

EL INE REALIZA UN ESTUDIO PILOTO SOBRE MOVILIDAD A PARTIR DE DATOS AGREGADOS DE TELEFONÍA MÓVIL

En el marco de preparación de los trabajos del Censo 2021, el INE está realizando un estudio a partir de datos de teléfonos móviles para incorporarlos a la información sobre movilidad que tradicionalmente se ofrece en esta operación. El proyecto, de indudable interés social, sirve para determinar movilidad cotidiana o estacional de cierta entidad y no contendrá desagregación por sexo, edad u otra variable de los portadores de los teléfonos.

Para el estudio piloto se han escogido aquellas compañías de telefonía móvil que se considera a priori que pueden tener suficiente nivel de representatividad en todo el territorio español. Para realizar este estudio se divide el territorio nacional en 3.220 celdas, cada una de ellas con al menos 5.000 residentes. Para cada celda, el INE recibirá información de los tres principales operadores de telefonía móvil de España sobre cuántos terminales se encuentran en dicha celda en varios momentos del día y cuántos se desplazan de una celda a otra. Los operadores no facilitarán datos individuales sobre números de teléfono, ni sobre los titulares de las líneas. Por lo que el INE no podrá rastrear, en ningún momento, la posición de ningún terminal. En concreto, el INE no recibe datos individuales, ni siquiera anonimizados.

El servicio de obtención de los datos agregados que se contrata con los operadores se sitúa en torno a 500.000 euros. Si el INE tuviera que hacer mediante cuestionarios

una encuesta para ofrecer datos estadísticos sobre cuántas personas se desplazan entre 3.200 áreas geográficas, sería necesario entrevistar a centenares de miles de hogares, lo que supondría un coste de varios millones de euros. De esta manera se consigue reducir tanto el coste económico como la tarea que supone para los hogares la cumplimentación de los cuestionarios.

La operación, al formar parte de los Censos de Población y Viviendas, está recogida en el Plan Estadístico Nacional y enmarcada en los principios de la Ley de la Función Estadística Pública (LFEP, de 12 de noviembre de 1989). Cabe señalar que el proyecto se sometió a dictamen por parte de la Agencia de Protección de Datos en 2016, cuando se estaba trabajando en sus primeras fases y se contemplaba la posible recepción por parte del INE de datos individuales anonimizados. Recibió ya entonces opinión favorable, si bien se insistió en la necesaria protección de la confidencialidad, instando a que todos los datos individuales que el INE recibiera fueran anónimos. El proyecto, en su diseño actual, cumple con creces estas directrices en tanto que no se reciben datos individuales de ningún tipo.

Se publicarán los resultados en el primer semestre de 2020, como estadística experimental. La información se presentará en forma de tablas agregadas y mapas, con el porcentaje de personas que se desplazan entre las celdas del estudio.

DOBLE RECONOCIMIENTO AL INE POR SU PARTICIPACIÓN EN LOS PROGRAMAS DE PRÁCTICAS UNIVERSITARIAS DE LA UCM

En la I Edición del acto de Reconocimiento a Entidades colaboradoras con Programas de Prácticas, se entregaron distinciones por parte de los diferentes decanos a los organismos públicos y privados que más han colaborado con dicha Universidad en el desarrollo de los programas de prácticas de sus alumnos. Las facultades de Económicas y Estudios Estadísticos otorgaron su premio al Instituto Nacional de Estadística (INE).

La decana de la Facultad de Económicas, Begoña García Greciano, destacó la labor del INE como uno de los principales organismos que han hecho posible la realización de prácticas a los alumnos que han cursado el Máster Universitario en Estadísticas Oficiales e Indicadores Sociales y Económicos, organizado por dicha Facultad y acreditado como Máster EMOS (European Master of Statistics) de

Eurostat. Por su parte, el decano de la Facultad de Estudios Estadísticos, Cristóbal Pareja Flores, señaló la estrecha relación de dicha facultad con el INE, desde sus tiempos como Escuela Universitaria de Estadística, prestando atención en su formación a los futuros profesionales del INE, incluyendo ahora la realización de prácticas de los estudiantes del Grado en Estadística Aplicada en el Instituto. El decano también hizo mención a la colaboración de la Facultad y el INE, junto con la Sociedad de Estadística e Investigación Operativa, en la organización de la Competición Estadística Europea, que es coordinada por España.

El acto estuvo presidido por el Rector de la Universidad Complutense, Joaquín Goyache, y a él asistió el Presidente del INE, Juan Manuel Rodríguez Poo, quien recogió el premio del organismo.

EL PLENO DEL CONSEJO DE EMPADRONAMIENTO APRUEBA LA PROPUESTA DE CIFRAS OFICIALES DE POBLACIÓN A 1 DE ENERO DE 2019

El Consejo de Empadronamiento celebró reunión plenaria en Madrid el pasado 13 de noviembre. Los miembros de este órgano colegiado aprobaron la propuesta de cifras oficiales de población resultantes de la revisión anual del Padrón municipal, referidas a 1 de enero de 2019, así como el calendario del procedimiento de cifras para la revisión anual a 1 de enero de 2020. Tras la aprobación de las actas de las últimas sesiones celebradas por el Consejo de Empadronamiento, se presentó ante el Pleno el informe de las discrepancias entre el Instituto Nacional de Estadística (INE) y los Ayuntamientos sobre las cifras de población a 1 de enero de 2019.

Los asistentes a la reunión plenaria analizaron las posibles incidencias que puede originar el Brexit en la gestión padronal, estudiando las distintas casuísticas que se pueden presentar en este proceso. Además, fueron informados de la respuesta del Defensor del Pueblo al informe de la Comisión Permanente de 2 de julio, sobre el empadronamiento de menores por un solo progenitor separado o divorciado. Por último, el Pleno aprobó la propuesta de una nueva Resolución de la Presidencia del INE y de la Dirección

General de Cooperación Autonómica y Local, por la que se dictan instrucciones técnicas a los Ayuntamientos sobre la gestión del Padrón municipal, elaborada por el grupo de trabajo del Consejo creado para analizar la situación de la Resolución de 30 de enero de 2015, tras la entrada en vigor de la Ley 39/2015, del Procedimiento Administrativo Común de las Administraciones Públicas, con el fin de adaptarla a los requerimientos de la administración electrónica, a la que también se han incorporado otras actualizaciones necesarias como consecuencia de las cuestiones planteadas por el Defensor del Pueblo, la entrada en vigor de la nueva normativa de protección de datos, etc.

En la convocatoria participaron el Presidente del INE, Juan Manuel Rodríguez Poo, el Director general de Coordinación Estadística y de Estadísticas Laborales y de Precios, Miguel Ángel de Castro, así como representantes de la Unidad de Padrón del INE y de la Oficina del Censo Electoral. También asistieron miembros del Ministerio de Política Territorial y Función Pública, de la Federación Española de Municipios y Provincias (FEMP) y del Ministerio de Asuntos Exteriores, Unión Europea y Cooperación.

LA PRESIDENTA DEL CSE DESTACA LA SOLIDEZ DE LA ESTADÍSTICA ESPAÑOLA Y LA NECESIDAD DE UTILIZAR NUEVAS FUENTES DE DATOS

La ministra de Economía y Empresa en funciones, Nadia Calviño, señaló en su intervención en el Pleno del Consejo Superior de Estadística (CSE) la “importante misión” que desempeñan tanto el Instituto Nacional de Estadística (INE) como el resto de productores del sistema estadístico español de producir información “de la más alta calidad sobre la economía, la sociedad y el medio ambiente, bajo los principios de independencia profesional, imparcialidad y fiabilidad”. Para Calviño será principal la incorporación de las nuevas tecnologías a la estadística pública y la importancia de modernizar las metodologías y las técnicas para reflejar adecuadamente la realidad económica. Tras destacar que la producción estadística en España es sólida, la presidenta del CSE planteó como reto de futuro avanzar en el uso y tratamiento de los datos procedentes de las nuevas fuentes de información, manteniendo la calidad de las estadísticas y la privacidad de los informantes.

El presidente del INE, Juan Manuel Rodríguez Poo, clausuró la sesión plenaria del CSE reseñando las principales

líneas de futuro de la actividad del INE. Entre ellas, destacó las oportunidades que ofrece el uso de Big Data para la estadística oficial, aunque comentó que ello también plantea “grandes retos”. Si bien resulta “prioritario” que la estadística oficial incorpore estas nuevas fuentes de información, Rodríguez Poo insistió en que se debe hacer “siempre y cuando se continúen garantizando los estándares de calidad y el cumplimiento de todos los principios del Código de Buenas Prácticas de las Estadísticas Europeas”.

Entre ambas intervenciones, el Pleno del CSE aprobó el dictamen sobre el Anteproyecto de Real Decreto por el que se aprueba el PEN 2021-2024 y la Memoria anual del año 2018 de este órgano colegiado. Además, se presentaron para conocimiento ante la sesión plenaria los dictámenes de las operaciones estadísticas aprobados por la Comisión Permanente del Consejo durante este año, así como las memorias aprobadas por la Comisión Interministerial de Estadística (CIME) y el Comité Interterritorial de Estadística (CITE).

D. José Luis Maldonado

“Si en el momento de la creación de las instituciones ya la estadística y la geografía estaban íntimamente relacionadas, parece que esta relación aumenta día a día”

El público no experto no siempre tiene clara la distinción entre la geolocalización y la georreferenciación. ¿Cuáles son las diferencias entre una y otra?

Ambas son parte del proceso de transformación digital de la ciencia estadística y geografía. La georreferenciación es un proceso que permite crear y trabajar con gemelos digitales del territorio, mientras que la geolocalización hace referencia a un proceso que permite identificar la ubicación de un dispositivo en un entorno digital. Ambos procesos se benefician de las capacidades de la geografía digital para aportar contenido y contexto a los sistemas y tecnologías de la información. Hay que decir que para estos trabajos la capacidad de localizar de manera precisa las entidades geográficas es fundamental, tanto en la representación cartográfica como en las tareas de análisis espacial.

En resumen, la geolocalización nos permite consultar la información de nuestra posición. Además, disponer de servicios georreferenciados (estadísticas oficiales, por ejemplo) será de utilidad para analizar los datos que tenemos a nuestra disposición y conocer el contexto de esa posición.

Ambos procesos, tanto la geolocalización como la georreferenciación, son determinantes para varios de los servicios y productos desarrollados por el INE. ¿En qué sentido?

Como indicas, tanto la geolocalización como especialmente la georreferenciación son determinantes para varios productos y servicios desarrollados por el INE.

En cuanto a **georreferenciación** y como infraestructura base, el principal producto del INE es el

Marco de Direcciones Georreferenciadas (MDG) y su equivalente para la Administración General del Estado o Modelo de Direcciones de la AGE (MDAGE).

Estos dos productos tratan de mantener unos callejeros oficiales de España en los que todos los elementos geográficos de una dirección postal y de los huecos de esta dirección se encuentren georreferenciados.

Esto, dicho así, no parece mucho, pero a efectos prácticos es la pieza clave para lo que se viene denominando Geo-Estadística y también la pieza imprescindible para poder intercambiar datos geográficos con el resto de las organizaciones sin la ambigüedad que tienen (o pueden tener) los datos alfanuméricos de las direcciones. Solo un par de ejemplos:

1. En el MDG/MDAGE se pretende que todas las direcciones postales estén georreferenciadas, es decir, dispongan de sus coordenadas X-Y, o lo que es lo mismo, una posición física en el espacio que es invariable. Esto tiene dos consecuencias directas:

a. No importan los cambios de denominación del resto de los atributos de la dirección (provincia, municipio, distrito, sección, unidad poblacional, vía ni número), ya que esa dirección seguirá estando en el mismo punto geográfico, evitando con ello duplicidades, ausencia de información, etc.

b. Para una determinada X-Y los intercambios de información entre instituciones carecerían de esta ambigüedad, ya que si coinciden el resto de los datos serán siempre comparables.

2. Suponemos ahora que hemos realizado cualquier encuesta dirigida a hogares. Clásicamente, se obtienen datos que se publicitan a través de tablas numéricas poco atractivas para el usuario no avanzado.

La publicación de esta encuesta (o censo) a través de técnicas geográficas (Geo-Estadística) estaría ga-

rantizada, ya que también en la encuesta realizada se podría asignar sin demasiados problemas las coordenadas X-Y y, por lo tanto, se podrían representar estos resultados, obtener “mapas” por lo que de un simple vistazo el usuario final podría entender fácilmente los resultados de esta encuesta.

Además del MDG/MDAGE, no nos podemos olvidar del que posiblemente sea el producto más clásico del INE, que no es otro que la digitalización de las secciones censales que el INE pone a disposición de forma gratuita a través de nuestra página oficial.

Un paso más que el INE está dando en la georreferenciación de las unidades Estadísticas es la georreferenciación de la Unidades Poblacionales (Entidades Colectivas, Singulares y Núcleos).

Finalmente, en cuanto a georreferenciación, el INE ha emprendido sin marcha atrás la publicación de las denominadas estadísticas inteligentes, en todo comparables a la Geo-Estadística, con fuentes propias o de terceros que tratan de representar unos resultados numéricos en mapas interactivos.

En cuanto a la **geolocalización** en sentido estricto, de momento el INE hace un uso muy limitado de este tipo de técnicas.

Aunque ha sido extraordinariamente mediática la utilización de técnicas de geolocalización de los dispositivos móviles, en realidad de lo único que disponía el INE era de datos agregados del número de móviles que a una determinada hora estaban en una determinada celda telefónica.

Por el contrario, el INE ha utilizado con bastante éxito técnicas de geolocalización para que el entrevistador en campo pudiese orientarse en el terreno y según su situación y la de los informantes poder hacer los recorridos lo más eficientemente posible.

Actualmente nos encontramos inmersos en una revolución tecnológica y digital sin precedentes. ¿Qué áreas del INE se benefician

esencialmente de la innovación y qué estrategia siguen para capitalizar este nuevo horizonte de posibilidades? Prevemos que la georreferenciación es uno de los ámbitos que más radical y rápidamente está cambiando.

Esta es una pregunta con muy diversas aristas y que posiblemente se pueda aclarar con la diferenciación que hemos hecho previamente entre georreferenciación y geolocalización.

Desde el punto de vista técnico, la georreferenciación no ha cambiado de manera ni radical ni más rápidamente que otras técnicas. De hecho, en los Censos de Población y Viviendas del 2011 el INE utilizó ampliamente técnicas y procesos de georreferenciación en la novedosa difusión de los datos Censales, en la que el usuario final podía (y puede) acceder a los datos a través de mapas y seleccionar sin limitaciones parte del territorio a estudiar al margen de las divisiones administrativas, e incluso la magnífica publicación de los Censos de Población y Viviendas 2011: “¿Cómo es España? 25 mapas para descubrirla Km² a Km²”.

Es posible que fuésemos pioneros en el 2011 en la utilización de estas técnicas cuyos objetivos permanecen inalterables.

Si esto es así, ¿qué es lo que ha cambiado? Además de que la tecnología permite el procesamiento más rápido y que la información que precisa el INE está disponible casi de forma inmediata y con una calidad sin precedentes, lo que verdaderamente ha cambiado son las demandas sociales de disponer de datos más rápidos, más exactos, más accesibles y sobre todo más amigables.

El INE quiere dar satisfacción a estas demandas sociales y es por ello que la marcha emprendida en el 2011 se ha incrementado notablemente y que, como ya he indicado, no hay marcha atrás.

Por otra parte, me pregunta sobre “¿qué áreas del INE se benefician esencialmente de la innovación y qué estrategia siguen para capitalizar este nuevo horizonte de posibilidades?”

Desde mi punto de vista muy particular, la pregunta debería ser “¿Qué áreas del INE NO...?” y si seguimos hablando de georreferenciación, basta ver el IDE-INE (Infraestructuras de Datos Espaciales del INE) en el que solo una pequeña parte es producción del INE (la que directamente es competencia del INE) y el grueso de la información que allí aparece es utilizada por el INE, pero generada por los organismos competentes.

En la actualidad el impulso, no solo en el INE, sino en el resto de la administración que se está dando a la utilización de “otras fuentes”, la información georreferenciada es una de las más fáciles de utilizar y que el INE utiliza profusamente y tiene planificado utilizar todavía más, si es que cabe, precisamente para disminuir la carga al informante y abaratar costes sin perder una pizca de calidad, e incluso aumentarla.

El ejemplo de lo que digo es claro, los próximos Censos de Población y Viviendas del 2021 van a ser realizados fundamentalmente a través de cruces de los datos del INE con los de otras fuentes, en los que los procesos de georreferenciación están siendo una pieza muy importante para llevar a cabo los fines propuestos.

¿Qué es posible en el ámbito de la georreferenciación que no lo fuera hace diez años?

Las tecnologías habilitadoras de la Cuarta Revolución Industrial, como la nube, la Inteligencia Artificial, el internet de las cosas (IoT) o los SIG avanzados aplicados a los Gemelos Digitales del territorio, están provocando la caída de las barreras entre lo físico y lo digital. En comparación con hace 10 años, ahora tenemos una mayor capacidad de cómputo a un mejor precio, tenemos más datos y también mejores algoritmos, sin olvidarnos del acceso al conocimiento universal y capacidades productivas distribuidas en HUBs de innovación y educación a nivel mundial.

Restringiéndome al INE, hoy estamos en una posición envidiable con el

resto de instituciones similares europeas, ya que a través del MDG hemos conseguido armonizar los datos alfanuméricos de las direcciones con los datos cartográficos y, por ello, estamos en disposición de poder representar en mapas cualquier encuesta en los que la dirección sea un factor relevante.

