

Nuevos procesos de geocodificación de empresas y establecimientos para su publicación a través del Inventario y Caracterización de los Espacios Productivos de Andalucía

REVISTA **MAPPING**
Vol. 35, 222, 96-102
2026
ISSN: 1131-9100

New Geocoding Processes for Businesses and Establishments for Publication through the Inventory and Characterization of Productive Spaces in Andalusia

José Antonio Moreno Muñoz, José Ignacio Merchán Jiménez-Andrades

Resumen

El artículo describe el desarrollo de nuevos procesos de normalización y geocodificación de empresas y establecimientos para su integración en el proyecto ESPAND, orientado a la promoción y análisis de los espacios productivos de Andalucía. El sistema integra múltiples fuentes de información geográfica y administrativa, como el Directorio de Empresas, el Callejero Digital de Andalucía Unificado (CDAU), el Catastro y los Datos Espaciales de Referencia de Andalucía (DERA). El proceso se basa en la normalización de direcciones mediante varianteros, la herramienta aLink y revisión manual cuando es necesario. Posteriormente se aplican diferentes servicios de geocodificación (CDAU, ArcGIS, Catastro u OpenStreetMap) para obtener coordenadas fiables. El resultado mejora la localización espacial de empresas y la calidad de los datos disponibles para planificación, análisis territorial y aplicaciones de geomarketing.

Abstract

The article presents new address normalization and geocoding processes developed to publish business and establishment data within the ESPAND project, which aims to promote and analyze productive spaces in Andalusia. The system integrates several geographic and administrative datasets, including the Business and Establishment Directory, the Unified Digital Street Map of Andalusia (CDAU), cadastral data, and the Reference Spatial Data of Andalusia (DERA). Address records are first standardized through variant dictionaries, the aLink matching tool, and manual validation when necessary. After normalization, multiple geocoding sources—such as CDAU, ArcGIS, Cadastre databases, and OpenStreetMap—are used to obtain coordinates. This methodology improves spatial accuracy and data quality for territorial planning, logistics analysis, and geomarketing applications.

Palabras clave: normalización, geocodificación, ESPAND, espacios productivos, directorio de empresas.

Keywords: normalization, geocoding, ESPAND, productive spaces, business directory.

Subdirección del Área de Infraestructuras de Información. IECA
josea.moreno.munoz@juntadeandalucia.es
Servicio de Gestión de la Información. IECA
josei.merchan@juntadeandalucia.es

Recepción 15/01/2026
Aprobación 20/01/2026

1. INTRODUCCIÓN Y CONTEXTO DEL PROYECTO ESPAND

El proyecto **Inventario y Caracterización de los Espacios Productivos de Andalucía (ESPAND)**, gestionado por el Instituto de Estadística y Cartografía de Andalucía <https://www.juntadeandalucia.es/institutodeestadisticaycartografia/dega/>, ofrece mediante un visor web herramientas de navegación y filtrado por tipología de espacios productivos, parcelas libres e información del Directorio de empresas, proporcionando información individualizada tanto de las áreas industriales (incluidas infraestructuras y parcelas catastrales), como de los establecimientos ubicados en las mismas mediante una consulta interactiva. Por otra parte, la información de los establecimientos (razón social, actividad, estrato de empleo y dirección postal) puede descargarse en formato tabla para su tratamiento. Además, pueden filtrarse aquellos municipios adscritos a la Red de Ciudades Industriales de Andalucía.

ESPAND se ha concebido como una herramienta útil para la promoción de los Espacios Productivos de Andalucía ya que permite obtener información de solares vacantes (superficie y referencia catastral de los mismos), identificar las especializaciones de dichos espacios, o conocer aspectos técnicos sobre sus infraestructuras y equipamientos para planificar inversiones, además de las oportunidades que se abren en el ámbito del geo-

marketing, la logística o cualquier otra actividad pública o privada en la que se requiera la dirección postal de la empresa.

