



Consejería de Economía, Hacienda y
Fondos Europeos
Instituto de Estadística y Cartografía de
Andalucía

Informe Técnico



Geocodificación de las direcciones postales de la BDLPA para la generación de la Distribución Espacial de la Población en Andalucía

Actualizado a diciembre 2022





ÍNDICE

1. Introducción.....	3
2. Proceso de geocodificación con aLink. Asignación directa: enlaces.....	5
2.1 Primer proceso de enlace: Portal exacto.....	5
2.2 Segundo proceso de enlace: Portal cercano.....	9
2.3 Tercer proceso de enlace: Portal exacto sin cvia.....	11
2.4 Cuarto proceso de enlace: Portal exacto sin cun.....	12
2.5 Geocodificación al centro de vía.....	14
3. Geocodificación de la población mediante asignación indirecta.....	17
4. Distintos usos de las coordenadas (X, Y) generadas para la Distribución Espacial de la Población en Andalucía	20
ANEXO I.....	21



1. Introducción

Cada año, la generación de la distribución espacial de la población se realiza a partir de la geocodificación de las direcciones postales contenidas en la Base de Datos Longitudinal de Población de Andalucía (BDLPA) a fecha de 1 de enero del año de referencia. Estas direcciones postales localizan dónde reside cada una de los habitantes en Andalucía.

Teniendo en cuenta el volumen de datos de la BDLPA, cada año se geocodifican en torno a 2 millones y medio de direcciones postales y para realizar este proceso de geocodificación masiva se utiliza la aplicación de fusión de ficheros **aLink (software libre)**, desarrollada por el IECA, el sistema gestor de bases de datos **PostgreSQL** y su componente espacial **PostGIS**, que sirven tanto para tratamiento masivo de datos, procesos espaciales y como sistema de almacenaje de la información y dos fuentes de referencia principales, a partir de las cuales se obtienen las coordenadas X e Y que localizan a los habitantes en el territorio: el **Callejero Digital de Andalucía Unificado (CDAU)** y el **Catastro Inmobiliario** a último dato de referencia disponible en el IECA. Además, para aquellos casos en los que no se ha podido asignar una coordenada con las anteriores fuentes de referencia se han utilizado las coordenadas disponibles de la fuente **Censo de Edificios de 2011 (Censo 2011)**.



Las coordenadas de todas las fuentes utilizadas vienen dadas en el **sistema de referencia de coordenadas ETRS89 / UTM zona 30N** con código **EPSG 25830**. En el año 2020, ya se empleó de modo experimental, información cartográfica de Catastro y a partir del año 2022 se utiliza información de esta fuente procedente de los trabajos realizados en la actividad estadística y cartográfica desarrollada por el IECA y denominada **“06.03.03 - Caracterización y Distribución del Espacio Construido en Andalucía”**

Además, para el proceso de geocodificación se ha establecido un protocolo de trabajo en el que se recogen las diferentes fases del mismo y las cuales se detallan en este documento.



Antes de describir cada una de las fases es muy importante aclarar que toda la información de la BDLPA con la que se trabaja está anonimizada, es decir, cada dirección postal está identificada unívocamente mediante un código numérico, no pudiéndose identificar a través de la misma a la población localizada en dicha ubicación.



2. Proceso de geocodificación con aLink. Asignación directa: enlaces

El universo de direcciones postales a geocodificar procedente de la BDLPA lo componen las distintas escrituras identificadas de una misma dirección postal en las que reside la población andaluza desde 1996 hasta el año de referencia con el que se está trabajando.

Por otra parte, dado que desde el año 2022 se trabaja principalmente con dos fuentes de referencia distintas CDAU y Catastro Inmobiliario, se realizan dos procedimientos de geocodificación masivos completos usando ambas fuentes. Es decir, por un lado se geocodifican las distintas escrituras identificadas de una misma dirección de la BDLPA con CDAU y por otro lado con Catastro.

En ambos procedimientos, el primer proceso de geocodificación con aLink es lo más estricto posible, en el sentido de que deben coincidir de forma exacta una serie de variables, mientras que en los siguientes procesos de geocodificación se permite ser más flexible. A continuación, se describe cada uno de los procesos.

2.1 Primer proceso de enlace: Portal exacto

Este primer proceso consiste en enlazar con aLink el fichero que contiene las distintas escrituras identificadas de una misma dirección de la BDLPA y el fichero de CDAU que contiene la información alfanumérica de los portales de Andalucía con las coordenadas a cada portal.

