

# GUÍA DE GEOCODIFICACIÓN DE FUENTES DE INFORMACIÓN ADMINISTRATIVA



Versión 1.0.  
30/04/2015



## ÍNDICE

|  |                  |
|--|------------------|
| <b>1. INTRODUCCIÓN .....</b>   | <b>4</b>         |
| <b>2. NORMALIZACIÓN.....</b>   | <b>7</b>         |
| <b><i>2.1 RECOGIDA DE INFORMACIÓN GEOGRÁFICA .....</i></b>                             | <b><i>11</i></b> |
| 2.1.1 Dirección postal.....  | 12               |
| 2.1.2 Referencia catastral.....  | 19               |
| <b><i>2.2 HERRAMIENTAS DE NORMALIZACIÓN DE DIRECCIONES .....</i></b>                   | <b><i>21</i></b> |
| 2.2.1 aLink: Herramienta de Normalización .....  | 23               |
| 2.2.2 NorDir.....  | 27               |
| <b>3. GEOCODIFICACIÓN .....</b>  | <b>30</b>        |
| <b><i>3.1 SISTEMAS DE REFERENCIA DE COORDENADAS Y COORDENADAS GEOGRÁFICAS.....</i></b> | <b><i>31</i></b> |
| 3.1.1 Sistemas de referencia.....  | 31               |
| 3.1.2 Transformación de coordenadas.....   | 34               |
| 3.1.3 Expresión de las coordenadas y precisión .....                                   | 36               |
| <b><i>3.2 BASES CARTOGRÁFICAS E INFORMACIÓN DE REFERENCIA.....</i></b>                 | <b><i>38</i></b> |
| 3.2.1 Bases cartográficas .....  | 38               |
| 3.2.2 Información de referencia.....   | 39               |
| <b><i>3.3 PROCESO DE GEOCODIFICACIÓN.....</i></b>                                      | <b><i>43</i></b> |
| <b><i>3.4 HERRAMIENTAS PARA GEOCODIFICAR.....</i></b>                                  | <b><i>45</i></b> |
| 3.4.1 aLink: Herramienta de Enlace.....  | 45               |
| 3.4.2 GeoDir y Geocoder.....   | 47               |
| 3.4.3 Gestores de Bases de datos .....   | 49               |
| 3.4.4 Herramientas SIG.....  | 50               |
| 3.4.5 Telegeo.....   | 52               |
| <b>4. CONTROL DE CALIDAD DE LA INFORMACIÓN GEORREFERENCIADA .....</b>                  | <b>55</b>        |
| <b>5. TRATAMIENTO DE LA INFORMACIÓN GEOCODIFICADA.....</b>                             | <b>58</b>        |



|  |           |
|--|-----------|
| <b>5.1 Sistemas de Información Geográfica (SIG).....</b>   | <b>58</b> |
| 5.1.1 QGIS .....   | 58        |
| 5.1.2 ARCGIS .....   | 59        |
| 5.1.3 GVSIG .....  | 60        |
| 5.1.4 KOSMO .....  | 61        |
| <b>5.2 BASES DE DATOS ESPACIALES .....</b>   | <b>62</b> |
| <b>5.3 POSIBILIDADES ANALÍTICAS QUE OFRECE LA INFORMACIÓN GEOCODIFICADA .....</b>                      | <b>63</b> |
| 5.3.1 Tabla de atributos y selección de elementos.....   | 64        |
| 5.3.2 Obtención de información a partir de relaciones espaciales entre capas (Análisis espacial) ..... | 65        |
| <b>6. DIFUSIÓN DE LA INFORMACIÓN.....</b>  | <b>71</b> |
| <b>6.1 PUBLICACIÓN DE DATOS ESPACIALES.....</b>  | <b>71</b> |
| <b>6.2 SERVIDORES DE MAPAS .....</b>   | <b>72</b> |
| <b>6.3 SERVICIOS .....</b>   | <b>75</b> |
| 6.3.1 Visores de Servicios.....  | 76        |
| <b>ÍNDICES DE ESQUEMAS, CUADROS Y FIGURAS.....</b>   | <b>81</b> |
| <b>ESQUEMAS.....</b>   | <b>81</b> |
| <b>CUADROS.....</b>  | <b>81</b> |
| <b>FIGURAS.....</b>  | <b>81</b> |

## 1. INTRODUCCIÓN

La estadística y la cartografía tienen un papel insustituible en la consolidación de la sociedad de la información, proporcionando una información que resulta imprescindible para la Administración Pública, los agentes económicos y sociales y para la ciudadanía en general. En este sentido, la disponibilidad de datos estadísticos y cartográficos se encuentra entre las necesidades esenciales en la nueva sociedad del conocimiento. La integración de los datos estadísticos con los espaciales refuerza además el valor de ambos, enriqueciéndolos mutuamente y abriendo nuevas posibilidades de utilización, de tal forma que la estadística demanda la territorialización de la información, al tiempo que la cartografía ha ampliado su alcance hacia el nuevo concepto de datos espaciales, que implica la incorporación de información georreferenciada de origen estadístico.

El Plan Estadístico y Cartográfico de Andalucía 2013-2017 (aprobado por la Ley 3/2013, de 24 de julio) es el instrumento de ordenación y planificación de la actividad estadística y cartográfica de la Comunidad Autónoma para sus propios fines. Asimismo, contribuye a reforzar una estrategia de confluencia entre los datos cartográficos y los estadísticos mediante un tratamiento conjunto de ambos tipos de información, con la finalidad de conseguir que la cartografía se alimente de fuentes estadísticas y que las estadísticas avancen en su georreferenciación.

Además, este Plan 2013-2017 define como estrategias esenciales para la consecución de sus objetivos el aprovechamiento de las fuentes, registros e infraestructuras de información, la normalización y garantía de la calidad y la difusión, así como el acceso y reutilización de la información.

El Título IV del Plan establece que las fuentes, registros administrativos y sistemas de información de la Junta de Andalucía serán utilizados en la producción de datos estadísticos y cartográficos, para lo cual se garantizará la explotación e integración de sus datos a dichos efectos y se promocionará su uso por el resto de entidades y órganos del Sistema Estadístico y Cartográfico de Andalucía.

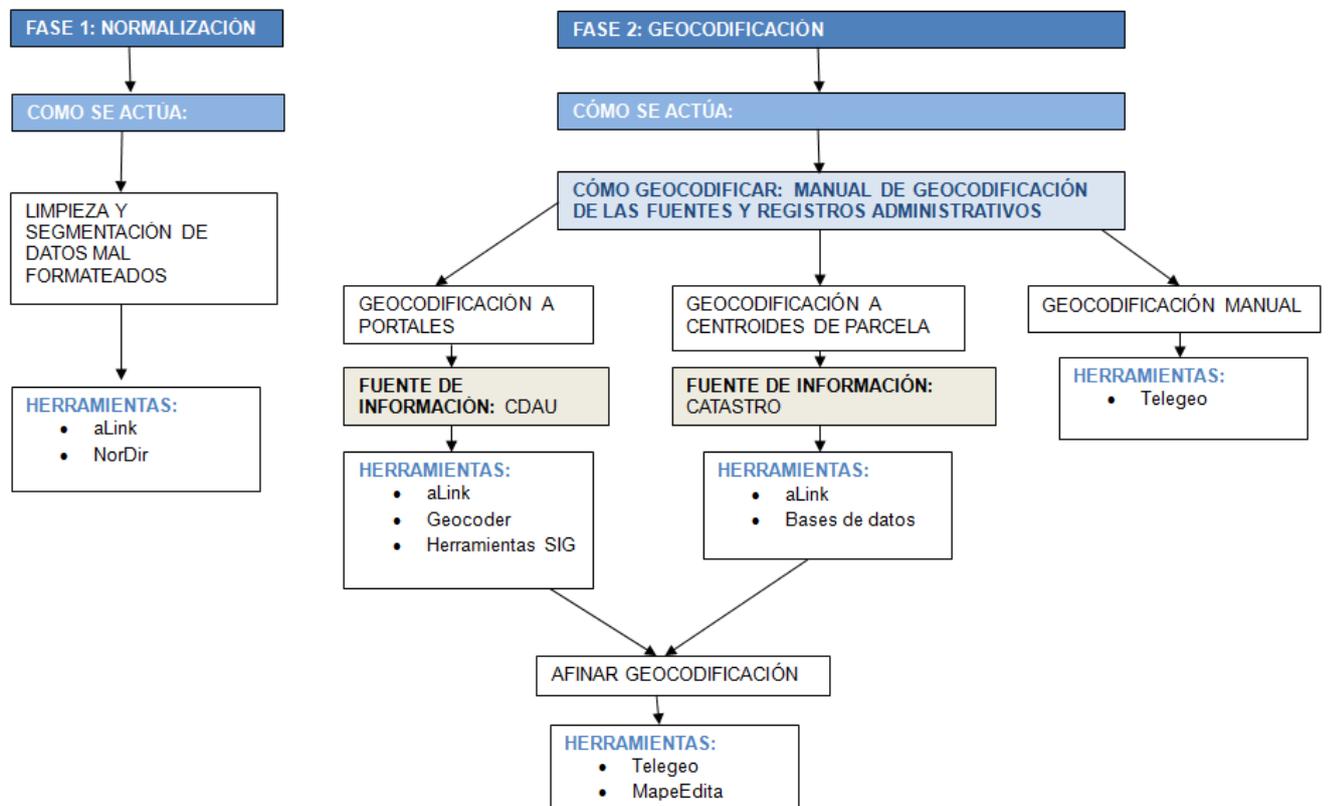
Por otra parte, el Instituto de Estadística y Cartografía de Andalucía (IECA) dispone de un *Inventario de Fuentes de información administrativa de Andalucía*, que ofrece las características de las fuentes administrativas que son competencia de la Administración Autonómica Andaluza. Más de un 90% de las fuentes que actualmente componen el Inventario disponen del domicilio entre los atributos recogidos en la misma.

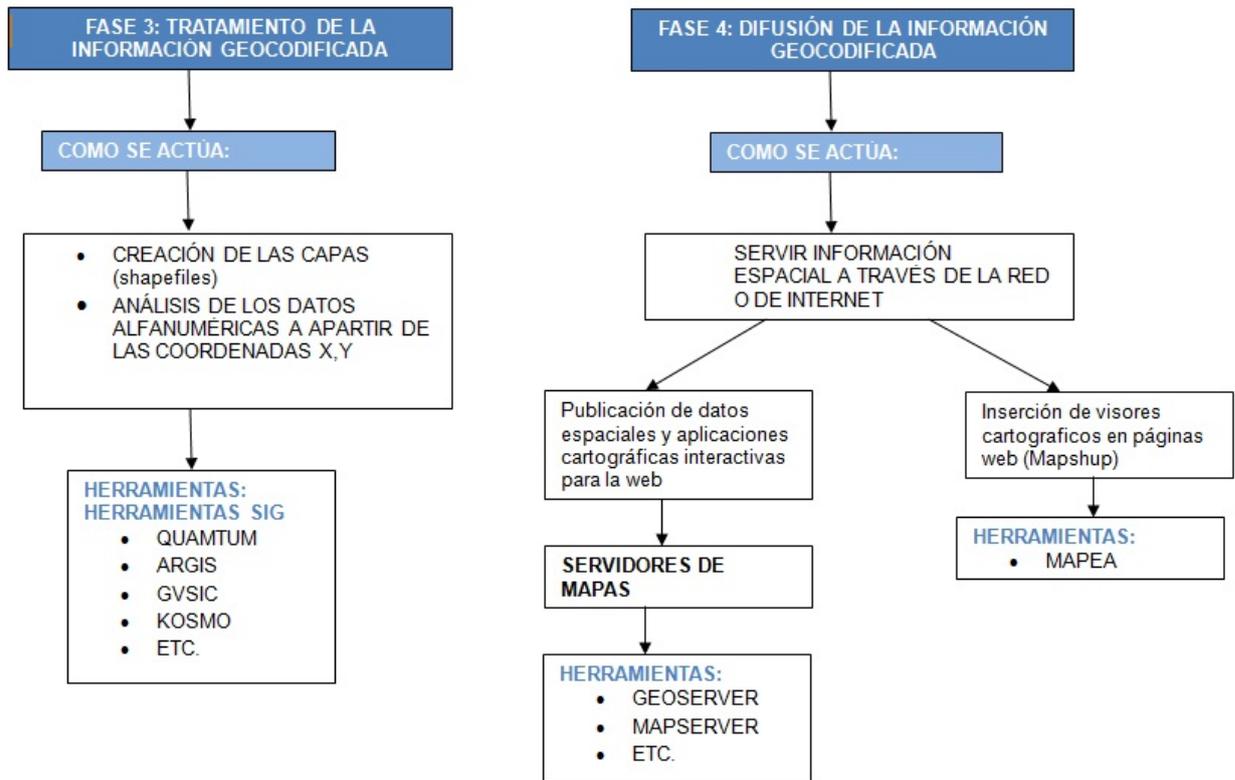
Este documento tiene como objetivo orientar en la geocodificación de fuentes de información administrativa a partir de la dirección postal o, en su caso, de la referencia catastral. El éxito de la geocodificación depende en gran medida de la estructura y de la calidad de la información que

compone la dirección postal, por lo que es fundamental que ésta esté bien estandarizada y normalizada.

En la guía se recogen los principales aspectos para realizar un proceso de geocodificación de fuentes de información desde la recogida hasta la difusión de la información. En concreto, se detallan las fases y herramientas que permiten realizar procesos de normalización y geocodificación, además de la información cartográfica de referencia necesaria para hacerlo. También se señalan los procedimientos de control de calidad y se expone la forma de visualizar y tratar la información geocodificada.

En el esquema 1 se señalan las fases, fuentes y herramientas de actuación para el proceso de geocodificación que se desarrollan en esta guía.





Esquema 1.- Proceso de geocodificación de fuentes de información: fases, fuentes y herramientas

## 2. NORMALIZACIÓN

En el actual contexto socioeconómico es esencial incidir en la mejora de la eficacia y el ahorro en los costes de la gestión. En este sentido, y en materia estadística y cartográfica, la utilización de registros y fuentes de información administrativa para el desarrollo de actividades estadísticas y cartográficas permite cubrir importantes lagunas de información a bajo costo y, por otra parte, sustituir o complementar la recogida de datos por muestreo, con su consiguiente ahorro.

En relación con estos registros y fuentes de información administrativa no debe perderse de vista que las mismas se crean para fines de gestión, pero si se consiguen introducir criterios normalizadores y buenas prácticas en el procedimiento regulador de la norma que lo crea o modifica, así como en el mantenimiento y actualización del sistema de información que contiene los datos recogidos, se podrá disponer de información mucho más fiable, comparable e integrada, con datos menos redundantes y previamente depurados y con alto potencial para su aprovechamiento estadístico y cartográfico.

En este sentido, el IECA presenta entre sus funciones, según el artículo 30 h) de la Ley 4/1989, de 12 de diciembre, de Estadística de la Comunidad Autónoma de Andalucía, la de “informar preceptivamente los proyectos de normas por las que se creen, modifiquen o supriman registros administrativos en lo relativo a su aprovechamiento estadístico”. Hasta la fecha se han informado más de 100 proyectos de normas en los que se ha propuesto la adopción de criterios normalizadores en relación con variables como el sexo, la fecha de nacimiento, la nacionalidad, los distintos componentes de la dirección y la utilización de clasificaciones estadísticas oficiales.

De igual forma, las Unidades Estadísticas y Cartográficas de las Consejerías y Agencias dependientes tienen entre sus funciones, según el artículo 35 c) de la citada Ley 4/1989, de 12 de diciembre, la de “participar en el diseño y, en su caso, en la implantación de registros o ficheros de información administrativa que sean susceptibles de posterior tratamiento estadístico, velando de manera especial por la compatibilidad de las clasificaciones utilizadas en aquellos con las clasificaciones estadísticas de uso obligatorio, así como organizar la incorporación de información de origen administrativo a la actividad estadística, garantizando la eficiencia, la integridad de su contenido y el respeto al secreto estadístico”.

Por su parte, el Plan Estadístico y Cartográfico de Andalucía 2013-2017 en su Título IV dedicado a fuentes, registros e infraestructuras de información establece que con la finalidad de asegurar la comparabilidad y facilitar la integración de las fuentes, los registros administrativos y los sistemas de información, el IECA elaborará y publicará reglas para la normalización en la codificación de variables, siguiendo estándares que están definidos de manera nacional o internacional y de acuerdo con el

código de buenas prácticas de las estadísticas europeas, los Reglamentos de desarrollo de la Directiva 2007/2/EC y el Esquema Nacional de Interoperabilidad.

En relación con esta función, el IECA propone reglas sobre la mayor parte de las “variables sociales nucleares” (denominadas así en Europa), elegidas por:

- Su relevancia y el uso potencial de la información recogida en las variables.
- La simplicidad de sus definiciones para la recogida de la información.
- La viabilidad de su aplicación.
- La armonización en el input que permita la recogida de la información de forma normalizada en los registros y otras fuentes de información administrativa.
- La utilización de definiciones ya existentes y suficientemente conocidas.
- El uso de estándares internacionales.

En estas variables se puede distinguir entre:

- Las que describen características sociodemográficas: sexo, edad, país de nacimiento, nacionalidad, estado civil y composición del hogar.
- Las que aportan información geográfica para la localización: país, región y provincia, localidad, dirección y coordenadas geográficas.
- Las que aportan información socioeconómica: situación laboral, situación profesional, ocupación, sector de actividad y nivel más alto de estudios terminados.

