



Islas Canarias
Del 15 al 19 de noviembre de 2021



LA ESTADÍSTICA ESPACIAL EN GALICIA PARA LA TOMA DE DECISIONES PÚBLICAS

Esther López Vizcaíno 1

Instituto Galego de Estatística

esther.lopez@ige.eu

Noa Veiguela Fernández 2

Instituto Galego de Estatística

noa.veiguela@ige.eu

Ana Andión Hermida 3

Instituto Galego de Estatística

ana.andion@ige.eu

1. INTRODUCCIÓN

En la actual organización administrativa del Estado español, los municipios constituyen las unidades administrativas menores en las que se divide el territorio y que tienen asignados lindes precisos. Por esta razón, los estudios cuyo objetivo es la localización de la población suelen descender hasta el nivel municipal (Goerlich y Mas, 2009). De todos modos, diversos autores (Reher, 1994; Rúa y otros, 2003) han recomendado que para el estudio del asentamiento de la población sobre el territorio esta división es claramente insuficiente y es necesario aumentar la resolución geográfica de análisis.

Además, el tratamiento histórico dado a las direcciones postales conducía a la asignación de divisiones territoriales, como es el caso de la población en municipios, entidades colectivas o secciones censales. El efecto final de esta asignación es que los estadísticos sólo pueden obtenerse respecto de estas zonas o sus agregados, produciéndose una pérdida de información no deseable.

Desde el punto de vista geográfico-estadístico, la división del territorio es relevante si los dominios de la partición son homogéneos, pero en el caso de los municipios y entidades colectivas, estamos ante una situación de falta de homogeneidad de superficie, población o concentración de unidades de producción. Teniendo en cuenta esta problemática, hace unos años, algunos institutos de estadística de la Unión Europea con el apoyo de Eurostat lanzaron el European Forum of Geography and Geostatistics (EFGS), con el objetivo de armonizar estadísticas europeas sobre la base de una cuadrícula de 1 km² de lado y en un sistema geodésico de referencia común.

Además, la creciente demanda de estadísticas con un elevado detalle territorial para un análisis espacial más preciso se concreta normativamente en el Reglamento (UE) 2017/2391 del Parlamento Europeo y del Consejo de 12 de diciembre de 2017 por el que se modifica el Reglamento (CE) num. 1059/2003 en lo que respecta a las tipologías territoriales (Tercet), publicado en el Diario Oficial de la Unión Europea de 29 de diciembre de 2017. Este reglamento señala ya la malla estadística, las tipologías de mallas y considera que debe aplicarse un sistema de mallas estadísticas para calcular y atribuir los tipos territoriales a las regiones y zonas en cuestión, ya que dichos tipos dependen de la distribución y la densidad de la población en celdas de malla de un kilómetro cuadrado. Este nuevo sistema zonal puede permitir combinar el máximo detalle territorial en la difusión con la preservación de la confidencialidad estadística.

De todos modos, las cuadrículas tampoco satisfacen a aquellos usuarios que necesitan información sobre territorios que no coinciden con divisiones administrativas o con la unión de cuadrículas. Por tanto, es necesario tener la información de la población georreferenciada para satisfacer las necesidades de información espacial.

El objetivo final de la georreferenciación es identificar mediante coordenadas geográficas dónde se sitúan las unidades estadísticas en el territorio. En los ficheros que se usan para georreferenciar es frecuente encontrar una proporción de registros que tienen información incompleta y que no pueden ser asignados a una localización precisa en el espacio. Esta situación es particularmente frecuente en zonas rurales donde las direcciones no están estandarizadas y la dispersión de los edificios lleva a que el error de posicionamiento sea más amplio. En esta situación en lugar de eliminar estos registros lo que se suele hacer, y lo que se hará en este trabajo, es imputar las coordenadas utilizando técnicas indirectas, dependiendo de la información disponible (Walter et al. (2013); Dilekli et al. (2018); Enrique (2016)).

Teniendo en cuenta lo anterior, el Instituto Galego de Estatística (IGE) comenzó en el año 2016 con la georreferenciación de la población disponible en los sucesivos Padrones Municipales de Habitantes (PMH), empleando para ello diversas técnicas estadísticas. Estos trabajos se completaron con la unión del PMH con otros registros administrativos disponibles en el IGE, con la finalidad de crear un sistema de información integrado con datos socioeconómicos de la población, las *bases de datos sociodemográficas*.

Con los trabajos desarrollados en los últimos años, en noviembre del 2020 el IGE publicó información estadística sobre la población de Galicia en la cuadrícula regular de 1km². En esta malla regular se presenta una aproximación de la población de Galicia según diversas características sociodemográficas: sexo, edad, lugar de nacimiento, afiliación a la Seguridad Social, pensionista...