Dado que se trabaja con un número bastante elevado de registros (más de 2 millones y medio de registros en cada fichero), la primera etapa de enlace con aLink, consiste en reducir, el número de comparaciones a realizar entre ambos ficheros, es decir, agrupar o realizar bloques de direcciones postales en ambos ficheros para comparar solo los grupos con información coincidente. En este primer proceso de enlace, como se es muy estricto, los grupos se forman obligando a que coincidan de forma estricta las variables que contienen el código INE de municipio (variable **cpm**), el código INE de núcleo de población (variable **cun**), el código INE de la vía (variable **cvia**) y el número de portal de las direcciones postales de la BDLPA y de la fuente de referencia CDAU portales (variable **numero**). Se obliga a que coincidan los números de portal porque se quieren geocodificar las direcciones a portal exacto.

Una vez que se tienen formados los grupos, **se comparan** de forma aproximada las variables que contienen



los nombres de las vías de ambos ficheros (variables **nvia**). Para ello se hace uso de la función de comparación de cadenas de caracteres **Edit-Dist** de aLink, la cual está basada en la distancia de edición o distancia de Levenshtein. En esta comparación se considera un peso de coincidencia de 1 cuando los nombres de las vías coinciden exactamente, 0 cuando no coinciden en nada y entre 0 y 1 cuando se parecen algo, siendo los nombres de vía más similares cuanto más cercanos están al valor 1 y más distintos cuanto más cercanos están al valor 0.

La comparación de los nombres de las vías devuelve automáticamente un peso de coincidencia asociado a cada par de nombres comparados. Este peso junto con dos **valores umbral (superior e inferior)** que el usuario introducirá en la aplicación aLink, son los que utiliza la misma para clasificar cada par de nombres comparados. En este caso los valores umbral indicados son: **0.000001 para el umbral inferior** y **0.000002 para el umbral superior**, de forma que todos los pares de nombres de vía comparados cuyo peso de coincidencia esté por encima del umbral superior en sentido estricto (>0.000002) se clasificaran como enlaces (es decir, son la misma dirección postal), los que estén por debajo del umbral inferior en sentido estricto (<0.000001) se clasificaran como no enlaces (es decir, no son la misma dirección postal) y los que estén entre el umbral inferior y el superior, incluyendo ambos valores, serán clasificados como posibles enlaces (podrían ser o no la misma dirección postal). El motivo por el que estos valores umbral se han establecido tan bajos en este proceso es porque en la etapa de agrupación se ha sido muy estricto, obligando a que coincidan en ambos ficheros una serie de variables entre las que se encuentra el código INE de la vía, que identifica a cada vía de forma unívoca. Por tanto, aquellos nombres de vía comparados y clasificados con un peso de coincidencia de 0 o cercano a 0, deberían corresponder a la misma dirección postal que ha cambiado su denominación o que contiene inconsistencias en los códigos de la vía.

Por último, tras la clasificación de los registros comparados en enlaces, no enlaces y posibles enlaces, aquellos pares clasificados como enlaces y que tienen un peso de coincidencia superior o igual a 0.7 se consideran automáticamente como verdaderos enlaces, es decir, se consideran que son la misma dirección postal y que por tanto las direcciones están geocodificadas a un portal exacto. El resto de registros requerirían una revisión manual, por si hubiera inconsistencias en las variables utilizadas en la etapa de agrupación de aLink. Esta revisión manual está ya también semiautomatizada reutilizando la información de las revisiones manuales realizadas de un año a otro.

Así pues, las direcciones postales obtenidas en este proceso de geocodificación y consideradas como direcciones geocodificadas a portal exacto se categorizan con el código de enlace (cod_enlace) siguiente:

- 1: Portal exacto

En el año 2022, dentro de este conjunto de direcciones enlazadas a portal exacto se ha realizado un nuevo proceso de enlace entre las direcciones de BDLPA y las de CDAU, utilizando los calificadores de numeración asociados al número de la vía, para mejorar la calidad y precisión de este primer proceso de enlace. En el caso de la BDLPA se ha tomado como conjunto de direcciones de partida, todas las direcciones



geocodificadas a portal exacto, mientras que en el caso del portalero de CDAU las direcciones que se han seleccionado son el conjunto de viales con números de vía coincidentes con los portales exactos encontrados en el primer proceso de enlace.

