



Islas Canarias  
Del 15 al 19 de noviembre de 2021

istac

INSTITUTO CANARIO  
DE ESTADÍSTICA

## INDICADORES DEMOGRÁFICOS DE CANARIAS EN MALLAS DE 250 METROS (2010-2020)

**Noelia Martín Morales**

Instituto Canario de Estadística, ISTAC  
nmarmor@gobiernodecanarias.org

**Rafael Betancor Villalba**

Instituto Canario de Estadística, ISTAC  
rbetvilb@gobiernodecanarias.org

### Introducción

Para atender la cada vez mayor demanda de datos a nivel inframunicipal se ha procedido a incorporar la publicación de datos estadísticos sobre la malla estadística de 250 metros de lado. Esta se deriva y es conforme a la malla de 1 kilómetro cuadrado definida por Eurostat. en el marco de GISCO (Geographic Information System of the COMmission).

Desde Eurostat se promueve la publicación de datos basados en una red uniforme y atendiendo a las recomendaciones de la Directiva Comunitaria INSPIRE. Esta malla es regular, el tamaño de la celda es constante y su posición es fija y conocida, además es continua en todo el territorio e invariable en el tiempo. Esto permite nuevas posibilidades de análisis como pueden ser los estudios a lo largo del tiempo frente a otras unidades territoriales que sí cambian (entidades, núcleos, distritos, secciones) o un mayor detalle territorial.

Para que esto sea posible se necesita la georreferenciación de la población, mediante un método que garantice dos objetivos importantes:

Las cifras obtenidas deben ser coherentes con las cifras oficiales de población del padrón municipal de habitante

Se preserva el secreto estadístico de los datos.

A lo largo de este trabajo se expone la solución desarrollada en el Instituto Canario de Estadística para la publicación de datos estadísticos en la malla estadística de 250 metros de lado.

## Objetivos

Los objetivos establecidos en el trabajo dan respuesta a los dos requisitos fundamentales que debe satisfacer la georreferenciación de los datos de población consolidados que fueron expuestos en la introducción. Los datos a los que hacemos referencia, el fichero SD-PMH, contiene datos anonimizados y por tanto no incluye direcciones. La primera tarea a abordar consiste en determinar la forma de georreferenciarlo. El camino seguido ha sido a través del fichero RD-PMH, sin anonimizar que incluye direcciones. Partiendo de esta información se definen dos objetivos específicos:

- Georreferenciar y geocodificar registros no anonimizados en RD-PMH
- Identificar registros equivalentes en ambos ficheros

Por otra parte, sobre el fichero resultante se deben llevar a cabo procesos para garantizar el secreto estadístico de los datos publicados.

## Metodología

### Georreferenciación y geocodificación

Previamente a la identificación se ha aplicado la georreferenciación de las direcciones del fichero del Padrón Municipal de Habitantes no anonimizado (georreferenciación de microdatos) para luego poder aplicar la geocodificación de dichas georreferencias.

El Sistema de Georreferenciación y Geocodificación está enumerado en el principio primero del Marco de Estadística Espacial de Canarias. La finalidad de este principio es poder dotar de representación espacial a las unidades de información estadística de manera que puedan ser integradas en análisis espaciales. Este proceso se denomina de forma genérica como georreferenciación o referenciación espacial.

- La georreferenciación, o referenciación espacial, es el proceso de referenciar datos contra un sistema de coordenadas geoespacial conocido, ajustándose a puntos conocidos en el sistema de coordenadas, de manera que los datos pueden ser visualizados, procesados, consultados y analizados junto con otros datos geográficos (*Figura 1*).
- La geocodificación es el proceso de asignación de un código geográfico a una unidad de análisis, asociándola a una línea o polígono georreferenciado. A efectos del Marco de Estadística Espacial de Canarias, la geocodificación se define como el proceso de dotar de dimensión espacial a las unidades de registro estadístico de manera que puedan ser utilizadas en análisis espaciales. El Sistema de geocodificación desarrollado en el Marco de Estadística Espacial de Canarias distingue entre la geocodificación de literales y la geocodificación de georreferencias. Este sistema facilita el análisis geoestadístico mediante el uso de polígonos (*Figura 2*).

La geocodificación de georreferencias asigna a un punto los geocódigos de los polígonos a los que pertenece dentro de una colección de polígonos almacenados en un esquema de información geográfica de referencia (IGR) o información geográfica de soporte (IGS). Los inputs de este proceso son dos tablas GEO, una con las referencias (x,y) asociadas a los microdatos, y otra con las geometrías de los polígonos almacenados en IGR o IGS. El

resultado es una tabla en la que se relacionan cada referencia con los polígonos a los que pertenece.

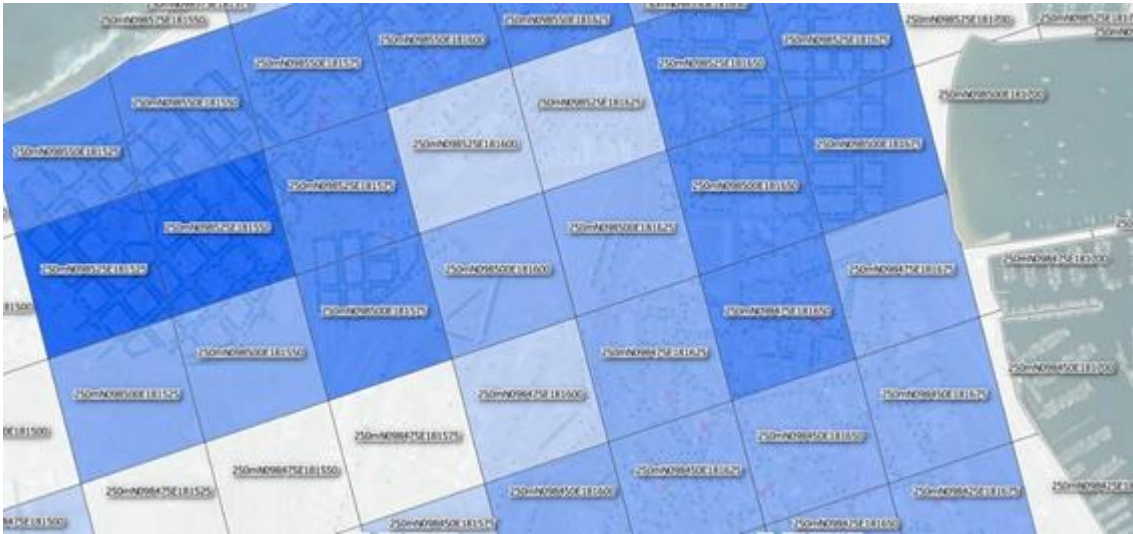


Figura 1. Georreferenciación de las direcciones.

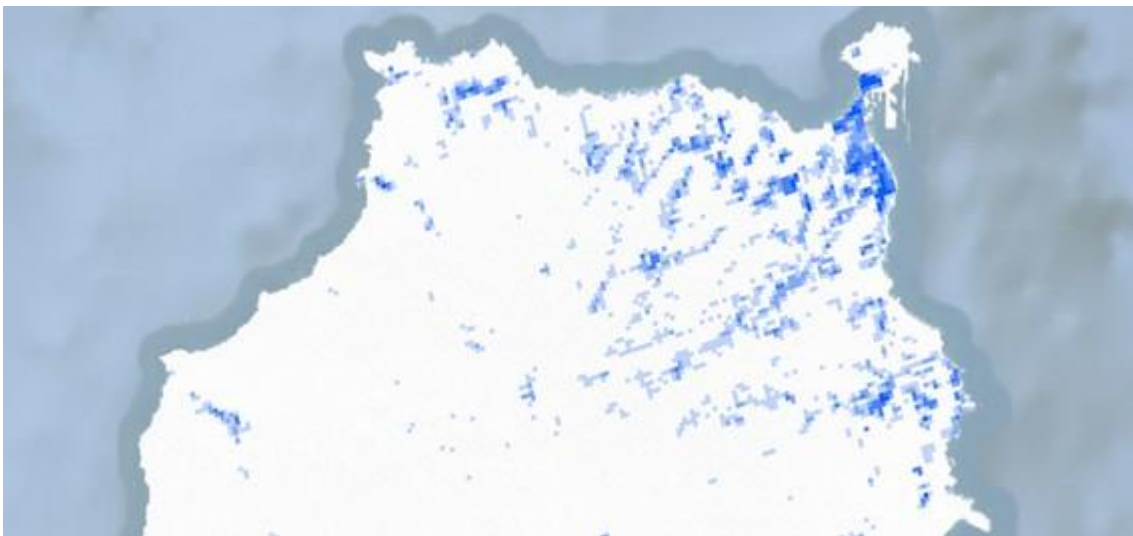


Figura 2. Geocodificación de las direcciones: intersección entre la capa de puntos de las direcciones georreferenciadas y los polígonos de la capa de celdas de 250 metros de lado.