Con esta publicación se inició una línea de peticiones a medida de agentes externos que van desde tener las características de la población en zonas diferentes a las divisiones administrativas hasta obtener el alumnado en las zonas de influencia de colegios.

2. OBJETIVO

La determinación de la localización de la población sobre el territorio, no directamente relacionada con los límites administrativos, es absolutamente esencial para numerosas cuestiones prácticas de organización social. El objetivo de esta ponencia es hacer un recorrido por los trabajos realizados por el IGE para llegar a construir el sistema de información integrado de la población y presentar diferentes casos reales de uso de esta información.

3. BASES DE DATOS SOCIODEMOGRÁFICAS

En el IGE comenzamos con la tarea de fusionar registros administrativos de los que disponemos, con la finalidad de crear unas *Bases de datos sociodemográficas* que contengan datos socioeconómicos de la población de Galicia. El objetivo en este punto es disponer de unas bases de datos donde estén todas las personas que en algún momento tuvieron relación con Galicia, constatada mediante los registros administrativos de los que disponemos en el IGE. Estas bases de datos permitirán seguir la evolución en el tiempo de esas personas.

Estos trabajos comenzaron en el año 2019 con la ejecución de la actividad de interés estadístico “Desarrollo de una base de datos sociodemográfica empleando diversas fuentes administrativas y Estadísticas”, disponible en el Decreto 165/2018, de 27 de diciembre, por lo que se aprueba el Programa estadístico anual de la Comunidad Autónoma de Galicia para el año 2019.

El registro de partida en la fusión es el Padrón estadístico del Instituto Nacional de Estadística (INE), que contiene información demográfica sobre toda la población que reside en algún ayuntamiento gallego en una fecha determinada. A priori, es lógico pensar que la variable de fusión en el caso de los registros administrativos es el DNI, ya que se trata de un documento que permite identificar “unívocamente” a cada persona. Pero, no siempre se puede recurrir a esta variable en el procedimiento. Puesto que no siempre se puede recurrir al DNI en el enlace de bases de datos, cruzaremos los distintos registros administrativos empleando un conjunto de variables, como el nombre y los apellidos de la persona, su fecha de nacimiento, el ayuntamiento de residencia, etc. esto es, variables comunes a los dos registros. En algunas ocasiones deberemos someter las variables de cruce a un procedimiento previo de depuración.

Las Bases de datos creadas se dividen en dos BD_ SOCIODEMOGAL y BD_SOCIO_ DIFUSION:

BD_ SOCIODEMOGAL cuenta con dos tablas que se describen a continuación:
1.- BD_ Personal

- Contiene información de todas las personas que tengan relación con Galicia o, mejor dicho, que se detecte a través de los registros administrativos que se manejan en el IGE que la tienen (residen en Galicia, trabajan en Galicia - sin residir-).

- Contiene la información personal básica de identificación de la persona; esto es, su nombre y apellidos, fecha de nacimiento, sexo, etc. Por regla general, contendrá una fila por persona, en la que se indicará de qué registro procede la información de la persona y cuando fue dada de alta en el sistema (así como de baja, acompañado del motivo, si la persona ya no figura en el sistema; por ejemplo, por fallecimiento). Tiene una columna contador como identificador único de las personas depurado y mantenido en el tiempo. Junto a él tiene los identificadores (identificador (DNI, NIE,...), nombre, apellidos, etc) y unas pocas variables estadísticas que ayuden a la identificación (sexo, etc). Estas variables estadísticas se depuran, se revisan y se mantienen.

2.- RELACIONS

Contiene una fila por persona de BD_ Personal y registro en el que la identificamos; esto es, si en el año t detectamos que esa persona aparece en el Padrón Municipal de Habitantes estadístico, en el continuo y en dos trimestres del registro de Afiliados a la Seguridad Social, en RELACIONS deberán figurar cuatro filas para esa persona y año. Recoge los enlaces entre el contador y la clave de cada fichero estadístico. Esto permite saber en qué registros figura cada persona de BD_PERSONAL. También variables de calidad del enlace, fecha, fichero enlazado y operación a la que pertenece.

En BD_SOCIO_DIFUSIÓN se depositan los ficheros estadísticos ya enlazados en BD_SOCIODEMOGAL, listos para su explotación estadística. Consta de las siguientes tablas:

1.-AfiliaciónsSegSocial

Contiene la información sin identificar de los ficheros de afiliaciones en alta laboral de la Seguridad Social que la TGSS envía al IGE con periodicidad trimestral/mensual.