Se dispone de un manual para promover la normalización en la recogida de la información administrativa, elaborado en desarrollo de esta función encomendada al IECA, que puede ser muy útil a las Unidades Estadísticas y Cartográficas en sus funciones de participación en el diseño e implantación de registros o ficheros de información administrativa. Este manual se denomina: “Manual de buenas prácticas para la normalización de fuentes y registros administrativos de la Junta de Andalucía” y se puede descargar desde la URL:

<http://www.juntadeandalucia.es/institutodeestadistica ycartografia/ieagen/instituto.html>

MANUAL DE BUENAS PRÁCTICAS  
PARA LA NORMALIZACIÓN DE  
FUENTES Y REGISTROS  
ADMINISTRATIVOS DE LA JUNTA DE  
ANDALUCÍA

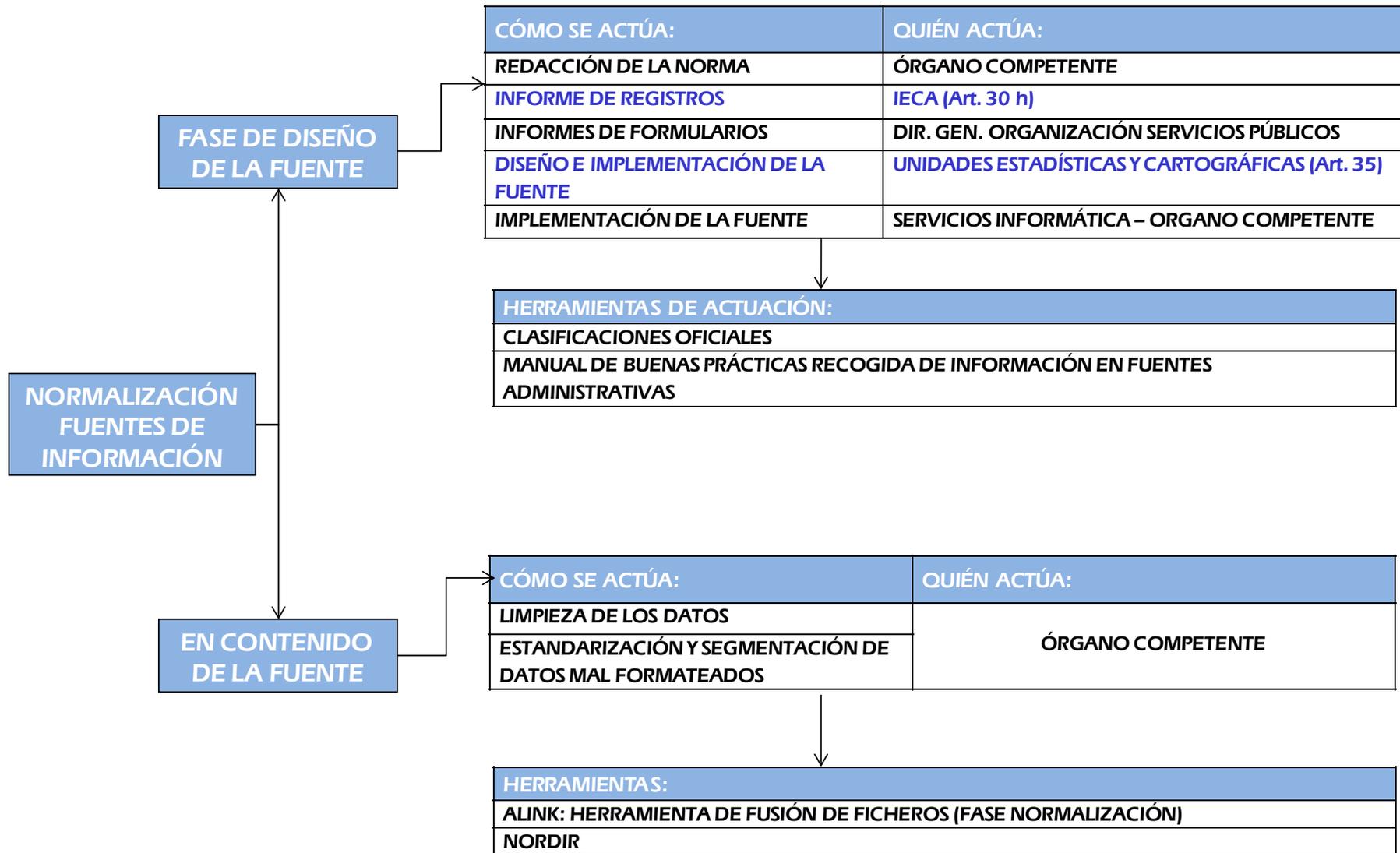


Una vez superada la fase de normalización en el diseño e implementación de la fuente debe tenerse presente que la información del mundo real, provenga o no de fuentes o registros administrativos, puede contener errores, estar incompleta o incorrectamente formateada. Por este motivo, es necesario transformar los datos originales brutos en otros datos con formatos consistentes y bien definidos, así como resolver las posibles inconsistencias sobre la forma en la que se representa y codifica la información.

En relación con la información contenida en las fuentes, suele ser preciso aplicar un conjunto de técnicas encaminadas a la obtención de datos consistentes (esta fase no será necesaria o precisará de menos recursos cuanto mejor haya sido el proceso de diseño e implementación de la fuente y la automatización en los procesos para incorporar información en las fuentes), que redundarán en una mejor calidad y fiabilidad en posteriores análisis de esos datos. En el caso de información de localización geográfica, este proceso es de suma importancia ya que su correcta ejecución ayudará a obtener mejores resultados con las distintas herramientas o procesos que se utilicen para la geocodificación, ya sea ante un proceso de enlace de registros o mediante el uso de otros procesos.

En el proceso de normalización de la información de direcciones contenida en las fuentes se establecen dos fases principales. Una primera *fase de limpieza* donde no importa el contenido semántico del fichero de datos, y al que se le realizan tareas de codificación así como de eliminación de abreviaturas y signos de puntuación en los datos contenidos en él. La segunda *fase es la de estandarización* del conjunto de datos, en la que se analiza el contenido semántico del fichero y se clasifica el contenido de éste según el valor de sus componentes. Debido a esa clasificación se realizará la segmentación de los datos en cada uno de los componentes que los forman.

En el esquema 2 se resumen las distintas fases del proceso de normalización, en las que el Sistema Estadístico y Cartográfico puede actuar a través de distintos mecanismos y utilizando una serie de herramientas según se describe en el desarrollo del presente apartado.



Esquema 2.- Normalización de fuentes de información: fases, posibilidades de actuación, agentes y herramientas

## 2.1 RECOGIDA DE INFORMACIÓN GEOGRÁFICA

Actualmente se están desarrollando tareas para la definición e implementación de un **gestor de direcciones postales** cuyo objetivo es obtener las direcciones que son recogidas en las fuentes de información administrativa a partir de una base de datos que almacene el conjunto de direcciones existentes en Andalucía. En este gestor se podrán incorporar aquellas direcciones aún no recogidas en el sistema, pudiéndose actualizar dicha base de datos.

Hasta que se disponga en un futuro de este gestor de direcciones postales es necesario recoger de manera normalizada las variables que aportan información geográfica, lo que favorecerá la obtención de datos fiables y simplificará el proceso de normalización de las direcciones, evitando y disminuyendo el número de errores, y a su vez, mejorando la comparabilidad e integración de la información del domicilio entre distintas fuentes o registros administrativos.

La localización de cualquier elemento (persona, actividad, evento, etc.) en el territorio puede hacerse, en general, de tres formas: mediante el empleo de la dirección postal, mediante la asignación de coordenadas geográficas o, si es el caso, aportando la referencia catastral de un bien inmueble (urbano o rústico). La información de las coordenadas geográficas permite tener georreferenciado el elemento, mientras que la información de la dirección postal y de la referencia catastral posibilita su geocodificación mediante su enlace con una fuente de información que disponga de las coordenadas de dichas direcciones.

Teniendo en cuenta que la inmensa mayoría de las fuentes de información administrativa recogen la dirección postal, en este documento se muestra cómo geocodificar una fuente de información a partir de ésta, con el objetivo de obtener sus coordenadas X e Y.

En el caso en que la dirección no sea correcta o no permita la asignación de una X e Y debe utilizarse la referencia catastral si se dispone de la misma. Por tanto, en esta guía se explica la geocodificación puntual a través de la dirección postal y, en su caso, de la referencia catastral. En otros casos, que quedan fuera del alcance de esta guía, la Administración precisa la realización de procesos de georreferenciación tomando directamente los datos sobre el terreno. Por otra parte, debe tenerse en cuenta la existencia de elementos geográficos que no pueden ser tratados de manera puntual: ríos, vías de comunicación, manzanas, zonas, etc.

A continuación se describe cómo se tiene que recoger la información **de la dirección postal y de la referencia catastral** para que sea susceptible de ser geocodificada cuando los datos de las coordenadas geográficas no figuren entre la información que ofrece la fuente.

### 2.1.1 Dirección postal

Una **dirección postal** se puede definir como una jerarquía de componentes tales como unidades administrativas, nombre geográficos con un creciente nivel de detalle (municipio, entidad singular, núcleo de población...), nombre de vía y finalmente localizador o número o nombre de la finca, incluyendo también algún identificador de tipo postal.

De acuerdo con INSPIRE, una dirección postal tiene una serie de componentes. A continuación se indica cómo están contemplados dichos componentes en el proyecto del Callejero Digital de Andalucía Unificado (CDAU):

- Nombre de las unidades administrativas: áreas delimitadas donde la administración nacional, regional o local ejerce su jurisdicción. En España se corresponderían con el país, la Comunidad Autónoma, la provincia y el municipio.
- Nombre de entidad inferior al municipio: este componente representa el nombre del área o localidad que, sin ser una unidad administrativa como tal, es necesario para la identificación inequívoca de una dirección. En España son las entidades colectivas, entidades singulares, núcleos de población y diseminados. En este componente se pueden diferenciar dos elementos: el tipo y el nombre.
- Nombre de la vía: nombre de la vía, que puede ser una carretera, calle, camino, plaza o incluso un canal de agua (por ejemplo, en los Países Bajos). Este componente posibilita el almacenamiento estructurado de su contenido, aunque no es obligatorio. No obstante, es conveniente desagregarlo en los siguientes elementos: tipo de vía y nombre de vía.
- Localizador: elemento de mayor detalle que permite identificar cada dirección inequívocamente. Este componente requiere desagregarlo en varios elementos: portal con número y extensión, bloque, escalera, piso y puerta y en su caso nombre por el que se conoce al edificio (por ejemplo: Edificio Sevilla 2, Nervión Plaza, Centro de Diagnóstico, etc.).
- Elemento Postal: subdivisión de direcciones originalmente creada para el reparto postal, pero que actualmente constituye un elemento habitual para muchas aplicaciones que necesitan direcciones. El modelo contempla tanto códigos como nombres (oficina de correos, localidad, por ejemplo: UK). En España se corresponde con el código postal. Los dos primeros valores hacen referencia a la provincia, siendo asignados siguiendo un orden alfabético, del 01 al 50, a excepción de las ciudades de Ceuta y Melilla, que al estar fuera de la división provincial tienen



asignados el 51 y el 52, respectivamente. Puede comprender sólo una parte de un municipio, un municipio en su totalidad o más de un municipio.

El orden de estos componentes en una dirección postal debe seguir el siguiente esquema: nombre de la vía, localizador, nombre de la entidad inferior al municipio, nombre de la unidad administrativa y elemento postal.

A continuación se va a describir en detalle **la propuesta de recogida de información de variables relacionadas con el territorio, como el país, la región, el municipio o la entidad de población, así como los distintos componentes de una dirección postal.**

## PAÍS

**Definición.-** Área geográfica bien delimitada, y una entidad políticamente independiente, con su propio gobierno, administración y leyes.

### **Propuesta de recogida de información.-**

La recogida de información sobre países en fuentes de información administrativa para facilitar su posterior explotación estadística y cartográfica y su comparabilidad se propone se realice siguiendo el estándar establecido por el INE al efecto. Dicho estándar se localiza en la URL: [http://www.ine.es/ss/Satellite?L=es\\_ES&c=Page&cid=1254735839296&p=1254735839296&pagename=MetodologiaYEstandares%2FINELayout#](http://www.ine.es/ss/Satellite?L=es_ES&c=Page&cid=1254735839296&p=1254735839296&pagename=MetodologiaYEstandares%2FINELayout#)

## REGIÓN Y PROVINCIA

**Definición.-** Esta variable indica la región de referencia de la persona u hogar, así como de otros datos que registren una ubicación. En el caso de España, en función de la desagregación que se precise, son las Comunidades Autónomas y/o las provincias.

### **Propuesta de recogida de información.-**

La recogida de esta información en fuentes de información administrativa, para facilitar su posterior explotación estadística y cartográfica y su comparabilidad, se propone se realice siguiendo el estándar establecido por el INE a tal efecto. Dicho estándar, en el que se distinguen, según el nivel de desagregación, Comunidades y Ciudades Autónomas y provincias, se localiza en la URL: [http://www.ine.es/ss/Satellite?L=es\\_ES&c=Page&cid=1254735839296&p=1254735839296&pagename=MetodologiaYEstandares%2FINELayout#](http://www.ine.es/ss/Satellite?L=es_ES&c=Page&cid=1254735839296&p=1254735839296&pagename=MetodologiaYEstandares%2FINELayout#)



Este estándar es compatible con la Nomenclatura de Unidades Territoriales para estadística (NUTS), con un nivel de desagregación de 2 dígitos (Comunidad Autónoma) o 3 dígitos (Provincias), regulado por el Reglamento (CE) N° 1059/2003 del Parlamento Europeo y del Consejo del 26 de mayo de 2003, [que establece una clasificación común de unidades territoriales para la estadística \(NUTS\), \[http://epp.eurostat.ec.europa.eu/cache/ITY\\\_OFFPUB/KS-CD-06-001-12/EN/KS-CD-06-001-12-EN.PDF\]\(http://epp.eurostat.ec.europa.eu/cache/ITY\_OFFPUB/KS-CD-06-001-12/EN/KS-CD-06-001-12-EN.PDF\)](http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-CD-06-001-12/EN/KS-CD-06-001-12-EN.PDF)

## MUNICIPIO Y ENTIDAD DE POBLACIÓN

**Definición.-** El municipio es una división territorial o administrativa que se define legalmente como la “Entidad básica de la organización territorial del Estado y cauce inmediato de participación ciudadana en los asuntos públicos, que institucionaliza y gestiona con autonomía los intereses propios de la correspondiente colectividad” y al que le confiere el artículo 11.1º de la Ley 7/1985 de 2 de abril, reguladora de las Bases del Régimen Local, personalidad jurídica y plena capacidad para el cumplimiento de sus fines.

También existen entidades de ámbito inferior al municipio que se recogen en el nomenclátor de unidades poblacionales que publica el INE. Este componente, se puede denominar *entidad de población*, representa el nombre del área o localidad que, sin ser una unidad administrativa como tal, es necesario para la identificación inequívoca de una dirección. En España: entidades colectivas, entidades singulares, núcleos de población y diseminados.

### Propuesta de recogida de información.-

La codificación a utilizar para los municipios se puede obtener de la relación de municipios y sus códigos que publica el Instituto Nacional de Estadística (INE). Esta relación puede consultarse en la dirección URL: <http://www.ine.es/daco/daco42/codmun/codmun11/11codmunmapa.htm> o en la siguiente [http://www.ine.es/ss/Satellite?L=es\\_ES&c=Page&cid=1254735839296&p=1254735839296&pagename=MetodologiaYEstandares%2FINELayout#](http://www.ine.es/ss/Satellite?L=es_ES&c=Page&cid=1254735839296&p=1254735839296&pagename=MetodologiaYEstandares%2FINELayout#)

La información sobre la entidad de población se puede obtener del nomenclátor publicado por el INE: <http://www.ine.es/nomen2/ficheros.do>

En el cuadro 1 se esquematizan las variables que hasta ahora se han descrito en relación con el territorio:

| DESCRIPCIÓN                                       | FORMULARIO                     | SISTEMA DE INFORMACIÓN |                   |
|---|--------------------------------|------------------------|-------------------|
|   |                                | Etiqueta               | Tipo dato         |
| Código de la entidad                              | -                              | ine_nucleo             | Texto (long. 11)  |
| Nombre de Entidad de población / Diseminado (INE) | ENTIDAD DE POBLACIÓN/LOCALIDAD | nom_nucleo             | Texto (long. 100) |
| Código del municipio                              | -                              | ine_mun                | Texto (long. 5)   |
| Nombre del municipio                              | MUNICIPIO                      | nom_municipio          | Texto (long. 100) |
| Código de la provincia                            | -                              | cod_pro                | Texto (long. 2)   |
| Nombre de la provincia                            | PROVINCIA                      | n_pro                  | Texto (long. 100) |
| Código postal                                     | CP                             | cod_postal             | Texto (long. 5)   |

Cuadro 1.- Recogida de información de variables territoriales en formularios y sistemas de información

Se propone que la información, en relación con estas variables, se recoja de acuerdo a lo indicado en la columna "formulario" del cuadro, y que los sistemas de información donde se implementen utilicen, además de las variables recogidas en el formulario, los códigos del núcleo o entidad de población, municipio y provincia, que permitirán un mejor y más eficaz tratamiento de la información. Estos códigos se pueden obtener, como ya se ha indicado anteriormente en este documento, en la página web del INE.

### **DIRECCIÓN POSTAL (Domicilio, domicilio social...)**

**Definición.-** Lugar en que legalmente está establecida una persona o sociedad para el cumplimiento de sus obligaciones y el ejercicio de sus derechos.