La gestión masiva de los datos multiplica nuestro conocimiento y nuestra capacidad de acción, lo que, previsiblemente, nos debería llevar a reforzar nuestro celo y nuestra cautela en el manejo de instrumentos tan potentes. Entendemos que la responsabilidad social y ética en esta gestión es una prioridad para el INE. ¿Qué desafíos enfrentan desde esta perspectiva?

Posiblemente la actual legislación sobre confidencialidad de los datos, para el INE, sea un aspecto superado desde que existen las estadísticas. Es posiblemente el aspecto mejor tratado de todo el proceso estadístico, desde la generación de muestras hasta la difusión de los datos finales.

No se trata de reforzar nuestro celo ante instrumentos tan potentes, sino de la utilización de todos los medios a nuestro alcance para preservar el secreto estadístico en todas y en cada una de las fases de producción, que es lo que se viene haciendo de forma más que palpable desde que el INE es INE.

En algunas instituciones tan prestigiosas como el INEGI convergen la estadística y la geografía. En nuestro país el IGN y el INE son dos instituciones separadas, pero con una intensa trayectoria colaborativa. ¿En qué forma se

integra la actividad de ambos institutos?

No solo es el INEGI como institución de referencia, sino que hay otras muchas instituciones, fuera y dentro de España, que han unido, siendo políticamente incorrecto, lo que nunca debería estar separado.

Hay que recordar que desde 1873 hasta 1945 el INE y el IGN eran un solo organismo, o lo que es lo mismo, llevamos más tiempo juntos que separados.

Si en el momento de la creación de las instituciones ya la estadística y la geografía estaban íntimamente relacionadas, parece que esta relación aumenta día a día.

A pesar de ser organismos diferentes, pero con una clara convergencia, la integración de las actividades se hace de forma natural por tener objetivos, proyectos, actividades comunes y en foros internacionales la postura es común entre estos dos organismos y con un tercero, que es la Dirección General del Catastro.

Proyectos como CARTOCIUDAD e incluso el MDAGE, ya mencionados anteriormente, son productos de la colaboración entre los tres organismos, bien como proyectos nacionales o como parte de proyectos europeos como INSPIRE, donde el MDAGE desarrollado por el INE forma parte de INSPIRE, cuyos representantes nacionales eran del IGN.

Las reuniones de coordinación de las actividades y los convenios, que son muy frecuentes y provechosas, están a la orden del día.

Las tecnologías habilitadoras de la Cuarta Revolución Industrial, como la nube, la Inteligencia Artificial, el internet de las cosas (IoT) o los SIG avanzados aplicados a los Gemelos Digitales del territorio, están provocando la caída de las barreras entre lo físico y lo digital

Aunque he indicado que en los Censos de 2011 ya se habían utilizado estas técnicas, lo cierto es que este es el punto de partida del MDG, sin el cual ahora no podría ser tan categórico cómo ha sido.

En resumen, ahora podemos hacer más proyectos, de manera más segura, en menos tiempo y a un menor coste.

JOSÉ LUIS MALDONADO CECILIA



Desde diciembre de 1986 presta servicios de forma ininterrumpida en el Instituto Nacional de Estadística, desempeñando diferentes puestos de trabajo en la Subdirección General de Tecnologías de la Información y las Comunicaciones.

Pertenece desde el año 1993 al Cuerpo Superior de Sistemas de Tecnologías de la Información y la Comunicaciones de la Administración General del Estado.

Desde Septiembre de 2009, ocupa el puesto Subdirector General de TIC.

Coordinador de Grupo de Trabajo de Direcciones y Callejero del Consejo Directivo de la Infraestructura de Información Geográfica de España (CODIGE).

2013: Premio DINTEL a Alta dirección.

2015: Premio Fundación Big Data / Asl@an

2014 y 2015: Premio TIC a la Administración General del Estado: SOCINFO.

GEORREFERENCIACIÓN Y DATOS ESTADÍSTICOS

Julià Talaya i López

*Subdirector General de Geodesia i Cartografia
Institut Cartogràfic i Geològic de Catalunya*

En los últimos años hemos asistido a una gran expansión de los datos abiertos. Estos conjuntos de información están producidos o recopilados por las administraciones públicas y puestos a disposición de los organismos públicos, de las entidades privadas y del ciudadano en general, para su utilización de una manera ágil y cómoda de forma gratuita. Entre ellos ocupan un lugar destacado los datos estadísticos y los datos geográficos. Un porcentaje muy alto de los datos abiertos no geográficos pueden ser georreferenciados, añadiendo así la componente espacial que permite aprovechar las capacidades de análisis que ofrecen los Sistemas de Información Geográfica y, a la vez, mejorar su comprensión a través de la publicación en visores geográficos y mapas temáticos.

Para aprovechar las posibilidades de georreferenciación, los organismos cartográficos están ofreciendo herramientas para transformar los sistemas de referenciación naturales, también llamados sistemas de referenciación indirecta, como las direcciones postales, los puntos kilométricos o los topónimos, a sistemas de referenciación por coordenadas, llamados sistemas de referenciación directa. Como ejemplo, el Institut Cartogràfic i Geològic de Catalunya inició la creación de una base de datos georreferenciada de calles, con el objetivo de dar respuesta a las necesidades de geocodificación de registros con direcciones postales en la Administración de la Generalitat de Catalunya. Dicho proyecto evolucionaría, por una parte, en el despliegue de una base de datos georreferenciada de direcciones, que cuenta actualmente con más de 1,6 millones de direcciones, y, por otra, en la publicación de servicios en línea de geocodificación directa, para obtener coordenadas a partir de direcciones, y de geocodificación inversa, para obtener direcciones a partir de coordenadas. En estos momentos se ha empezado un proceso de convergencia e integración de esta base de datos con otras bases de datos de direcciones elaboradas por algunas entidades municipales y supramunicipales, para obtener una única base sobre la totalidad del territorio, la Base

de datos municipal de direcciones de Catalunya (BDMAC). Otro elemento importante de georreferenciación es la base de topónimos que dispone de más de 300.000 topónimos georreferenciados, y que permiten la representación por coordenadas de la información del sistema de referencia natural basado en la toponimia.

La existencia y disponibilidad de bases de datos georreferenciadas de direcciones y de topónimos completas y de calidad es de vital importancia para los organismos oficiales que realizan la implantación de los actuales sistemas de información estadística, que incorporan como elemento fundamental la georreferenciación de los microdatos. Estas bases de datos son necesarias en la fase de preparación y recogida de la información, como soporte para extraer muestras para las encuestas. Son fundamentales también para la normalización y geocodificación de direcciones y toponimia en registros administrativos que puedan sustituir o complementar la recogida de datos por muestreo con un coste muy inferior, así como para georreferenciar las unidades estadísticas.

Así pues, los organismos oficiales de estadística incorporando la georreferenciación en los datos estadísticos, pueden sacar beneficio de ella en todas las fases, desde la recogida, pasando por la elaboración, y hasta la difusión. Pero, para ello, necesitan dotarse de un sistema de producción que permita almacenar y analizar los datos georreferenciados, con acceso a funcionalidades de normalización y geocodificación, y almacenar también las unidades territoriales a utilizar en la difusión de la información estadística.

Con respecto a la difusión, es importante considerar que los datos estadísticos georreferenciados a nivel de microdato permiten superar la tradicional limitación derivada de su asignación a sección censal. Así, se pueden agregar según unidades territoriales definidas por la propia administración o por el usuario, escogiendo las más adecuadas de acuerdo con el tipo de dato estadístico y el análisis o la explotación que se vaya a realizar. Entre ellas se pueden citar divisiones administrativas como la sección censal, el distrito, el municipio, el código postal o la región sanitaria, pero también otros tipos de unidades, como zonas de riesgo, por ejemplo, las zonas inundables, o cualquier otra zona de interés definida por el propio usuario. Mención

especial merece la zonificación de tipo rejilla, que divide el territorio en celdas homogéneas, por ejemplo, de 250 m, 500 m o 1 km de lado. Este tipo de zonificación permite disponer de datos estadísticos georreferenciados de gran detalle y de forma homogénea sobre todo el territorio. A resaltar que, aun utilizando unidades territoriales de mayor detalle, siempre se preserva el secreto estadístico.

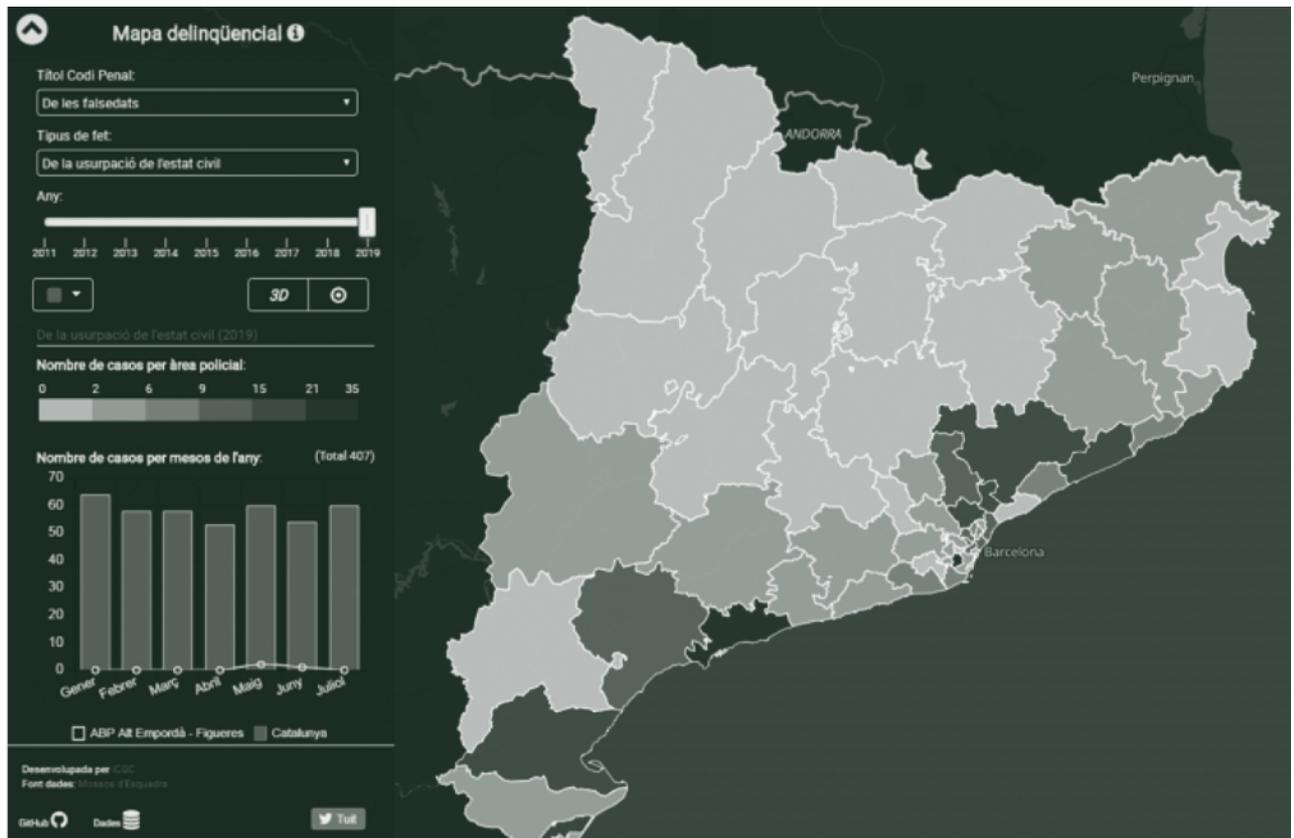
Las herramientas para la difusión se aprovechan también de la georreferenciación de datos estadísticos. Los visores permiten superponer estos datos con otros datos geográficos facilitando su interpretación y comprensión, ya que proporcionan la visión de la variable estadística sobre el territorio en cuanto a su situación y su distribución. A partir de una correcta estructuración y georreferenciación de los datos, los visores actuales permiten también avanzar desde la visualización de los datos a la modelización de geoinformación, combinando los diferentes conjuntos de datos estadísticos y permitiendo su utilización en modelos paramétricos. El Institut Cartogràfic i Geològic de Catalunya proporciona a los usuarios la plataforma Instamaps

diseñada para promover el uso y difusión de la geoinformación. Esta plataforma permite la creación de mapas de una manera sencilla a partir de conjuntos de datos georreferenciados y compartir el resultado mediante un visor específico.

Una consecuencia importante de la incorporación de la georreferenciación en los datos estadísticos es que conduce a su convergencia con los datos geográficos. De este modo, los datos geográficos pueden enriquecerse a su vez de la información contenida en las estadísticas. La utilización conjunta de datos específicos de la estadística y la geografía, conectados a través de la georreferenciación, crea sinergias entre ambas disciplinas que abren nuevas posibilidades en el uso de la información.

En la actual sociedad de la información, disponer de información geográfica y estadística diversa y de calidad conectada a través de la georreferenciación es imprescindible tanto para guiar la toma de decisiones por parte de la administración pública, para una correcta planificación y gestión del territorio o para proveer mejores servicios públicos, como para dar servicio a los usuarios en general.

Figura 1. El mapa delinqueñcial de Catalunya como ejemplo de georreferenciación de datos estadísticos.



Fuente: <https://visors.icgc.cat/mapa-delinqüencial/#755/41.602/1.239>

El registro estadístico de territorio: finalidades y calidad de sus fuentes

Eduard Suñé, Daniel Ibáñez, Anna Bernaus y Mireia Farré

Institut d'Estadística de Catalunya (Idescat)

El Institut d'Estadística de Catalunya (Idescat) está implementando un nuevo sistema de producción basado en registros administrativos. Este sistema está formado por una serie de subsistemas: el registro estadístico de población (REP), el registro estadístico de empresas y entidades (REE) y el registro estadístico de territorio (RET).

La idea es simple: si se gestiona adecuadamente la información que la administración conoce de la población y de las empresas, la obtención de resultados estadísticos resulta mucho más simple y económica que con el uso de métodos tradicionales que, generalmente, implican operaciones de campo.

Cuando hablamos de gestionar adecuadamente la información nos referimos a disponer de un sistema que sea capaz de mantener los estados por los que un individuo o empresa pueda pasar a lo largo de su existencia. En el caso de la población, como todo el mundo sabe, se nace, se obtiene una educación, se pasa por una vida laboral, se forma una nueva familia, etc. En todos estos cambios la administración interviene generando cierta información. Solo sería necesario gestionarla adecuadamente para obtener información estadística.

GEOLocalIZACIÓN Y VALIDACIÓN DE LAS DIRECCIONES POSTALES

El registro estadístico de territorio (RET) es el subsistema responsable de la gestión de las direcciones postales que aparecen en el resto de subsistemas, con una doble finalidad: geolocalizar los microdatos y validar las direcciones postales completas.

Hay que tener en cuenta que una dirección postal es una información compleja resultante de la combinación de una serie de campos: unos dan información a nivel horizontal y otros la dan a nivel vertical. Por ejemplo: calle Unión, 25, 4.º 3.ª de Vilanova i la Geltrú. En este caso concreto, el RET facilitaría las coordenadas (X, Y) del portal de la calle Unión nº 25 (geolocalización de la dirección horizontal) y comprobaríamos que existe un inmueble en la planta 4 puerta 3 (validación de la dirección vertical).

Al alcanzar esta doble finalidad, la información del resto de subsistemas tendrá la máxima preci-

sión espacial posible y será válida, con lo que desde el punto de vista territorial las estadísticas que puedan generarse serán de alta calidad.

La obtención de las coordenadas y la validación de la parte vertical de las direcciones postales se realizan utilizando sistemas de información externos al RET: por un lado, para la geocodificación se utiliza un servicio web del Instituto Cartográfico y Geológico de Catalunya (ICGC) y, por otro lado, para la validación de la dirección postal completa se utiliza la información alfanumérica de la Dirección General Catastro (DGC), considerándose una dirección válida cuando es posible encontrar en el Catastro un bien inmueble relativo a esa dirección.

En cuanto a la geocodificación, se obtienen las coordenadas del 96% de las direcciones postales cuando la información a tratar proviene del Registro estadístico de población (REP). Este porcentaje se explica por la alta calidad de las direcciones postales de entrada facilitadas por el Instituto Nacional de Estadística (INE), que incluye el código de vía. Para otras fuentes, los porcentajes pueden ser sensiblemente inferiores debido, entre otras razones, a la no disponibilidad de los códigos de vía o porque la dirección postal no está estructurada y se debe preprocesar antes de ejecutar el servicio web de geocodificación.

En cuanto a la validación de la dirección postal completa, que recordamos consiste en comprobar que existe un bien inmueble relativo a esa dirección, presenta la dificultad derivada de que los valores de campos como tipo de vía, código de vía, nombre de vía, bloque, escalera, planta y puerta presentan valores diferentes en los dos conjuntos de datos (la información del Registro estadístico de población (REP) procedente del INE y la información de la DGC). Por lo tanto, ha sido necesario construir un conjunto de tablas de equivalencias para cada uno de los campos que configuran la dirección, siendo el que se refiere a las vías el más problemático, dada su elevada cardinalidad. En efecto, en Catalunya existen aproximadamente unas 100.000 vías y solo un 50% de ellas tienen nombres exactamente iguales en los dos conjuntos de datos.

Para resolver este problema, evitando soluciones que impliquen el cálculo de métricas entre

literales, hemos utilizado métodos geométricos: una vez geocodificada la fuente de entrada (REP), obtenemos un conjunto de puntos que pueden ser agrupados según el código de vía. Así, para cada vía con más de dos portales geocodificados puede construirse un polígono mediante sus envolventes cóncavas obteniendo un polígono que es una buena aproximación a la geometría de la vía.

