto GIS”⁵.

A través de un contrato de vinculación tecnológica, se incorporaron al equipo de trabajo expertos en optimización del Departamento de Matemática y Computación de la Facultad de Ciencias Exactas y Naturales de la UBA.

En las siguientes secciones se presenta el caso de vinculación Universidad – Industria – Gobierno que permitió implementar el GIS para la Dirección Provincial de Estadística como apoyo a las actividades del Censo Nacional de Población, Hogares y Viviendas 2010 de la provincia de Buenos Aires.

2 Desafíos

El “Proyecto GIS” se inició en el mes de marzo de 2010, y desde sus inicios presentó diversos desafíos relacionados con la tecnología, los recursos humanos y los

⁵ El proyecto incluyó otras actividades como el operativo de listado de viviendas urbanas, la carga de los listados en una base de datos, su vinculación cartográfica, el redimensionamiento de radios y fracciones, el desarrollo de aplicaciones de impresión, todas tareas que dada la fecha límite del 27 de octubre se tuvieron que hacer solapando cada actividad.

procedimientos de trabajo⁶ para proveer del material necesario para el operativo censal del 27 de octubre de 2010.

Dado los plazos, no resultaba factible diseñar la especificación detallada de cada uno de los problemas a resolver para que luego sea tomada por un tercero que aporte una solución a cada problema como si fueran compartimentos estancos. Sino que se requería una solución integral que abarcara no solo cada una de las etapas sino que hiciera ágil el paso de una etapa a la otra.

Adicionalmente, desde el principio se sabía que muchas decisiones se iban a tener que tomar sobre la marcha sin contar con los plazos habituales de testeo fuera de las oficinas de la DPE.

3 Estructura de la vinculación

Ante los desafíos planteados, la manera de resolverlos fue mediante la interacción entre la DPE, Impronta y la UBA.

La DPE tuvo a su cargo el gerenciamiento integral del proyecto, la especificación funcional de los requerimientos, y la asignación de la infraestructura y recursos humanos necesarios.

IM estuvo a cargo de la coordinación técnica para el desarrollo e implementación del proyecto GIS mientras que el equipo de la UBA desarrolló el algoritmo de segmentación automática.

Esta estructura, al estar formada por tres actores que provienen de distintas tradiciones y dinámicas como el Gobierno, la Industria y la Universidad fue un desafío en sí mismo respecto a su gestión.

El mecanismo de gestión elegido consistió en la presencia permanente de los recursos apartados por IM en las oficinas de la DPE junto con reuniones semanales donde se tomaban las decisiones necesarias para realizar ajustes en forma concomitante con el diseño de las aplicaciones. Además de resultar en este caso una forma ágil, permitió adaptar los procedimientos a definiciones que tuvieron que ser tomadas con el proyecto en estado avanzado.

Desde el punto de vista de los recursos humanos, se conformó un equipo de trabajo multidisciplinario compuesto por:

- 50 operadores GIS contratados por la DPE que se sumaron a los 10 técnicos pertenecientes a la planta de la DPE. Este personal fue capacitado y entrenado en el uso de la tecnología GIS y los procedimientos para el tratamiento de información censal. Este equipo estuvo dedicado a la actualización cartográfica, la vinculación⁷ y el redimensionamiento de radios⁸ así como la impresión de

⁶ La DPE no contaba con la tradición de formalizar de manera escrita los procesos, por lo que su realización implicó un esfuerzo adicional.

⁷ Proceso que consiste en relacionar un listado de viviendas a su correspondiente frente de manzana en la base GIS.

⁸ El radio censal es una porción del territorio con aproximadamente 300 viviendas. El redimensionamiento consiste en subdividir radios de más de 500 viviendas en N radios de aproximadamente 300 viviendas.

más de 300.000 mapas y reportes y la logística para distribuir el material en toda la provincia.

- Equipo de 10 especialistas GIS encargados del diseño, desarrollo, instalación y configuración de software RDBMS⁹, Web y GIS para la gestión de las aplicaciones GIS.
- Equipo de 4 de especialistas en optimización con programación entera encargado del desarrollo del algoritmo de segmentación.