En la *Figura 3* y en la *Figura 4* se puede ver la población en Las Palmas de Gran Canaria y en Gran Canaria respectivamente.



*Figura 3. Análisis de la población sobre la capa de celdas de 250 metros de lado. Las Palmas de Gran Canaria.*



*Figura 4. Análisis de la población sobre la capa de celdas de 250 metros de lado. Gran Canaria*

La georreferenciación se ha aplicado utilizando:

- Fuentes propias: identificación de las direcciones del fichero del padrón no anonimizado con las direcciones del registro de direcciones.
- Fuentes Big Data: consultan a diferentes apis por las direcciones que no se encuentran en fuentes propias y aplicando el método desarrollado en el ISTAC para seleccionar la mejor respuesta [1].

Para la geocodificación publicada se ha utilizado la malla estadística de 250 metros de lado derivada y conforme a la malla de 1 kilómetro cuadrado definida en el marco de



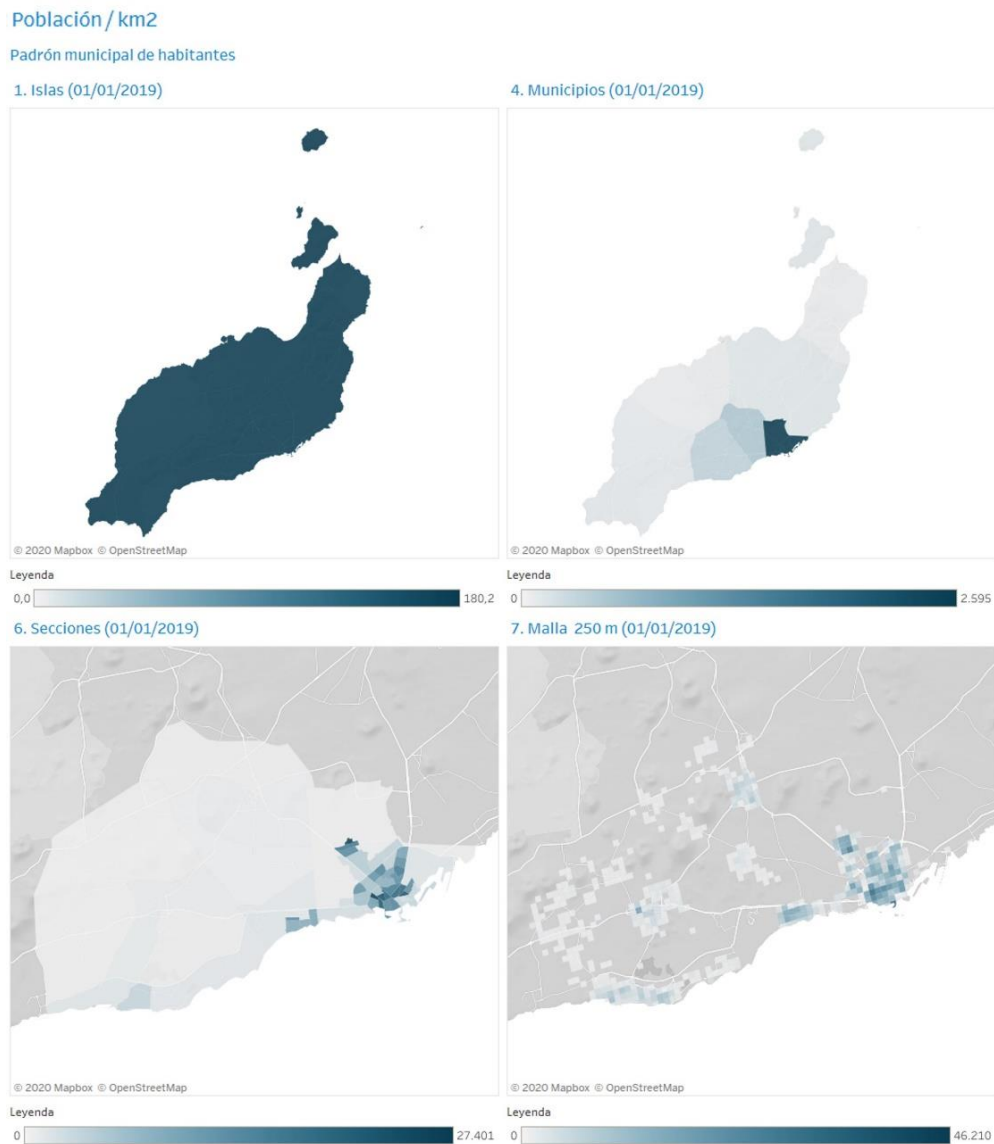
GISCO (Geographic Information System of the COMmission) por Eurostat [3], es cartografía básica no exacta utilizada para la representación simbólica de estadística geoespacial.

¿Cómo obtener la cartografía?:

- Versión 0.5 de la extensión del ISTAC para QGIS
  - <http://plugins.qgis.org/plugins/istacqgis/>
- Catálogo de datos abiertos del ISTAC
  - <https://datos.canarias.es/catalogos/estadisticas/dataset?q=cartografia>
  - <https://datos.canarias.es/catalogos/estadisticas/dataset/malla-estadistica-eurostat-adaptada-a-canarias-celdas-250-m-de-lado>

Este proceso de georreferenciación y geocodificación permite aproximar mejor las estadísticas obtenidas a la realidad del territorio.

Como ejemplo, en la *Figura 5* se puede ver la densidad de población por isla, municipio, sección y celdas de 250 metros de lado respectivamente para la isla de Lanzarote.



*Figura 5. Densidad de población por isla, municipio, sección y celdas de 250 metros de lado respectivamente para la isla de Lanzarote.*

## Identificación de registros

Con respecto a las cifras de población, los institutos de estadística disponen de dos tipos de ficheros:

- **RD-PMH:** se recibe casi inmediatamente, sin anonimizar y con direcciones completas. El problema que presenta este fichero es que no ha sido depurado, es decir, sin actualizar ciertas altas, bajas o modificaciones, sin eliminar algunos duplicados, sin aplicar cambios administrativos como la fusión, segregación de secciones, etc.
- **SD-PMH:** en este caso estamos ante un fichero coherente con las cifras oficiales de población, depurado de errores. Este proceso es costoso, lo que lleva a su publicación prácticamente un año después de la fecha a la que hace referencia. Por contra, el fichero está anonimizado y consecuentemente sin las direcciones.

La *Figura 6* presenta de forma comparada las características más relevantes de ambos.

RD-PMH: Sin anonimizar	SD-PMH: Anonimizado
Con identificación	Sin identificación
No es coherente con las cifras oficiales publicadas	Coherente con las cifras oficiales publicadas
sobran registros	los registros oficiales
errores: entidad, núcleo, sección, etc.	depuración de errores
Dirección completa y normalizada	Sin dirección
permite la georreferenciación	no se puede georreferenciar
población para cualquier territorio: GRID, barrios, ENP...	población sólo para los territorios definidos en el propio fichero: municipios, secciones y núcleos

*Figura 6. Comparación de las características de los ficheros del Padrón Municipal de Habitantes: anonimizado y sin anonimizar.*

Este fichero anonimizado nos permite obtener cifras consistentes con las cifras oficiales de población publicadas por el INE para todas las entidades territoriales definidas en el fichero: municipios, distritos y secciones, entidades y núcleos.

Por otra parte el fichero no anonimizado, al tener las direcciones completas permite la georreferenciación de las mismas, es decir, obtener las coordenadas geográficas de cada portal. Esta información con la que se complementa hace posible obtener estadísticas para cualquier entidad territorial que esté delimitada en una cartografía.