Por otro lado, dado que la información alfanumérica de las fincas de la DGC contiene sus centroides, puede realizarse también para esta fuente la operación descrita anteriormente, obteniendo así dos conjuntos de geometrías: las envolventes cóncavas de las vías INE y Catastro. La entidad física vía, tenga la descripción que tenga en las dos fuentes, tendrá asociados dos polígonos muy similares, es decir, dos polígonos con un elevado porcentaje de intersección y una orientación espacial (azimut) muy similar.

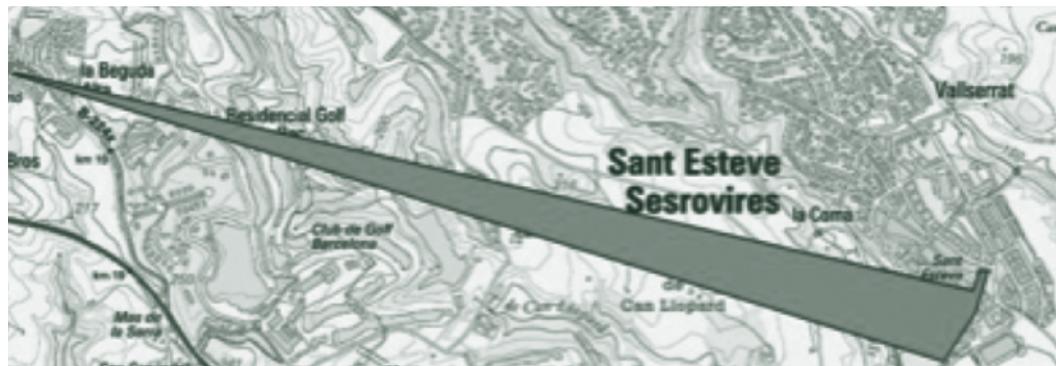
Con este criterio y bajo supervisión manual se logró construir una tabla de equivalencias entre vías INE y Catastro que cubría un porcentaje altísimo de portales, abriendo así el camino hacia la validación de las direcciones postales completas.

Así mismo, se construyeron las tablas de equivalencias del resto de campos de la dirección postal completa (horizontal y vertical) y en estos momentos hasta un 80% de las direcciones postales del RET son válidas, ya que han podido identificarse en la información de bienes inmuebles de la DGC.

MÉTODOS DE CONTROL DE CALIDAD DE LA DIRECCIÓN HORIZONTAL

Al generar las envolventes cóncavas de la fuente DGC observamos la presencia de puntos que sin duda eran erróneos debido a que su posición era anormalmente distante del resto, como se ilustra en la Figura 1.

Figura 1.



Interpretamos que, en la mayor parte de los casos, esto ocurre cuando el código de vía y alguna de las posiciones son incoherentes, es decir, cuando el código de vía asignado a una finca es incorrecto.

Para asegurar la calidad de las posiciones obtenidas en el RET era necesario disponer de algún indicador que permitiera detectar polígonos con algún punto erróneo, mediante métodos de control supervisados.

Del conjunto de envolventes cóncavas obtenidas con la información de fincas de la DGC (unos 80.000 polígonos), se extrajo una muestra discrecional de 1.500 polígonos para obtener los grupos de aprendizaje y test. Cada uno de los polígonos de la muestra se inspeccionó y etiquetó como correcto o erróneo como paso previo a la exploración de indicadores.

En primer lugar, se realizó un estudio univariante utilizando variables derivadas de los polígonos (número de puntos, longitud de los lados, etc.) usando como método de comparación las áreas bajo la curva COR¹.

Finalmente, se evaluaron indicadores multivariantes y de ellos el de mejor comportamiento fue el siguiente indicador de calidad del polígono (ICP_p):

$$ICP_p = CV_p^2 \cdot \max\{longitud_i - longitud_{i+1}\}_{i=1 \dots v_p - 1}$$

Donde "i" recorre los lados del polígono "p" según su longitud decreciente, "v_p" es su número de vértices y "CV_p" es el coeficiente de variación de las longitudes de los lados. Todo ello sin tener en cuenta el lado de mayor longitud.

¹ La curva COR (ROC en inglés, fórmula generalmente utilizada) es un gráfico que compara la sensibilidad (verdaderos positivos dividido entre los verdaderos positivos y los falsos negativos) y la especificidad (verdaderos negativos dividido entre los verdaderos negativos y los falsos positivos). En este estudio, se entiende como verdadero positivo un polígono etiquetado como erróneo y que realmente lo es.

La razón de excluir el lado más largo es clara: un polígono que se corresponda con una vía que solo tiene fincas a un lado, un paseo marítimo, por ejemplo, tendrá necesariamente un lado mucho más largo que el resto (el que cierra el polígono). Un polígono erróneo tendrá uno o más puntos alejados del resto, pero tendrá al menos dos lados anormalmente largos.

Para cada polígono puede evaluarse el indicador de calidad del polígono (ICP) y obtener un subconjunto con una probabilidad muy alta de que sean incorrectos, tal y como puede observarse en la Figura 2.

Figura 2.

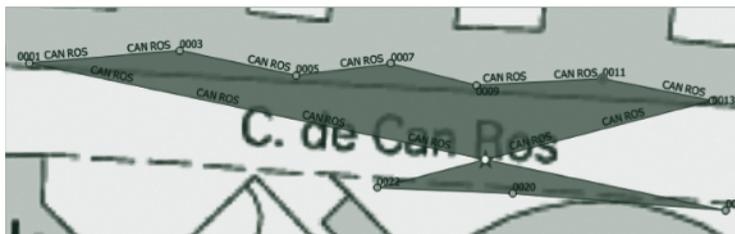


Los polígonos marcados en rojo son muy probablemente incorrectos, es decir, contienen algún punto incorrecto ya sea porque las posiciones sean erróneas o porque el código de vía asignado a la finca no es correcto.

Un método alternativo que puede utilizarse en relación al control de calidad de datos geocodificados consiste en detectar incoherencias entre el número del portal (número de policía) de una dirección postal y la posición de los portales. En efecto, en función del sistema de numeración de una calle, pueden construirse polígonos uniendo ordenadamente los puntos geocodificados. En principio, un polígono así construido debería ser geométricamente válido, es decir, que no se autointerseque. Debemos advertir aquí que este tipo de controles puede dar lugar a falsos positivos dependiendo de la distribución de portales en los dos lados de la calle y que hemos localizado casos de vías, como la calle de Can Ros en Barcelona, en los que la numeración es inconsistente pero correcta (Figura 3), crece en el lado de impares mientras que decrece en el lado de pares.

Por otra parte, en relación a la validación de las direcciones verticales, los resultados dependen del

Figura 3.



contenido de las tablas de equivalencias: tipos de vía, vías, bloque, escalera, planta y puerta. Para contrastar la calidad se construyeron tablas de equivalencias alternativas usando datos del modelo de direcciones de la Administración General de Estado (MDAGE), observando diferencias poco significativas entre las tablas de equivalencias construidas en Idescat y las que se derivan de los datos de la MDAGE.

La diversidad de las descripciones de las direcciones postales en las distintas fuentes utilizadas en el Registro Estadístico de Territorio dificulta la correcta identificación espacial de las unidades estadísticas

CONCLUSIONES

El Registro Estadístico de Territorio incorpora diferentes fuentes con el fin de asignar y validar un identificador territorial para cada una de las unidades estadísticas, como población, hogares, empresas, entidades, etc.

La diversidad de las descripciones de las direcciones postales en las distintas fuentes utilizadas dificulta la correcta identificación espacial de las unidades estadísticas. Para ello, el uso de métodos espaciales y geométricos ha demostrado ser una herramienta muy poderosa para establecer tablas de correspondencias entre las diferentes fuentes, así como para validar, detectar errores y establecer indicadores de calidad para las direcciones georreferenciadas.

Sobre la preservación del secreto estadístico en datos georreferenciados

Cristina Rovira y Eduard Suñé

Institut d'Estadística de Catalunya (Idescat)

Este documento describe una metodología desarrollada en el Instituto de Estadística de Catalunya (Idescat) para la preservación de la confidencialidad estadística en la difusión de los datos geocodificados procedentes de registros estadísticos.

Los registros estadísticos tienen en común la existencia de un conjunto de variables que expresan la ubicación espacial de los microdatos: la dirección postal (lugar de residencia, sede de una empresa, etc.).

Históricamente, el tratamiento dado a las direcciones postales condujo a la asignación de áreas administrativas, como en el caso de la población y las secciones censales. El efecto final de esta asignación es que los resultados estadísticos solo se pueden obtener con respecto a estas áreas o sus agregaciones, lo que ocasiona una pérdida indeseable de información.

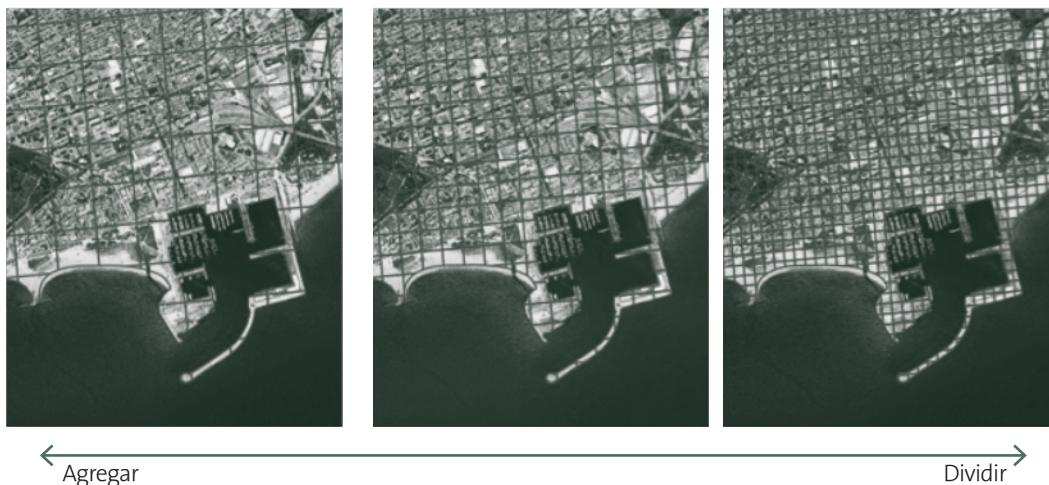
Los microdatos que contienen direcciones postales pueden transformarse mediante la geocodificación en otros que contienen, además de la información original, unas coordenadas que indican una posición en el espacio, es decir, se han transformado esos microdatos, utilizando la nomenclatura usada en los sistemas de información geográficos, en una capa de puntos.

No hace falta argumentar que desde el punto de vista de la preservación del secreto estadístico, esa capa de puntos no debe difundirse bajo ningún concepto ya que el riesgo de revelación es altísimo. Informar sobre la localización exacta de un microdato eleva el riesgo de revelación a unos niveles inasumibles. Tendremos que aplicar métodos de preservación que mantengan un cierto equilibrio entre seguridad y precisión.

Ante la problemática que acabamos de describir, Idescat decidió promover unos trabajos de investigación con la finalidad de encontrar soluciones metodológicas al problema de la difusión de datos georeferenciados. El grupo de trabajo analizó los dos grandes métodos de control del riesgo de revelación: la perturbación de coordenadas y la agregación espacial y determinó que la forma más adecuada de preservación del secreto estadístico era la agregación espacial, utilizando para ello un sistema multiresolución basado en *quadrees*.

Un *quadtree* es una estructura jerárquica que partiendo de una rejilla de elementos cuadrados se van dividiendo recursivamente por la mitad, de forma que cada elemento puede tener cuatro descendientes. Su construcción puede realizarse también por agregación, es decir, uniendo cuatro elementos contiguos para obtener un elemento del nivel superior (Figura 1).

Figura 1.



Si el proceso de división (o agregación) viene determinado por el número de puntos de la capa original dentro de los elementos, se optimiza la resolución adecuadamente, ya que en zonas de alta densidad tendremos resoluciones altas y en zona de baja densidad resoluciones menores.

Consideremos que nuestra capa de puntos es precisamente la población y que la regla de construcción del *quadtree* es que solo se procederá a la división de un elemento si el total de población asociado a los descendientes es mayor o igual que un cierto umbral, o si agregamos, se procederá a la unión espacial cuando alguno de los elementos contiguos tenga una población inferior a ese umbral.

Con esta estructura y regla conseguimos que la resolución espacial dependa de la densidad de puntos de la capa original, siendo mayor la resolución en zonas de alta densidad y menor en zonas de baja densidad. Implícitamente se optimizan también los riesgos de revelación y se minimizan los posibles errores que los usuarios de información puedan llegar a cometer en sus cálculos, cuestión que se aborda con detalle al final del artículo.

Por razones obvias, nuestra rejilla debe ser un estándar, de tal forma que sus resultados puedan incorporarse si es necesario a otros ámbitos superiores. En Idescat hemos partido de la rejilla

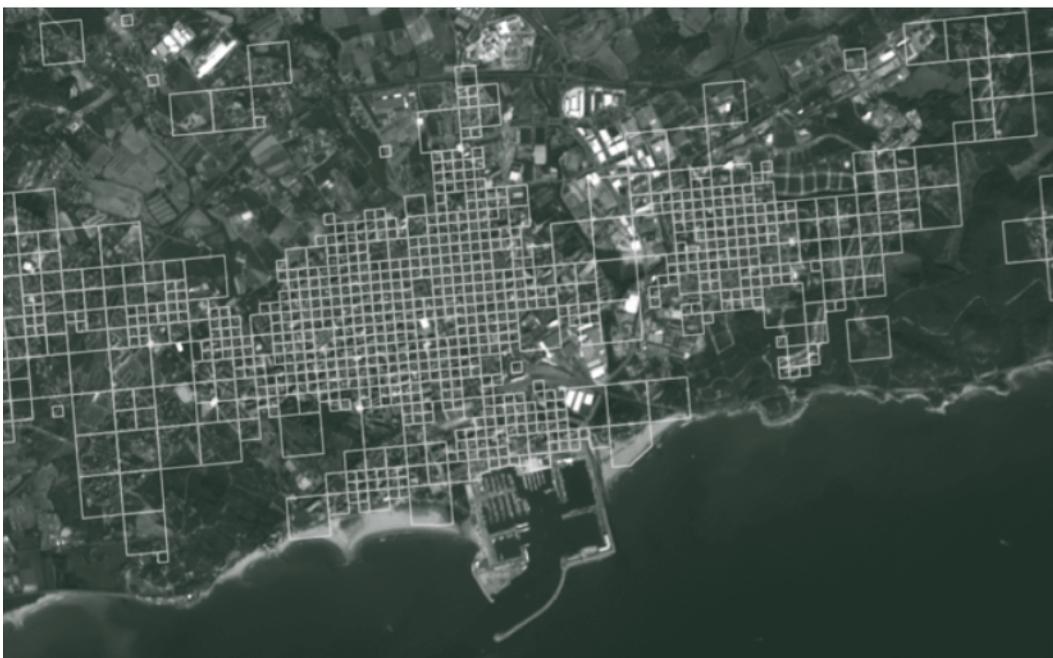
estándar europea de 1 Km para obtener las subdivisiones que el *quadtree* necesitaría, es decir, 500 m, 250 m, 125 m, 62,5 m.

Es necesario decidir, para unos microdatos en concreto, tres parámetros que determinarán finalmente nuestro *quadtree*: la resolución máxima, la resolución mínima y el valor del umbral que determina la agregación.

La resolución máxima es el grado de máxima subdivisión del espacio para el que consideramos que no hay riesgo de revelación. La resolución mínima se corresponde con el grado máximo de agregación que consideramos aceptable. Aunque viene en realidad determinado por el valor del umbral y los microdatos georeferenciados, hay zonas en las que la densidad de puntos es tan baja que deberíamos alcanzar niveles de agregación tales que esa información dejaría de ser útil para los usuarios. En los casos en que estamos por debajo de umbral se detiene la agregación y se asigna un valor nulo al total de efectivos para ese elemento.

Para la difusión de los datos georeferenciados del Registro estadístico de población, determinamos que las resoluciones máximas y mínimas serían de 62,5 m y 250 m y el valor umbral de 17 habitantes. Con estos parámetros, el *quadtree* resultante para una zona del municipio de Vilanova y la Geltrú da la imagen de la Figura 2.

Figura 2.



Se pueden observar los tres niveles de precisión: 62,5 m, 125 m y 250 m que quedan definidos por el umbral de 17 habitantes para el Registro de población a 1 de enero de 2016.

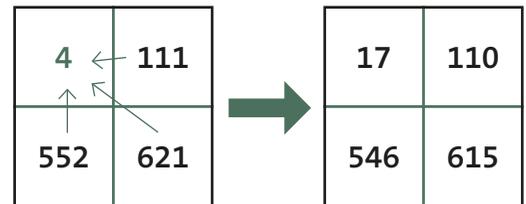
Dentro de cada elemento de esta jerarquía de geometrías se sitúan los microdatos georreferenciados y es posible calcular las matrices de frecuencias para diferentes variables como la edad, el sexo o la nacionalidad.

La generación del *quadtree* presenta un problema cuando la densidad de puntos decae muy rápidamente en un grupo de hermanos de la jerarquía (efecto frontera). Puede ocurrir que uno de los elementos tenga un valor inferior al umbral mientras que el resto tenga unos muy superiores. Si se construye el *quadtree* obtendríamos la agregación de estos cuatro elementos y desde el punto de vista del usuario sería equivalente a tener los cuatro elementos con un valor igual a la media de los cuatro.

En el siguiente ejemplo tenemos cuatro hermanos de la estructura con valores 4, 111, 552 y 621 de población respectivamente. El algoritmo de generación del *quadtree* realizará la agregación ya que hay un elemento que tiene un valor inferior al umbral y el resultado será un cuadrado con una población total de 1.288 habitantes y sería equivalente a cuatro con la322 (Figura 3).