En este proceso ha sido necesario revisar el contenido de los campos *cnumer*, *numers*, *cnumers*, *kmt*, *nbloq* y *nport* de BDLPA y de CDAU y realizar algunas actualizaciones en los mismos para que los campos estén contruidos de forma similar y tenga sentido su comparación. A continuación, se muestran las actualizaciones realizadas en cada campo de cada fuente:

BDLPA		CDAU	
Campo	Actualización	Campo	Actualización
cnumer	Valor '-' se actualiza a null	cnumer	Eliminación de los caracteres ., - y (que aparecen en el primer carácter Eliminación del carácter) Valor '' se actualiza a null Valor 0 y 00 se actualizan a null
numers	Valor 0 se actualiza a null	numers	No se realiza ninguna actualización
cnumers	Valor '-' se actualiza a null	cnumers	No se realiza ninguna actualización
kmt	Valor 0 se actualiza a null	kmt	No se realiza ninguna actualización
nbloq	Valor '-' se actualiza a null	nbloq	Eliminación de ceros iniciales en valores numéricos del tipo 002, 0001, 0089. Los casos 00dr se mantienen tal cual Valores 0 y 00 se actualizan por null Eliminación de cadenas de caracteres 'bloque ' , 'bloques ' , 'vloque ' y 'blouqe ' en cadenas del tipo 'bloque 3', 'bloque iii', 'bloque a', etc.
nport	Valor '-' se actualiza a null	nport	Eliminación de ceros iniciales en valores numéricos



BDLPA		CDAU	
Campo	Actualización	Campo	Actualización
			Eliminación de cadena de caracteres ' portal ' en cadenas del tipo 'portal 1' Valor 0 se actualiza a null

Una vez actualizados los campos en ambas fuentes, se han llevado a cabo dos procesos de enlace nuevos entre las direcciones de BDLPA y las de CDAU :

- El primero de los procesos es el más estricto, considerando en la comparación los campos *cpm*, *cvia*, *cun*, *numero*, junto con los calificadores de numeración *cnumer*, *numers*, *cnumers*, *kmt*, *nbloq*, *nport*. Todos estos campos tienen que coincidir de forma exacta, manteniendo en la coincidencia el orden en los campos *numero*, *cnumer*, *numers*, *cnumers*, *kmt*, *nbloq*, *nport*.

Un ejemplo de esta situación, que nos daría un enlace a portal exacto más exhaustivo, sería el siguiente:

BDLPA							CDAU						
numero	cnumer	numers	cnumers	kmt	nbloq	nport	numero	cnumer	numers	cnumers	kmt	nbloq	nport
1	A				13		1	A				13	

- El segundo de los procesos de enlace ha sido menos estricto o menos exhaustivo y compara los campos *cpm*, *cvia*, *cun*, *numero*, junto con los calificadores de numeración *cnumer*, *numers*, *cnumers*, *kmt*, *nbloq*, *nport*, pero en este caso los campos *cnumer*, *numers*, *cnumers*, *kmt*, *nbloq*, *nport* no tienen en cuenta el orden, por lo que se podría decir que esta comparación, aunque también es a portal exacto, es aproximada. En este caso, los pares de registros comparados coinciden de forma exacta en los campos *cpm*, *cvia*, *cun*, *numero* y en los campos *cnumer*, *numers*, *cnumers*, *kmt*, *nbloq* y *nport* pero sin mantener el orden de estos seis últimos campos.



Unos ejemplos de esta situación de enlace, serían los siguientes:

BDLPA							CDAU						
numero	cnumer	numers	cnumers	kmt	nbloq	nport	numero	cnumer	numers	cnumers	kmt	nbloq	nport
1	A				13		1	A					13
2	B						2					B	

- Para el resto de pares de registros que coinciden de forma exacta en los campos *cpm*, *cvia*, *cun* y *numero* pero no están en las dos situaciones anteriores, y enlazan hasta número de policía se sigue manteniendo la categorización de los registros con el código de enlace 1.

Este primer proceso de enlace se realiza de forma análoga utilizando la información de Catastro pero sin realizar los dos procesos de precisión máxima. La generación de la información del código INE de vía de Catastro utilizado en este proceso se ha llevado a cabo con geoprosos espaciales para asignar los códigos a las vías de Catastro.

2.2 Segundo proceso de enlace: Portal cercano

Para aquellos registros que no se han geocodificado con aLink a un portal exacto utilizando todas las variables y parámetros del proceso anterior, se ha realizado un segundo proceso de enlace a portal cercano, considerando como portal cercano aquel que está en el mismo lado de la acera de una vía y que se distancia del portal buscado como máximo en cinco portales. Al igual que en el primer proceso de enlace el fichero de CDAU utilizado es el que contiene la información alfanumérica de los portales de Andalucía con las coordenadas a cada portal.

En este caso en la etapa de agrupación de las direcciones postales con aLink, se obliga a que coincidan las variables que contienen los códigos INE de municipio (**cpm**), el código INE del núcleo de población (**cun**), el código INE de la vía (**cvia**) y el lado de la acera en el que se encuentra el portal (variable **tacera** construida a partir del número de portal: 1 para los portales impares, 2 para los portales pares y nulo para portales sin información). Como es lógico en esta etapa no se incluye el número de portal porque se está realizando una geocodificación a portal cercano.