Ejemplos de ellas son: celdas de 250 metros de lado, barrios, zonas básicas de salud, áreas de influencia de los centros escolares, espacios naturales protegidos. Por contra, los resultados obtenidos a partir de estos datos no coinciden con las cifras oficiales de población.

El objetivo ha sido combinar lo mejor de cada fichero, del anonimizado se requiere la coherencia con las cifras oficiales de población y del no anonimizado la georreferenciación aplicada a las direcciones. De esta forma se logra obtener información para nuevos análisis en diferentes entidades territoriales, más allá de los niveles que permite analizar el fichero anonimizado como se puede ver en la *Figura 7*.

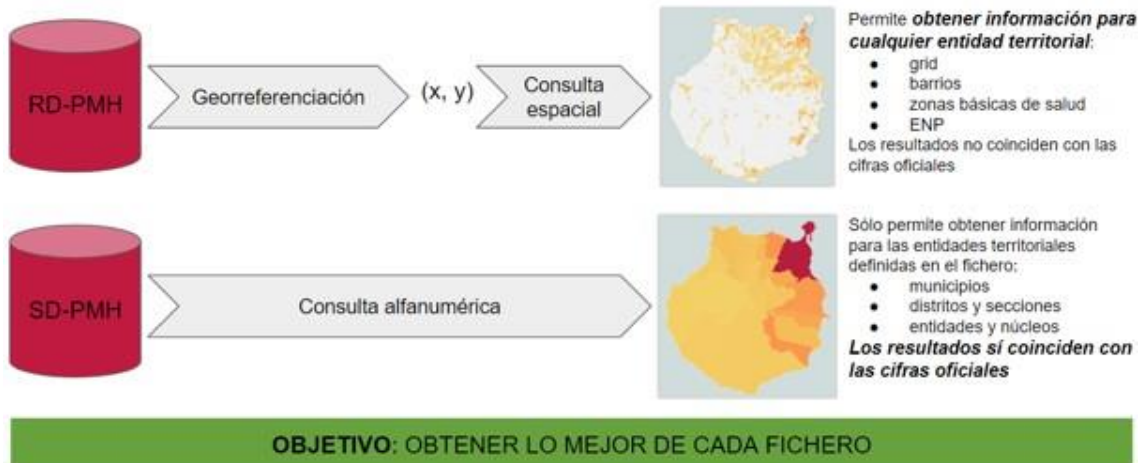


Figura 7. Comparación de los datos obtenidos a partir de cada fichero.

La dificultad se encuentra en que éste no cuenta con ninguna variable que permita relacionarlo con el no anonimizado y establecer una relación uno a uno entre ambos. En la *Figura 8* se puede ver la propuesta de relación *quasi-unívoca* entre los dos ficheros.

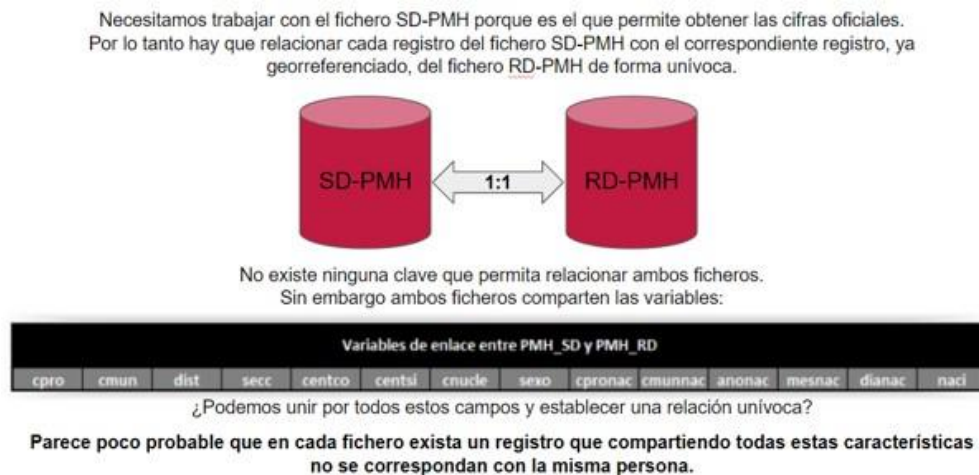


Figura 8. Propuesta de relación *quasi-unívoca* entre ambos ficheros.

La solución desarrollada se construye sobre las variables comunes en ambos ficheros, que denominamos variables de enlace. Se parte de la hipótesis de que es poco probable que registros en cada uno de los ficheros compartiendo todas estas características correspondan a diferentes personas.

Cabe destacar que no es necesario identificar a la persona, simplemente en el fichero anonimizado tenemos un registro con unas características que establecen el criterio de búsqueda de un registro con las mismas características en el otro fichero, que además cuenta con una variable dirección. En este proceso se establece la relación unívoca que permite obtener las coordenadas geográficas a las que fue transformada la dirección almacenada para ese registro en el fichero no anonimizado. Esto permite incorporar las coordenadas del fichero georreferenciado al fichero anonimizado.

Este proceso basado en las variables de enlace describe la forma en que se ha obtenido una georreferenciación válida para el fichero anonimizado. Dicha georreferenciación no se publica, pero sí se aplica para obtener resultados estadísticos para territorios para los que hasta ahora no se podían calcular.

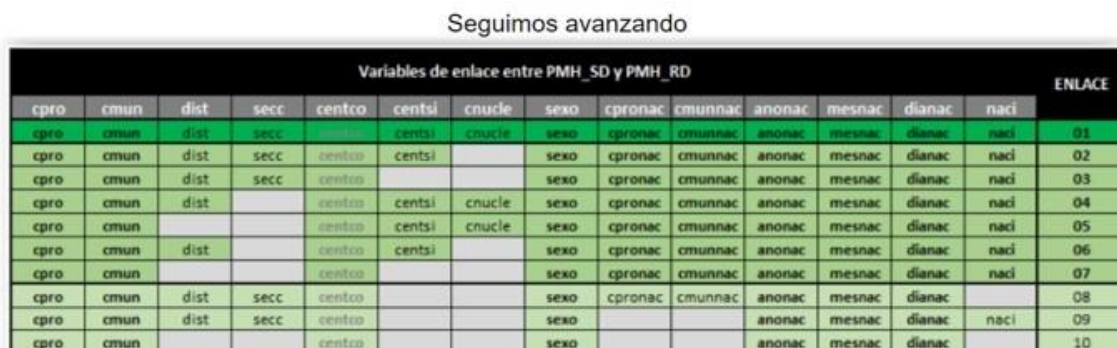
En la *Figura 9* se describe cómo con este primer enlace se ha podido relacionar el 95,22% con una relación [1:1], pero podemos llegar a relacionar el 95,59% del PMH anonimizado, ampliando a otras relaciones.



Los registros que se han conseguido enlazar se quitan de ambos ficheros para los pasos siguientes

*Figura 9. Primer enlace entre ambos ficheros y resultados.*

Tras aplicar este primer enlace, se quitan los registros unidos de ambos ficheros. A continuación se siguen probando otros enlaces en los que se relaja alguna de las condiciones. La *Figura 10* muestra los 10 primeros enlaces aplicados.



*Figura 10. Los 10 primeros enlaces aplicados.*



## Secreto estadístico

El proceso para garantizar el secreto estadístico consta de dos fases:

- FASE 1: Garantizar que ninguna celda tiene menos de 10 residentes (umbral  $t_1 = 10$ ).
  - Para reducir la pérdida de información se ha aplicado un algoritmo que reasigna la celda que inicialmente le corresponde a la dirección por la más cercana, intentando que los cambios afecten lo menos posible a los totales en las celdas padre (500 m x 500 m) y abuelas (1 km x 1 km).
- FASE 2: Para cada categoría de cada variable de clasificación no se publica la correspondiente cifra si es menor que 3 (umbral  $t_2 = 3$ ), y tampoco las de aquellas categorías que permitan obtener por diferencia el dato suprimido inicialmente.