ESPAND integra diferentes infraestructuras y conjuntos de datos como son el Directorio de Empresas y Establecimientos con Actividad Económica en Andalucía, el Callejero Digital de Andalucía Unificado (CDAU), información procedente de la Dirección General del Catastro, la capa de Espacios Productivos de los Datos Espaciales de Referencia de Andalucía (DERA), el sistema de información de poblaciones de Andalucía, así como información proporcionada por las empresas de suministros como son la infraestructura eléctrica (líneas y subestaciones), gasoductos y cobertura de redes ultrarrápidas. Finalmente, la propia Infraestructura de Datos Espaciales de Andalucía (IDEAndalucía) proporciona el soporte necesario para la publicación de los servicios web de los conjuntos de datos espaciales y la implementación de la tecnología de visores.

Este artículo detalla la metodología implementada para la normalización, geocodificación, selección y control de calidad de las coordenadas obtenidas a partir de las direcciones postales del Directorio de Empresas y Establecimientos de Andalucía mediante múltiples geocodificadores. El proceso combina la normalización de direcciones postales con una asignación secuencial de calidad (desde «Portal exacto» hasta «Inicio de vía») y la aplicación de rigurosos controles adicionales que aseguran la fiabilidad espacial de los datos.

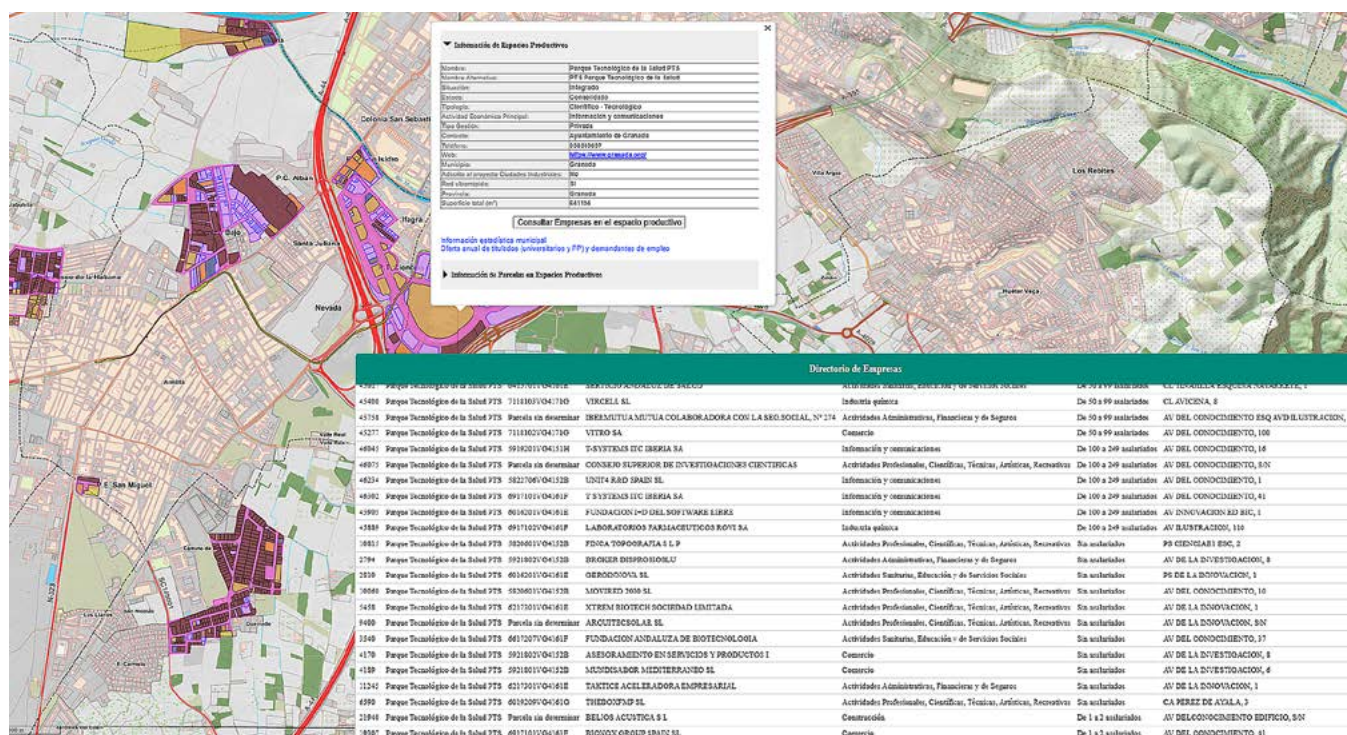


Figura 1. Visor de Espacios productivos de Andalucía. ESPAND.

<https://www.juntadeandalucia.es/institutodeestadisticaycartografia/dega/espacios-productivos-de-andalucia-espand>



Figura 2. Secuencia general de normalización y geocodificación.

2. PROCESO DE NORMALIZACIÓN Y GEOCODIFICACIÓN

El éxito del proceso de geocodificación está estrechamente ligado a la correcta normalización de las direcciones postales. El sistema desarrollado permite normalizar los registros incluidos en la base de datos para facilitar la obtención de datos de geolocalización.

2.1. Secuencia de normalización de direcciones

La normalización sigue una secuencia combinada de procesos:

1. Varianteros del IECA: se intenta localizar la dirección en los varianteros de direcciones postales del IECA. Si se encuentra, se recupera la forma normalizada.