Clave para la ejecución del proyecto fue dotar a la DPE de una plataforma de hardware y software GIS de última generación basada en software ArcGIS de ESRI¹⁰. Esta plataforma informática significó para la DPE un salto tecnológico de más de 10 años. Durante esta etapa se entrenó al personal afectado al proyecto y se proveyó un servicio de administración de infraestructura informática GIS denominado GeoDBA. En el siguiente diagrama se detalla la arquitectura informática utilizada.

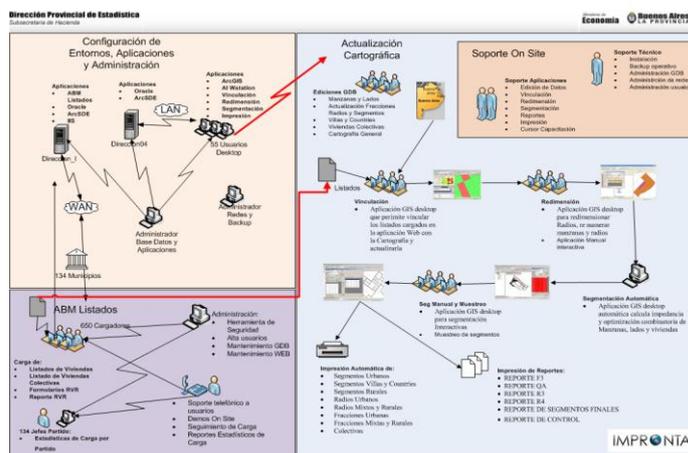


Fig. 1. Arquitectura GIS de la DPE [5]

La infraestructura informática incluyó:

- Wide Area Network (WAN) conectando 134 municipios con 600 operadores cargando viviendas a través de la aplicación Web ABM de Listados residente en un servidor dedicado.
- Red de Área Local (LAN) con 60 operadores GIS accediendo a una base de datos geográfica denominada Geodatabase. Incluyó equipamiento de impresión de alta velocidad.
- Aplicación Tablero de Control para el seguimiento de las actividades del proyecto

El desarrollo de aplicaciones se realizó sobre diferentes lenguajes de programación, que incluyeron ASP .Net, VB Forms, ANSI C, PL SQL y ArcObjects.

⁹ Relational Database Management System.

¹⁰ Environmental System Research Institute, www.esri.com.

4 Innovación: algoritmo de segmentación automática

El desafío más importante del proyecto estuvo relacionado con el proceso de segmentación de viviendas, el cual se describe en este apartado.

El área de trabajo de cada censista se denomina segmento y es el “conjunto de: a) una porción de territorio, b) las viviendas que en ella se asientan, y c) un promedio estimado de entrevistas a realizar durante el tiempo establecido para el operativo censal” [4].

La segmentación es la tarea que permite subdividir el radio censal en unidades operativas menores llamadas segmentos sujeta al cumplimiento de ciertas condiciones de contigüidad y exhaustividad¹¹. Se realiza a nivel de cada radio censal, con características diferentes según el tipo de radio: urbano, rural o mixto. Este procedimiento de segmentación, habitualmente se realiza manualmente o utilizando rudimentarios algoritmos de asignación.

Experiencias anteriores, y particularmente del Censo Nacional del año 2001, indicaban que la DPE necesitaba emplear 6.000 horas hombre, el equivalente a 25 operadores trabajando full-time durante 25 días, para realizar la segmentación de manera manual, y esto sin considerar los tiempos de entrenamiento, soporte y los riesgos asociados a los procesos manuales de asignación.

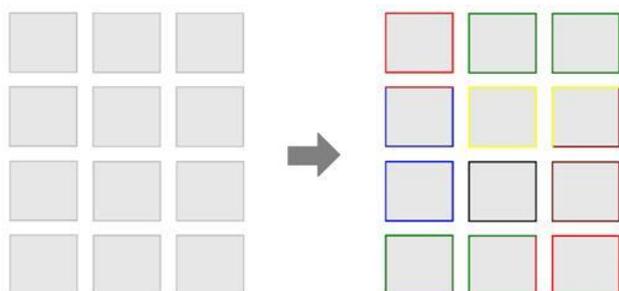


Fig. 2. Esquema de segmentación: lados de manzanas asignados a censistas. Cada color corresponde a un segmento asignado a un censista.