#### **Propuesta de recogida de información.-**

Se propone que la recogida de la información en relación con la dirección postal, se desagregue en los campos indicados en la columna "formulario" del cuadro 2. Una vez recogida dicha información, se recomienda la implementación del Sistema de Información utilizando las variables, etiquetas y tipos de datos indicados en el cuadro 3. En este cuadro se reproducen las etiquetas y tipos de datos utilizados por el Callejero Digital de Andalucía Unificado (CDAU) para estas variables. Su utilización facilitará el enlace de los registros del fichero con los registros del CDAU y con ello la posibilidad de obtener tanto las coordenadas X e Y para la geocodificación, como cualquier otra información útil para la fuente administrativa de las incluidas en el Callejero, como por ejemplo el código del INE, el código de portal o la referencia catastral.

| DESCRIPCIÓN                            | FORMULARIO            |
|--|-----------------------|
| <b>DESCRIPCIÓN A NIVEL DE VÍA</b>      |                       |
| Tipo de vía                            | TIPO VÍA              |
| Nombre de la vía                       | NOMBRE VÍA            |
| <b>DESCRIPCIÓN A NIVEL DE NÚMERO</b>   |                       |
| Número de portal/Km                    | NUMERO/Km             |
| Calificador del número (letra)         | CALIF. NUM            |
| Bloque                                 | BLOQUE                |
| Portal                                 | PORTAL                |
| Escalera                               | ESCALERA              |
| <b>DESCRIPCIÓN A NIVEL DE INMUEBLE</b> |                       |
| Planta                                 | PLANTA                |
| Puerta                                 | PUERTA                |
| Referencia Catastral Inmueble          | REFERENCIA CATASTRAL  |
| <b>INFORMACIÓN COMPLEMENTARIA</b>      |                       |
| Tipo de agrupación                     | COMPLEMENTO DOMICILIO |
| Nombre de agrupación                   |                       |
| Otros datos de ubicación (ODUB)        |                       |

Cuadro 2.- Recogida en el formulario



| DESCRIPCIÓN                                   | SISTEMA DE INFORMACIÓN |                   |
|---|------------------------|-------------------|
|   | Etiqueta               | Tipo dato         |
| <b>DESCRIPCIÓN A NIVEL DE VÍA</b>             |                        |                   |
| Código de vía                                 | id_vial                | Entero (long. 9)  |
| Código vía INE                                | ine_via                | Texto (long. 5)   |
| Código vía Catastro                           | dgc_via                | Entero (long. 5)  |
| Tipo de vía                                   | nom_tip_via            | Texto (long. 100) |
| Nombre de la vía                              | nom_via                | Texto (long. 100) |
| <b>DESCRIPCIÓN A NIVEL DE NÚMERO</b>          |                        |                   |
| Código de Portal                              | id_por_pk              | Entero (long. 9)  |
| Entidad inferior de numeración                | num_por_desde          | Entero            |
| Calificador de entidad inferior de numeración | ext_desde              | Texto (long. 20)  |
| Entidad superior de numeración                | num_por_hasta          | Entero            |
| Calificador de entidad superior de numeración | text_hasta             | Texto (long. 20)  |
| Bloque  | bloque                 | Texto (long. 20)  |
| Portal  | portal                 | Texto (long. 20)  |
| Escalera                                      | escalera               | Texto (long. 20)  |
| Referencia Catastral Parcela                  | refcatparc             | Texto (long. 14)  |
| <b>DESCRIPCIÓN A NIVEL DE INMUEBLE</b>        |                        |                   |
| Código de Hueco (Gesta)                       | id_hue                 | Entero (long. 10) |
| Planta  | planta                 | Texto (long. 3)   |
| Puerta  | puerta                 | Texto (long. 3)   |
| Referencia Catastral Inmueble                 | cos_hue_dgc            | Texto (long. 21)  |
| <b>INFORMACIÓN COMPLEMENTARIA</b>             |                        |                   |
| Tipo de agrupación                            | nom_tipo_agrupacion    | Texto (long. 100) |
| Nombre de agrupación                          | nom_agrup              | Texto (long. 100) |
| Otros datos de ubicación (ODUB)               | txt_app                | Texto (long. 200) |

Cuadro 3.- Implementación de los datos recogidos en el sistema de información

Integrando los datos recogidos en los formularios y su implementación en sistemas de información, se obtiene el cuadro que se muestra a continuación:

| DESCRIPCIÓN                                       | FORMULARIO                | SISTEMA DE INFORMACIÓN |                   |
|---|---------------------------|------------------------|-------------------|
|   |                           | Etiqueta               | Tipo dato         |
| <b>DESCRIPCIÓN A NIVEL DE VÍA</b>                 |                           |                        |                   |
| Código de vía                                     | -                         | id_vial                | Entero (long. 9)  |
| Código vía INE                                    | -                         | ine_via                | Texto (long. 5)   |
| Código vía Catastro                               | -                         | dgc_via                | Entero (long. 5)  |
| Tipo de vía                                       | TIPO VÍA                  | nom_tip_via            | Texto (long. 100) |
| Nombre de la vía                                  | NOMBRE VÍA                | nom_via                | Texto (long. 100) |
| <b>DESCRIPCIÓN A NIVEL DE NÚMERO</b>              |                           |                        |                   |
| Código de Portal                                  | -                         | id_por_pk              | Entero (long. 9)  |
| Número de portal/Km (1)                           | NUMERO/Km                 | -                      | -                 |
| Calificador del número (letra)                    | CALIF. NUM                | -                      | -                 |
| Entidad inferior de numeración (1)                | -                         | num_por_desde          | Entero            |
| Calificador de entidad inferior de numeración (1) | -                         | ext_desde              | Texto (long. 20)  |
| Entidad superior de numeración (1)                | -                         | num_por_hasta          | Entero            |
| Calificador de entidad superior de numeración (1) | -                         | ext_hasta              | Texto (long. 20)  |
| Bloque  | BLOQUE                    | bloque                 | Texto (long. 20)  |
| Portal  | PORTAL                    | portal                 | Texto (long. 20)  |
| Escalera  | ESCALERA                  | escalera               | Texto (long. 20)  |
| Referencia Catastral Parcela                      | -                         | refcatparc             | Texto (long. 14)  |
| <b>DESCRIPCIÓN A NIVEL DE INMUEBLE</b>            |                           |                        |                   |
| Código de Hueco (Gesta)                           | -                         | id_hue                 | Entero (long. 10) |
| Planta  | PLANTA                    | planta                 | Texto (long. 3)   |
| Puerta  | PUERTA                    | puerta                 | Texto (long. 3)   |
| Referencia Catastral Inmueble                     | REFERENCIA CATASTRAL (2)  | cos_hue_dgc            | Texto (long. 21)  |
| <b>INFORMACIÓN COMPLEMENTARIA</b>                 |                           |                        |                   |
| Tipo de agrupación (3)                            | COMPLEMENTO DOMICILIO (3) | nom_tipo_agrupacion    | Texto (long. 100) |
| Nombre de agrupación (3)                          |                           | nom_agrup              | Texto (long. 100) |
| Otros datos de ubicación (ODUB) (3)               |                           | txt_app                | Texto (long. 200) |

Cuadro 4.- Recogida y normalización de la dirección postal

(1) Un aspecto a tener en cuenta es que el número puede ser del tipo 28-30 e incluso del tipo 28A-30B. Para resolver estas direcciones es preciso incluir en el sistema de información (no en el formulario de recogida, que complicaría excesivamente el modelo) dos campos para el número (inferior y superior) y dos para la letra (letra superior e inferior). En el caso de que una dirección venga expresada como "autovía a-92 km 25" en el sistema de información 'autovía' será el tipo de vía, 'a-92' se incluirá como nombre de vía y el valor 25 se incluye dentro del campo de entidad inferior de numeración.

- (2) El campo referencia catastral será optativo en los formularios y dependerá del objeto de la fuente y la viabilidad del conocimiento de este dato por los informantes. En cualquier caso, en el sistema de información es conveniente incorporar este campo.
- (3) Tanto en el modelo de datos de la Administración General del Estado para direcciones como en el Callejero Digital Unificado de Andalucía se han añadido campos para recoger otros datos de ubicación o el nombre alternativo para los casos en los que la dirección esté en una pseudovía, por ejemplo, "Edificio Sevilla 2", o las direcciones que se pueden localizar en diseminado, o las que por el crecimiento de un núcleo ahora están dentro (un ejemplo puede ser un cortijo o hacienda que hace 20 años estaba en las afueras y ahora sigue conservando el nombre que tenía antes como la Hacienda el Marquesado que ahora está en la Avenida de América). Sin embargo, aunque toda la información complementaria se recopile en otros datos de ubicación, es recomendable que el sistema de información recoja en un campo diferente la información relativa a diferentes agrupaciones. El campo 'tipo de agrupación' recoge aquellos grupos de vías que generan una agrupación de algún tipo concreto, y que aportan mayor información a la dirección. En este campo se contemplan: polígonos industriales, barrios, barriadas, urbanizaciones, etc.

Teniendo en cuenta, las variables descritas en relación con el territorio y la dirección postal, la propuesta de recogida de información del domicilio en un formulario sería la que se muestra en la figura 1:

| DOMICILIO A EFECTOS DE NOTIFICACIONES: |             |   |            |              |                 |         |           |           |            |
|--|-------------|---|------------|--------------|-----------------|---------|-----------|-----------|------------|
| TIPO DE VÍA*:                          |             | NOMBRE VÍA*:  |            |              |                 |         |           |           |            |
| TIPO DE NUMERACIÓN*:                   |             | NÚMERO  | NÚMERO/KM: | CALIF. NÚM.: | BLOQUE:         | PORTAL: | ESCALERA: | PLT/PISO: | PTA/LETRA: |
| COMPLEMENTO DOMICILIO:                 |             | <div style="border: 1px solid black; padding: 2px;">           NÚMERO<br/>           KILÓMETRO<br/>           SIN NÚMERO         </div> |            |              |                 |         |           |           |            |
| PROVINCIA*:                            | MUNICIPIO*: | LOCALIDAD*:   |            |              | CÓDIGO POSTAL*: |         |           |           |            |

Figura 1.- Propuesta de recogida del campo domicilio en un formulario

## 2.1.2 Referencia catastral

La **referencia catastral** es el identificador oficial de los bienes inmuebles. Consiste en un código alfanumérico que es asignado por la Dirección General de Catastro de manera que todo inmueble debe tener una única referencia catastral que permita situarlo inequívocamente en la cartografía catastral.

Gracias a la referencia catastral se sabe con exactitud de qué bien inmueble se trata en los negocios jurídicos (compra-ventas, herencias, donaciones, etc.), no confundiéndose unos bienes con otros. Asimismo, la consignación de la referencia catastral en el Registro de la Propiedad facilita la identificación física del inmueble frente a terceros, mejorando así la agilidad y la seguridad de las transacciones inmobiliarias.

El Callejero Digital de Andalucía Unificado (CDAU) dispone de la referencia catastral asociada a cada uno de los portales, por lo que en fuentes que dispongan sólo de la dirección postal se puede añadir la referencia catastral de la parcela donde se encuentra el portal a través del CDAU.

La recogida de información a través de la referencia catastral debe realizarse del siguiente modo:

### 1.- Para fincas urbanas

La información de la referencia catastral de un inmueble en una finca urbana se compone de 20 dígitos agrupados en cuatro grupos según el siguiente modelo:

- Código de parcela: 7 dígitos
- Hoja del plano: 7 dígitos
- Código local: 4 dígitos
- Código control: 2 dígitos

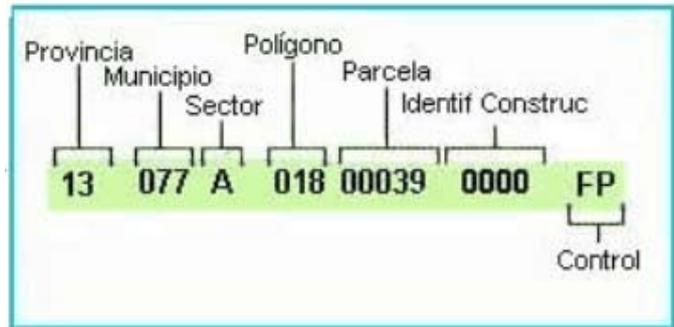


De los 20 dígitos totales, los 14 primeros corresponden a la parcela catastral.

### 2.- Para fincas rústicas

La información de la referencia catastral de un inmueble en una finca rústica se compone de 20 dígitos agrupados en siete grupos:

- Código de provincia: 2 dígitos
- Municipio: 3 dígitos
- Sector: 1 dígito
- Polígono: 3 dígitos
- Parcela: 5 dígitos
- Identificador de construcción: 4 dígitos
- Control: 2 dígitos



El uso de los primeros 14 dígitos del inmueble hace referencia a la parcela exacta donde se encuentra ubicado el inmueble y de la que puede obtenerse información sobre el punto central o centroide de dicha parcela.

## 2.2 HERRAMIENTAS DE NORMALIZACIÓN DE DIRECCIONES

Prácticamente toda la información, de cualquier índole, se puede localizar en un punto concreto de la superficie de la tierra. De toda la información que se puede localizar espacialmente un alto porcentaje de la misma se puede localizar mediante direcciones postales, ya sea en forma de calle y número, o en forma de carretera y kilómetro.

De ahí que el primer paso para geocodificar datos consiste en estudiar cómo se recoge la información relacionada con la dirección postal. Por este motivo un paso previo necesario para la geocodificación de las direcciones postales es **la normalización de dicho campo**.

El proceso de normalización de la dirección postal consiste en corregir y transformar los datos originales en otros datos con formatos consistentes y bien definidos según el estándar que se ha descrito anteriormente.

La normalización y depuración de los datos de las direcciones postales es un proceso complejo, ya que existen distintas posibilidades en la recogida de esta información. Así, ésta puede venir recogida en un solo campo, habitualmente de manera imprecisa y heterogénea, o desagregada en los campos recomendados, presentando por tanto un nivel elevado de normalización.

Para agilizar el proceso de normalización es necesario utilizar herramientas que ayuden a optimizar y normalizar la información de la dirección postal en aquellos ficheros que no lo estén, ya que una buena estructura y calidad de la información determinan en gran medida el éxito de la geocodificación.

Las herramientas de normalización se basan en dos fases principales:

1. Una fase de estandarización en la que se corrigen y se limpian los datos, y se asignan valores normalizados. Por ejemplo, en las direcciones cuyo identificador de vía venga descrito con el símbolo "c/" se podrá asignar un valor estandarizado que será "calle".
2. Una segunda fase en la que se segmenta la información en cada uno de los componentes que forma la dirección postal.

Dependiendo de la herramienta que se utilicen, ambas fases pueden realizarse por separado o en un solo paso.

Es importante e imprescindible estudiar cómo se recoge la información de la dirección postal y conocer el fichero de datos para poder determinar qué herramienta de normalización es la más adecuada o la que más interesa utilizar.

Las distintas aplicaciones informáticas que se van a describir permitirán realizar de forma sencilla un proceso de normalización de direcciones postales a partir de un fichero cuya estructura de datos sea conocida. Ambas herramientas devuelven un fichero con los datos de la dirección postal normalizados y desagregados en varios campos (véase figura 2).

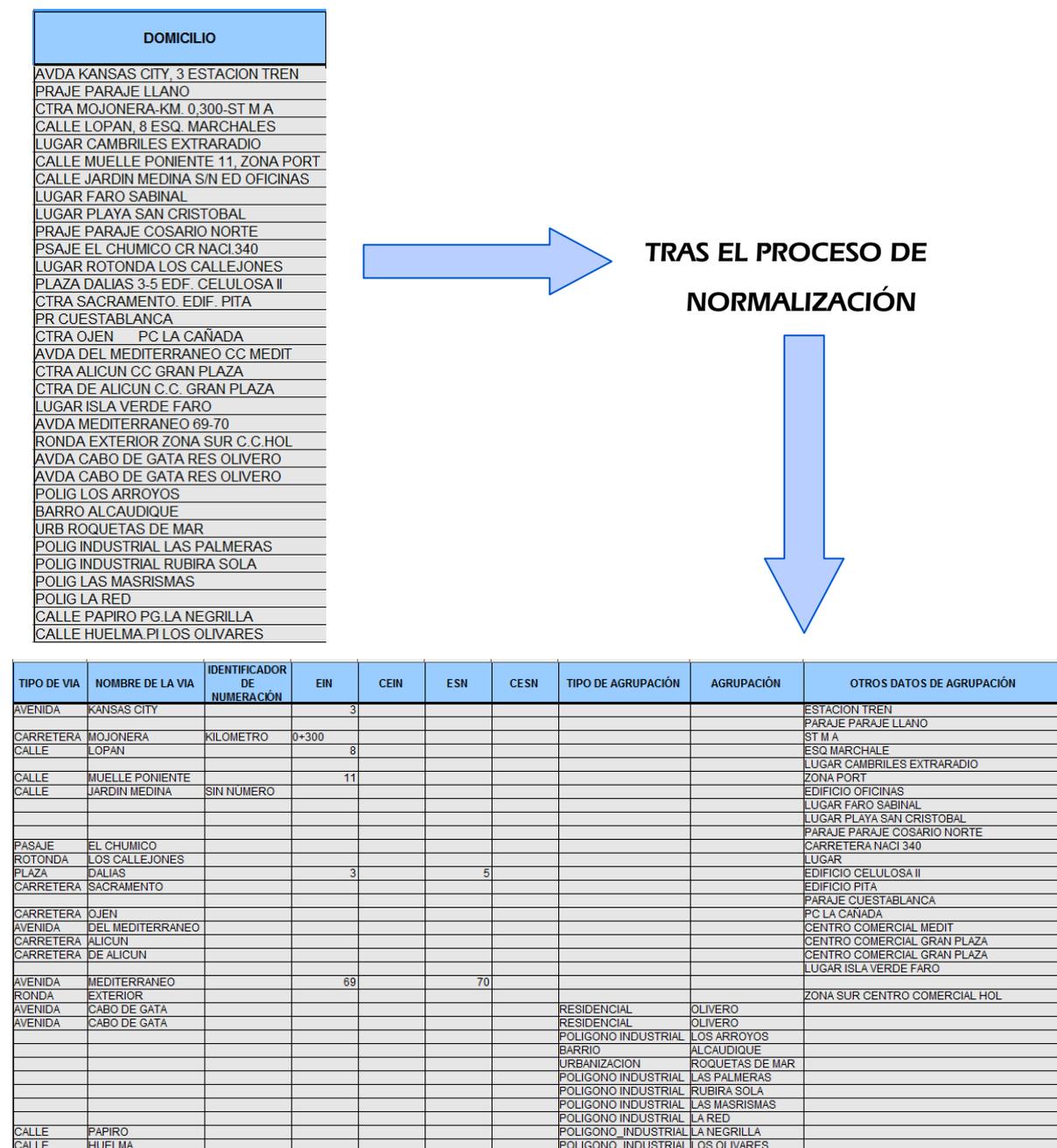


Figura 2.- Resultado del campo domicilio tras su normalización

### 2.2.1 aLink: Herramienta de Normalización

Constituye una de las herramientas de la aplicación aLink: Herramienta de Fusión de Ficheros y puede utilizarse tanto para normalizar direcciones postales como para nombres de personas e identificadores de personas físicas y jurídicas.

aLink es una aplicación de software libre desarrollada por el Instituto de Estadística y Cartografía (IECA) a partir de Febrl (desarrollo de software libre de la Universidad Nacional de Australia), accesible a todos los usuarios y descargable desde la página web del IECA a través de los Servicios de descarga de software:

<http://www.juntadeandalucia.es/institutodeestadisticaycartografia/ieagen/otrosServidores/software/index.htm#uno>

aLink permite trabajar en varios sistemas operativos de Windows y Linux, y una vez instalada se accede a sus diferentes funcionalidades a través de la interfaz principal (ver figura 3).