2. Herramienta aLink: si la dirección no se encuentra en los varianteros, se recurre a la herramienta de normalización y enlace «aLink», desarrollada por el IECA (basada en el *software* libre FEBRL de la Universidad Nacional de Australia).
3. Normalización Manual: si «aLink» tampoco consigue normalizar la dirección, se recurre a la normalización manual.

Este trabajo de normalización también ha permitido

enriquecer los propios varianteros mediante la actualización de identificadores de vías existentes o la incorporación de nuevas vías.

2.2. Implementación del sistema de geocodificación

El proceso de normalización y geocodificación se ejecuta de forma automática bajo la herramienta Pentaho Data Integration (PDI).

- El proceso inicia con el trabajo *Set Variables* y la eliminación de ficheros intermedios.
- El *Trabajo Obtención datos variantero 2* implica el corte de direcciones en tipo de vía y nombre de vía para su comparación, creando tablas y vistas auxiliares.
- El *Trabajo Script* búsqueda variantero utiliza un script Python (*obtencion_datos.py*) que devuelve registros encontrados y no encontrados (*datos.csv*).

2.3. Fuentes de geocodificación empleadas

Una vez normalizadas, las direcciones se geocodifican utilizando diversas fuentes:

Para la geocodificación con ArcGIS y Catastro, se realiza un tratamiento específico, incluyendo el mapeo del tipo de vía y municipio, para adecuar la cadena de texto a los requisitos de cada geocodificador, mejorando el porcentaje de éxito.

Fuente	Tipo de Búsqueda	Normalización Requerida
CDAU (Callejero Digital de Andalucía Unificado)	Directa a BBDD (vía id_vial)	Dirección normalizada por variantero y aLink
Geocoder	Servicio web (vía CDAU)	Dirección normalizada (si falla, dirección original)
ArcGis	Servicio web	Tipo y nombre de vía normalizados y no normalizados
Catastro	Base de datos (requiere número de vía)	Tipo de vía normalizado, nombre de vía combinado (normalizado o original)
OSM Dirección	Servicio web	Dirección normalizada (si falla, dirección original)
OSM Razón Social	Servicio web	Búsqueda por razón social y entidad (solo para establecimientos sin coordenada previa)

3. CARACTERÍSTICAS DE LOS GEOCODIFICADORES Y VARIABLES DE CALIDAD

Para cada geocodificador utilizado, se identifican las características que permiten medir la calidad de la georreferenciación, especificando la información utilizada para la localización.