La cantidad de viviendas asignada a cada segmento, comúnmente denominado “tamaño” del segmento, varía en función de la cantidad de preguntas del formulario censal a utilizar, el nivel socioeconómico de la población, la presencia de familias numerosas y las particularidades de contexto geográfico. Típicamente el “tamaño” de los segmentos varía entre 12 y 40 viviendas.

De acuerdo a la metodología establecida por el INDEC para el Censo 2010 hubieron dos tipos de formularios: un Formulario Ampliado (A) para ser aplicado en aquellas localidades con una población inferior a 50.000 habitantes y en el 10% de los segmentos de localidades con 50.000 habitantes o más; y un segundo Formulario

¹¹ Federico Ferella, como Coordinador del Departamento de Cartografía Estadística, aportó la descripción de los procedimientos que tradicionalmente se realizaban de manera manual y que sirvieron de base para el desarrollo de la segmentación automática.

Básico (B) para aplicar en el 90% de los segmentos de localidades con 50.000 habitantes o más.

De esta manera, y debido a la diversidad provincial en los 134 Municipios, existieron distintas situaciones:

- Todo el Municipio con muestra
- Todo el Municipio sin muestra, por no tener localidades que superen mas de 50.000 habitantes
- Municipios con localidades con muestra y resto de localidades sin muestra

En un segundo nivel existió diversidad de densidad de viviendas en cada partido, por niveles altos de pobreza, familias numerosas (información base proveniente del Censo del año 2001) asentamientos y villas de emergencia (con información proveniente del Operativo Listado Viviendas 2010¹²). Áreas de country, barrios cerrados y demás urbanizaciones que implicaron establecer diferentes tamaños de segmentos.

También el tamaño del segmento en áreas urbanas estuvo condicionado por su ubicación geográfica de acuerdo a la localidad a la que pertenezca, por el hacinamiento en hogares por radio censal según Censo 2001 y por el criterio de selección muestral utilizado.

De esta manera, la existencia de diversos tamaños de segmentos en una misma localidad generó mayor complejidad para resolver la asignación de áreas de trabajo a cada censista. En la siguiente tabla se detallan los ocho casos de segmentación utilizados y la cantidad de viviendas (tamaño de segmento).

Tamaño de Localidad	Hacinamiento en hogares 2001	Country / Villa E.	Muestra	Formulario	Tamaño de Segmento
Más 50.000 hab.	NO	NO	NO	B	e/ 32 y 40 Viv.
Más 50.000 hab.	NO	NO	SI	A	e/ 16 y 20 Viv.
Más 50.000 hab.	SI	NO	NO	B	e/ 26 y 32 Viv.
Más 50.000 hab.	SI	NO	SI	A	e/ 12 y 18 Viv.
Más 50.000 hab.	NO	SI	NO	A	e/ 15 y 25 Viv.
Menos 50.000 hab.	NO	NO	NO	A	e/ 16 y 24 Viv.
Menos 50.000 hab.	SI	NO	NO	A	e/ 14 y 20 Viv.
Menos 50.000 hab.	NO	SI	NO	A	e/ 15 y 25 Viv.

Tabla. 1. Tamaño del segmento según el tipo de localidad a la pertenece cada radio

¹² El Operativo Listado de Viviendas se realizó con anterioridad a la ejecución del Censo y consistió en un relevamiento “puerta a puerta” de todas las viviendas del territorio en planillas papel. Estas planillas fueron luego cargadas en el GIS y sirvieron de entrada a los algoritmos de segmentación.

El problema de segmentación de viviendas puede relacionarse a los problemas de "redistricting"¹³ [6, 7, 8], sin embargo, los algoritmos para redistricting existentes no se adaptaban fácilmente al problema particular de segmentación. Además, no se encontraron evidencias de la existencia de algoritmos automáticos de segmentación en la literatura.

El equipo de trabajo de la UBA, conformado por especialistas en optimización, trabajó en el desarrollo de un algoritmo de segmentación automática basado en programación entera. Este algoritmo fue integrado dentro del flujo de trabajo de la DPE dentro de la infraestructura GIS existente, permitiendo que todos los equipos de trabajo trabajaran armoniosamente y a máxima capacidad.

En el siguiente diagrama se detalla una típica salida del algoritmo de segmentación.