A continuación, tras la formación de los grupos de direcciones postales **se comparan el número de portal y**



los nombres de las vías. Para la comparación de los nombres de vía se utiliza como en el proceso de enlace anterior la función de comparación **Edit-Dist**, estableciendo un peso de coincidencia de 2 para los nombres de vía que coincidan exactamente y de 0 para los que no coincidan en nada y entre 0 y 2 para los que se parecen algo, siendo los nombres de vía más similares cuanto más cercanos están al valor 2 y más distintos cuanto más cercanos están al valor 0. Por otro lado, para comparar el número de la vía se utiliza la función de comparación de aLink: **Num-Abs** estableciendo el parámetro *diferencia máxima permitida* igual a 10 y un peso de coincidencia igual a 1 para cuando los números de portal coincidan y 0 cuando la diferencia entre números de portal sea mayor a 10 o no exista número de portal y entre 0 y 1 cuando los números de portal se diferencien 1, 2, 3, 4 o 5 portales. Se utiliza esta diferencia de 10 para obtener los enlaces con una diferencia máxima de hasta 5 portales en el mismo lado de la acera.

La comparación de los nombres de las vías y de los números de portal devuelve automáticamente un peso de coincidencia total asociado, que es la suma de los pesos de coincidencia individualizados de cada variable comparada. Como en el primer enlace, este peso total junto con los dos **valores umbral: 0.000001 para el umbral inferior y 0.000002 para el umbral superior** establecidos por el usuario en aLink, son los que utiliza la misma para clasificar cada par de registros comparados.

Tras la clasificación de los registros comparados en enlaces, no enlaces y posibles enlaces, aquellos pares clasificados como enlaces y que tienen un peso de coincidencia total superior o igual a 0.7 se consideran automáticamente como verdaderos enlaces, es decir, se consideran que son la misma dirección postal y que por tanto las direcciones están geocodificadas a un portal cercano. Y el resto de registros, al igual que en la primera fase de geocodificación, se revisarían manualmente por si hubiera erratas en las variables utilizadas en la etapa de agrupación de aLink. De igual forma esta revisión manual está semiautomatizada reutilizando la información de las revisiones manuales realizadas de un año a otro.

Así pues, las direcciones postales obtenidas en este proceso de geocodificación y consideradas como direcciones geocodificadas a portal cercano (en el mismo lado de la acera) se categorizan con el código de enlace (cod_enlace) siguiente:

- 2P1: portal inmediatamente contiguo
- 2P2: portal a una distancia de dos portales
- 2P3: portal a una distancia de tres portales
- 2P4: portal a una distancia de cuatro portales
- 2P5: portal a una distancia de cinco portales

Este segundo proceso de enlace se ha realizado en el año 2022 con el portalero de CDAU y de Catastro, en



este último caso se ha llevado a cabo un proceso de enlace a portal cercano pero sin considerar la variable código INE de la vía (**cvia**) en la fase de agrupación con aLink. Para la **comparación** de los **nombres de vía** de **BDLPA y Catastro**, se usa la función de comparación aproximada de aLink: **Edit-Dist**, estableciendo el mismo peso de coincidencia y no coincidencia que el de la segunda fase de geocodificación a portal cercano con el portalero de CDAU. En cuanto a la comparación del número de la vía se utiliza la función de comparación de aLink: **Num-Abs** estableciendo el parámetro *diferencia máxima permitida* igual a 10 (5 portales en el mismo lado de la acera) y un peso de coincidencia igual a 1 cuando los números de portal coincidan, 0 cuando la diferencia entre números de portal sea mayor a 10 o no exista número de portal y entre 0 y 1 cuando exista una diferencia de 1, 2, 3, 4 o 5 portales. En lo que se refiere a los **valores umbral** para la clasificación de los pares de registros en enlaces, no enlaces, y posibles enlaces, se han incrementado, puesto que ahora estamos comparando nombres de vías de fuentes muy distintas y no se dispone de la información del código INE de la vía (**cvia**) dentro de la etapa de agrupación de aLink. Por tanto, se están comparando todos los nombres de vía de todas las direcciones postales de los mismos núcleos de población (mismo código **cun**) sean parecidos o no. Exactamente, en este proceso de enlace se establece el valor **0.70 para el umbral inferior y 0.70 para el umbral superior**.

2.3 Tercer proceso de enlace: Portal exacto sin cvia

Tras llevar a cabo los dos procesos de enlace anteriores, siguen existiendo direcciones postales no geocodificadas ni a portal exacto ni cercano. En esta geocodificación se va a realizar de nuevo un proceso de enlace a portal exacto con aLink pero va a ser menos exhaustivo que el primer proceso de geocodificación, ya que no se va a tener en cuenta la variable código INE de la vía (**cvia**) dentro de la etapa de agrupación con aLink. El motivo por el que se toma esta decisión es porque podrían existir inconsistencias en los códigos INE de la vía que podrían estar haciendo que no se llevara a cabo el proceso de geocodificación de forma correcta. El fichero de CDAU que se utiliza en este proceso de enlace para esta tercera geocodificación es el mismo que en los dos enlaces anteriores, CDAU portales.