2.- PADRON_ estadístico

Contiene la información sin identificar del fichero final del Padrón estadístico con la población ajustada a las cifras oficiales a 1 de enero de cada año.

3.-PADRON_continuo

Contiene la información sin identificar de los microdatos procedentes de las descargas nominales a 1 de enero y 1 de julio de cada año del Padrón continuo.

4.- Defuncions

Contiene información sin identificar de las defunciones anuales de residentes en Galicia.

5.- Pensions

Contiene información sin identificar de los pensionistas residentes a 31 de diciembre de cada año en Galicia que reciben una prestación contributiva de la Seguridad Social.

El esquema por el que se interrelacionan todas las tablas se muestra en la Figura 1.

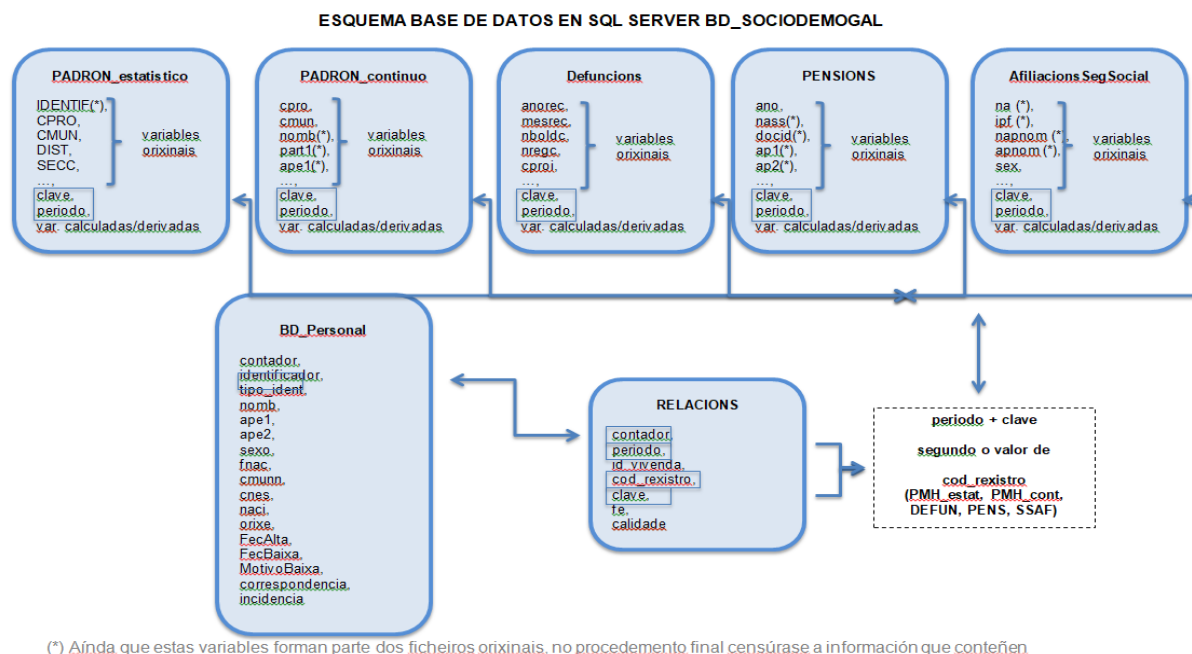


Figura 1.- Esquema de relaciones en la *Base de datos sociodemográfica*

La información disponible en las Bases de datos sociodemográficas se georreferenció teniendo en cuenta el portal en el que reside la persona. A cada uno de los portales se le asignó las coordenadas UTM en el Huso 29. La fuente principal que se emplea para georreferenciar es la base de datos proporcionada por el INE “Modelo de Direcciones de la Administración General del Estado” (MDAGE) y los datos del Catastro proporcionados por la Dirección General de Catastro. En MDAGE están disponibles las aproximaciones postales (portales) de Galicia con sus coordenadas UTM X, Y en el HUSO 30 y en el Catastro están disponibles las coordenadas UTM de la finca en la que se ubica el inmueble. Hay que resaltar que la base de datos MDAGE y los datos del Catastro no son completos, existen portales/fincas que no tienen unas coordenadas X, Y válidas o que no están georreferenciados. De todas formas, estas serán las fuentes de información principal, aunque después se complete con otras fuentes auxiliares: google, OpenStreetMap, Here ... En este procedimiento de georreferenciación se echó mano de diversas técnicas estadísticas, del software libre R y de los diversos paquetes cartográficos que ofrece

La estrategia que se empleará para georreferenciar la población será realizar un procedimiento ayuntamiento a ayuntamiento. Primero se georreferenciará las vías que tienen asignado un código de vía y que se corresponde (aproximadamente) con la parte más urbana de Galicia y a continuación se georreferenciarán las vías que no tienen asignado un código de vía, que se corresponde (aproximadamente) con la parte más rural de Galicia.