Fig. 3. Esquemas de segmentación en áreas urbanas y mixtas

El éxito del algoritmo de segmentación se basó en la capacidad de resolver, con un mismo modelo de programación entera, situaciones de distribución urbana muy distantes:

- más de 160 viviendas por manzana, típico caso de “listado”¹⁴
- manzanas de 40 viviendas: densidad media
- manzanas con menos de 5 viviendas: baja densidad

El algoritmo de segmentación permitió construir automáticamente el 96% de los segmentos urbanos y mixtos de la provincia en aproximadamente 320 hs de procesamiento (menos de un día con un cluster de 15 PCs).

¹³ Redistricting es el proceso de cubrir completamente un territorio con un conjunto mínimo de N polígonos que satisfagan alguna restricción (por ejemplo que la cantidad de habitantes de cada polígono difiera en menos de un 5%).

¹⁴ Se denomina “listado” al proceso de segmentar manzanas cuyos lados tienen más viviendas de la capacidad máxima del segmento, resultando en la asignación de N censistas por lado de manzana, incluso dentro de un mismo edificio.

Como referencia se detalla el modelo de programación entera [9]:

$$\begin{aligned} & \max \sum_{s \in S} val(s) \times x_s \\ & \text{sujeto a} \\ & \sum_{s \in S_v} x_s = 1 \quad \forall v \in V \\ & \sum_{s \in S_l} x_s = 1 \quad \forall l \in L_0 \\ & x_s \in \{0,1\} \quad \forall s \in S \end{aligned}$$

Donde:

- S : conjunto de todos los segmentos posibles s
- x_s : variable binaria que indica si el segmento s es parte de la solución
- $val(s)$: valuación del segmento s
- V : conjunto de viviendas
- S_v : conjunto de segmentos que incluyen a la vivienda v
- L_0 : conjunto de lados sin viviendas
- S_l : conjunto de segmentos que incluyen al lado sin viviendas l

La función de valuación maximizada en el modelo, toma en cuenta la proporción de lados por manzana contenidos en un segmento. Este valor proporciona una idea de qué tan compacto es el segmento en cuestión, y este aspecto es de alguna manera una buena métrica para comparar los segmentos entre sí. Por ejemplo, un segmento que recorre una manzana completa, tendrá una proporción de lados por manzana de 4 (si la manzana tiene 4 lados) mientras que un segmento que recorre cuatro manzanas visitando sólo un lado por manzana tendrá una proporción de 1. Según los criterios de segmentación definidos, la configuración del primer segmento es más deseada que la del segundo, y esto mismo queda reflejado en la función de valuación propuesta.

El modelo de programación entera descrito anteriormente resuelve el problema de segmentación planteado. Sin embargo, cabe aclarar que la cantidad de variables del mismo (i.e., la cantidad de segmentos posibles en un radio) puede resultar intratable en la práctica. Para resolver este inconveniente, se desarrolló un algoritmo que parte desde un conjunto reducido de segmentos $S_1 \subseteq S$ y ejecuta el modelo con dicho conjunto en busca de una solución. Si no se encuentra solución, el algoritmo genera más segmentos y reitera la ejecución del modelo con este nuevo conjunto de segmentos. El algoritmo continúa repitiendo este proceso hasta hallar una solución o bien hasta alcanzar algún límite predefinido de iteraciones.

El siguiente pseudocódigo describe el algoritmo implementado:

```
- i ← 1
- S1 ← segmentos no excedidos que no cruzan ninguna calle.
- Ejecutar el modelo de PE usando los segmentos válidos de S1.
- Mientras no haya solución e i < límite de iteraciones hacer
  - Si+1 ← Si
  - Para cada (si, s) en Si × S1 hacer
    - Si (si ∪ s) es conexo y no excedido entonces
      - Si+1 ← Si+1 ∪ {(si ∪ s)}
    - fin si
  - fin para
  - Ejecutar el modelo de PE usando los segmentos válidos de Si+1.
  - i ← i+1
- fin mientras
```

donde un segmento “no excedido” es un segmento de tamaño no mayor al tamaño máximo permitido (aunque no necesariamente no menor que el mínimo) y un segmento “válido” es un segmento no excedido de tamaño no menor que el mínimo.