Por tanto, en este caso, en la etapa de agrupación de las direcciones postales con aLink, se obliga a que coincidan las variables que contienen los códigos INE de municipio (**cpm**), el código INE del núcleo de población (**cun**) y el número de portal de las direcciones postales de la BDLPA y de la fuente de referencia CDAU portales (variable **numero**). Como en el primer proceso de enlace se obliga a que coincidan los números de portal porque se quieren geocodificar las direcciones postales a portal exacto.

En cuanto a la comparación solo **se comparan los nombres de las vías** de forma aproximada utilizando la función de comparación de aLink: **Edit-Dist** y estableciendo el mismo peso de coincidencia y no coincidencia que el de la primera fase de geocodificación.

Y en lo que se refiere a los valores umbral en este caso se han subido puesto que ahora no se tiene la variable **cvia** incluida dentro de la etapa de agrupación de aLink y por tanto se están comparando todos los



nombres de todas las direcciones postales de los mismos núcleos de población (misma variable **cun**) sean parecidos o no. Exactamente, en este proceso de enlace se establece el valor **0.29 para el umbral inferior y 0.30 para el umbral superior**, y estos valores son los que se utilizan para clasificar cada par de registros comparados en enlaces, no enlaces y posibles enlaces.

En este caso, se revisan manualmente los registros clasificados como enlaces, de forma que si el fichero contiene demasiados registros solo se revisan aquellos pares de registros con peso de coincidencia igual o superior a 0.6 y si el fichero es más pequeño se revisan todos, dando siempre prioridad a aquellos pares de registros que tienen igual tipo de vía o tipos de vía similares, por ejemplo, residencial y urbanización. Las direcciones postales obtenidas en este tercer proceso de geocodificación y consideradas como direcciones geocodificadas a portal exacto pero sin tener en cuenta la variable **cvia**, se categorizan con el código de enlace (cod_enlace) siguiente:

- 3: portal exacto con cun sin cvia

Este tercer proceso de enlace se realiza de forma análoga utilizando la información de Catastro, de manera que se **comparan los nombres de las vías de la BDLPA y Catastro de forma aproximada**, usando la función de comparación de aLink: **Edit-Dist** y estableciendo el mismo peso de coincidencia y no coincidencia respecto de la primera fase de geocodificación con el portalero de CDAU. En lo que se refiere a los valores umbral para la clasificación de los pares de registros en enlaces, no enlaces, y posibles enlaces, se han incrementado, puesto que ahora estamos comparando nombres de vías de fuentes muy distintas y no se dispone de la información del código INE de la vía (cvia) dentro de la etapa de agrupación de aLink. Por tanto, se están comparando todos los nombres de las vías de todas las direcciones postales de los mismos núcleos de población (mismo código **cun**) sean parecidos o no. Exactamente, en este proceso de enlace se establece el valor **0.70 para el umbral inferior y 0.70 para el umbral superior**.

2.4 Cuarto proceso de enlace: Portal exacto sin cun

En este cuarto proceso se va a realizar de nuevo un enlace a portal exacto con aLink pero al igual que el tercer proceso de enlace va a ser menos exhaustivo que el primero, ya que no se va a tener en cuenta la variable código INE del núcleo de población (**cun**) dentro de la etapa de agrupación con aLink. El motivo por el que se toma esta decisión es porque podrían existir inconsistencias en los códigos INE de los núcleos de población o podrían existir desajustes espaciales al definirse los límites de las divisiones administrativas. Ambas situaciones podrían estar haciendo que el proceso de geocodificación no se llevara a cabo de manera adecuada. El fichero de CDAU que se utiliza en este proceso de enlace es el mismo que en los tres procesos anteriores, y el de direcciones postales de la BDLPA es el mismo fichero de partida del segundo



proceso de enlace pero eliminando los registros clasificados como enlaces en el tercer proceso.

Por tanto, en este caso, en la etapa de agrupación de las direcciones postales con aLink, se obliga a que coincidan las variables que contienen los códigos INE de municipio (**cpm**), el código INE de la vía (**cvia**) y el número de portal de las direcciones postales de la BDLP y de la fuente de referencia CDAU portales (variable **numero**). Como en el primer enlace se obliga a que coincidan los números de portal porque se quieren geocodificar las direcciones postales a portal exacto.

En cuanto a la comparación solo **se comparan los nombres de las vías** de forma aproximada utilizando la función de comparación de aLink: **Edit-Dist** y estableciendo el mismo peso de coincidencia y no coincidencia que el primer proceso de enlace.