La solución que proponemos para evitar este efecto es la traslación proporcional de población desde los elementos que superan el umbral hacia los elementos que no lo superan con ciertas condiciones (Figura 4).

Figura 4.



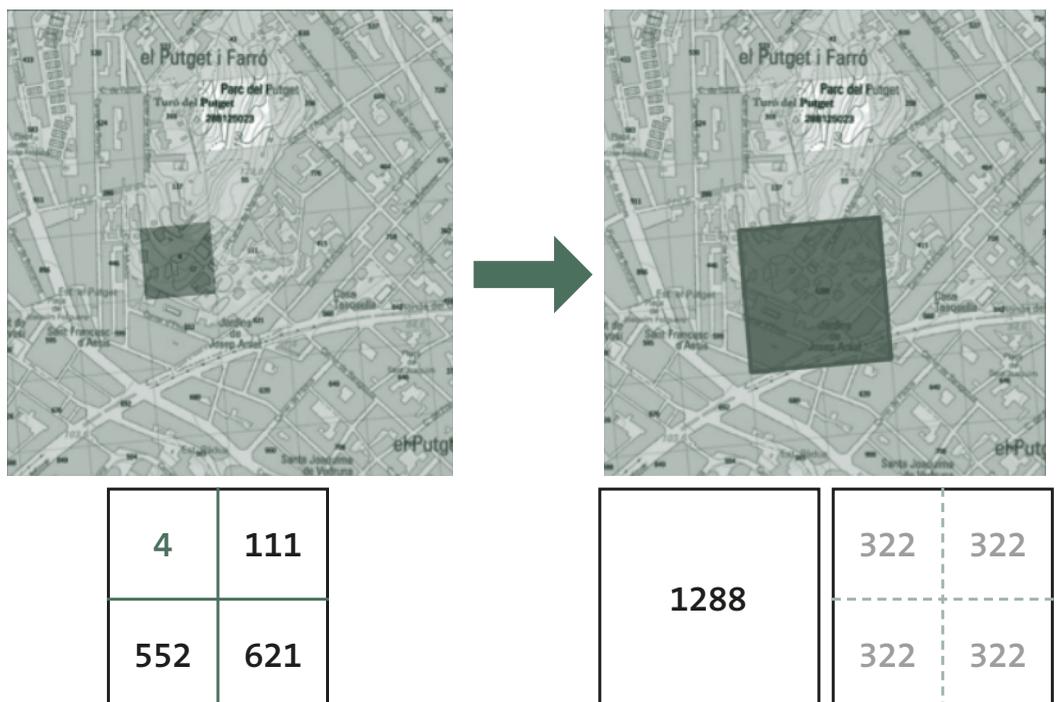
El error absoluto de las traslaciones es $e > 2N$ siendo N el número de elementos trasladados.

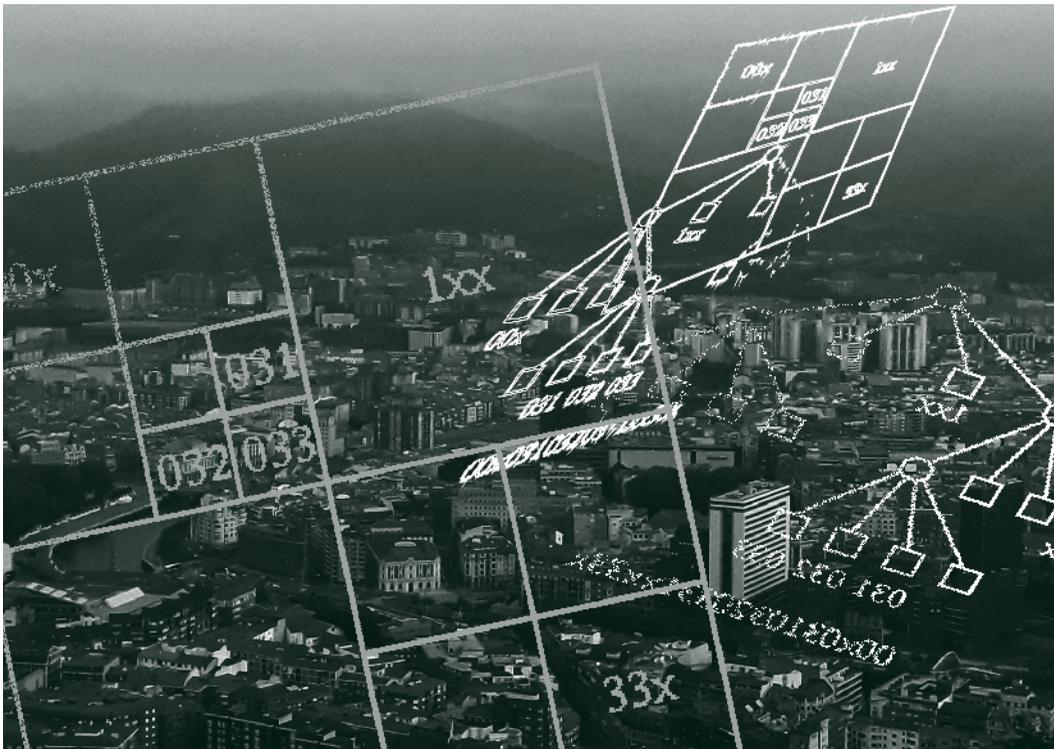
Por otro lado, el error absoluto derivado de la agregación es:

$$e = \sum p - \mu$$

Siendo p la población de cada elemento antes de la agregación y la media de las poblaciones de los cuatro hermanos.

Figura 3.





Pues bien, las traslaciones se llevan a cabo cuando $e > 2N$ siendo N el número de elementos a trasladar.

Esta solución complica la generación del *quadtree* ya que en cada paso de la recursión deben analizarse los casos en que se cumple la condición de traslación, pero de esta forma pueden evitarse más de la mitad de las agregaciones de un nivel al superior según nuestras observaciones.

Quizás la pregunta más importante que nos podemos hacer en relación a esta forma de difundir datos georeferenciados, que suponemos que ya son seguros desde el punto de vista de la preservación del secreto estadístico, sea la siguiente: ¿qué va a hacer un usuario con estos *quadtrees*?

Seguramente los usuarios interesados en este producto necesiten calcular agregados de esos microdatos para unas zonas concretas de su interés, zonas que no conocemos. Para hacer esos cálculos tendrán necesariamente que suponer que los microdatos se distribuyen uniformemente en cada elemento del *quadtree* y por lo tanto van a cometer errores derivados de esta suposición.

Mediante métodos de Monte Carlo y para los *quadtrees* derivados del Registro estadístico de población, hemos estimado esos errores.

Efectivamente, generando un conjunto aleatorio de polígonos (unos 50.000) hemos calculado

las poblaciones exactas mediante los microdatos georeferenciados originales y las poblaciones utilizando diferentes tipos de *quadtrees* (variando las resoluciones máxima y mínima, manteniendo el nivel umbral en 17 habitantes) llegando a la conclusión de que para el *quadtree* de 62,5 m, 250 m y umbral 17, la mediana de los errores relativos es del 5%.

Además, la distribución de los errores relativos y sus medidas de tendencia central dependen fuertemente del área de los polígonos: cuanto más grande sea el área menores son los valores de las medianas del error relativo, ya que estas áreas incluirán completamente muchos elementos del *quadtree* (recordemos que la población en cada elemento del *quadtree* es exacta, menos para aquellos pocos casos que han sufrido traslaciones para evitar el efecto frontera).

Así pues, la conclusión es la siguiente: la difusión de los datos mediante la estructura jerárquica en *quadtrees* minimiza el riesgo de revelación, pero los usuarios pueden cometer errores en el momento de calcular efectivos en las geometrías de su interés. Si ese error, que hemos estimado en un 5% de mediana, no es aceptable, la solución es que el usuario proporcione su geometría al Idescat y proceder entonces a realizar los cálculos devolviendo al usuario su geometría y los valores obtenidos, siempre y cuando ello no suponga ningún riesgo de revelación.

Estadísticas catastrales

Cristina de la Cierva Rodríguez de Rivas

Jefe de Área de Estadística. Subdirección General de Estudios y Sistemas de Información. Dirección General del Catastro

En las últimas décadas el Catastro se ha convertido en una gran infraestructura de información territorial puesta a disposición tanto de los ciudadanos como de las Administraciones Públicas. Su utilización es generalizada y ha contribuido notablemente a mejorar la seguridad jurídica y a dotar de transparencia al mercado inmobiliario.

La Dirección General del Catastro cuenta con información alfanumérica y cartografía catastral correspondiente a todos los bienes inmuebles de su ámbito de competencia. La descripción catastral de los bienes inmuebles comprende sus características físicas, económicas y jurídicas, entre las que se encuentran la localización y la referencia catastral, la superficie, el uso o destino, la clase de cultivo o aprovechamiento, la calidad de las construcciones, la representación gráfica, el valor de referencia de mercado, el valor catastral y el titular catastral, con su número de identificación fiscal o, en su caso, número de identidad de extranjero. Además, cuando los inmuebles estén coordinados con el Registro de la Propiedad se incorpora dicha circunstancia junto con su código registral. La información catastral comprende, por tanto, datos físicos, económicos y jurídicos de los inmuebles.

La Dirección General del Catastro cuenta con información alfanumérica y cartografía catastral correspondiente a todos los bienes inmuebles de su ámbito de competencia

Es posible acceder a esta información mediante servicios web y descargarla, y si se hace a través de la Sede Electrónica del Catastro es un producto gratuito, con la salvedad de que los datos de titularidad (nombre, apellidos, razón

social, código de identificación y domicilio de quienes figuren inscritos en el Catastro Inmobiliario como titulares) y el valor catastral (y los valores catastrales de suelo y construcción individualizados) tienen la consideración de datos protegidos, por lo que a estos son de aplicación las restricciones de acceso establecidas en la legislación.

La elaboración y el análisis de la información estadística contenida en las bases de datos catastrales y la relativa a la tributación de los bienes inmuebles es competencia de la Dirección General del Catastro; y la Subdirección General de Estudios y Sistemas de Información ejerce estas funciones, según se establece en el art.5 del RD 1113/2018.

Las estadísticas catastrales que publica la D. G. del Catastro están además sometidas a determinados requisitos al estar incluidas en el inventario de operaciones estadísticas de las Administraciones Públicas e integradas en el Plan Estadístico Nacional 2017-2020.

El Plan Estadístico Nacional es el principal instrumento ordenador de la actividad estadística de la Administración General del Estado. Comprende las estadísticas para fines estatales que han de efectuarse en el cuatrienio por los servicios estadísticos de la Administración General del Estado o cualesquiera otras entidades dependientes de ella. Para la ejecución temporal del mismo se elaboran programas anuales en los que se incluyen las actuaciones estratégicas y operativas que se vayan a efectuar.

En 2018 la D. G. del Catastro realizó una importante labor de análisis de las estadísticas que estaban incluidas en el Plan, y propuso una revisión y reestructuración de las mismas, que fue aprobada y es de aplicación desde 2019. Las operaciones estadísticas que actualmente se realizan y publican son:

- Estadística del Impuesto sobre Bienes Inmuebles
- Estadística de Ordenanzas Fiscales del Impuesto sobre Bienes Inmuebles
- Estadística del Catastro Inmobiliario Urbano
- Estadística del Catastro Inmobiliario Rústico
- Estadística del Catastro de Bienes Inmuebles de Características Especiales (BICE)
- Estadística de Titulares Catastrales

Las dos primeras estadísticas se generan a partir de la información del Impuesto que nos remiten las entidades locales: número de recibos, bases imponibles, bases liquidables, cuotas, etc. Los tipos impositivos son fijados anualmente por los Ayuntamientos y es información oficial y pública. Sin embargo, los datos relativos a exenciones, bonificaciones y cuotas, en la medida en que el impuesto es gestionado unas veces por las Diputaciones, otras por los Ayuntamientos y otras por entidades gestoras, en ocasiones son transmitidos con diferencias de criterio, lo que conlleva trabajos añadidos de depuración y control de calidad.

Las cuatro últimas operaciones se realizan a partir de los datos obrantes en el Catastro Inmobiliario, en base a los padrones que se generan anualmente. El Catastro como fuente de información, al no ser una fuente estadística como tal, sino un registro administrativo susceptible de ser tratado como fuente estadística, presenta ciertas necesidades de procesos de homogeneización y

control de la calidad que se han ido acometiendo en los últimos años.

El Catastro, siendo un registro administrativo susceptible de ser tratado como fuente estadística, presenta ciertas necesidades de procesos de homogeneización y control de la calidad que se han ido acometiendo en los últimos años

En el Portal de Catastro, en el apartado de “Estadísticas” se ofrece este amplio abanico de información. La estructura de los datos que publicamos se adapta a los compromisos con el Plan Estadístico Nacional y se ha diseñado una pantalla inicial clara, ordenada, que facilita la búsqueda, y que supone una clara mejora respecto a la anterior, más textual.

Como se puede apreciar en la figura 1, en los botones de acceso ya se indican los ámbitos territoriales a los que están referidas cada una de las estadísticas, el periodo temporal y las variables de las que se ofrece la información.

La aplicación permite seleccionar además del año al que están referidos los datos (por de-

Figura 1.



fecto se ofrece el último), el ámbito territorial, ofreciéndose la información a nivel nacional, de comunidad autónoma, de provincia y de municipio, y en el caso del Anuario, también por distrito censal.

Es posible personalizar las consultas y seleccionar uno, varios o todos los indicadores que se ofrecen, así como elegir el formato de salida y disposición en columnas y/o filas de las variables elegidas. Además, la vista con los resultados de la consulta según la selección realizada se puede descargar en un amplio abanico de formatos.

En el caso de la información ofrecida sobre valores catastrales existe una limitación que nace de su carácter de dato protegido. Es posible publicar datos agregados, pero no individualizados. En el caso de los Bienes Inmuebles de Características Especiales (BICE) nos podemos encontrar con que en un municipio haya ninguno o muy pocos. Por ello, solo se publican estadísticas de valores catastrales cuando en el municipio existen más de tres inmuebles de este tipo.

La información catastral constituye una herramienta para garantizar la seguridad jurídica y fortalece la transparencia del mercado inmobiliario

Uno de los grandes potenciales que tiene la información catastral es precisamente su georreferenciación. Todos los inmuebles disponen de una referencia catastral que les identifica de forma unívoca y se encuentran georreferenciados. Ello posibilita que mediante cruces gráficos se pueda facilitar la información por el ámbito seleccionado. Actualmente se está trabajando en esta línea y en un futuro próximo se quiere facilitar la información por secciones censales, lo que permitirá hacer un análisis más preciso y valioso de las características y estructura de los inmuebles, los valores y los datos jurídicos y tributarios.

La D. G. del Catastro siempre ha sido puntera en desarrollos tecnológicos y cuenta con numerosas aplicaciones tanto de uso interno como externo, además de un Portal y una Sede electrónica, con la que se facilita el acceso de los ciudadanos, empresas, organismos y administraciones a la infor-

mación. El programa SIGCA (Sistema de Información Geográfica Catastral) permite realizar mapas temáticos con la información alfanumérica. Las Ponencias de Valores catastrales, con sus mapas de valor de suelo, son un claro exponente de cómo es posible publicar información con base cartográfica.

Actualmente se está realizando un estudio de las diferentes aplicaciones de Business Intelligence existentes en el mercado con el fin de modernizar la publicación de las estadísticas, mejorar y sistematizar los estudios, la usabilidad de los formatos y los servicios de difusión de la información catastral.

El objetivo es implantar una aplicación que facilite la información de una forma mucho más visual, apoyada en cartografía, garantizando a su vez la calidad del proceso y el producto.

Con todo, hay que diferenciar las estadísticas bajo el paraguas del Plan Estadístico Nacional, absolutamente normalizadas, que son las que se pretenden ofrecer georreferenciadas, de las estadísticas que surgen en el día a día y que son generadas con *queries* y consultas a medida, para lo cual es necesario un conocimiento profundo de la estructura de la base de datos, sus tablas, campos y procedimientos.

La información es mucha y las posibilidades que ofrece su estudio y explotación cada vez mayor con la utilización de Machine Learning e inteligencia artificial. En la D.G. del Catastro convive la estadística en su doble vertiente, descriptiva y diferencial, que permite no solo explicar la estructura inmobiliaria sino también realizar predicciones de comportamiento futuro.

En la sociedad digital la información territorial ocupa cada vez un papel más relevante. La información catastral es mucho más que un conjunto de datos al servicio de las políticas tributarias, constituye una herramienta para garantizar la seguridad jurídica y fortalece la transparencia del mercado inmobiliario.

Finalicemos destacando la vocación de difusión, accesibilidad, transparencia y calidad de los datos, gráficos y literales, y recordando que Naciones Unidas hace ya años que puso de manifiesto la función decisiva que desempeña la información estadística oficial para el análisis y la adopción de decisiones en materia de políticas públicas y desarrollo sostenible. A través de un buen análisis estadístico es posible detectar carencias y en consecuencia implantar mejoras en la calidad de los servicios públicos. La georreferenciación de las estadísticas y su vinculación al territorio facilitan sin lugar a dudas esta tarea.

Estadísticas y Georreferenciación en las ayudas PAC

Miguel Ángel Riesgo

Presidente del Fondo Español de Garantía Agraria (FEGA)
Ministerio de Agricultura, Pesca y Alimentación (MAPA)

ORIGEN DE LOS DATOS

Con el paso de los años, las prioridades de la Política Agraria Común (PAC) hacia cuestiones cada vez más transversales y de índole medioambiental han ido modelando cada uno de los periodos de referencia. Esto ha provocado que la cantidad de datos generada en el ámbito de las administraciones y en particular en el “Fondo Español de Garantía Agraria, FEGA O.A.” (Organismo de Coordinación en España) vaya en aumento.