Figura 3.- Interfaz principal de aLink: Herramienta de Fusión de Ficheros

Ésta muestra tres botones, dos de los cuales permiten al usuario acceder a las principales herramientas de la aplicación:

- **Herramienta de Normalización:** permite realizar la normalización de direcciones postales, nombres de personas e identificadores de personas físicas y jurídicas. Es la evolución de la herramienta de normalización **ADYN**.
- **Herramienta de Enlace:** permite comparar los registros de dos ficheros de datos con información común, para identificar qué pares de los registros comparados corresponden a verdaderos enlaces.

La herramienta de normalización de aLink (ver figura 4) limpia y estandariza la información del campo a normalizar, además de ofrecer al usuario la posibilidad de desagregar la misma en una serie de campos de salida.

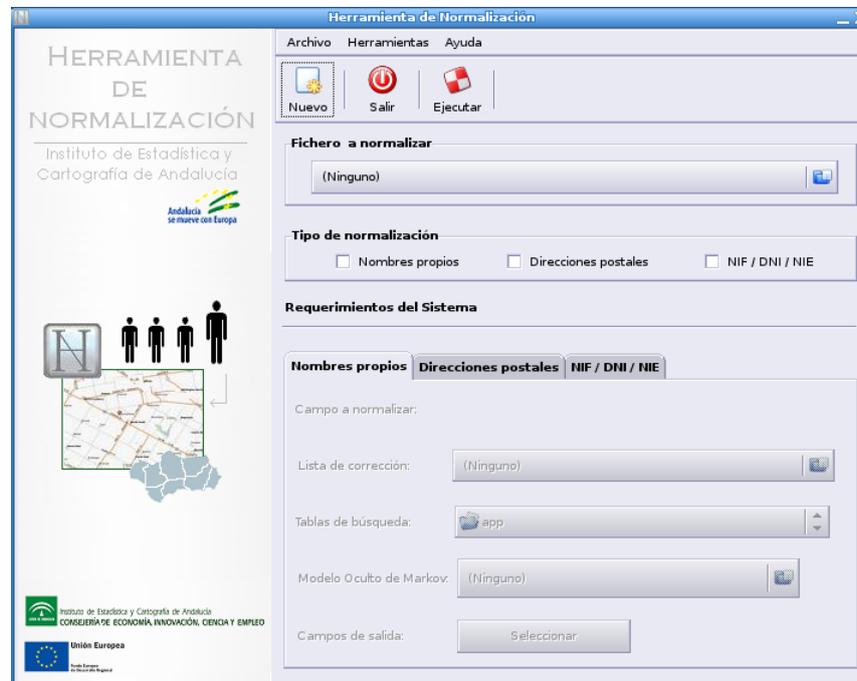


Figura 4.- aLink: Herramienta de Normalización

En el caso particular de las direcciones postales, para normalizar un campo dirección se hará uso de los distintos instrumentos que aporta la herramienta aLink. En concreto, para las tareas de limpieza y estandarización o asignación de valores normalizados, aLink cuenta con dos instrumentos:

- La lista de corrección de direcciones, que permite limpiar el fichero de datos de aquellos caracteres o cadenas que el usuario considere oportunos eliminar o sustituir. Por ejemplo, se eliminan caracteres extraños ( '|', '\$', ... ) o se sustituyen cadenas de caracteres del tipo 'hnos' por 'hermanos', 'ntra' por 'nuestra', etc.

La lista de corrección es editable por el usuario y siempre se le pueden incluir nuevos elementos, modificar o eliminar los existentes, según las necesidades de cada fichero de datos.

- Las tablas de búsqueda de direcciones, que son ficheros con listas de valores que hacen referencia a un mismo elemento común, como por ejemplo tipos de vías, municipios, provincias, etc., junto con las cadenas o valores estandarizados por los que se desean reemplazar. Así, por ejemplo, si la aplicación encuentra en el campo a normalizar el elemento "avda" ó "avd" se sustituirá por su valor normalizado que en este caso es "avenida". Todos los elementos de cada tabla de búsqueda tienen asociada una misma etiqueta común; por ejemplo, los elementos de la

tabla de búsqueda de tipos de vía tienen asociada la etiqueta "TV". Esta etiqueta jugará un papel importante a la hora de segmentar la información de la dirección.

Al igual que pasaba con la lista de corrección, estos listados de valores a normalizar y sustituir son editables por el usuario para que éste las adapte a sus necesidades.

En cuanto a la tarea de segmentación de la información, aLink dispone de varios mecanismos, como las ya mencionadas tablas de búsqueda y un instrumento basado en la metodología de los Modelos Ocultos de Markov (en inglés, Hidden Markov Models o HMM).

Estos modelos tratan de identificar y reconocer el patrón que siguen los datos de los ficheros a normalizar a través de una muestra extraída de dichos datos. De esta manera se extrapola el conocimiento de la muestra a la totalidad de los datos. Estos modelos son imprescindibles para segmentar cada una de las direcciones en los distintos elementos que las componen.

Más concretamente, esta metodología permite desagregar la información en tantos campos como elementos se encuentren en la dirección postal, o bien de una forma similar a cómo se recogen las direcciones en el CDAU. Para facilitar el trabajo al usuario, la herramienta de normalización de aLink incorpora por defecto dos modelos HMM con distintas estructuras para segmentar las direcciones, tanto a medida como según la desagregación que ofrece el CDAU. Haciendo uso de estos modelos incorporados en la herramienta se consigue normalizar la mayor parte de los registros. No obstante, si los modelos proporcionados con la aplicación no responden a la estructura de las direcciones del campo a normalizar, aLink permite crear un modelo HMM propio para dichas direcciones.

Para determinar el tipo de desagregación de la dirección aLink cuenta con una ventana de campos de salida, en la que se ofrece al usuario la posibilidad de realizar una segmentación de la dirección postal en los campos que éste decida teniendo en cuenta su información de partida, es decir, puede realizar una desagregación a medida (ver figura 5), o bien segmentar la dirección postal en una estructura similar a la estructura oficial del Callejero Digital de Andalucía Unificado, haciendo posible, por tanto, que el usuario segmente los datos de la dirección postal tal y como se recomienda en el "Manual de buenas prácticas para la normalización de fuentes y registros administrativos de la Junta de Andalucía" y que ésta pueda ser fácilmente geocodificada si fuese el caso (véase figura 5).

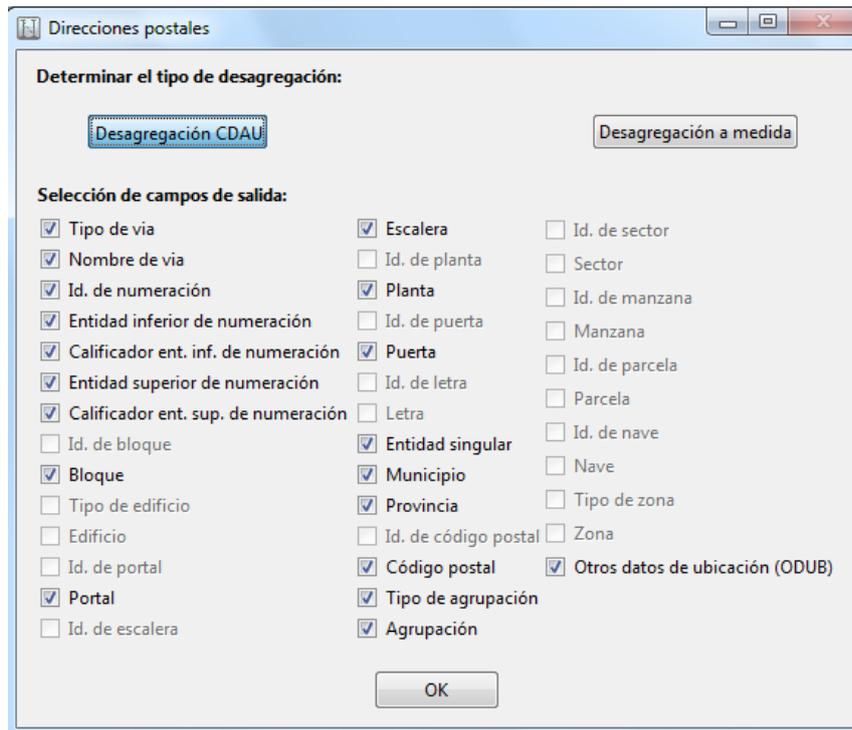


Figura 5.- Desagregación de campos de salida con aLink: a medida o desagregación según los campos del CDAU

A continuación se describen los pasos a seguir en el proceso de normalización con la herramienta aLink:

**Paso 1: Tratamiento previo** del fichero. Este paso es obligatorio y se realizará de manera automática en la opción establecida en la herramienta para tal efecto.

**Paso 2: Normalización** haciendo uso de los instrumentos proporcionados por la aplicación.

En este paso se usará la interfaz de normalización. En ella habrá que especificar el fichero de entrada, el campo que se quiere normalizar, la lista de corrección y las tablas de búsquedas de direcciones para realizar las estandarizaciones y el Modelo de Oculto de Markov para desagregar la dirección postal, que trae la herramienta por defecto.

**Paso 3: Validación de la normalización**. En este paso se comprueba cómo de bueno ha sido el proceso de normalización. Tras el paso anterior, se obtiene un fichero de datos con la dirección postal normalizada y segmentada en los campos indicados. Además, la herramienta crea una nueva columna en este fichero de resultados que ayuda a identificar aquellos registros cuya estructura en la dirección postal se ha normalizado correctamente de los que no lo hicieron. Así, la columna "*VALIDACIÓN*" devuelve un 0 si considera que el registro está bien normalizado o un 1 si considera que el registro no está bien normalizado. El hecho de que un registro no haya sido normalizado puede deberse a que la estructura o patrón de la dirección sea bastante heterogénea y no estuviera recogida en el Modelo Oculto de Markov que proporciona aLink.

Aquellos registros que no se han normalizado deben ser eliminados del fichero de resultados e incluidos en uno nuevo para tratar de normalizarlos en un siguiente paso. La herramienta permite crear modelos HMM propios y ajustados a estructuras de datos personalizadas.

**Paso 4: Creación de un Modelo Oculto de Markov.** En este paso se intenta identificar la estructura de las direcciones del fichero no normalizadas en el paso anterior.

**Paso 5: Normalización.** En este paso se utiliza el modelo creado anteriormente para normalizar los registros restantes, de igual modo que se hizo en el paso 2.

**Paso 6: Validación.** Se comprueba cómo de buena ha sido la normalización, de igual modo que en el paso 3.

Entre las **ventajas** de la herramienta aLink cabe resaltar que permite repetir el proceso tantas veces como sea necesario para normalizar de manera completa las direcciones postales de todos los registros. Así, para aquellos ficheros donde la estructura en el campo de dirección pudiera ser muy variable, se pueden crear tantos modelos de Markov como sean necesarios para completar la normalización de todo el fichero. Otra de las ventajas comentada anteriormente es que se adapta a las necesidades de los usuarios y a las características del fichero de trabajo.

### 2.2.2 NorDir

NorDir es la aplicación cliente del servicio web de normalización del Callejero, que fue desarrollada en el seno del proyecto del *Sistema de Información Geográfica Corporativo de la Junta de Andalucía (SIGC)* y que en la actualidad forma parte del proyecto del Callejero Digital de Andalucía Unificado.

El proyecto del SIGC se inició a finales del 2007 con el doble objetivo de definir una serie de pautas y normas que homogeneizaran el uso de este tipo de sistemas en la Junta de Andalucía, garantizando la interoperabilidad entre ellos, y de desarrollar herramientas y servicios horizontales de infraestructura en materia de tratamiento de Información Geográfica que permitiesen satisfacer necesidades no cubiertas por los SIG departamentales ya existentes en las diferentes Consejerías.

Con el fin de cumplir estos objetivos se identificó una necesidad horizontal presente en todas y cada una de las Consejerías de la Junta de Andalucía: la necesidad de normalizar y geocodificar direcciones postales. Para cubrir esta necesidad el proyecto desarrolló las herramientas NorDir y GeoDir (junto a la herramienta de escritorio GEOCODER) que se tratan en esta guía.

Como se ha visto anteriormente, de forma previa a la geocodificación es necesario realizar un proceso de normalización de los datos de partida para que éstos puedan ser consumidos posteriormente por las herramientas de geocodificación, que esperan direcciones postales adecuadamente estructuradas. En esta herramienta la normalización se plantea en dos niveles: la estructura y orden de los elementos en la dirección postal de entrada por un lado, y cómo se describen los mismos por otro.

En la figura 6 puede observarse cómo una misma dirección postal que sigue esta pauta puede tener muchas representaciones si se presentan variaciones en la estructura, orden y manera de describir los elementos (por ejemplo, la palabra 'Calle' puede escribirse de maneras diferentes).



Figura 6.- Estructura y orden de los componentes de una dirección postal con Nordir

Para tratar la amplia variedad de situaciones con las que el usuario se puede encontrar, es decir, las múltiples formas de escribir una dirección postal, la herramienta Nordir aplica una serie de filtros (**fase de estandarización y limpieza**): reemplazan en la dirección original abreviaturas comunes de palabras ('doc' por 'doctor', 'sta' por 'santa', etc.), elimina tildes y partículas que pueden interferir en la búsqueda posterior de la dirección en la base de datos, homogeneizando todos los datos hacia una misma política de puntuación, etc.

Una vez aplicados estos filtros, de tal forma que los datos han sido homogeneizados, identifica qué elementos componen la dirección postal no estructurada de partida, para convertir esta dirección en una dirección normalizada (**fase de desagregación**), siguiendo la estructura de normalización con la que trabaja la herramienta de geocodificación del proyecto.

La normalización llevada a cabo por la herramienta Nordir se realiza con el objetivo de geocodificar. Tanto la fase de estandarización como la de desagregación se realizan en un solo paso. No se debe

olvidar que el normalizador simplemente hace eso, filtra y reestructura. En caso de que el dato de entrada tuviera errores, por ejemplo el nombre de la vía mal escrito, este error no sería corregido.

Los tipos de archivos que se permiten cargar para llevar a cabo la normalización de direcciones son los archivos de extensión .txt o .csv; si el tipo de archivo no se corresponde con alguna de estas extensiones se muestra un mensaje de error.

Esta herramienta se encuentra en la siguiente web de la Junta de Andalucía:

<http://www.juntadeandalucia.es/servicios/mapas/geocoding-tools/clientSelector.jsp>

Para comenzar a utilizar la herramienta se deben indicar el fichero con direcciones a normalizar, el carácter usado como separador de campos por dicho fichero, los campos que componen la dirección y si la primera línea corresponde a la cabecera del fichero (ver figura 7).