3.1. CDAU (Callejero Digital de Andalucía Unificado)

Esta fuente busca el id_vial a partir de las direcciones normalizadas por variantero y aLink. En CDAU, solo se han considerado aquellas direcciones que van a portal exacto, por lo que los datos obtenidos de esta fuente tienen la máxima calidad esperada («Portal exacto»).

3.2. GEOCODER

La búsqueda se realiza a partir de direcciones normalizadas y, si no se encuentra localización, a partir de direcciones sin normalizar (con normalización de los tipos de vías), todas del CDAU. Las variables clave para la calidad son:

- Tipo de respuesta (res_tipo_geocoder): puede ser EXACT (campos coinciden con los proporcionados), PartialMatch (no coinciden en alguno de sus campos) o NoMatch (no encuentra nombre de vía que supere el umbral).
- Grado de similitud (similitud_geocoder): toma valores de 0 a 1.
- Nivel respuesta (nivel_geocoder): incluye StreetNumber, StreetName, StreetType, Unknown.

En búsquedas sin número de portal, si la vía tiene portales, devuelve la vía con portal central (coordenadas del centro de la vía). Si no existen vías con nombre exacto, se aplica un algoritmo de similitud. Si el algoritmo no encuentra un nombre que supere el umbral, la salida es No match.

3.3. ArcGIS

La búsqueda se realiza de manera similar a Geocoder, usando direcciones normalizadas y sin normalizar. Las variables clave de calidad son:

- Grado de similitud (similitud_arcgis): toma valores de 0 a 100.
- Nivel respuesta (nivel_arcgis):

PointAddress: dirección de calle basada en puntos que representan la ubicación de edificios; es el nivel de coincidencia más preciso espacialmente (representa la azotea de un edificio).

- StreetAddress: el número de casa se interpola a partir de un rango de números.
- StreetAddressExt: coincidencia interpolada donde el número introducido supera el rango.
- StreetName: similar a StreetAddress pero sin el número

de casa. Devuelve el centro de la vía.

En búsquedas sin número de portal, si la vía existe, las coordenadas corresponden al primer punto del primer tramo de la vía (nivel_arcgis='StreetName').

3.4. Catastro

Las búsquedas se realizan solo para aquellos casos en los que se disponga del número de vía. La normalización es necesaria para los tipos de vías y nombres de municipios. Las búsquedas exitosas proporcionan la referencia catastral de la parcela (parcela_catastro), de la cual se obtiene el centroide, resultando en coordenadas exactas.

3.5. OpenStreetMap (OSM) Dirección y Razón Social

OSM Dirección realiza la búsqueda a partir de direcciones (normalizadas o sin normalizar). Para la búsqueda sin número de portal, se obtienen las coordenadas correspondientes al inicio de vía. Si se encuentra vía y número de vía, se considera portal exacto.

OSM Razón Social realiza la búsqueda a partir de la razón social y la entidad del establecimiento. Esta búsqueda solo se utiliza para aquellos establecimientos que no han conseguido coordenada en ninguna otra fuente.

4. SELECCIÓN DEL GEOCODIFICADOR Y ASIGNACIÓN DE CALIDAD DE LA GEOCODIFICACIÓN

La selección de coordenadas es secuencial, de mayor calidad a menor calidad, y dentro de cada nivel de calidad, siguiendo un orden de prioridad entre las fuentes. Si un registro recibe una coordenada en un paso, es descartado para los siguientes.

4.1. Control de calidad inicial (Requisito Excluyente)

Antes de asignar la calidad, se realiza un control inicial:

- Se descartan aquellas coordenadas que no estén localizadas dentro del municipio que indica la información del establecimiento. Este chequeo es excluyente.
- Se obtiene el código postal donde se localizan las coordenadas para chequeo posterior.

4.2. Asignación del nivel de calidad (Orden de prioridad)

El algoritmo de selección determina la coordenada óptima. El orden de calidad (de mayor a menor) es: Portal Exacto, Portal Cercano, Centro Vía e Inicio de Vía.