Y en lo que se refiere a los valores umbral en este caso se establecen como los del primer proceso de enlace puesto que la variable **cvia** está incluida en la fase de agrupación realizada con aLink, por lo que estos son **0.000001 para el umbral inferior** y **0.000002 para el umbral superior**.

En este caso, tras la clasificación de los registros comparados en enlaces, no enlaces y posibles enlaces, aquellos pares clasificados como enlaces y que tienen un peso de coincidencia superior o igual a 0.7 se consideran automáticamente como verdaderos enlaces, es decir, se consideran que son la misma dirección postal y que por tanto las direcciones están geocodificadas a un portal exacto. Y el resto de registros, al igual que en el primer proceso de geocodificación, se revisarían manualmente por si hubiera inconsistencias en las variables utilizadas en la etapa de agrupación de aLink.

En este proceso de enlace se realiza además un geoproceto usando la componente espacial PostGIS de PostgreSQL. Exactamente se comparan los resultados finales obtenidos en el segundo proceso de geocodificación a portal cercano con los finales obtenidos en este cuarto proceso. El objetivo es detectar si los portales clasificados como portales cercanos en realidad lo son o si en realidad son portales exactos pero que debido a cómo están definidos los límites espaciales de las divisiones administrativas de los núcleos de población se han clasificado como portales cercanos.

Con este geoproceto se consiguen rescatar registros clasificados como portales cercanos y que en realidad son portales exactos. Por lo que a las direcciones postales obtenidas en este cuarto proceso de geocodificación y consideradas como direcciones geocodificadas a portal exacto pero sin tener en cuenta la variable **cun**, se categorizan con el código de enlace (cod_enlace) siguiente:

- 4: portal exacto sin cun con cvia

Mientras que a las direcciones postales clasificadas en un principio como geocodificadas a portal cercano pero rescatadas en este cuarto proceso como direcciones geocodificadas a portal exacto se categorizan con el código de enlace (cod_enlace) siguiente:

- 1e: portal exacto sin cun pero depurado haciendo un enlace espacial

Este cuarto proceso de enlace no se realiza utilizando el fichero de Catastro.



Llegados a este punto, estos cuatro procesos de geocodificación con aLink se engloban dentro de lo que se denomina **métodos directos de asignación de coordenadas X e Y**. Pero siguen existiendo direcciones postales de la BDLPA asociadas a población que no se han podido geocodificar mediante dichos métodos. Por tanto, hay que recurrir a otros **métodos indirectos de asignación de coordenadas X e Y**, que van a estar basados en aproximaciones espaciales y uno de ellos se basa en el siguiente proceso de geocodificación que se describe a continuación.

2.5 Geocodificación al centro de vía

Este proceso de geocodificación se realiza partiendo de aquellas direcciones postales de la BDLPA asociadas a la población y que no se han podido localizar en un portal exacto o cercano utilizando la fuente de referencia CDAU portales. Para llevarlo a cabo se enlazan dichas direcciones con el fichero de CDAU que contiene la información alfanumérica de las vías con las coordenadas puntuales al centro de la vía. Así pues, con este proceso se van a identificar las vías de CDAU en las que se ubica la población, aunque no se sepa el portal exacto o cercano donde se localizan los individuos.

En este caso la geocodificación obtenida es una geocodificación al centro de vía y va a servir de ayuda auxiliar en los métodos indirectos de asignación de coordenadas X e Y, para asignar a la población a lo largo de la vía.

Al igual que los cuatro procesos de geocodificación anteriores, este proceso de enlace se lleva a cabo utilizando aLink, de manera que en la etapa de agrupación se obliga a que coincidan las variables códigos INE de municipio (**cpm**) y el código INE de la vía (**cvia**) y **se comparan los nombres de las vías** de forma aproximada usando la función de comparación **Edit-Dist** y considerando como peso de coincidencia y no coincidencia los valores 1 y 0 respectivamente. Los valores umbral inferior y superior utilizados en este proceso para la clasificación de los pares de registros comparados son los que se indican a continuación: **0.000001 como umbral inferior y 0.000002 como umbral superior**.

Tras la clasificación de los pares de registros comparados, los que tienen peso de coincidencia superior o igual a 0.7 se dan por válidos directamente, es decir, se consideran direcciones postales geocodificadas al centro de la vía y el resto hay que revisarlos manualmente.

Hay que aclarar que en este proceso de enlace entre las direcciones de la BDLPA y las del fichero de CDAU viales no se obtienen coordenadas X e Y a un portal, sino al centroide de la vía. Luego, podría ocurrir que para vías de larga longitud como, por ejemplo, carreteras, se desvirtuaran los resultados obtenidos y se situara la población en lugares en los que realmente no lo están.