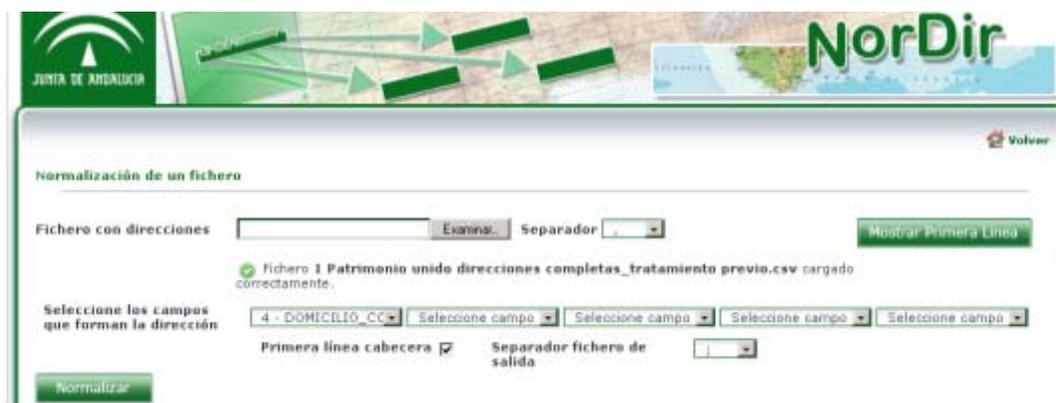


Figura 7.- NorDir

La herramienta NorDir devuelve un fichero de resultados en formato CSV, con información normalizada y desagregada en los siguientes campos:

- Tipo vía
- Nombre de vía
- Número portal
- Municipio
- Provincia

Esta herramienta es útil cuando la dirección a normalizar se basa en datos referentes a la vía y números de portales (o se quiera prescindir de otros datos que no sean éstos). Si se quiere normalizar el resto de la información que acompaña a la dirección postal (barriadas, urbanizaciones, polígonos industriales, bloque, portal, escalera, planta, puerta, etc.) es necesario acudir a otras herramientas.

### 3. GEOCODIFICACIÓN

Los elementos de la Tierra y la representación vectorial de entidades espaciales puede realizarse utilizando tres tipos de geometría: puntos, líneas o polígonos. Esta guía se ha desarrollado para orientar al usuario en la geocodificación de elementos mediante puntos concretos en la superficie terrestre.

La **geocodificación puntual** es el proceso de asignar coordenadas geográficas (X e Y) a puntos del territorio (direcciones, puntos de interés, etc.). Las coordenadas obtenidas posibilitan la integración de los elementos en un Sistema de Información Geográfica y éstos pueden ser almacenados, manipulados y analizados para poder resolver problemas y satisfacer unas necesidades concretas de información.

La localización de un punto sobre la superficie terrestre debe cumplir una serie de requisitos:

- Que el punto sea único.
- Que quede perfectamente identificado el sistema de referencia empleado al localizar el punto.

Entre las distintas utilidades que tiene la publicación de los datos geográficos, se destacan las siguientes:

- La mejora de la información que se ofrece al usuario. Por ejemplo, localizando la ubicación exacta de los servicios públicos.
- Es una herramienta fundamental para la mejora de la gestión y planificación de los servicios. La difusión de información localizada en el territorio multiplica las posibilidades de análisis y favorece los procesos de toma de decisiones basados en información estadística y cartográfica de calidad.
- La integración con otra información. La publicación de los datos geográficos siguiendo las normas que aseguran la comparabilidad de la información (sistema de coordenadas, nomenclaturas, clasificaciones, servicios interoperables etc.) permite su integración con otros datos. La publicación, la integración con otros datos y su puesta a disposición al conjunto de ciudadanos es un elemento fundamental para la implicación de éstos en la actividad del sector público.
- Garantizan el derecho al acceso a la información pública.

## 3.1 SISTEMAS DE REFERENCIA DE COORDENADAS Y COORDENADAS GEOGRÁFICAS

### 3.1.1 Sistemas de referencia

La posición de un punto al ser geocodificado se expresa en un sistema de coordenadas en relación a un sistema de referencia geodésico<sup>1</sup>.

La Tierra no es una esfera perfecta y contiene múltiples irregularidades provocadas por el relieve. Por ello, la forma geométrica más realista para representar la Tierra es el geoide. Sin embargo, para poder localizar un punto de forma absoluta en la superficie de la Tierra es necesario ajustarlo a una superficie geométrica que nos permita establecer un sistema de coordenadas. La superficie que se utiliza para localizar coordenadas X e Y es la de un elipsoide de revolución, es decir, una figura tridimensional que se genera al hacer girar una elipse sobre el eje de la Tierra. Sobre esta figura geométrica ya es posible establecer un sistema de coordenadas que permita conocer posiciones absolutas, calcular distancias, etc. Asociado a dicho elipsoide se establece un sistema de coordenadas cartesianas de tres ejes mutuamente perpendiculares con origen en el centro del elipsoide (sistema de referencia de tres dimensiones). El eje Z es coincidente con el eje de revolución (la recta imaginaria que une el Polo Norte y el Polo Sur), el eje X coincide con el meridiano origen (Greenwich) y el eje Y es el perpendicular a los dos anteriores.

Además, todos los sistemas de referencia geodésicos incluyen un DATUM geodésico que define su posición en la Tierra y que consiste en la “colocación” del elipsoide con respecto al GEOIDE en un punto coincidente. Debido a las diferentes irregularidades de la Tierra, cada país o región elige el elipsoide que más se acerca a su realidad y que mejor encaja en su relieve.

La forma de designar las coordenadas sobre este sistema de referencia vendrá dada por la situación del punto respecto del Ecuador (paralelo de mayor radio) y del meridiano de origen, es decir, la latitud y la longitud, respectivamente (ver figura 8).

La **latitud** proporciona la localización de un lugar en dirección Norte o Sur desde el Ecuador y se expresa en medidas angulares que varían desde los 0° del Ecuador hasta los 90°N del Polo Norte o los 90°S del Polo Sur. Es el ángulo que forma la recta que pasa por el punto hasta el centro de la esfera con el plano ecuatorial.

La **longitud** proporciona la localización de un lugar en dirección Este u Oeste desde el meridiano de referencia 0°, expresándose en medidas angulares comprendidas desde los 0° hasta 180°E y 180°W. Este ángulo mide la distancia angular del meridiano del punto con el meridiano 0° (meridiano de

---

<sup>1</sup> En la norma ISO 19111 se define un Sistema de Referencia de Coordenadas como un sistema de coordenadas que está referido a la Tierra a través de un Datum geodésico.

Greenwich). Es lo mismo medir este ángulo sobre el círculo del Ecuador que sobre el círculo del paralelo que pasa por el punto.

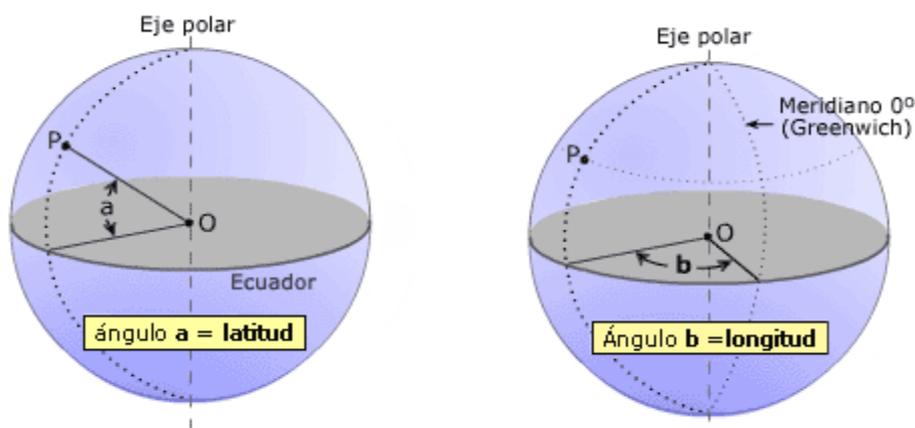


Figura 8.- Latitud y longitud

En la actualidad, en España existen tres elipsoides que se utilizan mayoritariamente, y que dan nombre a **sistemas de coordenadas**. Éstos son:

- **WGS84: World Geodetic System 84.** Este es el sistema de coordenadas que utilizan los dispositivos GPS por defecto. A nivel internacional es el sistema de coordenadas común más utilizado.
- **ED50: European Datum 1950.** Es un antiguo sistema de referencia geodésico empleado en Europa. En España ha sido el sistema oficial de la cartografía de la Península y Baleares hasta 2008, año en el que se adopta como oficial el ETRS89.
- **ETRS89: European Terrestrial Reference System 1989.** Su origen se remonta a la resolución de 1990 adoptada por EUREF (Subcomisión de la Asociación Internacional de Geodesia, AIG, para el Marco de Referencia Europeo) y trasladada a la Comisión Europea en 1999, por lo que está siendo adoptado sucesivamente por todos los países europeos sustituyendo al sistema ED50.

Los inconvenientes que tiene representar coordenadas sobre el globo terrestre (difícil de transportar, no permite ver toda la superficie terrestre completa, etc.) se resuelven utilizando un mapa, o lo que es lo mismo, pasando de una superficie esférica a una superficie plana. En este sentido, las *proyecciones* son la solución para pasar de la esfera al plano. Existen diferentes tipos de proyecciones, pero la más utilizada es la UTM (Universal Transversal de Mercator) que divide la Tierra en cuadrículas, donde las coordenadas métricas son las mismas (cuadrículas de 6 grados que se miden en metros/husos).

Por tanto, para saber cuál es el punto al que se hace referencia en coordenadas UTM es necesario conocer, además del valor X e Y de la coordenada, el cuadrante en el que se trabaja.

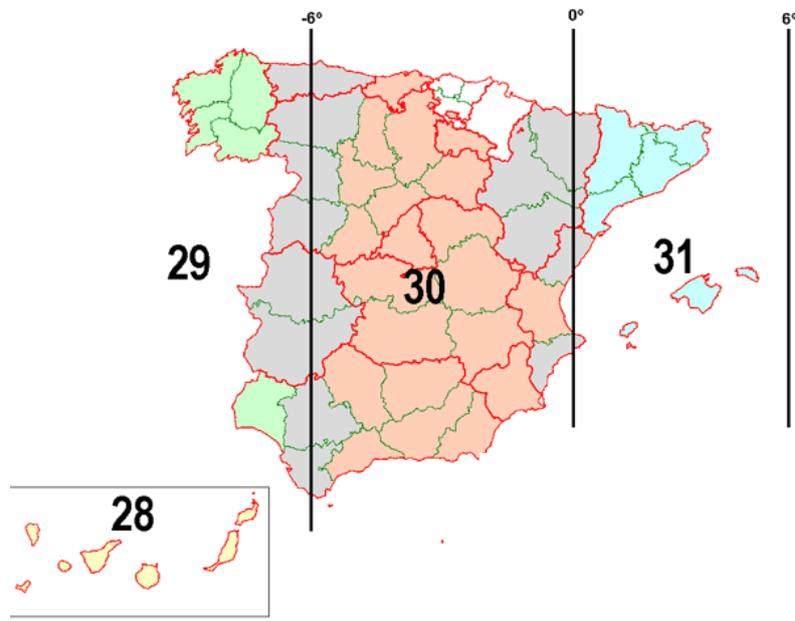


Figura 9.- Husos en España

En concreto, Andalucía se sitúa en el cuadrante 29N y 30N. La línea divisoria entre ambos husos pasa por la provincia de Sevilla y Cádiz como se ve en la figura 9; la provincia de Huelva está en huso 29N.

En cualquier proyecto cartográfico es fundamental definir, además del sistema de referencia, el huso en el que se trabaja. Toda la información de un mismo proyecto debe ser tratada siempre en un mismo sistema de referencia para que exista plena coincidencia espacial. En el caso de Andalucía se recomienda trabajar con el huso 30N.

Una vez definidos los sistemas de referencia, las coordenadas geográficas de un punto, ya sean en el elipsoide o proyectadas, se expresan mediante coordenadas X e Y. Cada Sistema de Referencia de Coordenadas (CRS) tiene asociado un código numérico denominado EPSG y que es usado ampliamente. Los más habituales son:

| Sistema de referencia | Sistema de coordenadas | Código EPSG |
|-----------------------|------------------------|-------------|
| ETRS89                | Geográficas            | 4258        |
|                       | UTM huso 30            | 25830       |
| ED50                  | Geográficas            | 4230        |
|                       | UTM huso 30            | 23030       |
| WGS84                 | Geográficas            | 4326        |
|                       | UTM huso 30            | 32630       |

Cuadro 5.- Sistemas de referencia, sistemas de coordenadas y códigos EPSG

### 3.1.2 Transformación de coordenadas

Muchos de los datos geográficos sobre el territorio andaluz han sido levantados en el sistema ED50. El [Real Decreto 1071/2007](#) establece el ETRS89 como sistema de referencia geodésico oficial en España para la referenciación geográfica y cartográfica en el ámbito de la Península Ibérica y las Islas Baleares. Entre ambos sistemas de referencia existen diferencias notables en el geoposicionamiento (del orden de centenares de metros).

Los cambios en los sistemas de referencia oficiales, el hecho de que hasta el 1 de enero de 2015 la información geográfica y cartográfica aún podía seguir referenciada en ED50, el disponer de información en distintos sistemas de referencia, y la existencia de varias proyecciones y husos oficiales en el territorio español y andaluz, hacen necesario disponer de herramientas que permitan reproyectar la cartografía entre las diferentes referencias geodésicas y proyecciones utilizadas comúnmente.

Existen diversas herramientas que facilitan las transformaciones entre distintos sistemas de referencia. En concreto, la Junta de Andalucía ha desarrollado la aplicación denominada **CALAR**, que facilita la transformación de la información espacial contenida en ficheros en los formatos más comunes entre los diversos sistemas de referencia oficiales en Andalucía.

La herramienta está disponible en la siguiente dirección web:

<http://www.juntadeandalucia.es/servicios/mapas/geodesiaclient/>

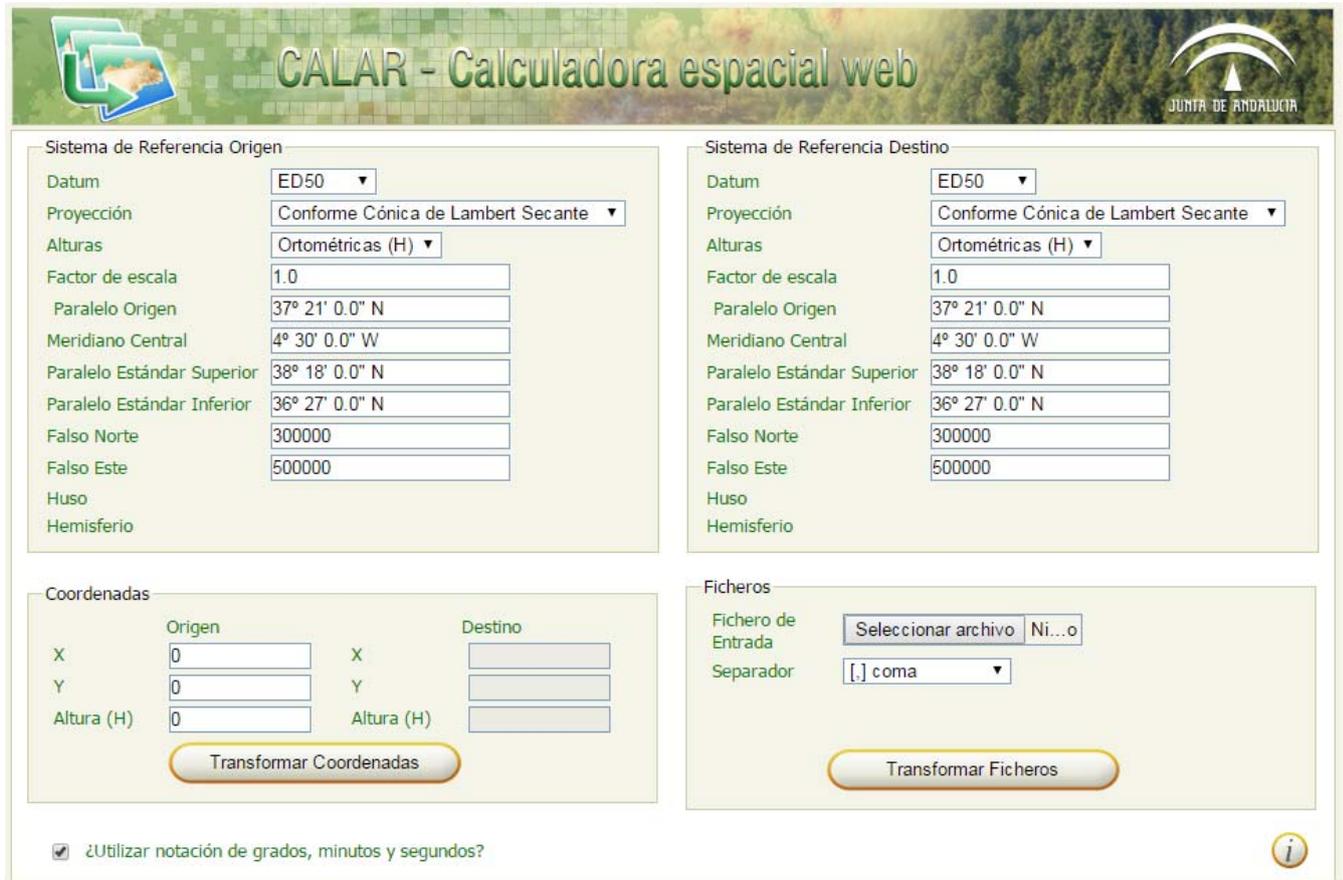


Figura 10.- Interfaz de Calar

En la pantalla principal (véase figura 10) se observan las siguientes funciones principales:

- Menú de selección de los Sistemas de Referencia de Origen y Destino. Consiste en seleccionar el Sistema de Referencia de Origen en el que se proporcionan las coordenadas y el Sistema de Referencia de Destino al que se quieren transformar las coordenadas. La herramienta calcula las transformaciones que se realizan para pasar de un sistema a otro y las utiliza para transformar coordenadas y ficheros.