4. INFORMACIÓN POR CUADRÍCULA DE 1 KM²

El objetivo en este punto es representar la distribución de la población de Galicia, teniendo en cuenta diversas características sociodemográficas, en un mapa con un amplio desglose territorial. Se emplea una malla regular formada por celdas de 1km de lado, en la que se difundirá una aproximación de la población gallega a partir de la explotación de las *Bases de datos sociodemográficas*.

Conocer el emplazamiento concreto de residencia de la persona permitirá analizar relaciones entre localizaciones geográficas concretas, áreas o zonas de influencia, que vayan más allá de las delimitaciones que se suelen emplear en la estadística pública (ayuntamientos o, bajando un nivel, entidades de población). Esta información también puede ayudar a los poderes públicos a trazar políticas de actuación focalizadas en territorios concretos.

A la información de las *Bases de datos sociodemográficas* georreferenciada se le asignó las celdas de la malla regular de 1 km² generada. En la operación de geocodificación se obtuvieron resultados que no llegan al 100% de la totalidad de la población, en concreto para el año 2018 se geocodificó el 99,9% de la población de Galicia. El resultado final es una capa vectorial compuesta por una malla regular de celdas cuadradas de 1km de lado. Para cada una de estas celdas se presenta la información estadística de carácter sociodemográfica que le corresponde. Entre la información estadística que se representa están las siguientes variables: Población total, Población de hombres (hombres), Población de mujeres (mujeres), Ratio de feminidad: Mujeres/ Hombres 100 (ratio_femi) , Porcentaje de población menor de 16 años (porc_menor_16), Porcentaje de población entre 16 y 64 años (porc_mayor15_me), ...

Aprovechando los trabajos anteriores, en noviembre del 2020 el IGE publicó en su web la actividad estadística “Distribución espacial de las características de la población de Galicia por cuadrícula de 1km²”. En esta actividad estadística, y con el fin de visualizar toda la información de una manera sencilla, se desarrolló un visualizador cartográfico de fácil manejo y muy intuitivo al que se puede acceder en la siguiente dirección y cuya página se presenta en la Figura 2:

[https://www.ige.gal/web/mostrar_activade_estadistica.jsp?codigo=0201001009&idioma=gl](https://www.ige.gal/web/mostrar_actividade_estadistica.jsp?codigo=0201001009&idioma=gl).