### FASE 1

Se pretende minimizar el número de celdas cuya población no llega al mínimo, para perder la menor cantidad de información en el siguiente paso: supresión de las celdas sensibles. El método vacía las celdas sensibles, modificando la geocodificación de cada portal, uno a uno, hacia otra celda, lo más cercana posible.

Tras georreferenciar y geocodificar las direcciones se han calculado los totales de población de cada celda de la malla de 250 metros de lado y se han señalado las celdas sensibles, es decir, aquellas cuya población no llega al límite fijado ( $t_1 = 10$ ). Para preservar el secreto estadístico se ha recodificado la celda a la que pertenece cada dirección con la finalidad de que ninguna celda tenga una población por debajo del límite propuesto. Esta recodificación se ha hecho procurando que la distancia entre la dirección original y la celda reasignada finalmente sea mínima.

El algoritmo usa los conceptos: padre, abuelo, hermano, para establecer la relación entre las celdas de dimensiones mayores que contienen a otras de menor dimensión. Por ejemplo, celdas grid de 250 m x 250 m contenidas en una misma celda de 500 m x 500 m se consideran hermanas. Todas las celdas de 250 m x 250 m contenidas en una misma celda de 1 km x 1 km serán nietas de esta. Se consideran primos, todas las celdas de 250 m x 250 m nietas de una misma celda de 1 km x 1 km y que no sean hermanas.

**Paso -1 (previo a este trabajo): georreferenciar**

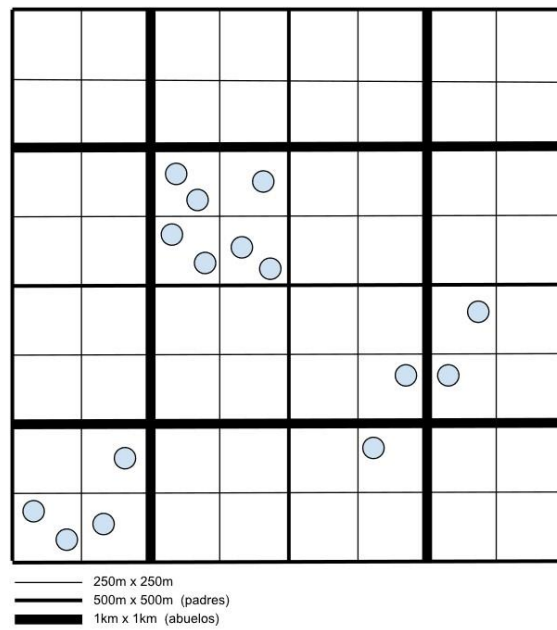


Figura 11. Georreferenciar.

**Paso 0 [detectar sensibles]: calcular la población por coordenadas XY y la población por celdas del GRID**

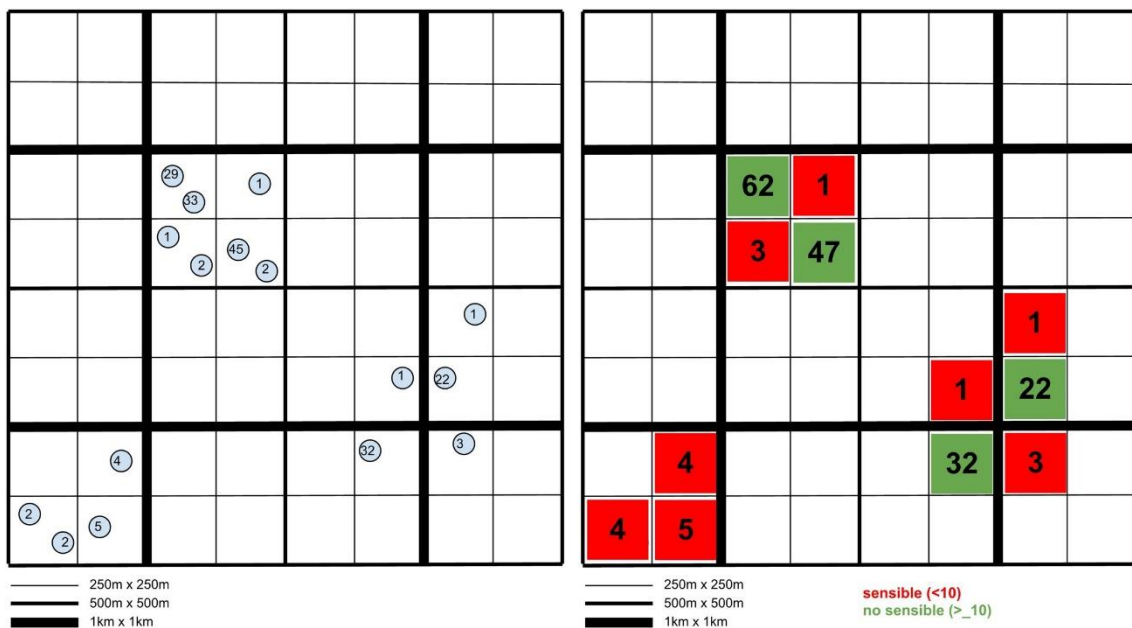


Figura 12. Detectar sensibles.

**Paso 1 [hermanos menores]:** todos los grid de 250m x 250m hermanos (hijos del mismo grid de 500m x 500m) son sensibles, sin embargo la suma no lo es: reasignar los hermanos a un único grid de 250m x 250m, el de mayor peso (reasignar a nivel de XY, no persona a persona).

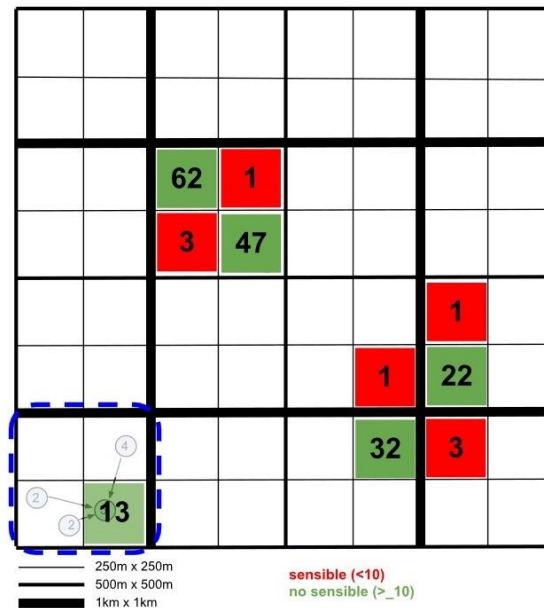


Figura 13. Hermanos menores.

**Paso 2 [hermano mayor]:** entre los grid de 250m x 250m hermanos (hijos del mismo grid de 500m x 500m) existen sensibles y no sensibles:

- se vuelve al nivel de XY
- se separan dos conjuntos, uno con los XY de los grid sensibles y otros de los XY' de los grid no sensibles
- para los XY de las celdas sensibles se calcula el XY' más cercano en una celda no sensible hermana
- se le reasigna el grid de 250m x 250m de dicho XY' no sensible
- los grid de 250m x 250m sensibles quedan vacíos
- y se reparten sus pesos entre los hermanos no sensibles según el XY' más cercano

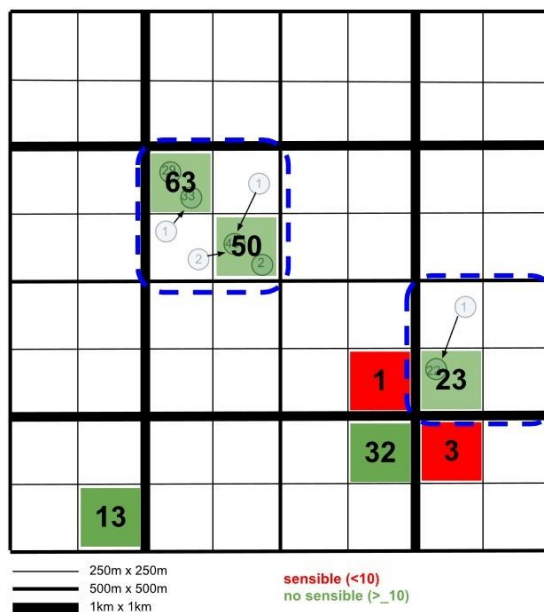


Figura 14. Hermanos mayores

**Paso 3 [primo]:** entre los grid de 250m x 250m primos (nietos del mismo grid de 1km x 1km, pero no hermanos, es decir no hijos del mismo grid 500m x 500m) existen sensibles y no sensibles:

- se vuelve al nivel de XY
- se separan dos conjuntos, uno con los XY de los grid sensibles y otros de los XY' de los grid no sensibles
- para los XY de las celdas sensibles se calcula el XY' más cercano en una celda no sensible prima
- se le reasigna el grid de 250m x 250m de dicho XY' no sensible
- los grid de 250m x 250m sensibles quedan vacíos
- y se reparten sus pesos entre los primos no sensibles según el XY' más cercano

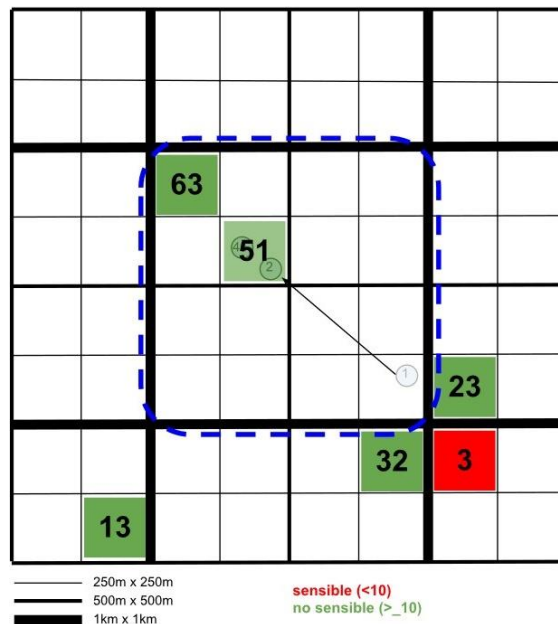


Figura 15. Primos.



**Paso 4 [suprimir sensibles]:** Se suprimen las celdas sensibles que no han sido procesadas anteriormente, perdiéndose mucha menos información, que si se hubiera aplicado la supresión de todas las celdas sensibles desde el inicio, sin aplicar esta metodología.

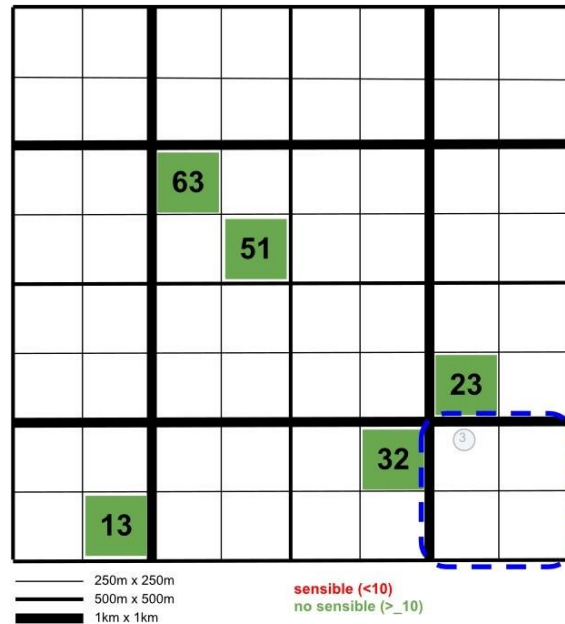


Figura 16. Suprimir sensibles.

Hasta aquí se ha explicado la primera versión del algoritmo implementado.

A partir de esta versión se implementaron las siguientes mejoras:

### Paso 3 [vecino]:

En lugar de suprimir, reasignar las celdas sensibles al portal más próximo aunque no sean de un grid de 250m x 250m ni hermano, ni primo.

Este paso se ha implementado entre el método del hermano y el del primo (entre los pasos 2 y 3), así se reasignaría a un portal más cercano, aunque no sean primos, y si ese método no funciona, entonces se buscaría entre los primos, que pueden estar más lejos. Así, en el diagrama del paso 3, se reasignaría al vecino del este o al del sur, y si esto no fuera posible, entonces sí se aplicaría el paso de los primos como se muestra en dicho diagrama. Quedaría:

1. hermanos menores
2. hermano mayor
3. vecino
4. primo.
5. suprimir sensibles

Tras valorar las distancias de Manhattan y la distancia euclídea, se ha implementado con la distancia euclídea (radio 500 m).

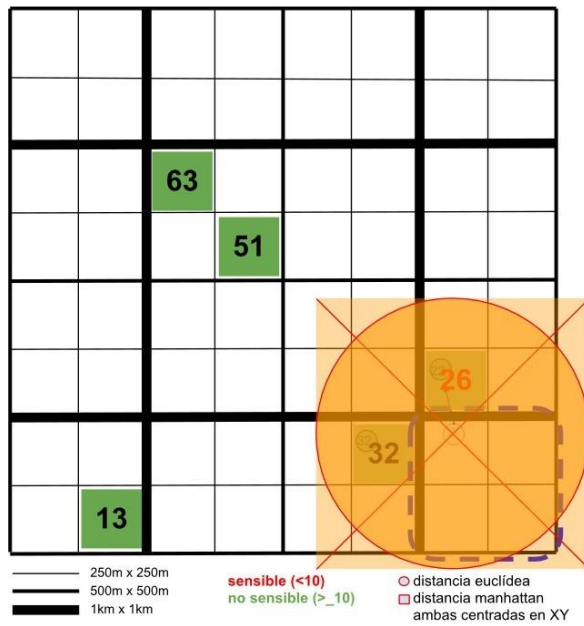


Figura 17. Vecino.

La otra mejora ha consistido en añadir restricciones como obligar a que sean hermanos o primos del mismo municipio o incluso del mismo distrito o sección.

Se ha implementado con la restricción del mismo municipio (el del centroide del grid 250 m).

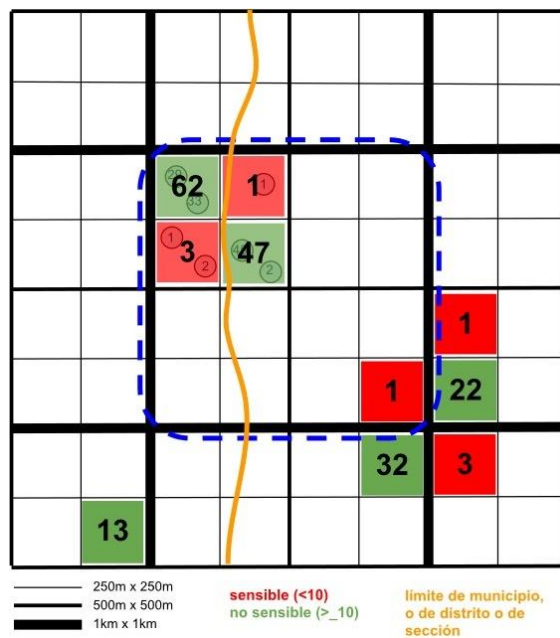


Figura 18. Restricción a la misma división territorial.

Comparación antes y después de las mejoras planteadas:

- Resultado (sin las posibles mejoras):