Cada vez que los beneficiarios realizan su petición anual de ayudas, facilitan a los Organismos Pagadores información acerca de sus cultivos y actividades agrarias para que la administración pueda realizar las correspondientes verificaciones previas a la concesión de las mismas. Esta petición anual se realiza a través de la denominada **Solicitud Única de la PAC (SU)**.

ANTECEDENTES

Tenemos que remontarnos a los años noventa, época en la cual las orientaciones de la PAC tienden a una disminución de los precios garantizados en favor de los pagos compensatorios (actuales ayudas directas del primer pilar de la PAC) por hectárea y cabeza de ganado. Con el **Reglamento comunitario 3508/92 del Consejo** se establece lo que conocemos actualmente como el **Sistema Integrado de Gestión y Control «SIGC»**. A raíz de la puesta en marcha de este sistema, la gestión de las diferentes ayudas se fue integrando y los elementos que lo integraban se hicieron cada vez más complejos.

El SIGC está formado por una Base de Datos Informática con la información de las ayudas, un **Sistema de Identificación Geográfica** que incorpora las parcelas de los beneficiarios a nivel nacional, un registro de los derechos de ayudas, las solicitudes de los agricultores (Solicitud Única) con la identificación de estos últimos y una serie de controles administrativos y sobre el terreno que se

realizan cada campaña. Todos estos elementos en su conjunto han generado con el paso de los años una información muy detallada y georreferenciada del sector agrario español.

Durante el periodo de referencia PAC 2014-20 ha cambiado la forma en la que el agricultor solicita sus ayudas. Se ha pasado de un formulario en el que la información facilitada por los agricultores era puramente alfanumérica, a una **declaración gráfica** de las superficies. Esto implica que desde la campaña 2018 tenemos, con carácter anual, un mapa bastante detallado de los cultivos de cada campaña con su delimitación gráfica. Esto último resulta muy útil a la hora de generar datos agregados y georreferenciados y facilitarlos a otras instituciones a través de convenios de colaboración.

FUTURA REFORMA PAC

Actualmente, se está trabajando en lo que será la futura reforma de la PAC en el periodo 2021-27. Dicha reforma se articula en **9 objetivos** específicos que van encaminados a impulsar un **sector resiliente**, atendiendo en todo momento al **medio ambiente y al clima** y reforzando el **medio rural**. En este sentido, se han agrupado estos objetivos por su carácter **económico, medioambiental y social**. Por otro lado, tenemos una serie de objetivos transversales encaminados a la **sostenibilidad, modernización, digitalización** y a la **simplificación** de esta política.

Dentro de los borradores de los futuros reglamentos podemos hacernos una idea del cambio de paradigma que propone la Comisión Europea (CE). Uno de los grandes cambios pasa por el modelo de gobernanza. Para la PAC post 2020 se propone el nuevo “*Delivery Model*”, un modelo orientado a los objetivos en los que la CE únicamente quiere verificar el cumplimiento de unos requisitos básicos y revisar que el Estado miembro (EM) ha diseñado una serie de intervenciones (ayudas) enfocadas a las necesidades detectadas por este último.

Para ello, se establece una serie de indicadores de realización, resultados e impacto que permitirán medir la efectividad de las intervenciones diseñadas por los Estados miembros y de la PAC en su conjunto. Muchos de estos indicadores se obtendrán de fuentes estadísticas oficiales, específicamente de EUROSTAT.

Otra de las novedades que introducen los futuros reglamentos es el **sistema de monitorización de superficies** como nuevo elemento del SIGC. Este sistema se basa en la verificación periódica y continua de la compatibilidad entre la actividad agraria declarada por el agricultor y la observada en la serie temporal de señales provenientes de los satélites **Sentinel** del programa **Copernicus** de la Agencia Espacial Europea. Esto permite un seguimiento continuo de las superficies y actividades agrarias para el 100% de las explotaciones agrícolas, y en todas sus parcelas.

La obligación de poner los datos del SIGC a disposición de las autoridades estadísticas es otra novedad que incluye el borrador del reglamento.

La obligación de poner los datos del SIGC a disposición de las autoridades estadísticas es otra novedad que incluye el borrador del reglamento. Los Estados miembros deben asegurarse de que el conjunto de datos recopilados mediante el sistema integrado que sean pertinentes para la **elaboración de estadísticas europeas**, tal y como se establece en el Reglamento (CE) nº 223/2009, puedan ser compartidos de forma gratuita con la autoridad estadística comunitaria, los institutos nacionales de estadística y, en su caso, con otras autoridades nacionales responsables de la elaboración de estadísticas europeas.

COLABORACIÓN FEAGA E INE (CENSO AGRARIO 2020)

El Reglamento (CE) Nº 1166/2008 del Parlamento Europeo y del Consejo establecía un marco para las estadísticas europeas sobre la estructura de las explotaciones agrícolas en el período 2009-2016. Para actualizar la información estructural de las

explotaciones estaba previsto efectuar un censo agrario cada diez años. El censo más reciente tuvo lugar en 2009/2010. La información del censo se actualizaba con dos encuestas de estructura en periodos intercensales.

Desde el año 2015, EUROSTAT, en colaboración con los Estados miembros, trabajó en la elaboración de una propuesta de **Reglamento (UE) 2018/1091 del Parlamento Europeo y del Consejo**, en el que se prevé la realización de un censo agrario con referencia a 2020. El Reglamento se aprobó por procedimiento de codecisión en julio de 2018. En noviembre del mismo año, se aprobaron los actos de ejecución que especifican el detalle de variables a investigar. Asimismo, durante el primer semestre de 2019, EUROSTAT ha elaborado un **manual metodológico** con la metodología a desarrollar en el próximo censo agrario.

Los Estados miembros han de recopilar y presentar «datos estructurales básicos» sobre explotaciones agrícolas correspondientes a los años de referencia **2020, 2023 y 2026**, según 186 variables que figuran en el anexo del reglamento. La recogida de datos básicos para el año de referencia 2020 se realizará mediante un **CENSO**.

Para obtener los datos contemplados en el presente Reglamento, los Estados miembros podrán recabar la información a través de encuestas estadísticas, obtener los datos de registros administrativos u otras fuentes. El INE ha acordado con el FEAGA la modificación del declarativo de la Solitud Única de ayudas de la PAC, a efectos de incluir las variables mínimas necesarias para dar respuesta a las variables que se exigen en el Reglamento Europeo del censo agrario. De esta forma, se dará respuesta a las exigencias del Reglamento, realizando encuestas solo en las explotaciones que no solicitan ayudas PAC.

Esta colaboración, a parte de dar cumplimiento a las futuras obligaciones mencionadas que introducen las propuestas de reglamentos PAC, reducirá la carga de las unidades informantes (agricultores) y contribuirá a la reducción del coste de la operación censal. Por otro lado, obtener esta información a través de la SU permite disponer de datos actualizados y continuos cada campaña.

SEN4STAT: SENTINELS PARA LAS ESTADÍSTICAS

El aumento de la productividad agrícola sostenible es el desafío número uno del informe "El futuro de la alimentación y la agricultura" de la Organización de las Naciones Unidas para la Agricultura y la Alimen-

tación (FAO) en respuesta al aumento previsto de la demanda de alimentos en un 50% entre 2012 y 2050.

Se requiere una amplia gama de datos e información oportuna sobre prácticas agrícolas y recursos naturales a escala nacional o mundial para abordar dicho desafío de manera efectiva, analizar y comprender tendencias, planificar, priorizar y diseñar soluciones.

La monitorización de los **Objetivos de Desarrollo Sostenible (ODS)** plantea una necesidad similar de información a nivel nacional para informar sobre la mejora de los diferentes objetivos, de los cuales los objetivos de ODS 2.4 y 6.4 son más relevantes para la agricultura sostenible.

En 2011, los ministros de agricultura del G20 lanzaron la **iniciativa GEOGLAM** para "fortalecer la **monitorización agrícola global** al mejorar el uso de herramientas de teledetección para las proyecciones de producción de cultivos y el pronóstico del tiempo".

Los datos para la monitorización agrícola nacional se recopilan en general mediante encuestas en granjas y hogares, pero recientemente ha sido reconocido el potencial que ofrecen los **satélites** para la elaboración de las **estadísticas agrícolas**.

Con la llegada del mencionado programa **Copernicus** y las observaciones recopiladas por sus satélites **Sentinel** se proporcionan un conjunto de datos abiertos y gratuitos sin precedentes relevante para la monitorización agrícola nacional a gran escala. Sin embargo, para los servicios nacionales encargados de realizar las estadísticas agrícolas será difícil sin una investigación y desarrollo de **algoritmos** que permitan su utilización.

A través del **proyecto SEN4Stat** se pretende, mediante productos de información adicionales y desarrollo metodológico, dar apoyo a la integración efectiva de este tipo de información en los flujos de trabajo para la elaboración de estadísticas.

Este proyecto tiene como objetivo facilitar la captación de información procedente de la **observación de la tierra** para apoyar a las oficinas nacionales en la elaboración de sus estadísticas agrícolas. Durante su fase de desarrollo, se prestará especial atención a la demostración de productos y mejores prácticas para la monitorización de la agricultura relevantes para la generación y presentación de este tipo de información.

Las actividades de Sen4Stat se organizarán en **dos fases principales de 18 meses**. La **primera fase** se centrará en la identificación de las necesidades de 5 países piloto en términos de integración de datos de observación terrestre dentro del flujo de trabajo de estadísticas agrícolas y en el desarrollo de soluciones prototipo. Otro aspecto clave en la asociación

entre las oficinas de estadística y el consorcio del proyecto será el intercambio de encuestas de campo, para desarrollar y evaluar las metodologías.

La **segunda fase** estará dedicada a la prueba y demostración de estos prototipos a escala nacional y en condiciones prácticamente reales para garantizar que las soluciones propuestas se ajusten a las expectativas del proyecto realizando los ajustes necesarios.

El **Comité Directivo**, compuesto por partes interesadas internacionales como la FAO, el Banco Mundial, el Programa Mundial de Alimentos, el CGIAR (red mundial para la innovación agrícola), la Comisión Europea y la USAID (Agencia de los Estados Unidos para el Desarrollo Internacional) seleccionó diferentes países para convertirse en socios de Oficinas de Estadística Nacionales del proyecto Sen4Stat incluyendo a España.

Debido a sus conocimientos en el uso de imágenes de satélite para la monitorización de las actividades agrícolas, el **FEGA**, en colaboración con el INE y la **Subdirección General de Análisis, Coordinación y Estadística** del Ministerio de Agricultura, Pesca y Alimentación (MAPA), ha facilitado al consorcio información relevante que les permitirá comprender el desarrollo de la actividad estadística en agricultura de nuestro país.

La obtención de datos georreferenciados y la monitorización de la superficie agraria es una clara apuesta por la digitalización del sector, encaminada a fomentar prácticas más eficientes en la gestión

Estos son varios ejemplos del abanico de posibilidades que ofrece el disponer de datos georreferenciados en un sector tan heterogéneo como es el sector agrario. La continua evolución de las tecnologías de análisis de datos, observación terrestre, etc. mejoran cada día el nivel de información disponible.

En definitiva, la obtención de datos georreferenciados y la monitorización de la superficie agraria es una clara apuesta por la digitalización del sector, encaminada a fomentar prácticas más eficientes en la gestión. Por otro lado, tiene un claro enfoque preventivo, mejora la información disponible para las administraciones, el propio sector y el público en general.

El Colegio de Geógrafos en España: trabajando por la profesión

María Zúñiga-Antón

Presidenta del Colegio de Geógrafos de España

En 2019, se celebra el 20 aniversario de la Ley 16/1999, de 4 de mayo por la cual se crea el Colegio, donde se reconoce que “el desarrollo científico de la Geografía en la segunda mitad del siglo XX debido a la ampliación de sus conocimientos, ayudado por la evolución de las técnicas, ha comportado una demanda profesional, por parte de la sociedad, desconocida hasta hace relativamente poco tiempo”. Esto justifica la creación del Colegio Profesional “como instrumento necesario para la regulación y ordenación de la profesión y para su mejor disposición al servicio de la sociedad”.

El Colegio de Geógrafos es una corporación de derecho público cuyos Estatutos están recogidos en el Real Decreto 377/2015, de 14 de mayo, en el que se definen sus fines principales: (1) ordenar el ejercicio de la profesión vinculada a la Geografía, (2) la representación institucional de la misma ante los poderes públicos, autoridades, empresas y particulares y (3) la propia defensa de los intereses profesionales tanto de las personas colegiadas como de los usuarios de los servicios de los colegiados.

Asimismo, son 17 las funciones que se llevan a cabo en el marco del Colegio de Geógrafos. El objetivo de este artículo es repasar las principales de ellas poniendo de relieve la importancia de la información geoespacial para nuestro colectivo.

La primera de las funciones a destacar es la que hace referencia a “la propuesta de medidas para el desarrollo y perfeccionamiento de la profesión, así como la emisión de estudios, informes y estadísticas relacionados con sus fines”.

En este sentido, a través del **Observatorio de la Profesión** se pone a disposición de la sociedad el **Informe de Perfiles Profesionales de la Geografía**. Es un informe público que se realiza de manera periódica cada 5 años y que tiene como objetivos visibilizar la labor de geógrafos y contrastar las diferentes tendencias que se observan en el mercado laboral actual, así como contribuir a la defensa del ejercicio de nuestra profesión. Son ya cuatro las ediciones que se han publicado y en la de 2018, por primera vez, se integra la perspectiva de género y estudia el peso del emprendimiento. Las conclusiones de esta última edición indican que las propias Tecnologías de la Información Geográfica

Figura 1.



(TIG) son el principal nicho de trabajo en el presente, especialmente para las generaciones más jóvenes de geógrafos. Por otro lado, ámbitos como la Planificación Estratégica, urbanística o el Desarrollo Local, aunque se consolidan, su peso relativo desciende a consecuencia tanto del descenso de inversión pública como de la crisis económica. Salidas profesionales como el Patrimonio cultural, la Educación Ambiental o el Comercio se encuentran en fuerte declive.

Dentro de esta función, se encuentra también la expresión que, como profesionales, podemos dar respecto a temas en los que somos especialistas. Este es el caso del **Manifiesto por la Nueva Cultura del Territorio** redactado en 2006 por sus promotores en el convencimiento de que el buen gobierno del territorio, de la ciudad y el campo, es responsabilidad de todos y un tema político de primer orden. El Colegio de Geógrafos y la Asociación Española de Geografía (AGE) convocan de manera bienal el **premio NCT**, con la finalidad de fomentar una cultura del territorio mediante el reconocimiento de personas, entidades e instituciones que hayan contribuido de forma relevante al impulso de una ordenación y gestión del territorio fundamentada en valores de sostenibilidad ambiental, eficiencia económica y equidad social. En la edición de 2019 se renovó el premio reflejando dos categorías en vez de una, concediendo el premio en la categoría de divulgación de los valores del territorio al escritor Julio Llamazares, y en la categoría de gestión del territorio al Observatorio del Paisaje de Cataluña.

Es también función del Colegio “**el mantenimiento del contacto con los centros docentes facilitando el acceso a la vida profesional de los egresados**”. En este contexto son varias las acciones que se realizan desde el Colegio de Geógrafos. Por un lado, la participación en la **Comisión de evaluación del Plan Bolonia en los Grados de Geografía** en colaboración con la AGE y los 26 departamentos universitarios en los que se imparte la titulación de Geografía.

Por otro lado, una segunda acción es el mantenimiento del **Portal de Empleo**, en el que se recopila información sobre ofertas de trabajo. En los últimos 10 años han sido 2.821 puestos los que se han publicado, con perfiles sobre todo provenientes de la docencia, la investigación, las TIG y el medio ambiente.

En esta función se enmarca también la organización de la **Olimpiada de Geografía**. Como sucede en otros países, la celebración de un evento como este es un estímulo de aprendizaje para los jóvenes, e incentiva el compromiso de los centros educativos con acciones que promueven el interés por la ciencia y el conocimiento. Las Olimpiadas se

Figura 2.



vienen celebrando desde 2010 anualmente. Para el Colegio de Geógrafos, las Olimpiadas son importantes porque movilizan a agentes diversos vinculados con la disciplina geográfica: profesionales comprometidos con la estructura organizativa del Colegio de Geógrafos; profesorado, tanto de la Universidad como de centros de Educación Secundaria; y, claro está, estudiantes de Bachillerato. Además, es una actividad que tiene su eco en los medios de comunicación. Tanto las fases locales como la estatal generan un volumen de noticias y de actividad en redes sociales que se traduce en la relevancia de la palabra “Geografía”, por unos días, en el conjunto de la sociedad española. Es, pues, una forma de promoción de nuestra disciplina.

La “*Defensa de los intereses de la profesión en todos los ámbitos*” es otra de las funciones funda-

mentales del Colegio. Para esto, está activo el **Servicio de Defensa de la Profesión**, que trabaja en la lógica de sumar, considerando que nuestra profesión debe poder optar a determinadas plazas, pero nunca solicitando la exclusión de otras disciplinas. De esta manera, son 657 los casos tramitados en los últimos 12 años, consiguiendo un 85% de éxito en los recursos presentados que principalmente se centran en temáticas vinculadas con el Medio Ambiente, el Desarrollo Local y la Ordenación del Territorio. Han sido cinco los dictámenes emitidos en temas como Catastro, Planes Generales de Ordenación del Territorio, Evaluación Ambiental, Riesgos Naturales y Planes técnicos de caza.