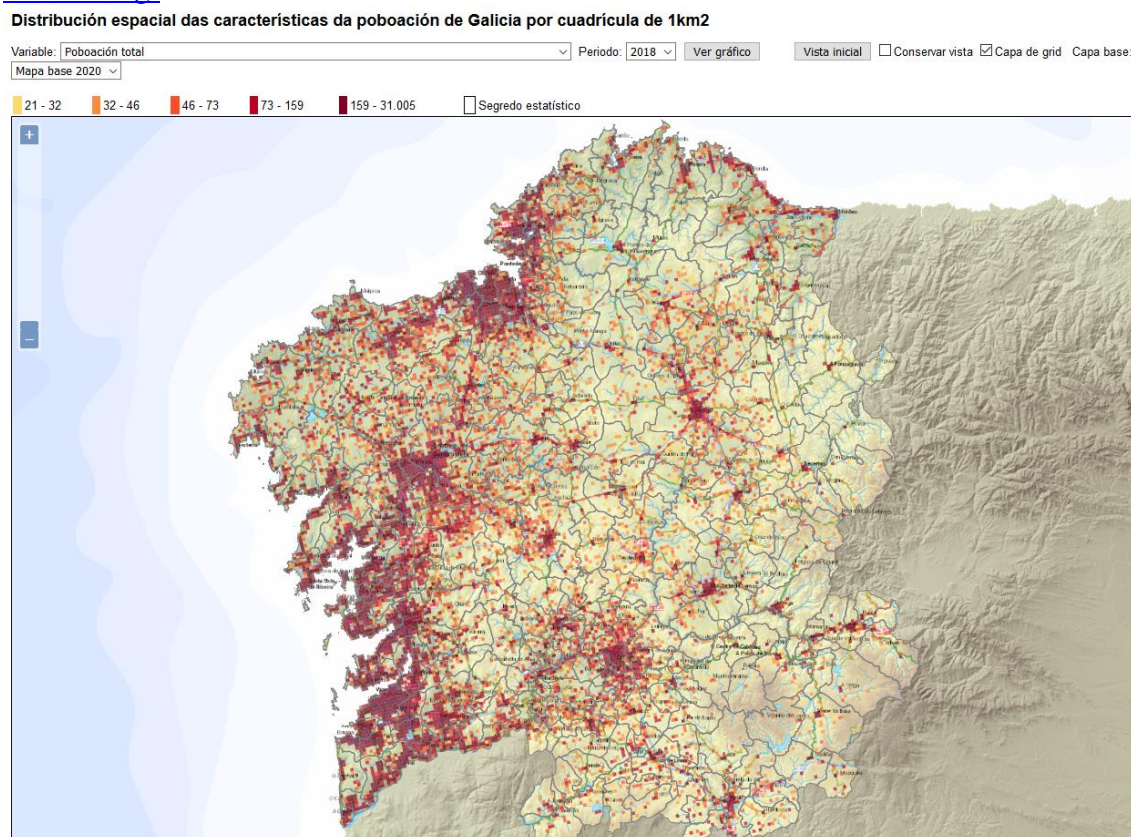


Figura 2.- Visualizador de la distribución espacial de las características de la población de Galicia en cuadrículas de 1km²

En el visualizador se puede consultar la información en forma de mapas. A cada celda se le asigna un color según el intervalo en el que se encuentre la variable sociodemográfica que se va a representar. Los datos de celdas con población menor de 21 habitantes no se representan y los restantes datos censurados aparecerán representados con una celda blanca. El visualizador permite realizar desplazamientos por todo el territorio, así como todos aquellos zooms que muestren al usuario la información en el nivel de detalle que desee. Estos niveles van desde lo más general que permite la visión global de toda Galicia, hasta escalas urbanas de gran detalle. Existe también la posibilidad de consultar los datos de cada celda, mediante el recurso de picar para que aparezca un cuadro emergente, en el que se muestra la información del ayuntamiento donde se localiza la celda, el número de personas residentes en la misma y la característica elegida para representar en el mapa: hombres, mujeres, número de personas receptoras de pensiones contributivas,... También se puede consultar la información en formato tabla, descargable en formato hoja de cálculo y a través de APIs.

Se garantizó la protección y la confidencialidad de la información mediante el método de eliminación de datos. Inicialmente se realizó una eliminación de los datos considerados sensibles y, posteriormente, cuando fue necesario, una supresión secundaria que impida la identificación exacta por deducción de los datos sensibles censurados, por diferencia con respecto al total. En concreto, los criterios empleados fueron los siguientes:

- Se eliminaron las celdas que tienen una población total menor de 21 habitantes.
- Se censuraron las celdas que tienen menos de 5 habitantes (>0) en alguna de las características sociodemográficas por las que se clasifica.
- Se censuraron celdas complementarias para evitar la deducción de datos sensibles censurados en el paso anterior por diferencia respecto al total. En este caso se eliminaron los datos teniendo en cuenta su valor, con la intención de minimizar el coste en términos de pérdida de información

Por lo tanto, los criterios aplicados impiden la identificación exacta de los datos considerados sensibles, no su aproximación. El proceso de eliminación de datos se realizó aplicando un algoritmo desarrollado ad-hoc que cumple los criterios de secreto estadístico establecidos.

El territorio de Galicia se cubre con un total de 30.776 celdas de 1 km². En el año 2018, 19.021 celdas tenían población. De estas, se difunden 10.775 celdas que tienen una población superior a 20 personas. Estas celdas concentran el 97,2% de la población de Galicia, por lo tanto, las 8.246 celdas que no se puede difundir por problemas de secreto estadístico, concentrarían el 2,8% de la población de Galicia. Este sería el coste en términos de pérdida de información que hay que asumir por el cumplimiento del secreto estadístico.

De todos modos, las cuadrículas tampoco satisfacen a aquellos usuarios que necesitan información sobre territorios que no coinciden con divisiones administrativas

o con la unión de cuadrículas. Por tanto, es necesario tener la información de la población georreferenciada para satisfacer las necesidades de información espacial.

5. ALGUNOS CASOS REALES DE USO DE INFORMACIÓN GEORREFERENCIADA

Con la publicación en la web del IGE de la actividad estadística “Distribución espacial de las características de la población de Galicia por cuadrícula de 1km²” se inició una línea de peticiones a medida de agentes externos necesitados de información estadística en zonas diferentes a las divisiones administrativas habituales o a la unión de celdas de 1km².