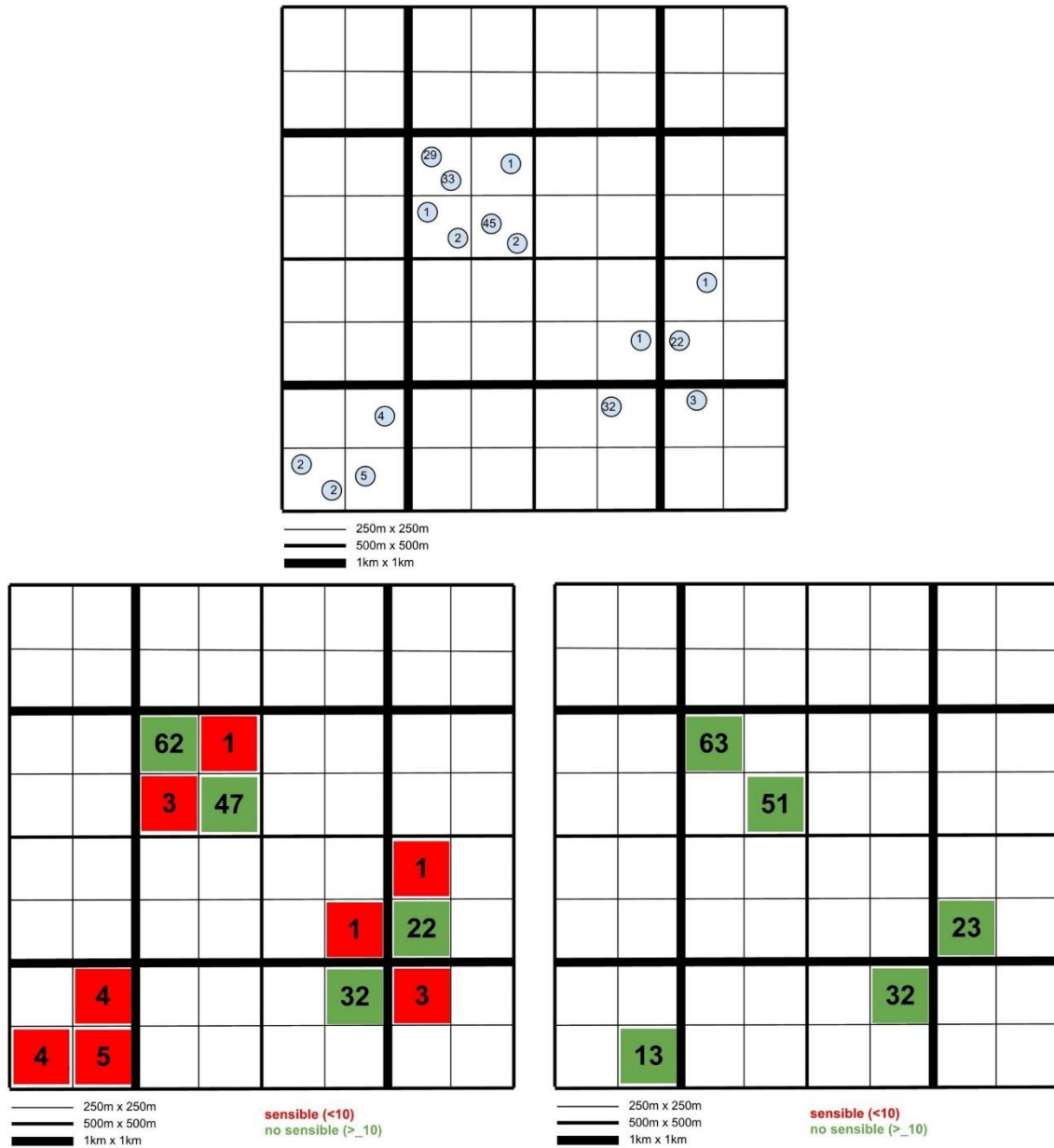


Figura 19. Resultados de la primera versión del algoritmo.

- Inicialmente hay 185
- Si suprimimos las sensibles se pierden 22
- Con el método propuesto se reasignan 19 y se pierden 3
- Con la mejora propuesta se podrían recuperar esos 3

	Sin secreto estadístico	Secreto estadístico supresión sensibles	Secreto estadístico propuesta ISTAC
No sensible	163	163	182
Sensible	22		
Suprimido		22	3
Publicado	185	163	182

Tabla 1. Resultados versión inicial del algoritmo.

- Resultado (con las mejoras):

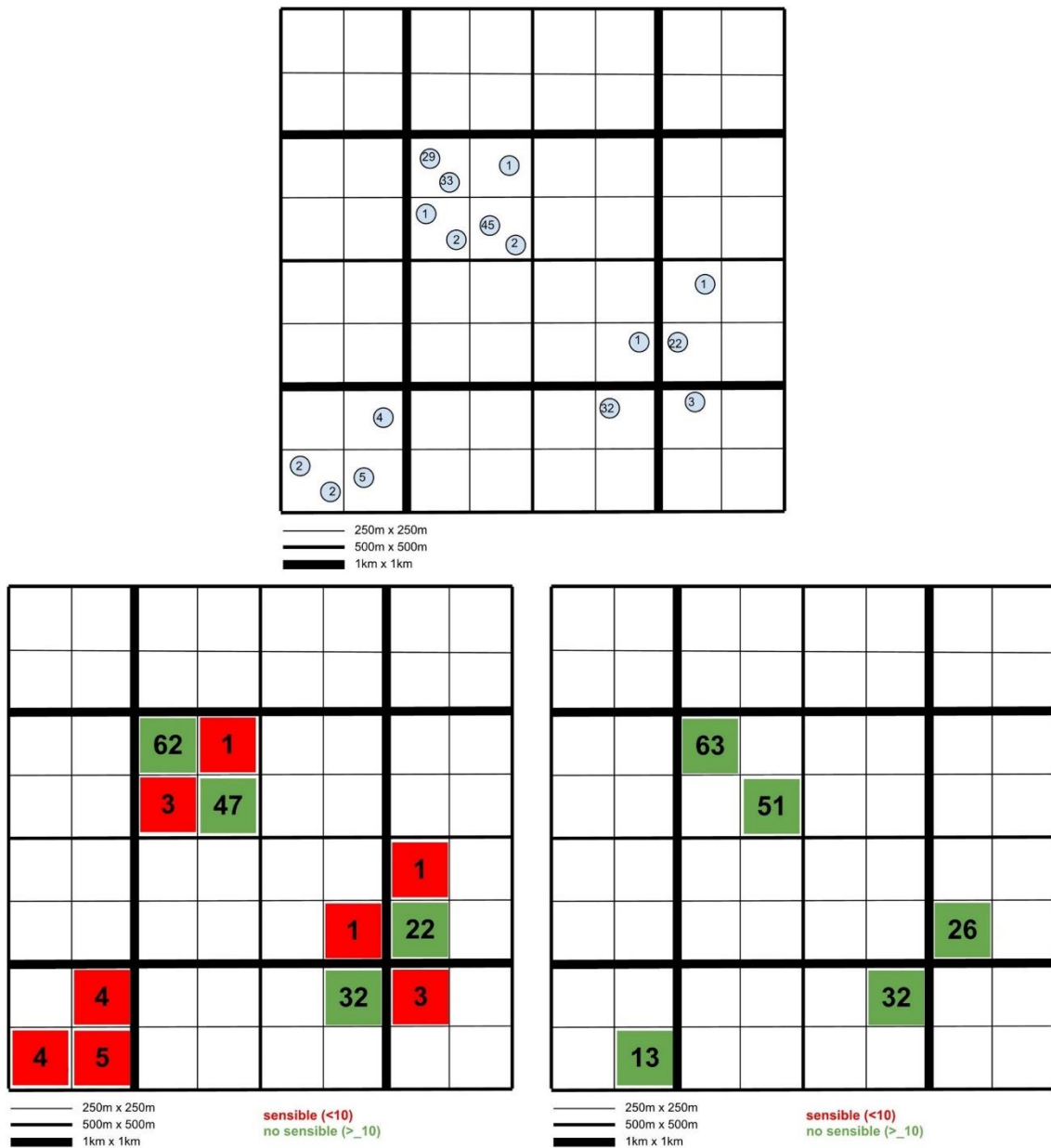


Figura 20. Resultados de la segunda versión del algoritmo.

- Inicialmente hay 185
- Si suprimimos las sensibles se pierden 22
- Con el método propuesto inicialmente se reasignaban 19 y se perdían 3
- Con la mejora propuesta se han recuperado esos 3

	Sin secreto estadístico	Secreto estadístico supresión sensibles	Secreto estadístico propuesta inicial ISTAC	Secreto estadístico propuesta mejorada ISTAC
No sensible	163	163	182	185
Sensible	22			
Suprimido		22	3	
Publicado	185	163	182	185

Tabla 2. Resultados versión mejorada del algoritmo.



En la *Tabla 3* se pueden comparar los resultados al aplicar los dos métodos a los datos -anonimizados- del Padrón Municipal de Habitantes a 01/01/2017.