La visualización y la utilización de contenidos georreferenciados por parte de, prácticamente, cualquier usuario, constituyen uno de los factores que ha favorecido el “renacer” reciente de lo espacial

Por supuesto, *“participar en los Consejos u Organismos Consultivos de la Administración en materia de su competencia profesional”* es otra de las funciones que se asume con la colaboración de miembros de las diferentes Juntas de Gobierno del Colegio a muy diferentes escalas territoriales. Ejemplo de esto a nivel nacional es la pertenencia al Comité Consultivo del Consejo Superior Geográfico, que es el órgano de dirección del Sistema Cartográfico Nacional o al Consejo Estatal para el Patrimonio Natural y la Biodiversidad. A nivel regional se participa, entre otros muchos, en la Comisión Regional de Ordenación del Territorio y Urbanismo de Cantabria (CROTU), en el Consejo Autonómico de Medio Ambiente, Urbanismo y Ordenación del Territorio de la Junta de Castilla y León, en el Consell Rector de l'Institut Geogràfic i Cartogràfic y en la Comissió de Política Territorial i

d'Urbanisme de Catalunya, o en la ponencia técnica del Consejo de Ordenación del Territorio de Aragón (COTA). A nivel local tenemos presencia en la Comisión de Medio Ambiente y Urbanismo de Segovia, en la Comisión Territorial de Urbanismo de la ciudad de Barcelona o en la Comisión Local de Patrimonio Histórico de Sevilla. Todos ellos son solo algunos ejemplos significativos. Por supuesto, se ha colaborado en procesos participativos y en la revisión de normativas como el foro de participación del Área Metropolitana Alicante-Elche, la Directriz de Política Demográfica de Aragón o la norma PNE 178601 “Territorios Rurales Inteligentes. Definición, atributos y requisitos” por parte de la Asociación Española de Normalización y Certificación.

Por último, debemos mencionar la función de *“impulsar y contribuir al progreso de los métodos y técnicas propias de la profesión y a la difusión de las mismas, ayudando a la investigación científica”* que se materializa con la firma de más de 40 convenios con diferentes organizaciones, entidades y empresas. Hay firmados acuerdos con entidades de gran relevancia en el ámbito de la Geografía como es el caso del Instituto Geográfico Nacional (IGN-CNIG), la Asociación Española de Geografía o la Asociación Interprofesional de Ordenación del Territorio de España (FUNDICOT), así como con algunas de empresas y asociaciones que lideran la innovación en TIG como ESRI-España o la Asociación QGIS-España.

Estas son solo 5 de las 17 funciones que tiene encomendadas el Colegio de Geógrafos. Parte de estas funciones se basan en el conocimiento del territorio, para lo que el acceso a la información georreferenciada generada en el marco del Sistema Cartográfico Nacional resulta esencial, toda vez que permite la realización de estudios específicos a escalas territoriales muy diversas y sin la que nuestra labor de servicio a la sociedad no sería posible.

La casi infinita posibilidad de crear contenidos georreferenciados, en ámbitos temáticos de enorme amplitud, constituye, en este sentido, una oportunidad extraordinaria para divulgar el quehacer geográfico, en sectores de población amplios, incluyendo a los jóvenes. La visualización y la utilización de contenidos georreferenciados por parte de, prácticamente, cualquier usuario, constituyen uno de los factores que ha favorecido el “renacer” reciente de lo espacial. Queda pendiente, eso sí, que la capacidad analítica propia de los geógrafos posibilite vincular este tipo de contenidos a prácticas territoriales diversas, y que abarcan a líneas de actividad múltiples que conciernen tanto a la administración como al ámbito empresarial y académico.

Para saber más...

- Colegio de Geógrafos www.geografos.org

La importancia de la georreferenciación y la geolocalización para las empresas

Jorge González
DataCentric

El uso de datos con una variable geográfica está siendo un importante activo para las grandes empresas tecnológicas de nuestro tiempo. Existen servicios arraigados en la sociedad, como la telefonía móvil o los navegadores GPS, que hacen uso de estos datos con una conveniencia reconocida por todos.

Así mismo, España es el país europeo, y del mundo en 2017, con mayor penetración de teléfonos inteligentes, con un 88% de penetración de mercado. En el camino para adaptarse a una sociedad y a un mercado móvil, las empresas multinacionales hacen esfuerzos por conectarse con sus clientes. Es común que se faciliten aplicaciones móviles para ofrecer un servicio complementario o facilitar la gestión del propio servicio desde un teléfono móvil. Sin embargo, salvo que sea una aplicación de pago por uso o transaccional (en la que se suele obtener una comisión por transacción) raramente se consigue monetizar esos datos.

EL VALOR DE LA GEOLOCALIZACIÓN Y LA VARIABLE TEMPORAL

Si reflexionamos sobre los datos que podemos obtener de una aplicación móvil y su utilidad a nivel de negocio nos encontramos normalmente con: un email de registro como canal de contacto, los

datos de comportamiento del usuario (se suelen analizar para la mejora de la experiencia de usuario) y algunas aplicaciones, además, solicitan acceso a datos internos del dispositivo, como la geolocalización, acceso al micrófono o a los bancos de imágenes. De estos datos, la localización del dispositivo, previo consentimiento del usuario, es el dato más potente al que las empresas pueden acceder. Además, a las variables x, y de localización se suma la variable temporal, que multiplica las capacidades analíticas pudiendo obtener un conocimiento del usuario y del entorno suficiente para posteriormente accionar estrategias comerciales basadas en datos.

Pensemos que si un usuario pasa cierto tiempo en un centro deportivo de pádel de forma recurrente, ¿no podríamos deducir que sea aficionado al pádel y que sea susceptible de recibir información sobre equipamiento deportivo específico? O si el usuario pasa varias noches en diferentes localizaciones de forma recurrente, ¿no podríamos deducir que es una persona que viaja y que recibirá con interés información sobre seguros de viaje? ¿O sobre un seguro del hogar para su vivienda vacacional si la localización se repite y está en la costa? Indudablemente las acciones comerciales que realizásemos tendrían mayor probabilidad de éxito.

A continuación, veremos cómo podemos extraer información de valor de una aplicación para

Figura 1.

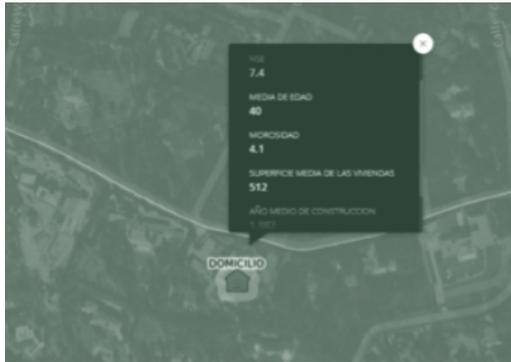


Ruta del usuario a partir de datos georreferenciados.



Keylogs identificados tras analizar el comportamiento del usuario.

Figura 2.



Información agregada de la sección censal.



Datos de la construcción de la vivienda. Área de la zona deportiva.

poder integrarlo en estrategias de negocio a través del caso de uso de Pyramid Datalake. Pyramid Datalake es la fuente de datos de DataCentric utilizada en el caso para complementar datos de geolocalización extraídos de una aplicación con datos que incluyen variables georreferenciadas.

POTENCIAR DATOS GEOLOCALIZADOS CON DATOS GEORREFERENCIADOS

En este caso de uso se analiza la serie histórica de los datos de geolocalización de un usuario en un municipio del norte de Madrid. Basándonos en lo comentado anteriormente y tras el análisis de los datos se identifican unas geolocalizaciones clave llamadas “keylogs”. Esas localizaciones son el lugar de trabajo, la vivienda, el colegio y el supermercado habitual del usuario.

El análisis determina el domicilio o vivienda habitual del usuario tras estudiar el comportamiento de la serie en el horizonte temporal que

abarca desde las 24 horas hasta las 8 horas dentro del histórico de datos. Después enriqueciendo el registro con información georreferenciada de catastro contenida en Pyramid se consigue perfilar al usuario a través de los datos correspondientes a su parcela. En ella se observan 10.000 m² con la vivienda principal y otro edificio de 95 m², una piscina, una pista de tenis y una pista de pádel. A continuación, se añaden datos agregados de la sección censal como el nivel socioeconómico, la media de edad, la tasa de morosidad, la superficie media de las viviendas, el año medio de construcción de la vivienda y el gasto en ocio.

Con estos datos ya podríamos comenzar a inferir que el usuario puede ser un **potencial cliente de marcas de lujo**: el usuario es de clase alta, su gasto en ocio es alto y la morosidad es baja. Además, por su vivienda, podría ser **potencial cliente también de productos y servicios asociados al tenis, pádel, la piscina, accesorios de jardín, sistemas de seguridad etc...**

En un segundo análisis del comportamiento de la serie en el horizonte temporal que abarca de las 9 ho-

Figura 3.



Vínculo de la georreferenciación física con la web.



Datos asociados a la georreferenciación de empresas.

ras hasta las 18 horas, se determina la identidad del lugar de trabajo del usuario, una empresa de transportes. Esto se debe al combinar la geolocalización de los datos del usuario con los datos georreferenciados de empresas contenidas en la fuente de datos.

En este caso simplemente se busca determinar la posible profesión del individuo, pero gracias a los datos del universo web contenidos en Pyramid, incluso se podría vincular datos georreferenciados de empresas con los datos contenidos en su web. Muy útil para realizar segmentaciones vía georreferenciación y vía característica web del universo de empresas. Por ejemplo, permitiría hacer una segmentación de todas las empresas en un radio de 15 km que además tengan linkedin en su web o que sean ecommerce.

Volviendo al caso, el análisis continúa identificando un lugar de compra habitual del usuario al analizar los logs de datos de la aplicación y combinarlos con los datos de empresas y POIs (o puntos de interés georreferenciados) de los que dispone Pyramid y que nos permite identificar la tienda física con precisión.

Conocer dónde compran los usuarios es de utilidad para saber qué *retailer* físico utilizar para servir productos y dónde invertir presupuesto y esfuerzos a la hora de realizar acciones comerciales. En este caso se trata de un Hipermercado Carrefour. Imaginemos que comprobamos que el 60% de nuestros usuarios compran en Carrefour. Podríamos, entonces, unirnos con esta cadena de supermercados para una acción conjunta con la finalidad de buscar nuevos clientes con las mismas características, o para hacer crecer el ticket de compra o fidelizar a aquellos que ya lo son.

El último punto georreferenciado clave detectado es un Colegio Internacional Bilingüe. El usuario para habitualmente a la hora de salida del colegio. Esto ayuda a seguir perfilando al usuario y su tipología de hogar, que serviría para accionar campañas de crecimiento dentro del usuario con ofertas para perfiles de familia con hijos. También podría ser de utilidad la información pensando en los negocios vinculados al sector de la educación. Por el perfil del colegio se deduce que, en este caso en concreto, la disposición de la familia a invertir en la educación de sus hijos es más alta que la media.

CONCLUSIONES

A partir de un email de registro y la aceptación del seguimiento de la geolocalización móvil por parte del usuario permite sacar información de georreferenciaciones identificadas:

Figura 4.



Punto georreferenciado clave identificado como lugar de compra habitual.

Figura 5.



Punto georreferenciado clave identificado como colegio bilingüe internacional.

- Vivienda habitual: Nivel socioeconómico alto, potencial cliente de marcas de lujo, posible cliente también de productos y servicios asociados al tenis, pádel, la piscina, accesorios de jardín, sistemas de seguridad.
- Lugar de trabajo: Empresa de transportes.
- Lugar de compra habitual: Compra en Carrefour habitualmente.
- Lugar de paso habitual identificado como colegio: Familia con hijos, colegio bilingüe internacional, dispuesto a invertir en educación.

Con este caso de uso se ratifica por tanto la hipótesis inicial de que al combinar datos geolocalizados de usuario de una aplicación móvil y datos georreferenciados con fuentes de datos externas como Pyramid Datalake*, es posible obtener información de valor que permite a las empresas accionar todo tipo de estrategias de comerciales.

* Los datos enriquecidos en este caso de uso son solo una muestra del potencial de Pyramid Datalake. Existen diferentes datos accionables no contenidos en el caso y se pueden crear nuevos a partir de KPIs a medida.

Sobre los datos y usos de los sistemas de información geográfica españoles e internacionales

Laura Maravall

Investigador postdoctoral en Historia Económica, Universidad de Tübingen

En los últimos años, los sistemas de información geográfica (GIS) han permitido un avance muy significativo dentro de las ciencias sociales. Gracias a ellos es posible obtener, guardar, analizar, manejar y presentar datos espaciales, y combinar dicha información con datos estadísticos. Esto ha dado lugar a grandes avances en la investigación científica. Por ejemplo, se ha logrado mejorar los modelos econométricos mediante la inclusión de variables geográficas que anteriormente no eran fáciles de obtener, tales como la fertilidad de la tierra para un determinado cultivo, las distancias entre todas las ciudades de un país, o la densidad de carreteras para distintos límites administrativos. La tecnología GIS también ha facilitado el acceso a la información geográfica a gran escala, posibilitando que cualquier usuario, incluso sin previo conocimiento estadístico, pueda jugar con los datos de manera sencilla y visualizar mapas que muestran todo tipo de información espacial, desde los resultados electorales por municipios hasta la distancia media de un infarto al desfibrilador más cercano. Por estas razones esta tecnología se ha ido expandiendo rápidamente y ya, en el 2014, se afirmaba que “todas las ramas de la investigación científica, económica y social se han subido al carro de la revolución GIS” y predecía que dicha tecnología suponía “una revolución científica”¹.

La creciente popularidad y divulgación de los GIS dentro de las Ciencias Sociales se debe en gran medida al esfuerzo realizado por parte de los organismos públicos y privados a la hora de proporcionar datos espaciales y servicios de localización, visualización y descarga. En particular, en el caso de España, los organismos se han coordinado a nivel nacional y europeo para impulsar la creación de proyectos de Infraestructuras de Datos Espaciales (IDE) cuya función, entre otras, es la de crear

geoportales en internet que permiten visualizar, buscar y manejar datos, metadatos y servicios relacionados con información espacial. El proceso es costoso y complicado, y requiere el seguimiento de una serie de requisitos y directrices que garanticen la compatibilidad e interoperabilidad a nivel nacional y comunitario². Por ello, y para dar una idea general de dicho esfuerzo, este artículo resume brevemente qué datos se necesitan a la hora de realizar un análisis espacial y presenta un panorama general sobre algunas de las fuentes de información geográfica nacionales e internacionales³.

En primer lugar –y de manera muy simplificada– si un usuario quiere manejar datos espaciales, sea cual sea el propósito, intentará encontrar y descargar dos tipos de ficheros: vectoriales y ráster. Los elementos geográficos en los archivos vectoriales (también conocidos como shapefiles) aparecen como puntos, líneas y polígonos y se almacenan en función de sus coordenadas, mientras que los ficheros ráster se guardan en formato imagen (ej., imágenes de satélite, fotografías aéreas digitales o mapas escaneados) y están formados por una matriz de píxeles o celdas. Ambos contienen información geográfica y se pueden utilizar para aplicar datos abstractos a posiciones espaciales exactas. Además, una de las grandes ventajas de GIS es que todos los datos se pueden extraer a nivel más micro, es decir, para uno o varios puntos concretos (municipios, celdas o polígonos personalizados), y a un nivel más agregado, como, por ejemplo, para un país entero. En segundo lugar, una vez obtenidos dichos ficheros, el usuario puede cargarlos en los sistemas de información geográfica y ejecutar varios tipos de análisis y cálculos espaciales. En términos estadísticos, se pueden

¹ https://elpais.com/elpais/2014/11/26/ciencia/1417005124_889841.html

² Dichos requisitos y directrices se basan en la Directiva Europea Inspire (Directiva 2007/2/CE, *Infrastructure for Spatial Information in Europe*). Para más detalle, ver <https://www.idee.es/europeo-inspire>.

³ Para un informe más detallado sobre el uso de GIS dentro de las ciencias sociales, ver Del Bosque's et al. (2012).

realizar análisis descriptivos y contrastar hipótesis. A modo de ejemplo, el usuario puede extraer datos sobre la precipitación media por kilómetro cuadrado y correlacionar dicha variable con otra que mida la capacidad de drenaje de la tierra y así incluso modelar la probabilidad de inundación. También se podría relacionar espacialmente el promedio de lluvias con la producción de naranjas en esa misma zona y analizar si existe una correlación significativa entre el promedio de lluvias en determinados municipios con el nivel de desarrollo económico, medido mediante la iluminación nocturna capturada por satélites.

Para realizar dicho análisis se debe disponer de fuentes de información geográfica digital en la web. En España, un referente es el Centro Nacional de Información Geográfica, un organismo autónomo adscrito al Ministerio de Fomento y bajo la dirección del Instituto Geográfico Nacional (IGN). En su página web se facilita un centro de descargas con todo tipo de información geográfica, desde datos topográficos y temáticos concebidos para su uso en GIS, hasta ficheros digitales creados a partir de documentación geográfica antigua (por ejemplo, mapas de población de finales de 1870). Además, la página web contiene información detallada sobre el proyecto de Infraestructura de Datos Espaciales de España (IDEE), encargado de integrar los distintos IDE producidos en España a nivel nacional, estatal y local, tanto de organismos públicos como de privados y bajo la coordinación del Consejo Superior Geográfico⁴. Gracias a este proyecto, la información geográfica de toda España se encuentra armonizada en Internet e incluye datos, metadatos y servicios, permitiendo un continuo desarrollo de las estadísticas de georreferenciación.