4.1. Información para el Plan de Impulso al Medio Ambiente (PIMA Adapta Costas)

El fin del PIMA Adapta Costas es contribuir al desarrollo del Plan Nacional de Adaptación al Cambio Climático y la puesta en marcha de la Estrategia de Adaptación de la Costa Española al Cambio Climático, contribuyendo a estudios de detalle por tramos de costa de dimensión autonómica, garantizando la uniformidad en la metodología aplicada a la generación, adquisición y recopilación de datos y a la evaluación de la vulnerabilidad de las costas españolas al cambio climático.

La labor del IGE consistió en elaborar y proporcionar al Instituto de Estudios del Territorio la distribución de la población de Galicia en una malla de 250m*250m en el ámbito PIMA. El ámbito PIMA se corresponde con una zona del litoral de Galicia, tal y como se muestra en la Figura 3. En esta extracción de información se tuvo en cuenta la información proporcionada a nivel de celdas de 1km² con el fin de no infringir el secreto estadístico.

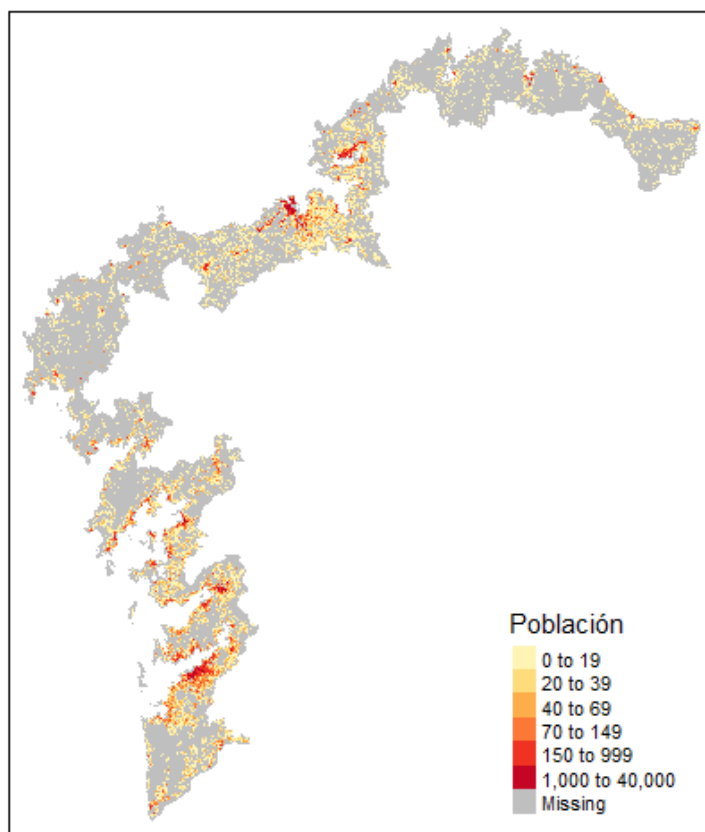


Figura 3.- Población en cuadrículas de 250mx250m en el ámbito PIMA

4.2 Población en el ámbito del Plan Especial de la Ciudad Histórica de Ourense

La sangría demográfica no es un fenómeno exclusivo de entornos rurales. En plena capital de Ourense, el casco viejo sufre una pérdida de población continuada. De ahí que el Concello de Ourense esté interesado en cuantificar la población actual y la pérdida continuada que está sufriendo. El ámbito del Plan Especial de la Ciudad Histórica de Ourense no se corresponde con ninguna de las divisiones administrativas disponibles hasta el momento, con lo que es necesario disponer, de la información de la población georreferenciada y la delimitación del ámbito en un formato GIS, para proceder al cálculo de la población residente en ese ámbito. Desde el municipio de Ourense nos proporcionaron la delimitación de la Ciudad Histórica y desde el IGE les aproximamos la población que vive en la zona que aparece en la Figura 4 según el sexo y el grupo de edad:

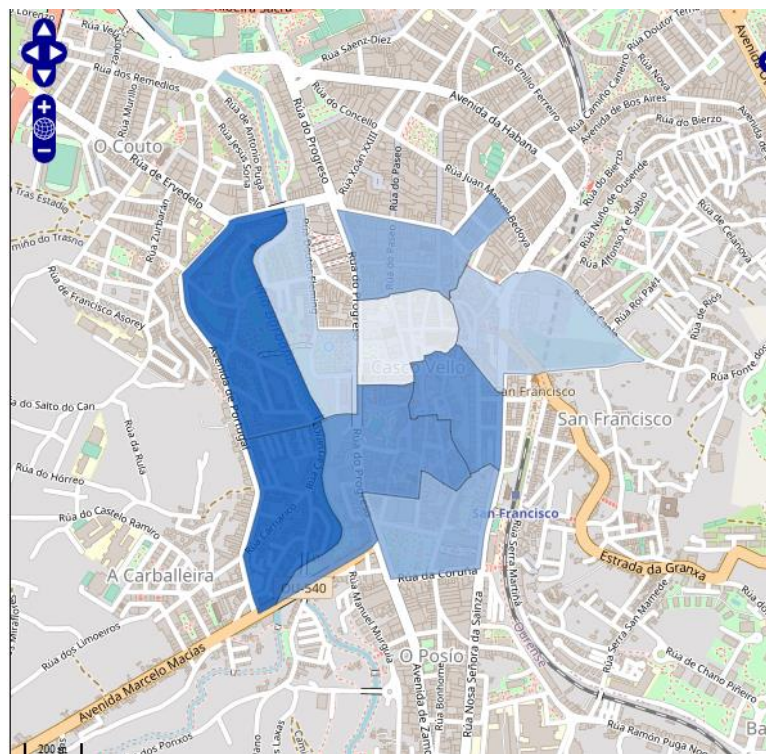
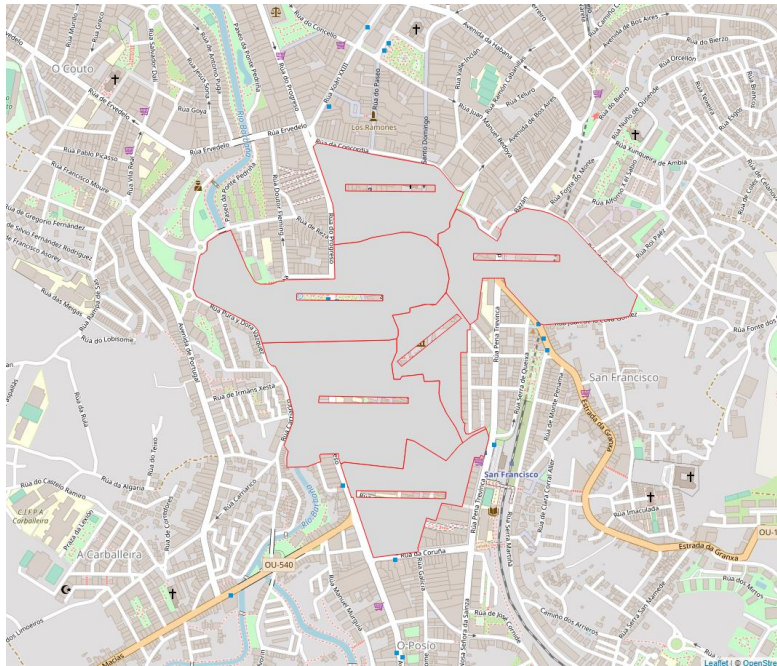


Figura 4.- Delimitación de la Ciudad Histórica de Ourense (superior) y delimitación de las secciones censales (inferior)

Esta zona tenía 5.793 habitantes en el año 2012 y en el 2018 ya eran 4.145, es decir, 1.648 menos en tan solo seis años. Las secciones censales que la contienen (Figura 4 inferior) tenían en el año 2018 una población de 13.481, cifra muy diferente a la población del ámbito. La pirámide de población se presenta en la Figura 5.