Padrón Municipal de Habitantes (PMH) a 01/01/2017	Sin vecino		Con vecino	
	Sin restricción municipio		Con restricción municipio	
TOTAL	2.108.121	100%	2.108.121	100%
0   Sensible antes del secreto estadístico	19.877	0,94%	19.877	0,94%
1   HERMANOS MENORES Grid 500 no sensible con todos los hijos sensibles	2.346	0,11%	2.323 es menor por la restricción al mismo municipio	0,11%
2   HERMANO MAYOR Portal más cercano en grid no sensible dentro del mismo grid de 500m	11.638	0,55%	11.427 es menor por la restricción al mismo municipio	0,54%
3   VECINO Portal más cercano en grid no sensible dentro de un radio de 500m, puede que no sea ni hermano ni primo			4.199	0,20%
4   PRIMO Portal más cercano en grid no sensible dentro del mismo grid de 1km	3.276	0,16%	418	0,02%
5   Sensible después del secreto estadístico	2.617	0,12%	1.508	0,07%

Tabla 3. Resultados al aplicar las dos versiones del algoritmo a los datos del Padrón Municipal de Habitantes a 01/01/2017.

## FASE 2

En una segunda fase, para cada categoría de cada variable de clasificación no se publica la correspondiente cifra si es menor que el límite propuesto ( $t_2 = 3$ ), ni alguna de las categorías complementarias que permita deducir dicho valor.

Por ejemplo si el número de hombres no llega al límite no se publica, pero tampoco se publica el número de mujeres y como consecuencia no se publica el índice de feminidad, como se puede ver en la *Figura 21*.

S.E. fase 1	S.E. fase 2				S.E. fase 1	
POBLACION	POBLACION_HOMBRES	POBLACION_MUJERES	POBLACION_INDICE_FEMINIDAD	POBLACION_HOMBRES	POBLACION_MUJERES	
10				2	8	
13	13	0	0	13	0	
11				9	2	
10				8	2	
13				12	1	

Figura 21. Secreto estadístico | Fase 2: variable sexo

Cuando hay más de dos valores posibles, si uno es sensible y los otros dos no, se suprimen el sensible y el menor de los otros dos. Por ejemplo si el número de menores de 15 años es sensible pero no lo son ni el número de 15 a 64 años, ni el de 65 o más años se suprimen el de menores de 15 años y el de menor valor de las otras dos categorías, la otra cifra sí se publica para minimizar la información perdida.

S.E. fase 1		S.E. fase 2			S.E. fase 1		
POBLACION	POBLACION_00A14	POBLACION_15A64	POBLACION_65MAS	POBLACION_00A14	POBLACION_15A64	POBLACION_65MAS	
11	0	11	0	0	11	0	
14		12		2	12	0	
30		19		9	19	2	
14		9		4	9	1	
15		13		1	13	1	

Figura 22. Secreto estadístico | Fase 2: variable grupos de edad.

Con esta segunda fase se pretende evitar obtener el perfil de una persona concreta.

Este proceso limita la información publicable para este ámbito territorial por lo siguiente: si quisiéramos publicar la población por grupos quinquenales de edad, aumentaría considerablemente el número de celdas en las que no se llegaría al mínimo, como consecuencia se tendría una mayor pérdida de información. Se puede considerar que la fase previa del secreto estadístico es el análisis de las variables y categorías que se publican.

### Variables de clasificación

Las variables del Padrón Municipal de habitantes que se han incluido en este análisis como variables de clasificación para la fase 2 del secreto estadístico son:

- Sexo: hombre, mujer
- Grupos de edad: menos de 15 años, de 15 a 64 años, de 65 y más años
- Nacionalidades: españoles, extranjeros
- Lugar de nacimiento: Canarias, resto de España, extranjero

Se ha intentado encontrar un equilibrio entre la información publicada y la información perdida en la fase 2. Por ejemplo, si se añaden más grupos de edad, tendríamos una mayor cantidad de celdas suprimidas, es decir, lo que parece una ventaja puede ser un inconveniente.

### Indicadores publicados

A partir de las variables detalladas en el punto anterior se han podido calcular los siguientes indicadores para todas las celdas de 250 metros con población (fase 1) o con población suficiente para su cálculo (fase 2):

- Población
- Edad media
- Hombres
- Mujeres
- Índice de feminidad
- Población nacionalidad española
- Población nacionalidad extranjera
- Población extranjera (%)
- Población de 0 a 14 años
- Población de 15 a 64 años
- Población de 65 o más años
- Población de 0 a 14 años (%)
- Población de 15 a 64 años (%)
- Población de 65 o más años (%)
- Índice de dependencia
- Índice de dependencia mayores
- Índice de dependencia menores
- Índice de vejez
- Índice de juventud
- Densidad de población
- Población nacionalidad extranjera por km<sup>2</sup>
- Población nacida en Canarias
- Población nacida en el resto de España
- Población nacida en el extranjero
- Población nacida en Canarias (%)
- Población nacida en el resto de España (%)
- Población nacida en el extranjero (%)

## Resultados

### Difusión

Los “Indicadores demográficos para la malla estadística de Canarias 250 m de lado” se han publicado en formato CSV en el catálogo de datos abiertos del ISTAC. Se pueden consultar desde:

- Página de la operación estadística “[Explotación Estadística del Padrón Municipal](#)”.

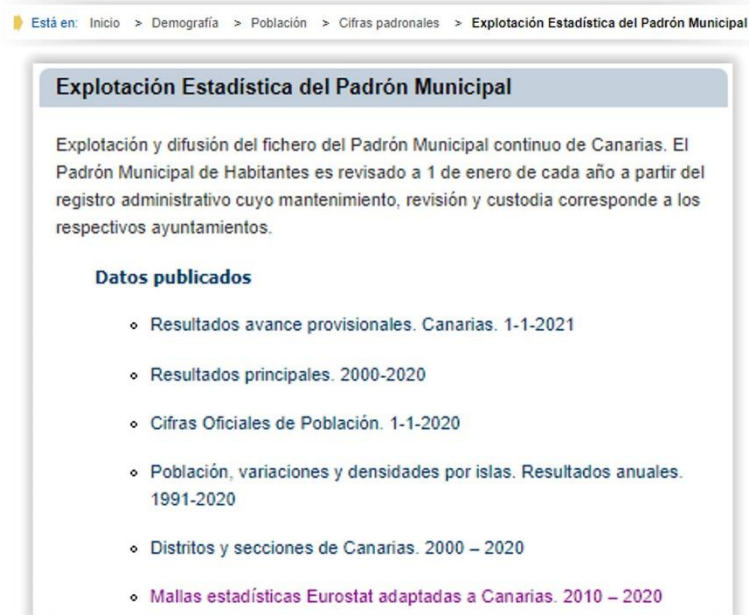


Figura 23. Página web de la Explotación Estadística del Padrón Municipal.

- Catálogo de datos abiertos: “[Indicadores demográficos para malla 250 m](#)”.

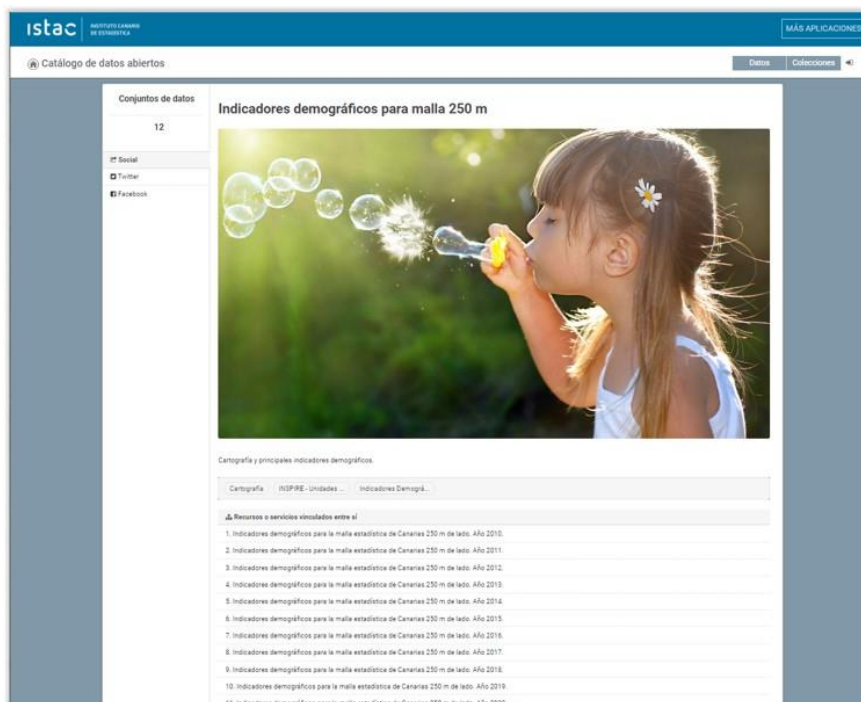


Figura 24. Página web del catálogo de datos abiertos del ISTAC..