El Instituto Nacional de Estadística (INE) también ha puesto a disposición del público datos estadísticos de libre acceso y un visor geográfico que permite representar su distribución espacial. Los datos censales se pueden visualizar para distintos límites administrativos (nación, comunidad autónoma, provincia, municipio) o para celdas de distinto tamaño y analizar variables como la densidad poblacional, la edad media o el porcentaje de población extranjera, haciendo posible entender los patrones de aglomeración o dispersión geográfica. Además, los interesados pueden perso-

nalizar sus mapas, crear leyendas, descargarlos e imprimirlos. No obstante, lo que resulta de mayor interés en términos científicos ha sido que desde el 2011 se pueden descargar los mapas digitales en formato shapefile a nivel de municipios y secciones censales (subdivisión de los municipios), de manera que todo aquel que quiera pueda realizar estudios basados en el padrón y las encuestas del INE y hacer sus propios cálculos geográficos.

Esta “revolución” con respecto a la disponibilidad de datos geográficos y el uso de SIG ha permitido dar un paso adelante en otras disciplinas dentro de las ciencias sociales, como son la economía del desarrollo, ciencias políticas e historia económica

Son bastantes más los organismos que han contribuido a la creación de datos geográficos y de servicios de red en España. Entre ellos la Dirección General del Catastro ha creado en su página web una sección dedicada a la difusión de datos catastrales y en la sede electrónica se pueden descargar múltiples capas en formato shapefile por municipios, como, por ejemplo, las parcelas catastrales y los límites administrativos de suelo urbano. Además, se proporcionan los datos transformados y metadatos junto con toda la información acerca de la metodología implementada para construir las bases de datos en cumplimiento con la directiva europea. El Ministerio de Agricultura, Pesca y Alimentación también ha puesto a disposición del público en su página web una sección de cartografía y GIS. En ella se pueden consultar mediante visores la información geográfica relacionada con cada uno de los campos de acción del Ministerio. Por ejemplo, se pueden superponer distintas capas con información geográfica acerca de la agricultura tales como el índice de aridez y los tipos de cultivos, el cartografiado relacionado con los recursos pesqueros, alimentación, e incluso se puede ver la evolución histórica de la evaluación de la calidad del aire. Por último, a la hora de buscar datos más específicos a nivel auto-

⁴ Por ejemplo, el IGN está encargado de gestionar los proyectos IDE como el Sistema de Información Geográfica Nacional (SIGNA), CartoCiudad e Infraestructura de Datos Espaciales de la Administración General del Estado (IDEAGE). Para más información, ver <https://www.idee.es/web/guest/inicio>.

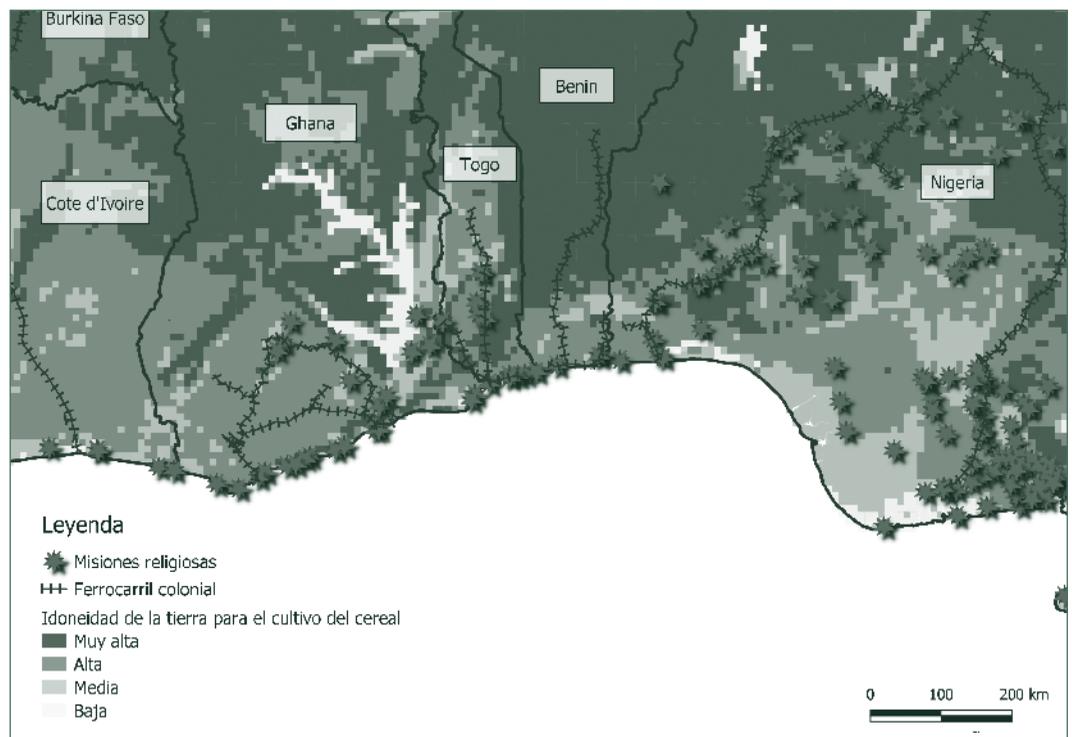
nómico, cabe destacar la página web del Instituto de Estadística y Cartografía de Andalucía (Junta de Andalucía), donde se dedica una sección a la georreferenciación y se pueden analizar con más detalle las distribuciones de variables de población y urbanización.

Esta “revolución” con respecto a la disponibilidad de datos geográficos y el uso de sistemas de información geográfica ha permitido dar un paso adelante en otras disciplinas dentro de las ciencias sociales, como son la economía del desarrollo, ciencias políticas e historia económica. En el 2012 se creó en el CSIC una Unidad Especializada de Sistemas de Información Geográfica y Humanidades Digitales (SIGyHD) aplicado a las Ciencias Sociales y, actualmente, el número de proyectos científicos en las universidades ha aumentado considerablemente. Entre otros, la historia se ha beneficiado de la digitalización de una multitud de mapas históricos. Por poner un ejemplo, la Figura 1 muestra cómo se pueden combinar en un mismo mapa datos geográficos, como son la calidad de la tierra, con datos obtenidos a partir de mapas históricos, como, por ejemplo, las líneas de ferrocarril coloniales o la distribución de las misiones religiosas. En el caso de España, en el 2014

en la página del IGN se pusieron a disposición del público 36.000 mapas de los Fondos Históricos. Dichos mapas, georreferenciados, permiten combinar datos históricos (ej. fronteras históricas o carreteras romanas) con variables geográficas y de desarrollo económico a largo plazo.

La expansión de GIS en España se ha visto reforzada por la creciente disponibilidad de bases de datos geográficas internacionales, tales como el European Data Portal, GAEZ-FAO e IPUMS-International, permitiendo ampliar el abanico de datos disponibles y realizar comparaciones internacionales. La cooperación a nivel europeo en la construcción de bases de datos geográficos se hace particularmente visible en el European Data Portal, financiado por la Unión Europea. De hecho, este geoportal unifica los proyectos nacionales realizados por instituciones como el IDEE, el IGN, IDEAndalucía, y el Instituto Mediterráneo de Estudios Avanzados (IMEDEA) (formado por el CSIC y Universitat de les Illes Balears). El geoportal consta de múltiples categorías (por ejemplo, medio ambiente, gobierno y sector público, ciencias y tecnología, transporte y sanidad) y permite filtrar los datos en función del país. En el caso de España,

Figura 1.



Nota: Imagen creada por el autor del texto con el programa qGIS. Fuentes utilizadas: Para las misiones religiosas y las líneas de ferrocarril coloniales, ver: <https://scholar.harvard.edu/nunn/pages/data-0>. Para la idoneidad de la tierra, ver <http://www.fao.org/nr/gaez/en/>.

se pueden extraer mapas estratégicos en formato shapefile creados por la IDEE para la evaluación y gestión de ruido debido a aeropuertos y líneas ferroviarias, estaciones de la calidad del aire, y también se pueden descargar ficheros ráster con información acerca del nivel de pastos o del tipo de árbol en los bosques.

Otra de las fuentes que destacan a nivel internacional es la base de datos espaciales GAEZ-FAO, creada por la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) y el Instituto Internacional para el Análisis de Sistemas Aplicados (IIASA). Dicha base, titulada "Global Agro-Ecological Zones", contiene una gran variedad de información geográfica a nivel mundial sobre los recursos agrícolas y su potencial, siendo utilizada con frecuencia en las investigaciones científicas. Dispone de datos espaciales como por ejemplo las distintas zonas termales (trópicos, templados, boreales, etc.) y los recursos de la tierra y agua en general, permitiendo extraer índices calculados de gran interés científico como son la idoneidad agrícola y los rendimientos agrícolas potenciales. Dichos índices, disponibles a nivel de celdas y extraíbles en formato ráster, se basan en estimaciones que combinan el clima promedio y una serie de cálculos realizados a partir de variables climáticas y agroclimáticas. A modo de ejemplo, un usuario interesado en cultivar en el sur de España puede extraer el índice de idoneidad para el trigo de secano dadas ciertas condiciones (por ejemplo, una gestión tradicional sin el uso de nutrientes o productos químicos) y compararlo con la idoneidad para el maíz dadas las mismas condiciones.

Por último, cabe señalar el proyecto IPUMS-International (Integrated Public Use Microdata Series), encargado de la construcción de la base de datos censales más grande del mundo y que ahora ya es posible georreferenciar. Gracias a la colaboración de más de 100 agencias estadísticas nacionales, el proyecto ha sido capaz de armonizar los datos censales y encuestas (aprox. 750) para más de 100 países e incluso para censos que datan desde 1850, permitiendo comparaciones a nivel micro (con más de mil millones de registros de personas) a través del tiempo y del espacio. La gran mayoría de estos datos se pueden georreferenciar gracias a los shapefiles disponibles en la web, mediante los cuales es posible analizar y visualizar los micro datos para distintas unidades administrativas en las que se enumeran los hogares censados, como son los municipios o provincias. De esta manera un investigador puede, por ejemplo, comparar los patrones de movilidad

intergeneracionales en educación o profesiones a nivel municipal para más de 10 países.

En conclusión, actualmente existen una multitud de fuentes de datos geográficos a nivel nacional e internacional de libre acceso para cualquier usuario, sea un académico escribiendo un artículo sobre la calidad del aire por municipios, un reportero analizando la distribución espacial de los resultados electorales, un agricultor que quiere saber dónde cultivar maíz, un ingeniero que quiere evitar zonas con una alta probabilidad de inundación o simplemente un ciudadano que quiere visualizar su parcela catastral. La expansión de dichas fuentes ha sido posible gracias a la colaboración de múltiples organismos públicos y privados y al creciente interés por la tecnología GIS en España, que no solo se percibe por su expansión dentro de las múltiples disciplinas, sino que también queda reflejado por el mayor número de ofertas de empleo y la creciente disponibilidad de cursos y servicios impartidos.

Para saber más...

- Instituto Geográfico Nacional (IGN), Centro Nacional de Información Geográfica
<https://www.ign.es/web/ign/portal/inicio>
- Dirección General del Catastro
http://www.catastro.meh.es/esp/acceso_infocat.asp
- Instituto Nacional de Estadística
http://www.ine.es/censos2011_datos/cen11_datos_detallados.htm
- Unidad Especializada de Sistemas de Información Geográfica y Humanidades Digitales (SIGyHD)
<http://unidadsig.cchs.csic.es/sig/>
- ESRI España
<https://www.esri.es/>
- Del Bosque, I., et al. 2012. Los sistemas de información geográfica y la investigación en Ciencias Humanas y Sociales, Madrid: Confederación Española de Centros de Estudios Locales (CSIC)
<http://digital.csic.es/bitstream/10261/64940/1/Los%20SIG%20y%20la%20Investigacion%20en%20Ciencias%20Humanas%20y%20Sociales.pdf>
- European Data Portal
<https://www.europeandataportal.eu/data/datasets?locale=es>
- FAO-GAEZ
<http://www.fao.org/nr/gaez/en/>
- IPUMS International
<https://international.ipums.org/international/>

La georreferenciación para el diseño y evaluación de políticas públicas

Enrique Muñoz Concen, José Netzahualcóyotl Salazar-Guzmán y Alfonso Rafael Aranda
Instituto Nacional de Estadística, Geografía e Informática de México (INEGI)

Actualmente la georreferenciación de la información relativa al diseño y evaluación de las políticas públicas es uno de los aspectos indispensables para entender el ámbito de la gestión pública contemporánea. Contar con registros estadísticos o administrativos que carezcan de georreferenciación conlleva el riesgo de plantear diagnósticos inacabados o parciales en los que el tema de análisis no se revele en toda su expresión. La componente espacial a la hora de diseñar y evaluar los resultados e impactos de la política pública resultan cruciales para el ejercicio de gobierno. Por lo tanto, se hace necesario no solo integrar un enfoque territorial y espacial sino también preparar la gestión de información para que integre de forma natural los atributos espaciales pertinentes.

INTRODUCCIÓN

Los métodos y técnicas relacionados con la georreferenciación de los rasgos geográficos han cobrado un auge significativo en el marco del desarrollo de las nuevas tecnologías, sobre todo en lo que se refiere a los productores o generadores de información geoespacial. Esta actividad permite conocer la ubicación de cualquier fenómeno que está referenciado a un sistema de coordenadas, a partir del uso de geotecnologías como los Sistema Globales de Navegación por Satélite (GNSS, por sus siglas en inglés) y los Sistemas de Información Geográfica, ya sea usando imágenes tipo ráster o bien objetos vectoriales (puntos, líneas y áreas que representan elementos geográficos).

Es evidente que en la gestión pública que se maneja bajo un enfoque territorial surge de forma natural la necesidad de contar con información georreferenciada a nivel de *domicilio geográfico*. Ante este escenario, desde hace más de una década en México se han coordinado tareas operativas en campo a nivel nacional para captar la información de los *números exteriores* a nivel manzana, en todas las localidades de país. Hasta el momento se cuenta con un universo de 10.915 localidades, lo que equivale a aproximadamente 1.700.000 manzanas que contienen cerca de 27 millones de números exteriores.

Estos operativos buscan contribuir a la comprensión del uso concreto de la información y del proceso de georreferenciación que apoya la elaboración de la política pública. Al respecto, se entiende por *uso concreto de la georreferenciación* aquella utilización de la información que influye en las

fases de diseño y evaluación de las políticas públicas y que enriquece el entendimiento del territorio que se busca intervenir, ya sea con la generación de indicadores, de estimadores o a partir de visibilizar los factores que intervienen en el modelado o simulación y permiten la construcción de alternativas de política pública.

El interés principal se centra en la georreferenciación que se utiliza en las fases de diseño y evaluación de políticas *desarrollo social*. La georreferenciación consiste en la localización de los beneficiarios de dichas políticas con la finalidad de contar con el correspondiente padrón referenciado al espacio y, por lo tanto, sujeto a la aplicación de técnicas del análisis espacial. Sin embargo, es importante precisar que en el ámbito de lo público la información georreferenciada no solo forma parte de un proceso técnico, sino que está sujeta a la complejidad de los procesos sociales y a los aspectos normativos involucrados en la gestión pública. Por lo tanto, se parte del supuesto de que la información georreferenciada se sitúa de manera imprecisa entre una dimensión técnica y la esfera de la política pública.

APROXIMACIÓN CONCEPTUAL

Los procesos de la gestión pública que demandan información georreferenciada son diversos, tanto en su contenido como en su método y estrategia. En este sentido no todos requieren —ni en todos los casos particulares— de información georreferenciada.

renciada o esta información no tiene el mismo nivel de pertinencia en todos los ámbitos de la gestión pública. Para el caso de México se distinguen dos grandes fases durante la elaboración de políticas públicas que hacen uso de este tipo de información georreferenciada: el diseño de política y la evaluación de las políticas ya establecidas.

En la fase de diseño, el uso de la información georreferenciada se utiliza principalmente para la construcción de escenarios y alternativas de política. Estas, a su vez, se basan en una serie de indicadores que permiten realizar estimaciones a partir de la *ubicación* y representación espacial de diferentes variables estadísticas. Un ejemplo de estos indicadores es el conjunto de variables de tipo socioeconómico que se representan sobre determinadas unidades territoriales (v.gr., superficie de áreas verdes o arboladas por habitante).

En la fase de evaluación de políticas públicas, por otro lado, el uso de la información georreferenciada se plantea más sobre el análisis de las series o las líneas de tiempo que facilitan la comparación y detección de cambios en el territorio intervenido por dichas políticas. La información georreferenciada acerca de la cobertura, resultados e impacto de una política pública a lo largo del tiempo permite la comparabilidad de ese territorio antes, durante y después de la intervención política, así como de las diferencias —si las hubiere— al interior de la zona de intervención o territorio, entendido este como un espacio que puede ser homogéneo o heterogéneo respecto a la respuesta social a la política ejecutada.

La georreferenciación como proceso alude asimismo a dotar de manera explícita con atributos espaciales a los fenómenos o eventos que ocurren en el territorio. Desde un punto de vista operativo, se trata de incorporar los atributos espaciales a aquellos fenómenos que generalmente no se analizan desde un punto de vista territorial. La incorporación del territorio en el análisis de los fenómenos sociales o políticos puede mostrar nuevas relaciones entre variables o complementar las ya encontradas. Visto desde un punto de vista teórico, la incorporación de los atributos espaciales a los eventos socio-políticos implica utilizar un modelo matemático del que surge un sistema de referencia geográfico, para describir y analizar los fenómenos territoriales. Tal sistema de referencia, definido además al aplicarlo en un marco de referencia, se basa en una serie de criterios geodésicos que deben ser adecuados al territorio en cuestión. Esto significa que, una vez reconocido el territorio en dónde se va a trabajar, se procede al establecimiento de dicho marco de referencia, a partir de coordenadas y de un *datum* determinado, sobre los que se locali-

zarán los objetos espaciales representados mediante puntos, líneas y polígonos. Una vez hecha esta geolocalización del evento o fenómeno territorial, la información georreferenciada está en posibilidades de aportar una serie de elementos de valor, pertinentes al diseño y la evaluación de políticas públicas o al trabajo de los tomadores de decisiones.