4.2.1. Portal exacto (Máxima calidad)

Las fuentes utilizadas en esta primera fase son: CDAU, Geocoder, ArcGIS, Catastro, OSM Dirección y OSM Razón Social.

Nota: CDAU, Catastro y OSM Razón Social se utilizan únicamente en este primer paso, obteniendo de ellas solo coordenadas de «Portal exacto».

Geocodificador	Condiciones de Asignación de "Portal Exacto"
CDAU	Siempre se selecciona como «Portal exacto».
GEOCODER	Tipo de respuesta: 'EXACT' o 'PartialMatch'. Grado de similitud > 0,74. Número de dirección a geocodificar igual al número de vía devuelto por Geocoder, y mayor que 0.
ARCGIS	Tipo de respuesta: 'PointAddress', 'StreetAddress' o 'StreetAddressExt'. Grado de similitud > 86. Número de dirección a geocodificar igual al número de vía devuelto por ArcGis, y mayor que 0.
CATASTRO	Se selecciona como «Portal exacto» al proporcionar el centroide de la parcela catastral.
OSM Dirección	Código postal del establecimiento coincide con el CP de las coordenadas. Número de dirección a geocodificar mayor que 0 e igual al número de vía devuelto por OSM.
OSM Razón Social	Código postal del establecimiento coincide con el CP de las coordenadas.

4.2.2. Portal cercano (Segundo nivel de calidad)

Solo se utilizan Geocoder y ArcGIS en este nivel.

- GEOCODER: tipo de respuesta 'PartialMatch'. Grado de similitud > 0,79. La diferencia absoluta entre el número de dirección a geocodificar y el número de vía devuelto por Geocoder debe ser menor que 6.
- ARCGIS: tipo de respuesta 'PointAddress', 'StreetAddress' o 'StreetAddressExt'. Grado de similitud > 80. La diferencia absoluta entre el número de dirección a geocodificar y el número de vía devuelto por ArcGis debe ser menor que 6.

4.2.3. Centro vía (Tercer nivel de calidad)

Solo se utilizan Geocoder y ArcGIS en este nivel.

- GEOCODER: tipo de respuesta 'EXACT' (similitud > 0,69) o 'PartialMatch' (similitud > 0,64). Número de dirección a geocodificar y número de vía devuelto mayores que 0. No clasificado previamente como «Portal exacto» ni «Portal cercano».
- ARCGIS: tipo de respuesta 'StreetName'. Grado de similitud > 82. Número de dirección a geocodificar y número de vía devuelto mayores que 0. No clasificado previamente como «Portal exacto» ni «Portal cercano».

4.2.4. Inicio de vía (Cuarto nivel de calidad)

Se utilizan Geocoder, ArcGIS y OSM Dirección.

- GEOCODER: se cumplen los criterios de similitud y tipo de respuesta del Centro Vía, pero el número de dirección a geocodificar es igual a 0 o nulo, o el número de vía devuelto por Geocoder es igual a 0.

- ARCGIS: tipo de respuesta 'StreetName'. Grado de similitud > 82. El número de dirección a geocodificar es igual a 0 o nulo, y el número de vía devuelto por ArcGis es nulo.
- OSM Dirección: Se considera «Inicio de vía» si geocodifica y no ha sido clasificado previamente como 'Portal exacto'.

5. CONTROLES ADICIONALES DE CALIDAD Y CRITERIOS DE DESCARTE

Una vez seleccionado el geocodificador y asignada su calidad, se implementan controles adicionales para comparar fuentes y descartar geolocalizaciones que no cumplan requisitos estrictos.

5.1. Comparación Inter-Geocodificadores

Se comparan aquellos geolocalizadores que tienen la misma calidad que el geocodificador seleccionado, midiendo la distancia entre ellos. Las distancias se clasifican en rangos: menor a 1 metro, de 1 metro a < 5 metros, de 5 metros a < 10 metros, de 10 metros a < 15 metros, y de 15 metros a < 20 metros.