5 Resultados alcanzados

Luego de más de 500 meses hombre de esfuerzo, 8 meses de trabajo, y culminando con el operativo Censal del 27 de octubre de 2010, se proveyó en tiempo y forma de todo el material necesario para la ejecución del Censo Nacional de Población, Viviendas y Hogares en la provincia de Buenos Aires. Durante el proyecto, se lograron las siguientes metas:

- Confeccionar la Cartografía Actualizada de 134 partidos residente en una única base de Datos Multiusuario del Tipo Geodatabase
- Proveer a la DPE de una plataforma tecnológica GIS moderna y escalable
- 5.500.000 viviendas listadas durante el operativo precensal
- Listados de viviendas y cartografía vinculados en una única base de datos
- 600 fracciones censales redimensionadas
- 2.600 radios censales redimensionadas
- 2.900 radios rurales segmentados manualmente¹⁵ con 10.300 segmentos
- 16.700 radios urbanos y mixtos segmentados automáticamente con 200.500 segmentos
- Más de 300.000 mapas y reportes impresos¹⁶
- Material para censista, jefe de radio y fracción entregados en tiempo y forma

¹⁵ Los radios rurales no estaban incluidos en el algoritmo de segmentación.

¹⁶ Si se apilaran por la base todas las resmas de papel utilizadas, se alcanzaría la altura de un edificio de 10 pisos.

6 Consideraciones finales

El proyecto GIS para el Censo Nacional de Población, Hogares y Viviendas 2010 de la provincia de Buenos Aires representó para sus directores un gran desafío: conformar un equipo multidisciplinario, con alta disponibilidad y gran capacidad para ejecutar en menos de 8 meses los algoritmos y aplicativos GIS necesarios para proveer al operativo del material censal necesario. Particularmente era necesario acortar los tiempos asociados a los procesos de segmentación de viviendas para poder alcanzar los objetivos para el 27 de octubre de 2010.

El trabajo realizado a través de la vinculación entre Universidad- Industria- Gobierno permitió desarrollar y utilizar un algoritmo de segmentación automática de viviendas del cual en ese momento no había antecedentes exitosos de utilización para una realidad de la complejidad de la provincia de Buenos Aires. Así, la aplicación de este algoritmo junto con el desarrollo del resto de las actividades precensales permitió que la carga de trabajo de cada integrante de la estructura censal estuviera asignada de manera de cumplir con el requisito de cobertura el día del operativo censal.

El rol de la innovación en este proyecto, que encuentra su máxima expresión en el desarrollo de un algoritmo de segmentación de viviendas inédito para la industria, demuestra la importancia de trabajar creativamente en la resolución de los problemas concretos de nuestras empresas y organizaciones como una forma de colaborar con el desarrollo del país y su pueblo.

7 Referencias

1. Dirección Provincia de Estadística de la provincia de Buenos Aires. Hallado en: <http://www.ec.gba.gov.ar/Estadística/institucional.html>. Consultado el 16 de marzo de 2011.
2. Dirección de Geodesia, Subsecretaría de Obras Públicas, Ministerio de Infraestructura de la provincia de Buenos Aires
3. INDEC, 2010. Censo Nacional de Población, Hogares y Viviendas 2010 Resultados provisionales de la provincia de Buenos Aires. Hallado en <http://www.censo2010.indec.gov.ar>. Consultado el 5 de enero de 2011.
4. Departamento de Cartografía y SIG-INDEC, 2010. Manual de Segmentación.
5. Presentación “Memorias del proyecto GIS para el censo 2010”, Fernando Aliaga, diciembre de 2010
6. Martin D, 2001. Developing the Automated Zoning Procedure to Reconcile Incompatible Zoning Systems. , Proceedings, 6th International Conference on GeoComputation, held at Brisbane, Australia, 24-26 September 2001. CD-ROM Publication.
7. Altman M, 1997. Is Automation the Answer: The Computational Complexity of Automated Redistricting. *RC and Law Technology Journal*, 23(1), pp: 81-141.
8. Altman M, MacDonald K. and McDonald M.P., 2005. From Crayons to Computers: The Evolution of Computer Use in Redistricting. *Social Science Computer Review*, 23(3), pp: 334-346.
9. Delle Donne D, Durán G y Marengo J, 2011. Solving the segmentation problem for the 2010 Argentine census with integer programming. *Electronic Notes in Discrete Mathematics*, en prensa. doi: 10.1016/j.endm.2011.05.