La selección de un sistema de referencia se realiza configurando los siguientes parámetros:

Datum: soporta actualmente dos opciones:

- ED50: European Datum 1950
- ETRS89: European Terrestrial Reference System 1989

Proyección: hay varias proyecciones disponibles, pero interesan sólo dos:

- Coordenadas geográficas (Latitud/Longitud)
- Coordenadas UTM (X,Y)

Alturas: Se deben seleccionar alturas ortométricas

Al configurar un Sistema de Referencia de Origen cuya proyección sea Latitud/Longitud es posible introducir la coordenada bien en notación decimal o sexagesimal. Esto queda reflejado en la casilla "¿Utilizar notación de grados, minutos y segundos?".

- Panel de conversión de coordenadas. Este panel de conversión de coordenadas permite transformar manualmente coordenadas desde el Sistema de Referencia Origen al Sistema de Referencia Destino configurados previamente.
- Panel de conversión de ficheros (individuales y por lotes). Calar permite la conversión de ficheros completos.

Toda la documentación que haya sido geocodificada debe estar acompañada de la información referida al sistema de proyección de los datos en la forma que el formato de almacenamiento tenga prevista al efecto.

Por su parte, el Ministerio de Fomento, a través del Instituto Geográfico Nacional (IGN), ha desarrollado también una herramienta específica para la transformación de coordenadas: **Programa de Aplicaciones Geodésicas**.

Esta herramienta puede descargarse en la siguiente dirección web:

<http://www.ign.es/ign/layoutIn/herramientas.do>



El Ministerio incorpora también un servicio Web de transformación de coordenadas, que se encuentra disponible en el siguiente enlace:

<http://www.ign.es/wcts-app/>

### 3.1.3 Expresión de las coordenadas y precisión

Las coordenadas geográficas son la expresión de un punto sobre la superficie terrestre, o lo que es lo mismo, un conjunto de valores que definen exactamente la posición de puntos sobre un sistema de referencia determinado, pudiendo ser geodésico o proyectado.

La expresión de las coordenadas depende del sistema de referencia que se use:

1. Sistemas de coordenadas geodésicos o geográficos: (Latitud, Longitud) para coordenadas elipsóidicas. Esta expresión de la posición de un punto determina la distancia a la que se encuentra del Ecuador en la dirección N-S (latitud) y la distancia a la que se encuentra de un meridiano de referencia en la dirección E-W (longitud). Los valores vienen expresados en ángulos (grados, minutos, segundos) o en grados decimales. Andalucía está situada en longitud Oeste y latitud Norte.

La expresión de una coordenada en este sistema se expresa en grados sexagesimales en formato decimal, en el CRS geodésico ETRS89 latitud/longitud con una precisión mínima de cinco decimales, utilizando el carácter punto como separador. La dirección Oeste quedará expresada mediante signo negativo.

2. Sistema de coordenadas proyectados: son los utilizados para un sistema proyectado (en el plano), la expresión de las coordenadas vendrá dado por un valor X e Y en un sistema de coordenadas cartesianas. Las magnitudes en el sistema de proyecciones UTM se expresan en metros.

Las coordenadas expresadas en Universal Transversal de Mercator (UTM) deberán incluir un mínimo de dos decimales.

Visto todo lo anterior, resulta evidente que dar un par de coordenadas sin hacer referencia al datum no es lo suficientemente preciso. En un datum todo punto tiene un par de coordenadas único, mientras que el mismo punto tendrá diferentes coordenadas en diferentes datums, o lo que es lo mismo, un par de coordenadas puede corresponder a diferentes puntos en diferentes datums.

En definitiva, las consideraciones necesarias para la expresión de coordenadas son:

- a. Indicar siempre el Sistema de Referencia al que vienen referidas las coordenadas. En caso de desconocimiento del sistema de referencia indicar la herramienta de donde provienen las coordenadas.
- b. Preferiblemente usar el Sistema ETRS89.
- c. Expresar las coordenadas geográficas (latitud/longitud) en grados decimales con 5 decimales de precisión o en su defecto proyectadas (X,Y) en UTM huso 30N con dos decimales de precisión.

La Norma Técnica Cartográfica del Sistema Estadístico y Cartográfico de Andalucía, "NTCA\_01008\_Módulo Geodésico de Referencia y Altitudes", en su artículo 24.BIS, especifica que la localización espacial por coordenadas de un elemento para su inscripción en registros oficiales de la Junta de Andalucía seguirá los siguientes criterios: "Debe expresarse, al menos, en grados sexagesimales en formato decimal en el CRS geodésico ETRS89 bidimensional latitud/longitud con una precisión mínima de cinco decimales. La dirección oeste quedará expresada mediante signo negativo."

## 3.2 BASES CARTOGRÁFICAS E INFORMACIÓN DE REFERENCIA

### 3.2.1 Bases cartográficas

La cartografía es la ciencia que se encarga de crear mapas geográficos con la finalidad de representar e ilustrar diferentes elementos. La cartografía básica es aquella que se obtiene por procesos de observación directa y por medición de la superficie terrestre a partir de la que pueden obtenerse otros mapas derivados. Ésta se utiliza para representar dónde se encuentran los elementos geocodificados, ayudando a la comprensión de los mismos y a la contextualización geográfica de la información.

Según se recoge en el Decreto 141/2006, de 18 de julio, por el que se ordena la actividad cartográfica en la Comunidad Autónoma de Andalucía, se considera cartografía básica oficial el Mapa Topográfico de Andalucía 1:10.000 (MTA10) y la Ortofotografía Digital de Andalucía. Para interpretar esta disposición es necesario tener en cuenta a qué escala se quiere trabajar y el grado de actualización de la cartografía básica. En este sentido, a continuación se señala el estado en que se encuentran estos proyectos así como sus fechas de actualización:

- **Ortofotografía Digital de Andalucía:** esta base de referencia se podrá utilizar en un abanico de escalas comprendido entre 1:5.000 y 1:20.000 aunque visualmente, y debido a la herramienta de zoom de la que disponen todos los visores y sistemas de información geográfica, se puede ampliar un ráster hasta escalas mucho mayores. Sin embargo, la posibilidad de uso viene determinada por la precisión de la propia ortofotografía, que en las series más actuales oscila entre 1 metro y 25 centímetros. Se dispone de series bianuales desde 2004-2005 a 2010-2011 servidas por la Infraestructura de Datos Espaciales de Andalucía, se puede acceder al servicio de la ortofoto 2010-2011 a través de la siguiente dirección <http://www.ideandalucia.es/wms/ortofoto2010?> y de la ortofotografía de máxima actualidad del Plan Nacional de Ortofotografía Aérea (PNOA) ofrecida por el Instituto Geográfico Nacional, a cuyo servicio se puede acceder a través de la siguiente URL: <http://www.ign.es/wms-inspire/pnoa-ma>.
- **Mapa Topográfico de Andalucía (MTA10):** se dispone de las series de 2001 y 2007 completas y actualmente se trabaja en la producción de la información de 2010 a partir de la Base Cartográfica de Andalucía 1:10.000, versión que es la serie cartográfica territorial de referencia para toda la Comunidad.

La **Base Cartográfica de Andalucía (BCA)** está en proceso de producción. Ésta se concibe como una base de datos de carácter topográfico a escala 1:10.000, que recoge la localización espacial de los elementos estructurantes del territorio (relieve, infraestructuras de comunicación,

red hidrográfica, construcciones y poblamiento, usos del parcelario aparente, usos del suelo, toponimia, etc.), y está orientada para su explotación mediante sistemas informáticos.

Se trata de un producto de carácter horizontal de referencia para cuantos proyectos, análisis, estudios, etc. requieran una modelización espacial o la posición geográfica de los elementos que conforman el territorio o, simplemente, para conformar una base geométrica sobre la que situar y contextualizar elementos (por ejemplo: para labores de planeamiento; estudios previos de ingeniería; determinación de parcelas agrarias, usos del suelo, definición de red de carreteras, de caminos, de hidrografía, etc.)

Sus precisiones planimétricas y altimétricas corresponden a una cartografía escala 1:10.000. Su rango de aplicación abarca desde escala 1:5.000 a 1:20.000.

Se puede acceder a la Base Cartográfica de Andalucía 1:10.000 (BCA10) en el siguiente enlace:

<http://www.juntadeandalucia.es/institutodeestadisticaycartografia/bca/index.htm>

Pueden consultarse el estado del proyecto y la situación del mismo a través de la siguiente URL:

<http://www.juntadeandalucia.es/institutodeestadisticaycartografia/bca/index.htm>

Para cualquier levantamiento de elementos georreferenciados en el territorio andaluz, éstas constituyen las series cartográficas de referencia y las más recomendables. Estas series se encuentran disponibles mediante servicios interoperables a través de la Infraestructura de Datos Espaciales de Andalucía (IDEA) en la dirección web: <http://www.ideandalucia.es>

A partir de la cartografía básica, y siguiendo criterios de reutilización de datos y el uso de los Sistemas de Información Geográfica, se producen y difunden numerosas series (*cartografía derivada y temática*) dirigidas a muy distintos usuarios: mapas de ámbitos comarcales, aglomeraciones urbanas, callejeros, atlas, parques naturales, etc.

### 3.2.2 Información de referencia

#### **Callejero Digital de Andalucía Unificado**

El Callejero Digital de Andalucía Unificado (CDAU), desarrollado por el Instituto de Estadística y Cartografía de Andalucía, es un proyecto cuyo objetivo es mantener actualizada la información de datos alfanuméricos y geográficos de las vías y los portales de todos los municipios de Andalucía, por lo que es útil para usar como información de referencia para los procesos de geocodificación.



En concreto, para geocodificar diferentes registros haciendo uso de la dirección postal, la fuente de referencia que se precisa es el CDAU. Éste dispone de forma actualizada de la información relativa a las dos entidades básicas de su modelo de datos: las vías (urbanas, interurbanas y rurales) y las aproximaciones postales o portales (números de vía, puntos kilométricos, diseminados), de forma que cualquier objeto geográfico que posea dirección postal pueda ser situado en el territorio, es decir, geocodificado con una precisión a nivel de aproximación postal.

En relación al CDAU, la competencia sobre el mantenimiento de las entidades básicas (vías y portales) corresponde a las Entidades Locales, conforme establece el Real Decreto 2612/1996 (artículo 75.1), que señala que los Ayuntamientos mantendrán actualizadas la nomenclatura y rotulación de las vías públicas y la numeración de los edificios, informando de ello a todas las Administraciones Públicas interesadas. En este sentido, el IECA desarrolla numerosas actuaciones en colaboración con Diputaciones y Ayuntamientos, que se han integrado de forma mayoritaria en el proyecto del CDAU.

Se puede acceder al visor del CDAU a través de la siguiente URL (ver figura 11):

<http://www.callejerodeandalucia.es/>

El Callejero Digital de Andalucía Unificado tiene usos y aplicaciones en campos muy diversos (generación de una base cartográfica municipal para distintos usos, registro de la población del municipio y la edificación, recaudación, actividad del ayuntamiento, actividad económica del municipio, emergencias, urbanismo, etc.) pero, fundamentalmente, se trata de una herramienta básica para poder mantener georreferenciada sin ambigüedades cualquier información sobre personas, actividades o eventos que lleve asociada una dirección postal y, de esa forma, poder analizar y tomar decisiones teniendo en cuenta la dimensión espacial, con el mayor nivel de detalle posible.

En el CDAU el sistema de referencia geodésico es ETRS89 correspondiéndole el código EPSG 4258.

El nivel de desagregación de la información desciende a las entidades básicas del CDAU, es decir, vías y aproximaciones postales, de forma que cualquier variable estadística asociada a una dirección postal pueda geocodificarse.

Puede accederse al portal con información del CDAU pulsando "Información" o en la siguiente URL: <http://www.callejerodeandalucia.es/portal/web/cdau/>, donde puede obtenerse información básica y técnica sobre el proyecto, se puede acceder al visualizador del CDAU y a un amplio abanico de servicios interoperables: WMS, WMTS, WCS y WFS.

Adicionalmente, en el apartado "Recursos" se puede descargar la siguiente información:

- a) información geográfica con los planos callejeros de los municipios a escala 1:2.000 y en formato PDF.
- b) información alfanumérica de los portales con las coordenadas X e Y exactas a cada portal, y la relativa a las vías con las coordenadas X e Y al centro de las mismas. La información es de utilidad para la geocodificación, ya que puede ser utilizada con aplicaciones que permiten automatizar el proceso de asignación de coordenadas geográficas a direcciones postales mediante enlaces de registros, tales como la herramienta aLink: Herramienta de Fusión de Ficheros o los gestores de bases de datos.

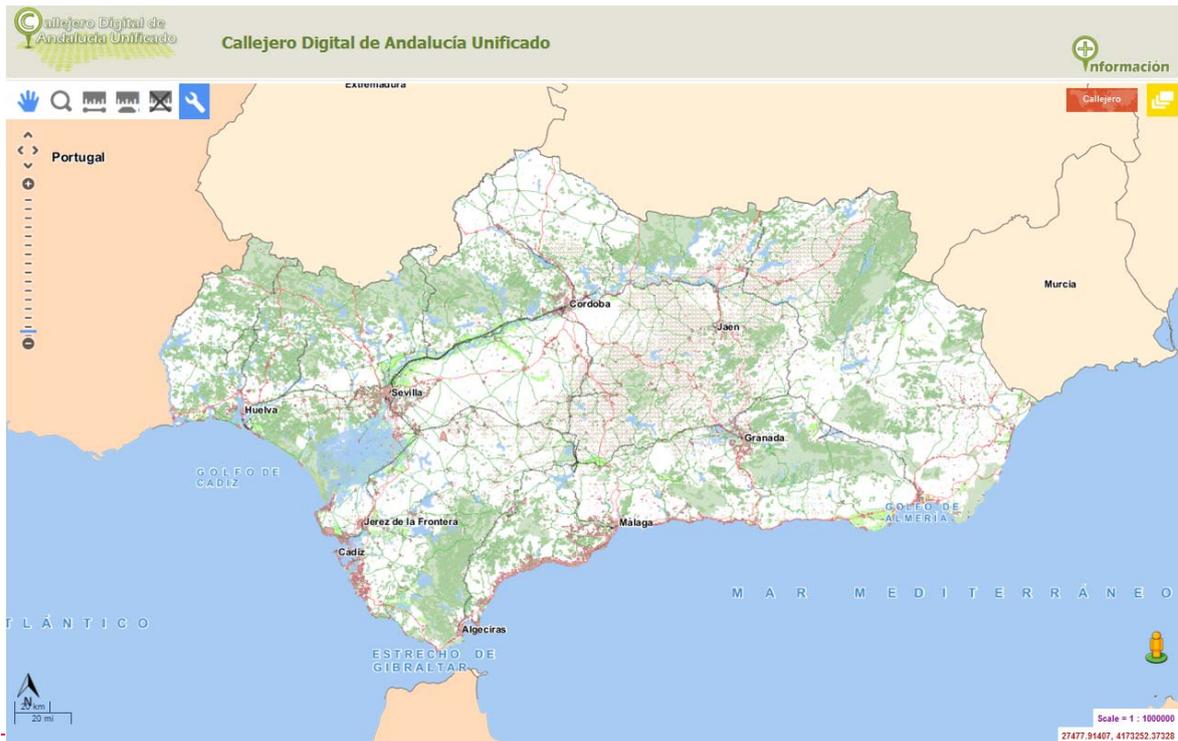


Figura 11.- Visualizador del Callejero Digital de Andalucía Unificado (CDAU)

## **Cartografía Catastral**

La cartografía catastral, consolidada por una larga tradición aunque con un objetivo (recaudación tributaria) y un planteamiento distinto a los restantes proyectos de callejeros, está constituida por un conjunto de capas (en formato SHAPE) y tablas que contienen información sobre distintas entidades, destacando el parcelario catastral (PARCELA, SUBPARCE y CONSTRU).

Para realizar la geocodificación a través de la referencia catastral se debe hacer uso de las bases de datos de la Dirección General de Catastro. En la siguiente URL se puede descargar diferente información y datos sobre cartografía catastral:

<https://www.sedecatastro.gob.es/OVCFrames.aspx?TIPO=TIT&a=masiv>



Las tablas descargadas contienen información de las parcelas catastrales de los municipios. Entre otros campos se pueden encontrar la referencia catastral (que será el campo de comparación con el fichero a geocodificar) y las coordenadas X e Y del centroide de las parcelas.