Los resultados se documentan en la tabla de establecimientos mediante dos campos:

- geocodificadores: array que contiene el geocodificador seleccionado y todos aquellos que cumplen con la misma calidad (calidad_g).

- tipo_distancia: indica la clasificación de distancia de la geolocalización del conjunto de geocodificadores.

5.2. Control de calidad para descartar geolocalizaciones (Requisitos excluyentes)

Existen tres requisitos clave que la geolocalización de los establecimientos debe cumplir:

1. No deben estar ubicados en organismos públicos.
2. No deben estar en inmuebles con uso catastral de «Solar» si el establecimiento tiene actividad económica.
3. El número de establecimientos en una parcela no debe superar al número de construcciones susceptibles de albergar actividad económica («Establecimientos Apilados»).

5.2.1. Distancia a organismos públicos

Se mide la distancia entre la geocodificación seleccionada y la ubicación de los organismos públicos de la Junta de Andalucía (Tabla tablaEstructuraOrganica).

- Un establecimiento se considera ubicado en un organismo público cuando la distancia es menor a 10 metros.
- Si esta condición se cumple, la geocodificación es descartada (Control Excluyente).
- El resultado se documenta en el campo booleano en_estru_organica.

5.2.2. Chequeo con Catastro

Se realizan varios controles utilizando la información catastral:

- Ubicación en Parcela Productiva: se comprueba si el establecimiento se ubica dentro de una parcela catastral

considerada espacio productivo o a una distancia inferior a 10 metros. Este chequeo es excluyente. Los resultados se documentan en refcat (referencia catastral) y dist_pc (distancia a la parcela).

- Uso Solar: se comprueba si la parcela asignada es un solar y si el establecimiento tiene actividad económica. Este chequeo es excluyente. La información se documenta en el campo pc_es_solar.
- Bienes Inmuebles: se migra de la parcela catastral el número de bienes inmuebles emplazados (pc_num_bi).

5.2.3. Chequeo de establecimientos apilados

Este control identifica establecimientos que comparten las mismas coordenadas, lo que se considera un apilamiento.

1. Se considera que dos establecimientos están en el mismo lugar si tienen las mismas coordenadas, siempre que la geocodificación sea «Portal exacto».
2. Se cuenta el número de establecimientos que cumplen esta condición (n_apilados).
3. El establecimiento se considera apilado (y por tanto excluyente) si cumple una de las siguientes condiciones:
 - El número de apilados en el establecimiento (n_apilados) es mayor que el número de bienes inmuebles de la parcela catastral (pc_num_bi).
 - No se ha encontrado parcela catastral (pc_num_bi es nulo) y el número de apilados es mayor que 1.

5.3. Resumen de controles de calidad

La siguiente tabla resume los controles de calidad aplicados y su implicación en el descarte de la geocodificación obtenida:

Control de Calidad	Tipo de Búsqueda	Normalización Requerida
Geocodificación ubicada en el municipio informado	VERDADERO	Excluyente
Geocodificación ubicada en el código postal informado	VERDADERO	No excluyente
Geocodificación ubicada a una distancia menor de 10 metros de un organismo público	FALSO	Excluyente
Geocodificación ubicada a una distancia menor de 10 metros de una parcela catastral de un espacio productivo	VERDADERO	Excluyente
Geocodificación ubicada a una distancia menor de 10 metros de una parcela catastral tipo solar, teniendo el establecimiento actividad económica	VERDADERO	Excluyente
Geocodificación considerada dentro de un conjunto de establecimientos apilados	VERDADERO	Excluyente

6. REQUERIMIENTOS TÉCNICOS PARA LA EJECUCIÓN DE CONTROLES DE CALIDAD (ESQUEMAS DE DATOS)

Los controles de calidad se implementan mediante consultas SQL sobre la Base de Datos (BD) que alberga la tabla a normalizar y georreferenciar, implicando tablas con información alfanumérica y espacial. Las tablas necesarias se ubican principalmente en los esquemas `esquemaEspand` y `esquemaCatastro`.