- Catálogo de datos del [Atlas estadístico de Canarias](#):

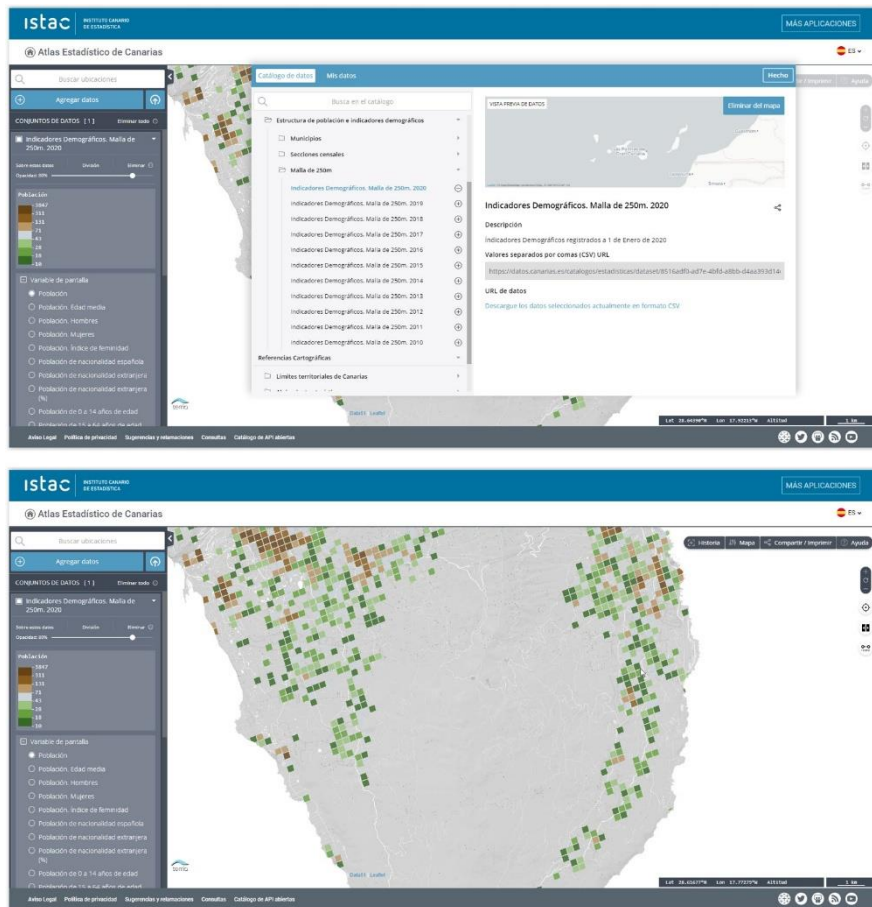


Figura 25. Página web del Atlas estadístico de Canarias.

Ejemplo de reutilización:

Desde QGIS, con el plugin [Open Data \(CKAN\) Browser](#) es posible conectar con Open Data Gobierno de Canarias, ISTAC, usando la url:

<https://datos.canarias.es/catalogos/estadisticas/api/3>



Figura 26. Complemento Open Data (CKAN) Browser.



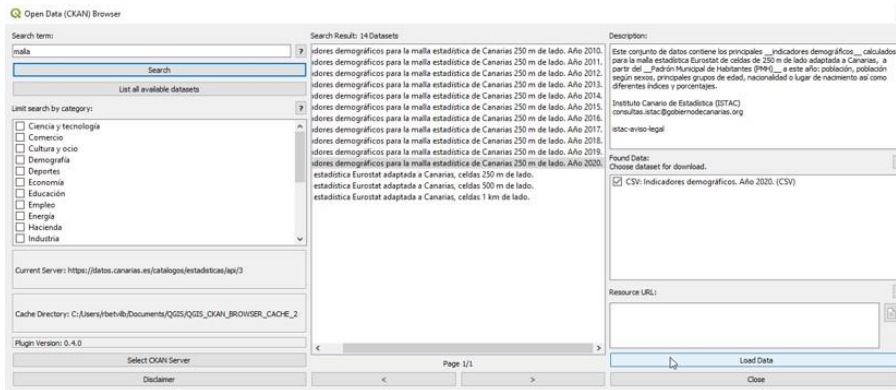


Figura 27. Conexión Open Data (CKAN) Browser al Open Data ISTAC.

## Conclusión

Debido a la demanda de datos a nivel inframunicipal se han publicado datos estadísticos demográficos sobre la malla estadística de 250 metros de lado, derivada y conforme a la malla de 1 kilómetro cuadrado definida por Eurostat, teniendo presente siempre el doble objetivo de que las cifras aportadas sean coherentes con las cifras oficiales de población y de preservar el secreto estadístico.

En un primer momento se procedió a georreferenciar el fichero RD-PMH (no anonimizado) y a su enlace con el fichero SD-PMH (anonimizado PMH), lo que permite incorporar las coordenadas de georreferenciación y la geocodificación para todas las filas del SD-PMH. Posteriormente y de cara a la publicación de resultados, se procedió a analizar el número de casos de cada grid y su posible identificación. Mediante diversos algoritmos se consiguió reasignar la población de manera que se preserve el secreto estadístico minimizando la información suprimida.

El proceso de georreferenciación y geocodificación permite aproximar mejor las estadísticas obtenidas a la realidad del territorio, e implantar decisiones de política pública más afinadas, independientes de los límites territoriales administrativos ya definidos.

Los resultados se han publicado en el Catálogo de Datos Abiertos del ISTAC, con un acceso desde la página web de la operación estadística. Además se han incorporado al Atlas Estadístico de Canarias.

## Principales referencias bibliográficas

- [1] JA González Yanes, R Betancor Villalba, E Torres Medina, M Sanz Gil, MS Hernández García. (2018, octubre). Marco de estadística espacial de Canarias. Instituto Canario de Estadística. XX JECAS 2018 La Rioja.
- [2] JA González Yanes, N Martín Morales, A Rodríguez González, DJ Lorenzo Díaz, R Betancor Villalba, E Torres Medina. (2018, octubre). Marco de desarrollo del Sistema de Datos Integrados de Canarias (iDatos). Instituto Canario de Estadística. XX JECAS 2018 La Rioja.
- [3] Reglamento de Ejecución (UE) 2018/1799 de la Comisión, de 21 de noviembre de 2018, relativo al establecimiento de una medida estadística directa temporal para la difusión de temas seleccionados del censo de población y vivienda de 2021 geocodificados en una malla de 1 km<sup>2</sup>
- [4] [Memoria técnica de la actividad. Distribución espacial de la población en Andalucía. Instituto de Estadística y Cartografía de Andalucía](#)
- [5] [Proyecto de los Censos Demográficos 2011. Instituto Nacional de Estadística](#)
- [6] [Censos de Población y Viviendas 2021. Proyecto técnico. Instituto Nacional de Estadística](#)
- [7] de Jonge, E., & de Wolf, P. P. (2016, September). Spatial smoothing and statistical disclosure control. In International Conference on Privacy in Statistical Databases (pp. 107-117). Springer, Cham.
- [8] de Wolf, P. P., & de Jonge, E. (2018, September). Safely Plotting Continuous Variables on a Map. In International Conference on Privacy in Statistical Databases (pp. 347-359). Springer, Cham.
- [9] [Suñé, E., Rovira, C., Ibáñez, D., Farré, M. \(2017\). Statistical disclosure control on visualising geocoded population data using a structure in quadtrees, NTTTS 2017](#)
- [10] [Suñé, E. \(2016\) Consideraciones sobre la utilización de quadtrees en la difusión de datos geocodificados y la preservación del secreto estadístico. JECAS 2016 Madrid](#)