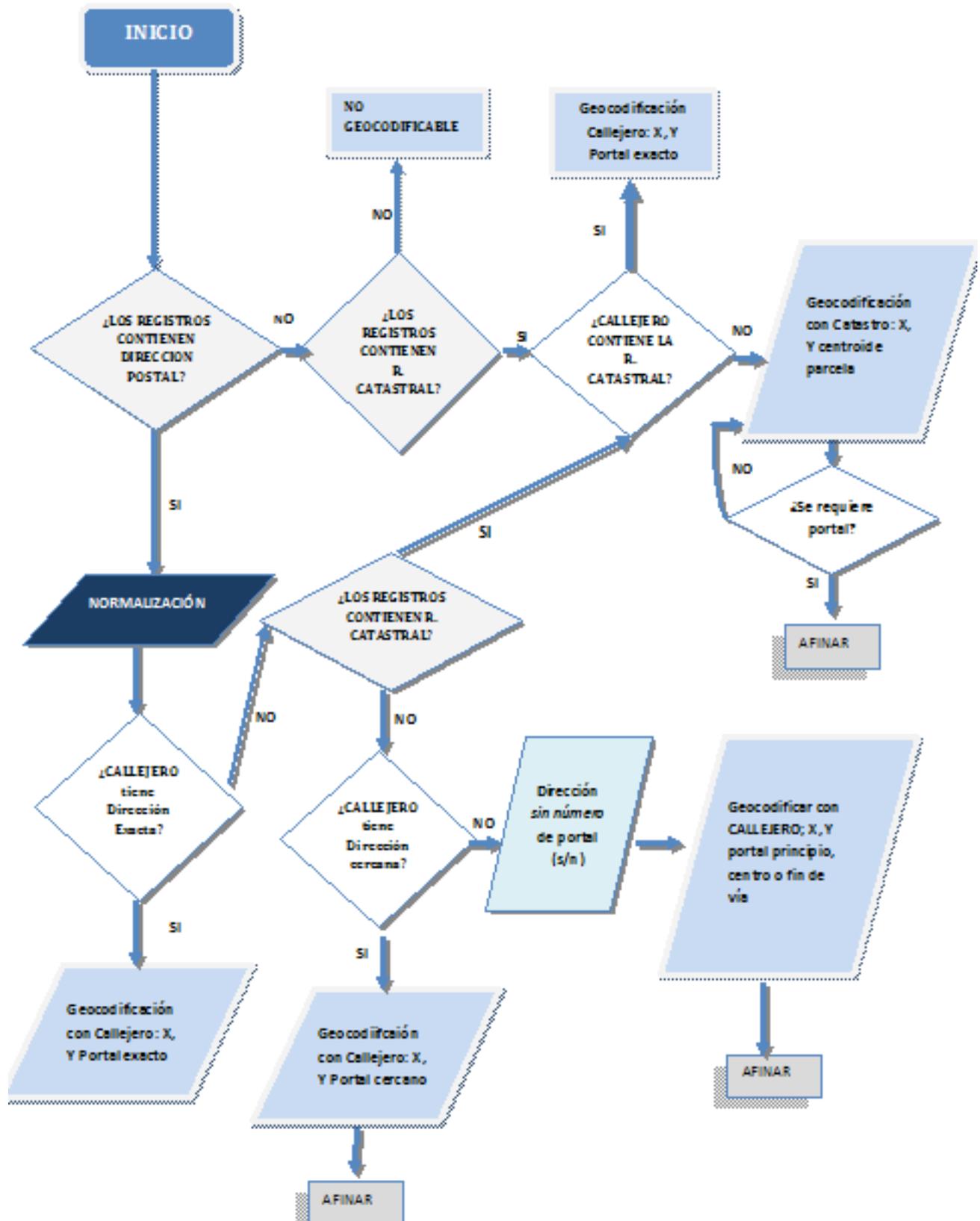
### 3.3 PROCESO DE GEOCODIFICACIÓN

La geocodificación de los registros de un fichero de datos depende en gran medida de la información alfanumérica que indica su localización. Como ya se ha comentado en epígrafes anteriores, la dirección postal permite, en la mayoría de los casos, la localización territorial de los registros y fuentes administrativas de la Junta de Andalucía. En su defecto, o de forma complementaria, se puede utilizar el campo de la referencia catastral como un indicador de la localización de un inmueble o una parcela urbana o rústica.

La geocodificación a un punto, en el caso de direcciones postales, se realiza a un punto exacto de un número de vía, mientras que en el caso de referencias catastrales, la geocodificación de un inmueble se realiza al centroide de la parcela.

El proceso de geocodificación que se describe en este apartado tiene por objeto ayudar a realizar la mejor geocodificación posible, que en cualquier caso estará condicionada por la información de la localización geográfica que aporte la fuente, registro o fichero que se quiera georreferenciar.

El proceso propuesto para la geocodificación de fuentes y registros administrativos, siempre y cuando sea posible y según la información o herramientas de las que se disponga, se esquematiza a continuación (esquema 3):



Esquema 3.- Proceso de geocodificación de fuentes y registros administrativos

### 3.4 HERRAMIENTAS PARA GEOCODIFICAR

Para realizar la geocodificación de un fichero existen distintas herramientas cuya utilización depende del modelo de datos que se tenga y de la información de localización.

#### 3.4.1 aLink: Herramienta de Enlace

La herramienta de enlace de aLink tiene por finalidad comparar registros de dos ficheros de datos para detectar aquellos registros que se identifican como el mismo en ambos ficheros, de tal modo que pueda recogerse información de un fichero e incorporarla en el otro.

En particular, si uno de los ficheros que se va a enlazar contiene direcciones postales y el otro también, pero además incluye las coordenadas geográficas X e Y asociadas a la dirección postal, el proceso de enlace se convierte en un proceso de geocodificación. Es decir, se identifican las direcciones postales coincidentes en ambos ficheros y se asocian las coordenadas X e Y a las direcciones del fichero que se quiere geocodificar.

Esta herramienta puede descargarse desde la página web del IECA, en el siguiente enlace:

<http://www.juntadeandalucia.es/institutodeestadisticaycartografia/ieagen/otrosServidores/software/index.htm#uno>

Para realizar un proceso de geocodificación con aLink son necesarios dos ficheros:

- a) Fichero que contiene las direcciones postales que se quieren geocodificar. Estas direcciones pueden estar normalizadas o no. Cuando no lo estén se recomienda normalizarlas con la herramienta de normalización de aLink, ya que de esta forma aumentará el grado de éxito del proceso de geocodificación.
  
- b) Cualquier fichero de referencia que contenga direcciones postales junto con las coordenadas X e Y que permitan localizar la dirección en el territorio. Entre los distintos ficheros con los que se puede trabajar se encuentran los ficheros de referencia descritos en el punto 3.2.2 (CDAU y Catastro).

Siguiendo el proceso de geocodificación descrito en el esquema 3, se utiliza, por orden de prioridad, el fichero de portales del CDAU, el fichero de parcelas de Catastro y el fichero de vías del CDAU.

La metodología bajo la que se desarrolla el proceso de geocodificación llevado a cabo con la herramienta de enlace de aLink se divide en una serie de fases que se describen a continuación (ver esquema 4):

### **Fase 1: Agrupación de registros.**

Esta fase consiste en formar grupos de tal forma que se reduzca el número de comparaciones a realizar.

Ello resulta muy útil, ya que, por ejemplo, si se necesitan comparar las direcciones postales que existen en ambos ficheros resulta lógico hacerlo únicamente con aquellas que corresponden al mismo municipio.

Por lo tanto, habrá que utilizar una variable de agrupación que contenga el nombre o código INE del municipio. En caso de tener los dos se recomienda usar el código INE, ya que el literal del municipio puede que no se recoja de igual forma en ambos ficheros y se obtengan peores resultados.

### **Fase 2: Comparación de campos de registros.**

En esta fase se indica al programa qué campo o campos se quieren comparar, la función de comparación que se va a utilizar para ello, así como el peso máximo, mínimo y de valor perdido que se le otorga a cada una de las comparaciones realizadas. El peso total será la suma máxima de todas las comparaciones establecidas.

Siguiendo con el ejemplo anterior, para todas aquellas direcciones postales que se encuentran en el mismo municipio se comparará, por ejemplo, el tipo de vía, el nombre de la vía y el número de vía, utilizando para ello alguna de las diversas funciones de comparación disponibles y estableciendo la ponderación que el usuario considere lógica.

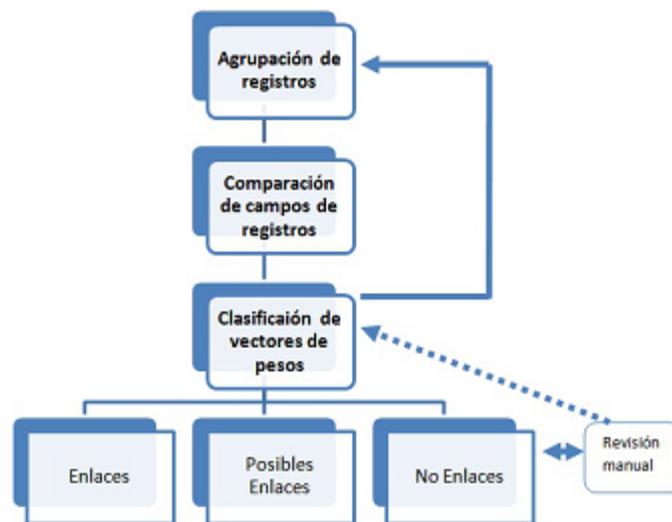
### **Fase 3: Clasificación de los vectores de pesos.**

En esta última fase el usuario establece los valores a partir de los cuales los registros comparados son clasificados como enlaces, no enlaces o posibles enlaces; si bien la herramienta de enlace también provee de un método de clasificación para realizar esta asignación de forma automática.

Tomando el ejemplo anterior nuevamente, se establecerá que aquellas direcciones cuyo peso total de comparación supere cierto límite superior se clasificarán como enlaces, las que estén por debajo de un límite inferior se clasificarán como no enlaces y en otro caso se clasificarán como posibles enlaces.

Por último, y a modo de control del proceso de enlace, es recomendable realizar una revisión manual de los enlaces obtenidos.

Un resumen general del proceso se presenta en el siguiente esquema:



Esquema 4.- Etapas del proceso de enlace de registros

Una de las principales características de la herramienta de enlace de aLink es que el usuario puede intervenir en cada una de las fases que conforman el proceso de enlace, lo que conlleva a una mejor interpretación de los resultados.

Finalmente, y a modo de conclusión, se especifican las ventajas de esta herramienta:

- Es libre y gratuita.
- Permite el enlace o georreferenciación de ficheros con grandes volúmenes de datos.
- Transparencia y flexibilidad. Durante el proceso de enlace el usuario establecerá el tipo de proceso a realizar, así como los parámetros a utilizar en cada una de sus etapas. Además, para aquellos registros que no se hayan enlazado en una primera fase el usuario podrá seleccionarlos y aplicar un nuevo proceso de enlace.

### 3.4.2 GeoDir y Geocoder

**GeoDir** es la aplicación cliente del servicio web de geocodificación del Callejero. Esta herramienta, al igual que NorDir, se desarrolló dentro del proyecto SIG Corporativo y en la actualidad se encuentra dentro del proyecto del CDAU del Instituto de Estadística y Cartografía de Andalucía. Realiza la geocodificación a través de la información de las direcciones postales haciendo uso de los campos tipo de vía, nombre de la vía, número y municipio.

La aplicación puede usarse a través de la siguiente URL (ver figura 12):

<http://www.juntadeandalucia.es/servicios/mapas/geocoding-tools/clientSelector.jsp>



Figura 12.- Geodir

GeoDir, al ser una aplicación web, presenta algunas limitaciones para trabajar con ficheros de grandes volúmenes de datos. Por este motivo, dentro del SIG Corporativo también se ha desarrollado Geocoder, una herramienta por lotes para la geocodificación masiva de miles de direcciones, disponible para su descarga en el Repositorio de Software Libre de la Junta de Andalucía:

[http://www.juntadeandalucia.es/repositorio/usuario/listado/fichacompleta.jsf?linkDummyForm:\\_idcl=\\_id153&idProyecto=679](http://www.juntadeandalucia.es/repositorio/usuario/listado/fichacompleta.jsf?linkDummyForm:_idcl=_id153&idProyecto=679)

Tanto en GeoDir como en Geocoder el **proceso de geocodificación** se basa en un algoritmo que tiene por objetivo unir las direcciones del fichero a geocodificar con las direcciones almacenadas en el fichero de referencia. De forma interna, el proceso de geocodificación tiene en cuenta una serie de situaciones particulares:

- Si no se encuentran vías con nombre igual al buscado (en su totalidad o parcialmente), se aplica un algoritmo de similitud de manera que se buscan vías con nombres parecidos, habiendo establecido previamente un umbral de similitud mínimo que debe cumplirse.
- Si se busca una vía sin especificar número de portal, el sistema devolverá el portal central de la vía, a menos que no tenga portales, en cuyo caso se devuelve el comienzo de la misma.
- Si el portal buscado se sale del rango de portales existentes para la vía, se devuelve el del extremo correspondiente más cercano.
- Si se busca un portal sin número en una vía en la que hay varios portales de este tipo, el sistema los devuelve todos como resultados exactos.

Para que el proceso busque y enlace un registro del fichero introducido con uno de la base de datos del Portadero, es obligatorio indicar a la herramienta el municipio al que pertenece la dirección. Para establecer este dato pueden usarse dos caminos diferentes:

1. Indicar el nombre del **municipio y provincia**

## 2. O bien, el **código INE** del municipio

El proceso de geocodificación con Geocoder es automático y ofrece un fichero de resultados en formato CSV con los siguientes atributos de la dirección: tipo de vía, nombre de la vía, número del portal, código INE del municipio y coordenadas según el tipo de resultado (del portal, del centro de la vía, etc.)

Además, esta herramienta añade, junto a las coordenadas asignadas a cada registro, un campo con el **Peso** de la geocodificación. Éste estará comprendido entre 0 y 1 según la similitud del resultado obtenido con el buscado. Para ello, se hace una comparación elemento a elemento de la dirección (tipo\_de\_vía, nombre\_de\_vía y portal) y se puntúa la similitud existente en base a un peso asignado a cada elemento. Este indicador varía dependiendo de la exactitud del tipo de vía, nombre de la vía y número del portal.

Finalmente, se especifican las ventajas de esta herramienta:

- Es gratuita.
- Con ficheros de tamaño pequeño se puede realizar un proceso más rápido a través de la web con GeoDir.

### 3.4.3 Gestores de Bases de datos

La utilización de gestores de bases de datos para geocodificar es aconsejable cuando un fichero contiene entre su información alfanumérica las coordenadas X e Y. En este caso se pueden realizar enlaces entre el fichero a geocodificar y el que contiene la información de las coordenadas.

Estos enlaces, denominados determinísticos, se realizan a través de un campo común en ambos ficheros, es decir, un campo que contiene valores idénticos y por tanto sólo tiene un único resultado coincidente. Esta opción es adecuada cuando los registros que se quieren geocodificar contienen el campo de la *referencia catastral*. Este campo suele ser común y exacto a diferencia, por ejemplo, de la dirección postal que puede contener errores o el nombre de una misma vía puede estar escrito de forma diferenciada entre un fichero y otro.

Para realizar la georreferenciación a través de la referencia catastral se debe hacer uso de las bases de datos de la Dirección General de Catastro. En la siguiente URL se puede descargar diferente información y datos sobre cartografía catastral:

<https://www.sedecatastro.gob.es/OVCFrames.aspx?TIPO=TIT&a=masiv>

Las tablas descargadas contienen información de las parcelas catastrales de los municipios. Entre otros campos se encuentran la referencia catastral (que será el campo de comparación con el fichero a geocodificar) y las coordenadas X e Y del centroide de las parcelas.

El proceso de geocodificación con Bases de Datos consiste en unir los registros del fichero a geocodificar con el fichero de la Dirección General de Catastro a través de la referencia catastral a 14 dígitos. Cuando se enlazan los registros se incluye la información que contienen los ficheros de Catastro, por tanto se añade la coordenada X e Y del centroide de la parcela.

### 3.4.4 Herramientas SIG

A continuación se especifican una serie de herramientas de los Sistemas de Información Geográfica (SIG) que permiten geocodificar a partir de direcciones postales:

#### ARCGIS

El software ArcGIS constituye un sistema integrado completo que comparte la misma arquitectura de componentes (ArcObjects) con el fin de poder manipular, distribuir, crear y analizar la información geográfica. Usa estándares abiertos: COM, XML, SQL... para comunicarse con bases de datos y servidores. Está disponible para plataformas Windows.

Es el Sistema de Información Geográfica (SIG) de mayor difusión en la actualidad. Es un software propietario no gratuito. Bajo el nombre de ArcGIS Desktop (software de escritorio) se comercializan tres licencias: ArcInfo, ArcEdit y ArcView, que comparten un mismo núcleo y un número de funciones que varía de la versión más completa (ArcInfo) hasta la más simple (ArcView).

Mediante la caja de herramientas denominada "Geocodificación" integrada en ArcToolBox se pueden realizar las funciones de geocodificación de direcciones postales. Para geocodificar se requiere una tabla que almacene las direcciones que se desean geocodificar y un localizador de direcciones o un localizador de direcciones compuesto.

Esta herramienta hace concordar las direcciones con el localizador de direcciones y guarda el resultado para cada registro de entrada en una nueva clase de entidad de punto.

En el siguiente enlace se encuentra un tutorial para Geocodificación con ARCGIS:

[http://help.arcgis.com/es/arcgisdesktop/10.0/pdf/tutorial\\_geocoding.pdf](http://help.arcgis.com/es/arcgisdesktop/10.0/pdf/tutorial_geocoding.pdf)

#### QGIS

QGIS es un software libre y opera bajo la licencia GNU GPL. El software QGIS puede ser modificado libremente de tal manera que pueda realizar diferentes y más especializadas funcionalidades. Es compatible para plataformas GNU/Linux, Unix, Mac OS y Microsoft Windows. QGIS fue un proyecto desarrollado por la fundación OSGeo.

Existe una extensión denominada “Geocoding” que se instala en QGIS como un complemento y que utiliza servicios web de Google para realizar la geocodificación o la geocodificación inversa (convierte coordenadas en direcciones).

La geocodificación de las direcciones con esta aplicación se realiza introduciendo la dirección de manera manual, por lo que no permite realizar geocodificación de direcciones de forma masiva.

## **GEOMEDIA**

GeoMedia es una plataforma de gestión de SIG que permite agregar datos a partir de una variedad de fuentes y analizarlos en conjunto para extraer información. Proporciona acceso simultáneo a los datos geoespaciales, en casi cualquier forma, uniéndolos en una sola vista de mapa para un procesamiento eficiente, análisis, presentación y distribución. Además, cuenta con algunas funciones especializadas que lo hace ideal en el caso que se necesite extraer información de un conjunto de datos en constante cambio para apoyar decisiones más inteligentes.