Esquema `esquemaEspand` (Geometrías y Variables Administrativas):

- Tabla `municipios`: contiene los municipios de Andalucía con sus geometrías. Campos obligatorios incluyen `codprov`, `inemun`, `nom_municipio`, `municipio_catastro` y `geom` (MultiPolygon, 25830).
- Tabla `municipios_cambios`: contiene la correspondencia del código municipal antiguo y nuevo (para municipios desagregados o con código cambiado).
- Tabla `codigo_postal`: contiene los códigos postales de Andalucía y sus geometrías (MultiPolygon, 25830).
- Tabla `tablaEstructuraOrganica`: contiene la localización (geometría de punto) de los organismos públicos de la Junta de Andalucía, utilizada para el control de distancia a estructuras orgánicas.

Esquema `esquemaCatastro` (Información Catastral):

El usuario `espan` debe tener permisos de lectura sobre las tablas consultadas en el esquema catastro.

- Tabla `tablaParcelasUnificadas`: parcelas catastrales identificadas como espacios productivos de Andalucía con sus

geometrías. Contiene `refcat` y `cdmunicipio`.

- Tablas de Solares - Parcelas Catastrales: utiliza `tablaParcelasBimR14` (contiene información detallada sobre construcciones y destinos, incluyendo `numero_bi` y `numero_cons`) y `t_num_r15_par` (conteniendo `numero_bi`).
- Tabla `tablaParcelasNumBi`: parcelas catastrales identificando el número de bienes inmuebles (`numero_bi`) y la superficie destinada a cada uso.

7. CONCLUSIÓN

La metodología de geocodificación implementada para el proyecto ESPAND se distingue por su enfoque integral y jerárquico, priorizando la calidad geométrica de la georreferenciación. Al emplear un algoritmo de selección secuencial que valora la proximidad al portal (Portal Exacto > Portal Cercano) y la precisión de la fuente (CDAU > Geocoder > ArcGIS > Catastro > OSM), se asegura que la coordenada final asignada sea la óptima disponible. La precisión del proceso es reforzada por una batería de controles de calidad de carácter excluyente (municipio, organismos públicos, uso solar y apilamiento), que mitigan errores espaciales inherentes a la normalización de direcciones postales administrativas. Los resultados obtenidos han sido muy positivos, permitiendo una geocodificación amplia y precisa de las empresas y establecimientos, fundamental para la consulta interactiva y el análisis espacial del tejido productivo andaluz.

El proceso de selección de coordenadas actúa como un filtro de alta resolución, asegurando que solo las localizaciones que superan múltiples capas de verificación (similar a un conjunto de tamices de diferente grosor) sean utilizadas para la caracterización territorial, garantizando así la exactitud de la información publicada por el IECA.

Sobre los autores

José Antonio Moreno Muñoz

Experto en estadística y sistemas de información geográfica. Subdirector del Área de Infraestructuras de Información del Instituto de Estadística y Cartografía de Andalucía (IECA). Licenciado en Ciencias y Técnicas Estadísticas (Universidad de Sevilla) y diplomado en Estadística (Universidad de Granada). Ha liderado proyectos como BADEA y el Callejero Digital de Andalucía, además de desarrollar bases de datos espaciales y marcos geográficos históricos. Ha sido docente, publicado artículos científicos y participado en congresos internacionales.

José Ignacio Merchán Jiménez-Andrades

Es geógrafo y Jefe de Servicio de Gestión de la Información del Instituto de Estadística y Cartografía de Andalucía. Licenciado en Geografía e Historia por la Universidad de Sevilla y Experto en Evaluación de Impactos Ambientales del Planeamiento Urbanístico. Ha trabajado en las Consejerías de Cultura, Urbanismo y Ordenación del Territorio y Turismo. Desde 2012 trabaja en el IECA, donde coordina el Callejero Digital de Andalucía Unificado y es responsable del Banco de Datos Estadísticos de Andalucía.