Por este motivo, no todos los viales se dan por válidos en este proceso de geocodificación y en la siguiente tabla se indican los considerados como NO válidos. El resto de viales no incluidos en esta tabla se consideran válidos para este proceso.

Tipo de vía

Autopista
Carretera
Cerro
Diseminado
Edificios
Espalda
Explanada
Extrarradio
Jardín
Loma
Parque
Partido
Playa
Polígono
Puente
Ramal
Rincón
Rinconada
Sector
Terrenos
Transversal
Trasera
Zona

Para finalizar este proceso, indicar que las direcciones postales obtenidas en este proceso de geocodificación al centro de la vía se categorizan con el código de enlace (cod_enlace) siguiente:

- 5: centro de vía

Por último, hay que indicar que estos cinco procesos de geocodificación se realizaron de la forma descrita anteriormente en el año 2016 para las fuentes de referencia CDAU portales y Censo 2011 y que a partir del mismo solo se realiza el primer proceso de geocodificación a portal exacto, con la fuente CDAU portales, para los 2 millones y medio aproximados de las distintas escrituras identificadas de una dirección contenidas en la BDLPA. El motivo es el de intentar actualizar las direcciones postales de la BDLPA geocodificadas a portales cercanos o centros de vías ya que CDAU está en continua actualización y podría darse el caso de que de un año a otro se incluyan viales o portales nuevos. Los restantes cuatro procesos de



geocodificación con CDAU portales o CDAU viales solo se llevan a cabo con las direcciones postales de la BDLPA que entran nuevas en el año de referencia. Por otra parte a partir del año 2022, para la generación de la Distribución Espacial de la Población en Andalucía 2021, se ha utilizado la fuente de referencia Catastro Inmobiliario. No obstante, para aquellas direcciones que no se han conseguido geocodificar con alguna de estas dos fuentes y se geocodificaron en años anteriores con la fuente Censo 2011, se mantienen las coordenadas de ésta última.

Porcentaje de la geocodificación de la población a **1 de enero de 2021** en Andalucía a través de métodos directos.

Tipo de enlace	Porcentaje
1. Portal exacto	94,09 %
2. Portal cercano	1,29 %
3. Portal exacto sin cvia	1,16 %
4. Portal exacto sin cun	0,30 %
Total	96,84 %

Tipo de enlace	Porcentaje
5. Centro de vía (1)	1,20 %

Esto supone un **98,04%** de población residente en Andalucía ubicadas en **coordenadas seguras**.

(1) en el siguiente apartado a través de métodos indirectos se busca el portal a lo largo de la vía



3. Geocodificación de la población mediante asignación indirecta

Una vez que aLink ha geocodificado al 96,84% de la población, quedaría un 3,16% de la población por geocodificar. aLink ha realizado el proceso de geocodificación a través de lo que se denominan métodos directos, ahora este 3,16% de la población se va a geocodificar a través de lo que se denominan métodos indirectos, es decir a través de información espacial auxiliar se van a realizar aproximaciones utilizando sistemas zonales de carácter administrativo, concretamente las secciones censales.



El objetivo de los métodos de asignación indirecta, es asignar un juego de coordenadas a esa población que no se ha podido geocodificar a través de métodos directos (3,16%). Existe población que no se ha podido ubicar de forma exacta a través de la dirección, pero sí se sabe por dónde empezar a buscar, como se ha dicho anteriormente se acota la zona de búsqueda a través de la sección censal y por medio de la propia dirección se puede además saber, que la población está agrupada en determinados edificios. A partir de aquí se realizan distintos procesos de búsquedas para encontrar los edificios dentro de la sección censal según la información espacial auxiliar de la que se dispone.



El primer proceso de búsqueda está relacionado con el enlace 5 de la asignación directa, en este enlace aunque no se encuentra portal exacto, si se dispone de la vía a la que pertenece, mediante la información auxiliar se recorre la vía buscando el mejor candidato.

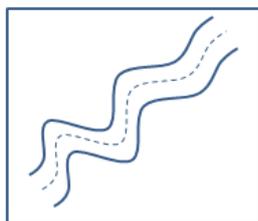
La información auxiliar **espacial** que se utiliza para el proceso de búsqueda es la siguiente:

- Cartografía Vectorial Catastral: Parcelas (polígonos)
- Secciones Censales (polígonos)
- Portalero de CDAU (puntos)
- Portalero de CENSO (puntos)

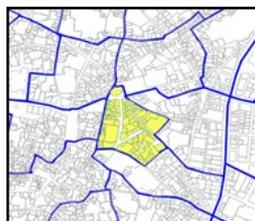
La información auxiliar **alfanumérica** que se utiliza en el proceso de búsqueda es la siguiente:

- Catastro. Registro de Inmueble. Tipo 15. Identifica cada uno de los bienes inmuebles dentro de una parcela catastral
- Base de Datos Longitudinal de Población de Andalucía
 - Creación de edificios a través de la dirección postal de la población
 - Número de huecos
 - Sección Censal
- Censo: número de huecos por edificio

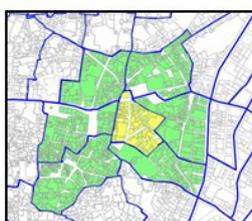
Se buscan entidades puntuales (portales) en parcelas residenciales con similares características al número de huecos, en aquellos portales que no han sido habitados, según los siguientes tipos de búsquedas:



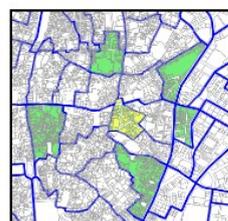
A lo largo de la vía



Dentro de la misma sección



Secciones contiguas



Sección más cercana

Los porcentajes de la geocodificación de la población a 1 de enero de 2021 a través de la asignación indirecta son los siguientes:



Tipo de búsqueda	Porcentaje
1. A lo largo de la vía	1,20%
2. Dentro de la misma sección	1,74%
3. Sección contigua	0,18%
4. Sección más cercana	0,04%

El éxito de las búsquedas depende en gran medida del número de candidatos de las entidades puntuales que se presentan en la información auxiliar espacial, en algunos casos esta información es incompleta o nula, porque se trata de zonas de difícil acceso y no se encuentran bien informadas. Para intentar solventar este déficit, se realizó en el año 2020 un primer acercamiento de la información cartográfica de Catastro, utilizándola como otra fuente de información tanto para la asignación directa con la herramienta aLink como en la asignación indirecta con nuevas entidades puntuales. A partir de estos primeros resultados experimentales obtenidos con Catastro y teniendo en cuenta los trabajos realizados en la actividad “**06.03.03 - Caracterización y Distribución del Espacio Construido en Andalucía**”, la fuente Catastro Inmobiliario se ha consolidado como una nueva fuente de referencia en los procesos de georreferenciación de la población.

El resultado de la asignación directa muestra una geocodificación del 96,84% de la población (de enlaces a portal exacto o cercano), y un 1,20% de la población geocodificada a lo largo de la vía, lo que supone un **98,04%** de población residente en Andalucía ubicadas en **coordenadas seguras**. Pero este 98% no se reparte de forma uniforme por todo el territorio, y como se ha comentado anteriormente existen zonas que pueden no encontrarse bien informadas. Esto conlleva que el porcentaje de población geocodificada de manera indirecta en algunas secciones censales sea alto y que, por lo tanto, la distribución espacial de la población resultante en esa zona presente mayor incertidumbre. En este sentido en el Anexo I se muestra el porcentaje de población geocodificada mediante asignación directa e indirecta según sección censal.



4. Distintos usos de las coordenadas (X, Y) generadas para la Distribución Espacial de la Población en Andalucía

Las coordenadas (X, Y) obtenidas a partir de la geocodificación de las direcciones de la BDLPA se utilizan para la difusión anual en formato malla de la actividad "Distribución espacial de la población en Andalucía" pero además, se usan para georreferenciar a malla o seccionado censal la información de los siguientes productos de difusión:

- [Estadísticas Longitudinales de Supervivencia y Longevidad en Andalucía 2002-2016](#)
- [Estadísticas Longitudinales de Biografías Reproductivas en Andalucía, 2002-2019](#)
- [Estadísticas de población de Andalucía basadas en registros administrativos](#)

asociados a las dos actividades llevadas a cabo por el Instituto de Estadística y Cartografía de Andalucía:

- 01.01.03 Estadísticas de itinerarios vitales de los residentes en Andalucía
- 01.01.02 Estadísticas transversales de la Base de Datos Longitudinal de Población de Andalucía

Estos datos asisten igualmente en el proceso de diseño y selección de muestras para la realización de encuestas oficiales, así como en la atención a necesidades de información específicas planteadas a través del servicio de peticiones de información por parte distintos centros directivos y organismos, por ej:

- La Agencia de Servicios Sociales y Dependencia de Andalucía (ASSDA) para mejorar los servicios de asistencia a domicilio de personas dependientes.

- Universidades y grupos de investigación que evalúan accesibilidad a infraestructuras y exposición de la población a riesgos y catástrofes



ANEXO I

Porcentaje de la geocodificación de la población a 1 de enero de 2021 en Andalucía a través del método de asignación directa e indirecta según sección censal.

La información se puede consultar en la siguiente hoja de cálculo:

https://www.juntadeandalucia.es/institutodeestadisticaycartografia/distribucionpob/seccion_censal_geo.xls