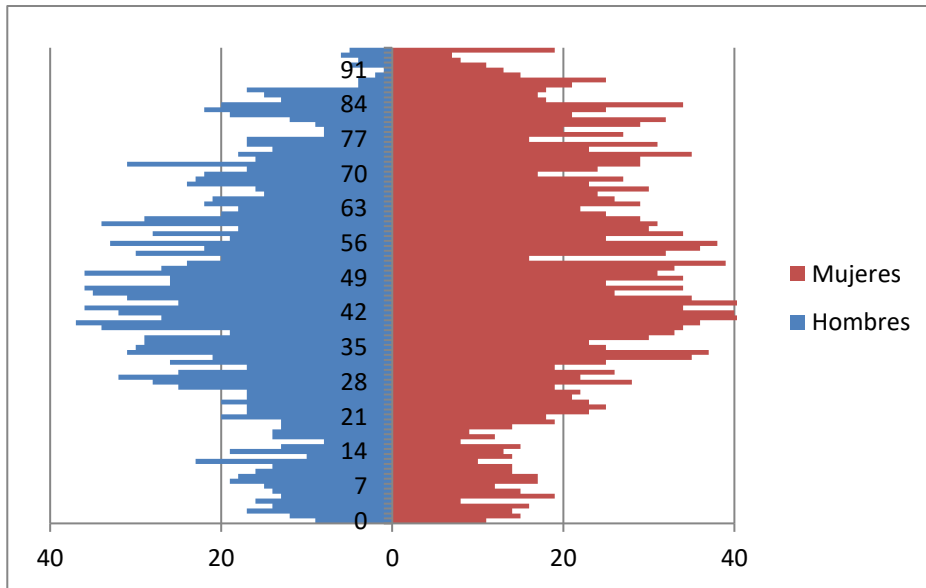


Figura 5.- Pirámide de población de la Ciudad Histórica de Ourense

4.3 Proyección de matrícula en un colegio de Educación Primaria a dos años vista

A partir de la base de datos sociodemográfica se puede aproximar la población nacida en los años 2018 y 2019 en las áreas de influencia de los colegios. Esto podría servir como proyección de altas en los colegios a dos años vista: para el curso 21/22 y para el curso 22/23. Para esto es necesario tener el área de influencia de los colegios delimitada en formato GIS.

En la Figura 6 se puede observar el área de influencia digitalizada de un colegio de infantil y primaria de Santiago de Compostela. A partir de la base de datos sociodemográfica y de los datos a 1/1/2020 se puede saber cuántos niños se prevé que se matricularán en el colegio en el curso 21/22 y en el 22/23. Los resultados se presentan en la Tabla 1.



Figura 6. Área de influencia de un centro de enseñanza de infantil y primaria en Santiago de Compostela

Edad	Niños
0	55
1	91
2	95

Tabla 1.- Población por edad en la zona de influencia del colegio a 1/1/2020

6. CONCLUSIONES

La determinación de la localización de la población sobre el territorio, no directamente relacionada con los límites administrativos, es absolutamente esencial para numerosas cuestiones prácticas de organización social (Goerlich y Cantarino, 2012). Disponer de las unidades estadísticas georreferenciadas permitirá una planificación social, económica o ambiental más fácil de diseñar y de evaluar, así como, una ordenación territorial más adecuada, menos dependiente de la subdivisión administrativa. Procurando que uno de los problemas más acuciantes de Galicia, el territorio, se convierta en una oportunidad.

7. BIBLIOGRAFÍA

DILEKLI, N., JANITZ, A. E. AND CAMPBELL, J.E. and BEURS K.M.. (2018): "Evaluation of geospatial strategies in a large case study". *International Journal of Health Geographics*. <https://doi.org/10.1186/s12942-018-0151-y>

ENRIQUE, I. (2016): "Official grid-based statistics: a new approach to regional statistics in Andalusia". *European Forum for Geography and Statistics*, Paris

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS (EFGS) (2017): "Grid". Recuperado de: <http://www.efgs.info/data/>

GARCÍA-PALOMARES, J.C., SALAS-OLMEDO, M.H., MOYA-GÓMEZ, B., CONDEÇO-MELHORADO, A. M. and GUTIÉRREZ, J. (2017): "City dynamics through Twitter: relationships between land use and spatiotemporal demographics". *Cities*, V 72, Part B, 310-319.

GOERLICH, F.J. y CANTARINO, I. (2011): "Cartografía y Demografía: una grid de población para la Comunitat Valenciana". *Estadística Española*. Vol. 53, Núm. 176, 127-161.

GOERLICH, F.J. y CANTARINO, I. (2012): Una "grid" de densidad de población para España. Bilbao, Fundación BBVA

GOERLICH, F. J. and MAS, M (2009): "Drivers of agglomeration: Geography versus History". *The Open Urban Studies Journal* 2. Recuperado de: <http://www.bentham.org/open/tousj/openaccess2.htm>

REHER, D. (1994): "Ciudades, procesos de urbanización y sistemas urbanos en la Península Ibérica 1550-1991". En M. Guardia, F.J. Monclús, J. Oyón (dirs.). *Atlas*

histórico de ciudades europeas. Barcelona: Centre de Cultura Contemporànea de Barcelona y Salvat, 1.29.

RÚA, A., REDONDO, R. y DEL CAMPO C. (2003): “Distribución municipal de la realidad socioeconómica gallega”. *Revista Galega de Economía*, vol 12, núm 2, pp 243-262

WALTER, S.R. and ROSE, N. (2013): “Random property allocation: A novel geographic imputation procedure based on a complete geocoded address file”. *Spatial an Spatio-temporal Epidemiology*, vol 6, pp. 7-16