ENFOQUE TERRITORIAL

En la actualidad la gestión pública contemporánea recurre de forma natural a la información georreferenciada. Busca obtener mayor certidumbre respecto a los lugares que interviene y al impacto referente a las personas que habitan esos lugares, por ejemplo, en cuanto a los recursos públicos que les han asignado en manera de apoyos en el marco de una política social contra la pobreza o para el desarrollo social.

El enfoque territorializado requiere la información georreferenciada para ayudar a focalizar y evaluar los instrumentos de política que están diseñados para mejorar el bienestar de la sociedad

El enfoque territorializado requiere la información georreferenciada para ayudar a focalizar y evaluar los instrumentos de política que están diseñados para mejorar el bienestar de la sociedad. Este enfoque puede ser utilizado para una diversidad de fines. Tiene, por ejemplo, una dimensión diagnóstica en la que se pretenden identificar y analizar espacialmente la ocurrencia de problemas y la selección de proyectos o instrumentos prioritarios. El énfasis está puesto en identificar espacialmente aquellas zonas de población marginada o con altos problemas de inseguridad o de violencia a partir de la información georreferenciada. La ausencia de un enfoque territorial en el análisis de problemas sociales presenta el riesgo de no estar en posibilidades de focalizar los instrumentos de política pública, lo que puede derivar en la utilización ineficiente e ineficaz de los recursos públicos, o simplemente en no contar con adecuadas capacidades diagnósticas de los problemas sociales.

ESTUDIO DE CASO

Existen actualmente diversos productos de información georreferenciada de los que se han podido obtener diferentes componentes de acuerdo con el tipo de política que se busca desarrollar. Tal es el caso del diseño y evaluación de políticas de desarrollo social que requieren identificar y dar seguimiento territorial al conjunto de sus beneficiarios.

En este caso, la información geoespacial de *números exteriores* de las viviendas del país y de los establecimientos económicos se torna necesaria y fundamental. La información georreferenciada a nivel de domicilios puede ser visualizada a través de la información vectorial relativa a localidades amanzanadas, que contiene variables socio demográficas asimismo georreferenciadas.

Resulta indispensable realizar procesos técnicos, no solo de georreferenciación sino también de estructuración e higienización de la información en torno a los domicilios geográficos

Es de resaltar que la comunicación eficiente con los organismos públicos responsables de las políticas es una tarea transversal durante el proceso de georreferenciación. La finalidad es afinar lo mejor posible los requerimientos para georreferenciar los diversos listados socioeconómicos y demográficos que se vincularán con los registros de los domicilios. Este tipo de acompañamiento constante del organismo público es crucial dado que la información, en la mayoría de los casos, carece de estructuras homogéneas y consistentes, además de que refleja ausencia o inconsistencia de datos por abreviaturas, espacios en blanco, diferencias en formato de almacenamiento, entre otras. Por ello, resulta indispensable realizar procesos técnicos, no solo de georreferenciación sino también de estructuración e higienización de la información en torno a los domicilios geográficos.

En este sentido, nuestro país cuenta con una Norma Técnica sobre Domicilios Geográficos (NTDG) publicada en el Diario Oficial de la Federación el 12 de noviembre de 2010. Esta tiene como finalidad minimizar este tipo de inconsistencias en las bases de datos y prever los vacíos en la infor-

mación que generan las diferentes unidades del estado mexicano (secretarías de estado, congreso federal y estatales, dependencias del gobierno federal, organismos autónomos, entre otros).

Contar con datos normalizados de los domicilios geográficos puede facilitar la planeación y ejecución de actividades de geocodificación sobre los registros de los diferentes padrones de beneficiarios de políticas de desarrollo social. Con ello, se está en posibilidad de proporcionar elementos territoriales no identificados al inicio del proceso para afinar y focalizar la toma de decisiones, tales como: determinación, asignación, seguimiento y auditoría de los recursos a los programas sociales, contar con insumos para marcos muestrales, facilitar las tareas de las unidades del estado respecto a la ubicación de domicilios geográficos, efectuar estudios mercadológicos, estadísticos, sociodemográficos, científicos, etc.

Existen metodologías y herramientas informáticas con las que es posible georreferenciar todos aquellos padrones de beneficiarios públicos (escuelas, estancias infantiles, clínicas, mercados, instalaciones deportivas o recreativas, oficinas gubernamentales, etc.), sin embargo, no son suficientes mientras no se cuente con estructuras de datos geoespaciales adecuadas para fines de geocodificación. Ejemplo de tales estructuras son aquellas generadas con la metodología del *Dual Independent Map Encoding* (DIME/TIGER) que aplican la codificación y la referencia geográfica bajo un marco topológico integrado.

La actividad de *georreferenciar* los registros de los padrones de beneficiarios de políticas públicas a partir de los domicilios geográficos es una tarea que continúa en vías de mejora y desarrollo. Hasta el día de hoy se ha podido lograr en una primera aproximación atender a dichas necesidades de georreferenciación a través de los archivos vectoriales de localidades, que cuentan con una traza definida y sobre la que se pueden plasmar los puntos de cada uno de los domicilios ubicados, dependiendo aún de servicios propietarios de geocodificación.

Una de las características de la información acerca de números exteriores georreferenciados es que tienen asociados una serie de atributos estadísticos importantes. Entre otros se tiene la clave geoestadística del número exterior compuesta por numerales que hacen referencia al estado, el municipio, la localidad, el Área Geoestadística Básica y la manzana, la clave del frente de manzana, el tipo y nombre de la vialidad, el tipo y nombre de las "entre calles", así como el tipo y nombre del asentamiento (colonia, barrio, fraccionamiento, etc.), que permiten vincular en un primer nivel con los registros de los padrones a un domicilio geográfico geolocalizado.

El propósito principal que se busca en torno al tema abordado en esta publicación es poder contar con un servicio de código abierto para georreferenciación de domicilios, en este caso, provenientes de los diversos padrones de beneficiarios de políticas públicas de desarrollo social. De la misma manera, se busca georreferenciar cualquier inventario de registros administrativos de algunas unidades del estado, de usuarios en general y de instituciones diversas. El objetivo es poder determinar la ubicación geográfica de estos y brindar información y transparencia a los procesos, resultados e impactos de la política social.

En virtud de que se busca alcanzar un proceso automatizado y sistematizado de georreferenciación, la estructura lógica de los datos geoespaciales será el principal reto a mejorar, principalmente con el objetivo de poder geolocalizar grandes números de domicilios en lapsos cortos de tiempo, dejando en última instancia aquellos que tienen información incompleta y para los que se requieren procesos semiautomatizados o incluso manuales. La intención es recuperar la mayor cantidad de domicilios posibles.

CONCLUSIONES

La información georreferenciada es un insumo fundamental para los trabajos que realizan instituciones públicas y privadas de diferentes sectores, entre ellos el económico y de competitividad, el de infraestructura y turismo, el de gobierno y de políticas de seguridad pública y social, pero también los de investigación y de desarrollo. Asimismo, es factible incorporar los datos obtenidos en un Sistema de Información Geográfica que permita no sólo conocer la ubicación espacial de la aplicación de la política sino también llevar a cabo otras actividades de análisis espacial.

Los trabajos de georreferenciación contribuyen a la vinculación de los datos y los fenómenos con el espacio geográfico, lo que aporta en correspondencia mayor certidumbre a la información generada en las diversas instituciones y sectores. La entrega de información georreferenciada a los diseñadores y evaluadores de políticas públicas no garantiza su uso y aplicación durante el ciclo de vida de la política pública, sin embargo, se enfatiza la importancia de visualizar la aplicación, resultados e impactos de la política en el territorio. La georreferenciación de la evaluación de la política pública tiene tal pertinencia que pueden llegar a cambiar el rumbo de la aplicación de los instrumentos en el futuro.

Para entender el desarrollo y uso de la información georreferenciada es crucial tener una visión de las características específicas de cómo se gestiona y de los procesos de política que alimenta desde un enfoque sistémico. Lo anterior redundará en el hecho de que se hace fundamental contar con un enfoque sistémico-territorial que permita obtener una visión clara y global de las características específicas y dinámicas de la política pública en las que este tipo de información georreferenciada resulta ser pertinente o de utilidad. A fin de identificar los momentos en los que la política pública utiliza información georreferenciada, es necesario contar con una definición operativa que permita distinguir sus fases y los dominios de conocimiento en los que se enmarca. En cada uno de estos dominios deberá reflejarse el nivel de pertinencia de la información georreferenciada.

Referencias

- Aguilar, Luis. 2000. "La hechura de las políticas públicas". Colección Antologías de Política Pública. Tercera Antología. México. Porrúa. 284 p
- Aranda, Rafael. 2015. "La trascendencia de la georreferenciación de padrones de beneficiarios de programas sociales, a través de domicilios geográficos, como un mecanismo de certidumbre en la asignación de recursos públicos." México. IPGH.
- Brewer y De León. 1983. "Foundations of Policy Analysis". Illinois. Dorsey Press. Homewood, 476 p.
- Diario Oficial de la Federación. "Norma Técnica sobre Domicilios Geográficos", publicada por el Instituto Nacional de Estadística y Geografía, el 12 de noviembre de 2010.
- Instituto Nacional de Estadística y Geografía. "Manual para la Georreferenciación de Domicilios provenientes de diferentes Padrones". Aguascalientes, México. Marzo de 2015.
- Muñoz, Enrique. 2015. "Geomática y Política Pública: un modelo teórico de comunicación territorial para una gobernanza efectiva". México.
- <https://www.inegi.org.mx/temas/mg/>
- <http://gaia.inegi.org.mx/mdm6/?v=bGF00JlzLjMyMDA4LGxvbjotMTAxLjUwMDAwLHo6MSxsOmMxMTFzZXJ2aWNpb3N8dGMxMTFzZXJ2aWNpb3M=>
- <https://books.google.com.mx/books?hl=es&lr=&id=FWvBWV2OngC&oi=fnd&pg=PA1&dq=georreferenciaci%C3%B3n&ots=EVdp9fAb5k&sig=BMXaYFqOEB6gRcjRty5IKyJB2RE#v=onepage&q=georreferenciaci%C3%B3n&f=false>

CONTABILIDAD NACIONAL ANUAL DE ESPAÑA. 2016. TABLAS DE ORIGEN Y DESTINO

Descarga gratuita a través de la web del INE

La Contabilidad nacional anual de España: Tablas de Origen y Destino es una estadística de síntesis que proporciona una descripción exhaustiva del proceso productivo y del equilibrio recursos-empleos de la economía nacional a nivel de producto. Ofrece, por tanto, una estimación robusta de los agregados contables de los enfoques de oferta y demanda del PIB a un nivel muy detallado de productos y por ramas de actividad para la economía nacional.

Esta operación ofrece la medición de la producción, consumos intermedios, valor añadido bruto (VAB), remuneración de asalariados, excedente de explotación y renta mixta y otros impuestos netos sobre la producción, desglosados por ramas de actividad, impuestos y subvenciones sobre los productos, márgenes de comercio y transporte, gasto en consumo final, formación bruta de capital, exportaciones e importaciones. Todas las variables se presentan desglosadas por productos excepto los agregados de empleo, rentas y VAB.

Esta operación estadística adopta la metodología contenida en el Sistema Europeo de Cuentas Económicas (SEC-2010) en cumplimiento de lo dispuesto en el Reglamento (UE) Nº 549/2013 del Parlamento Europeo y el Consejo, de 21 de mayo, al igual que el resto de operaciones estadísticas que integran el la Contabilidad nacional de España, lo que garantiza la comparabilidad internacional de sus resultados.

CUENTAS AMBIENTALES: CUENTAS DE EMISIONES A LA ATMÓSFERA. SERIE 2008-2017 Y AVANCE 2018

Descarga gratuita a través de la web del INE

Las Cuentas de emisiones a la atmósfera presentan los datos sobre las emisiones contaminantes a la atmósfera, de manera compatible con el Sistema de Cuentas Nacionales, registrando los agentes emisores desagregados por ramas de actividad económica y hogares como consumidores finales.

Las estimaciones de la Cuenta de emisiones a la atmósfera se realizan a partir de los Inventarios Nacionales de Emisiones a la Atmósfera, elaborados por el Ministerio para la Transición Ecológica, que utilizan la metodología IPCC y EMEP/EEA, con la nomenclatura NFR/CRF (Nomenclature for Reporting/Common Reporting Format), que agrupa las emisiones en sectores, categorías y subcategorías.

DIRECCIONES Y TELÉFONOS DE INTERÉS

INE-Pº de la Castellana, 181 y 183 - 28046 Madrid.
www.ine.es

Atención a usuarios

Tfno.: 91.583.91.00
Fax: 91.583.91.58
Consultas: www.ine.es/infoine
Lunes a jueves de 9 a 14 y de 16 a 18 horas
Viernes de 9 a 14:30 horas

Índice-Librería del INE

Tfno.: 91.583.94.38
Fax: 91.583.45.65
E-mail: indice@ine.es
Lunes a viernes de 9 a 14:30 horas
Biblioteca
E-mail: biblioteca@ine.es

PUBLICACIONES EDITADAS POR EL INE DE OCTUBRE A DICIEMBRE DE 2019

INEbase. Noviembre 2019

Descarga gratuita a través de la web del INE

Contenido:

Contabilidad Nacional Anual de España. 2016. Tablas de origen y destino
Cuentas ambientales: Cuenta de flujos físicos de la energía. Serie 2014-2017
Cuentas ambientales: Cuentas de emisiones a la atmósfera. Serie 2008-2017 y avance 2018
Cuentas ambientales: Impuestos ambientales. Serie 2008-2017 y avance 2018
Cuentas de los Residuos en España. Serie 2015-2017
Demografía Armonizada de Empresas. 2017
EPA. Decil de salarios del empleo principal. Serie 2006-2019
Estadística del Procedimiento Concursal. Serie 2004-2018
Estadística del Taxi. Serie 1994-2019
Estadística sobre actividades de I+D. 2018
Estadística sobre Generación de Residuos en el Sector Servicios y Construcción. 2017
Estadísticas sobre Recogida y Tratamiento de Residuos. Residuos Urbanos. Serie 2010-2017
Estadísticas sobre Recogida y Tratamiento de Residuos. Tratamiento de Residuos. Serie 2010-2017
Indicadores sobre residuos urbanos. Serie 2010-2017

INEbase. Octubre 2019

Descarga gratuita a través de la web del INE

Contenido:

Encuesta de Comercio Internacional de Servicios. Serie 2014-2018
Encuesta de Turismo de Residentes. 2º Trimestre 2019
Encuesta sobre el uso de TIC y Comercio Electrónico en las empresas. 2018-2019
Encuesta sobre Equipamiento y Uso de TIC en los hogares. 2019
EPA. Flujos de la población activa. Serie 2005-3º Trimestre 2019
EPA. Resultados trimestrales. 3º Trimestre 2019
Estadística de Ejecuciones Hipotecarias. Serie 2014-2018
Estadística de Hipotecas. 2018
Estadística de Transmisiones de Derechos de la Propiedad. 2018
Indicadores de Calidad de Vida. 2019
Indicadores de Confianza Empresarial. 4º Trimestre 2019
Sociedades mercantiles. 2018

INEbase. Septiembre 2019

Descarga gratuita a través de la web del INE

Contenido:

Contabilidad Nacional Anual de España. Serie 1995-2018. Principales agregados
Contabilidad Nacional Anual de España. Serie 1995-2018. Agregados por rama de actividad
Contabilidad Nacional Anual de España. Serie 1995-2018. Cuentas anuales no financieras de los sectores institucionales
Contabilidad Nacional Trimestral de España. Serie 1º trim. 1995 - 2º trim. 2019
El Salario de las Personas con Discapacidad. Serie 2010-2017
Elecciones a Cortes Generales. De 28 de abril de 2019
Elecciones al Parlamento Europeo. De 26 de mayo de 2019
Elecciones municipales y autonómicas. De 26 de mayo de 2019
Encuesta Coyuntural sobre Stock y Existencias. 2º Trimestre 2019.
Encuesta de Comercio Internacional de Servicios. 2º Trimestre 2019.
Encuesta de Ocupación en Albergues. 2018
Encuesta de Ocupación en Alojamientos de Turismo Rural. 2018
Encuesta de Ocupación en Apartamentos Turísticos. 2018
Encuesta de Ocupación en Campings. 2018
Encuesta de Ocupación Hotelera. 2018
Encuesta sobre Centros y Servicios de Atención a Personas sin Hogar. 2018
Encuesta sobre Innovación de las Empresas. 2017
Encuesta Trimestral de Coste Laboral. Serie 1º trim. 2008 - 2º trim. 2019
Estadística de Condenados: Adultos. Serie 2013-2018
Estadística de Condenados: Menores. Serie 2013-2018
Estadística de Ejecuciones Hipotecarias. 2º Trimestre 2019.
Estadística de Filiales de Empresas Españolas en el Exterior. 2017
Estadística de Filiales de Empresas Extranjeras en España. 2017
Estadística de Movilidad Laboral y Geográfica. Serie 2010-2019
Estadística de Nulidades, Separaciones y Divorcios. Serie 2013-2018
Estadística sobre Transporte Ferroviario. Serie 1º trim. 2010 - 2º trim. 2019
Índice de Coste Laboral Armonizado. ICLA. Serie 1º trim. 2000 - 2º trim. 2019
Índice de Precios de Vivienda (IPV). 2º Trimestre 2019.
Índice de Precios del Sector Servicios. 2º Trimestre 2019.