El software de escritorio que incluye GEOMEDIA, GeoMedia Viewer, es una aplicación gratuita pero no libre. Entre sus funcionalidades, existe una extensión denominada “Geocoding” que permite realizar diferentes tipos de geocodificación (por ejes de calles, números de portal, bordes de manzana, referencias catastrales de parcelas, etc).

Para realizar la geocodificación mediante esta herramienta se deben especificar los siguientes atributos: dirección o intersección, ciudad, estado y código postal.

Los resultados obtenidos muestran, además de las variables introducidas, una columna con el peso de la valoración de la geocodificación que va de 0 (coincidencia perfecta) a 999 y NULL para las que no se han podido enlazar. También muestra información como el lado de la calle que contiene la geocodificación de la dirección del registro (derecha, izquierda e intersección entre dos o más calles).

No obstante, parece que esta geocodificación de direcciones que sigue el modelo americano no es la más óptima para las direcciones de España.

## **MAPINFO**

Es un software creado por Pitney Bowes Software (PBS) en la década de 1980 en Estados Unidos. Permite producir mapas en formato digital. MapInfo está diseñado en torno a una tarjeta de motor de edición que permite la superposición de capas digitales. Pueden realizarse representaciones mediante un sistema de capas de información geolocalizada: puntos, polígonos, imágenes de trama, etc. Asimismo, incorpora un gran número de formatos de datos, funciones cartográficas y gestión de los datos.

Entre sus herramientas, Mapinfo incorpora "Geocode", que es muy similar a la utilizada por GeoMedia.

La geocodificación o asignación de coordenadas que realiza Mapinfo se realiza mediante la comparación de la información geográfica de una tabla de su base de datos con otra tabla que ya tiene asociadas las coordenadas X e Y. La tabla que posee las coordenadas se denomina tabla de búsqueda. La tabla a geocodificar es la tabla de destino, ya que en ésta se colocarán los objetos de punto, y la tabla que contiene las coordenadas geográficas es la tabla de origen. Para realizar la geocodificación deben tenerse en cuenta las siguientes consideraciones: abrir al menos una tabla representada en un mapa que será la tabla fuente y que contiene las coordenadas X e Y; entre la tabla (a geocodificar) y el mapa debe haber un campo común y en la tabla debe estar el campo en común indexado.

La geocodificación puede ser automática, interactiva y manual.

Cuenta con el mismo problema que GeoMedia, ya que al ser un programa americano, sólo geocodifica bien las direcciones americanas.

### 3.4.5 Telegeo

La aplicación Telegeo (Tramitación Electrónica Georreferenciada) es una aplicación Web que permite la búsqueda y localización de un punto determinado del territorio de la Comunidad Autónoma obteniendo sus coordenadas normalizadas. A través de su visor web un usuario no especializado puede ubicar un elemento puntual y obtener su georreferenciación de diversas formas, entre las que se encuentran un fichero de texto con las coordenadas normalizadas, la publicación de un mapa web con la ubicación del elemento que puede ser visualizado automáticamente a través de una URL, la creación de este mapa en formato PDF, o la generación de un fichero en formato KML que puede ser incorporado al visor de Google Earth.

A ese respecto se requiere que estas coordenadas se expresen de forma normalizada en el conjunto de Sistemas de Información Geográfica de la Junta de Andalucía, asegurándose así su interoperabilidad, es decir, la posibilidad de ser integradas y comparadas con el resto de sistemas que siguen los estándares europeos (Directiva INSPIRE), nacionales (LISIGE) y los propios de la Comunidad Autónoma (Decreto141/2006 y Norma Técnica\_01008\_Módulo Geodésico de Referencia y Altitudes).

Esta herramienta permite realizar búsquedas de una localización y de manera manual permite ubicar el lugar en el punto de interés exacto para obtener así sus coordenadas correspondientes. Así mismo, la herramienta ofrece información adicional, como información catastral o una URL codificada, para acceder a un visor cartográfico con las capas seleccionadas y además permite visualizar la localización realizada sobre la ortofotografía y la cartografía catastral.

Telegeo proporciona coordenadas geográficas en el sistema de referencia espacial oficial ETRS89.

Esta aplicación del Instituto de Estadística y Cartografía de Andalucía está accesible en la URL (ver figura 13):

[www.juntadeandalucia.es/institutodeestadisticaycartografia/telegeo](http://www.juntadeandalucia.es/institutodeestadisticaycartografia/telegeo)

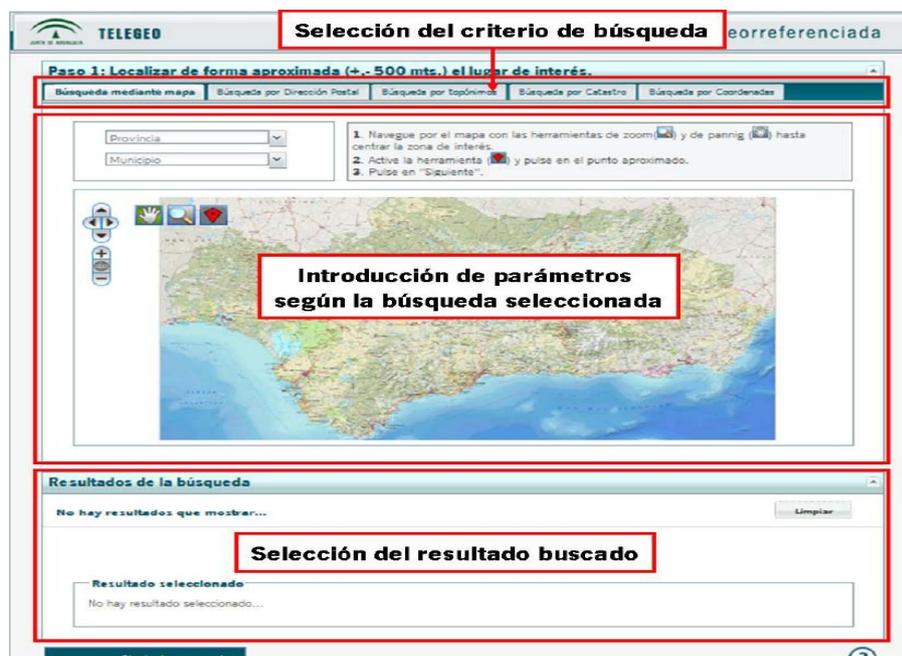


Figura 13.- TELEGEO

El proceso de localización mediante Telegeo se realiza en tres pasos:

### **Paso 1: Ubicación aproximada del punto de interés**

Durante esta etapa el usuario realiza una primera aproximación al lugar de interés, mediante alguno de los siguientes criterios:

- búsqueda mediante mapa: se realiza cuando se conoce dónde está ubicado el registro y se puede localizar en la ortografía del mapa
- búsqueda por dirección postal: se realiza cuando tenemos información sobre esta variable
- búsqueda por topónimos
- búsqueda por catastro
- búsqueda por coordenadas

La finalidad de este paso es que el usuario ubique el visor en las proximidades del punto de interés, antes de realizar la localización precisa en el paso siguiente, es decir, obtener una primera aproximación en torno a 500 metros de distancia, no situar la ubicación exacta.

### **Paso 2: Localización de precisión (afinación del punto)**

En este paso el usuario sitúa con exactitud la ubicación del elemento sobre la ortofotografía de 0,5 metros de resolución y la cartografía catastral 1:500, a través de un visor cartográfico.

El punto que se localiza en el primer paso es desplazado por el usuario de manera manual hasta el lugar exacto donde se quiere ubicar.

### **Paso 3: Generación de la coordenada afinada, mapas e informes**

A partir de la ubicación definida en la etapa anterior, el usuario, además de las coordenadas normalizadas, puede generar una serie de salidas relacionadas con el lugar de interés localizado: un archivo de texto; un informe con información adicional de las coordenadas buscadas; un enlace directo a un visor cartográfico donde está la información utilizada y generada durante el proceso y un fichero en formato KML útil para la aplicación Google Earth.

Es conveniente usar Telegeo en aquellos casos que se quiera aumentar la exactitud de una localización inicial o **afinar** mucho más la coordenada obtenida mediante otras herramientas o procesos. Por ejemplo, pasar la coordenada del centroide de la parcela al portal de la misma.

La principal **ventaja** de esta herramienta radica precisamente en la facilidad de uso y en la exactitud que se puede conseguir en la ubicación de registros. Además, con Telegeo se obtienen las coordenadas de un punto normalizadas usando el CRS oficial indicando además el tipo de coordenadas y su precisión.

Otra ventaja es que ayuda a mejorar la exactitud posicional de un elemento cuando se tiene geolocalizado a un portal cercano o al centroide de la parcela y se quiere obtener la coordenada exacta de entrada al portal del inmueble.

Con Telegeo también es posible realizar una geocodificación inversa, si se conoce la coordenada puede obtenerse la dirección postal, referencia catastral, provincia y municipio.

Los procesos para georreferenciar a través de esta herramienta se realizan de manera manual y de uno en uno, por lo que la herramienta es útil en casos en los que se quiera afinar o cuando el fichero de datos contenga pocos registros.

## 4. CONTROL DE CALIDAD DE LA INFORMACIÓN GEORREFERENCIADA

El control de calidad es la determinación, de manera estadística y objetiva, de las cualidades del producto ya acabado, sin ocuparse del proceso de producción. Esta calidad viene determinada por la medida en que el producto satisface las necesidades requeridas por el usuario.

De manera general, debe garantizarse que los procesos de trabajo y los productos generados cumplan con los objetivos realizando un control de calidad.

La evaluación de la calidad de la información geográfica está sometida a estándares internacionales, establecidos en el marco de la ISO (International Organization for Standardization). En concreto, son de aplicación las normas ISO nº 19113, 19114, 19131 y 19138.

En el caso de la georreferenciación puntual de registros, la garantía de que el producto final es el adecuado y que se ajusta a las necesidades del usuario depende, por un lado, de que la información que se utiliza como referencia sea lo más óptima posible, y por otro, que el resultado tras el proceso de georreferenciación ofrezca los resultados adecuados. Por ello, el control de calidad viene determinado por dos procesos:

**1.-** La información georreferenciada depende de la fuente de referencia que se utiliza para la obtención de las coordenadas X e Y. La fuente de donde se obtienen las coordenadas debe ser óptima y debe, por tanto, pasar un control de calidad para que se pueda garantizar la idoneidad de la información que se utiliza como base para la georreferenciación.

En esta guía ya se han descrito las dos fuentes de información de donde se obtienen las coordenadas X e Y para la geocodificación de registros. Para ambas fuentes, CDAU y Catastro, el control de calidad queda garantizado como tal por los procedimientos exhaustivos a los que se someten ambas fuentes en sus procesos de elaboración.

En el caso de la utilización del CDAU como información de referencia, queda garantizada la calidad de la información a nivel de portal y vía, ya que éste se somete a un exhaustivo control de calidad en el que se evalúan todos los aspectos a tener en cuenta para garantizar que la información es adecuada:

- **Completitud:** se controla el exceso o ausencia de datos.
- **Consistencia lógica de formato, de concepto, de dominio y topológica:** se analiza el grado de cumplimiento de las reglas lógicas de la estructura de datos, atributos y relaciones; se comprueba que los datos que se han entregado se ajustan a las características del modelo de datos aprobado

para la realización del CDAU, que los datos alfanuméricos que forman parte del CDAU siguen los parámetros determinados en la elaboración del modelo de datos y si éstos cumplen con los requisitos establecidos; se lleva a cabo una evaluación de aquellos valores que entran a configurar parte de los datos en forma de dominios ajustándose a las características del modelo de datos y se realiza un análisis de las relaciones topológicas comprobando que los datos geográficos de las capas se ajustan a las normas de calidad establecidas previamente.

- Exactitud posicional: se realizan distintas verificaciones mediante la delimitación de zonas de muestreos, asegurando que la información geométrica cumple los márgenes de error previstos en los pliegos técnicos.
- Exactitud temática: se lleva a cabo una evaluación completa del nivel de calidad de los atributos alfanuméricos asociados a cada una de las tablas que forman el CDAU.

Por tanto, el control de calidad y la adecuación de la información de base para la georreferenciación quedan garantizadas al utilizar las fuentes de datos descritas en esta guía.

2.- En segundo lugar, deben someterse a un control de calidad los datos resultantes de la georreferenciación para garantizar que el proceso ha dado resultados óptimos.

En este caso los controles de calidad a aplicar, así como la exigencia de los mismos, dependen del uso que se vaya a dar a la información cartográfica. Por tanto, los procedimientos de control de calidad deben establecerse específicamente para cada caso, además de las exigencias requeridas por cada usuario para establecer como buena la geocodificación.

El dato geográfico se caracteriza por una posición espacial, sus atributos y el tiempo en el que suceden. En el caso que nos ocupa son especialmente relevantes las características de exactitud posicional y completitud, ya que las demás han quedado ya establecidas en el control de calidad de la fuente de base utilizada.

En la siguiente tabla se recogen los elementos y subelementos a medir en una base con datos geográficos procedente de un proceso de georreferenciación puntual.

| ELEMENTOS DE CALIDAD        | DESCRIPCIÓN DE LOS ELEMENTOS                                  | SUB-ELEMENTOS                | DESCRIPCIÓN DE LOS SUBELEMENTOS  |
|-----------------------------|---|------------------------------|--|
| <b>Exactitud Posicional</b> | Exactitud de la posición de los elementos                     | Exactitud absoluta o externa | Proximidad entre los valores de coordenadas reportados y los valores verdaderos aceptados como tales |
| <b>Completitud</b>          | Presencia y ausencia de fenómenos, sus atributos y relaciones | Comisión                     | Datos excedentes   |
|                             |   | Omisión                      | Datos ausentes   |

La **exactitud posicional** hace referencia a la exactitud planimétrica y altimétrica; en el caso que nos ocupa se trata de exactitud planimétrica absoluta, es decir, la proximidad entre los valores de coordenadas reportados y los valores aceptados como verdaderos. Es importante señalar que el valor aceptable es definido siempre por el usuario. Se determina mediante el análisis de una muestra significativa de puntos. Para ello, existen numerosos test para medir la exactitud posicional teniendo en cuenta el tamaño de la muestra, la distribución espacial de la misma y la exactitud de los puntos de control.

El tamaño de dicha muestra dependerá de la capacidad de revisión que se tenga, aunque será suficiente con muestras de entre el 5% y el 10%. Se debe tener en cuenta que las recomendaciones establecen siempre muestras de al menos 20 puntos, independientemente del tamaño del área a controlar o del tamaño de la base de datos geográfica. La muestra debe estar repartida de manera homogénea en la zona a controlar.

El método a utilizar es definido por el usuario dependiendo de la disponibilidad de información de referencia, o según el conocimiento en análisis de información geográfica. Un método simple de control de la exactitud posicional está basado en la observación directa de la información georreferenciada. Sobre una base cartográfica se observarán los puntos escogidos de la muestra y se comprueba si éstos están situados en una posición aceptable.

Otro método basado en información de referencia consiste en realizar un análisis geográfico para comprobar si los puntos georreferenciados están situados en la posición adecuada con respecto a la información de referencia. La información de referencia pueden ser las capas (shp) con geometrías comparables a los puntos de nuestro estudio y que, por tanto, pueden usarse para analizar si la información es correcta o no. Por ejemplo, se puede observar si todos los puntos que se han georreferenciado coinciden con puntos del portal del Callejero o están contenidos dentro de una parcela de Catastro. En este ejemplo, si algún punto no se corresponde con alguna de esas geometrías, entonces debería estudiarse para comprobar si es un error (a no ser que se traten de los puntos afinados).

La **completitud** indica en qué medida de omisión o de comisión los datos presentes cumplen con las prescripciones establecidas, es decir, la presencia de elementos que no deberían estar presentes o la ausencia de otros que si deberían estarlo. En este caso concreto, todo elemento que es georreferenciado debe tener la coordenada X e Y (a no ser que no sea geocodificable) y no tiene que haber elementos o coordenadas que no pertenecen a la fuente a geocodificar.

Siempre que se realice un control de calidad deben especificarse los procedimientos seguidos para establecer la calidad de la geocodificación.

## 5. TRATAMIENTO DE LA INFORMACIÓN GEOCODIFICADA

### 5.1 Sistemas de Información Geográfica (SIG)

Los Sistemas de Información Geográfica (SIG) son herramientas que permiten visualizar información geográfica, realizar consultas interactivas, editar datos o analizar relaciones espaciales, posibilitando la generación de resultados de carácter cartográfico. Se pueden distinguir tres tipos de SIG muy definidos:

- a) Los SIG de escritorio (clientes pesados). Son aquellos diseñados para ser utilizados en nuestros ordenadores particulares.
- b) Los SIG web (clientes ligeros). Un cliente, es decir, el usuario, accede a la información mediante internet o una red local. Dependiendo de la complejidad del SIG, se puede o bien añadir información (capas), o incluso algunas operaciones de análisis espacial.
- c) SIG móviles. Los SIG móviles son aquellos que están diseñados para usarlos en dispositivos móviles para poder trabajar en campo.

Este apartado se centra en los software de escritorio, que son los que ofrecen un mayor número de funcionalidades. Son muchos los existentes en el mercado que permiten cargar la información geocodificada para poder realizar análisis de datos alfanuméricos asociados a una componente espacial y/o realizar operaciones sobre imágenes existentes. El tratamiento de la información geocodificada y los análisis geográficos o espaciales realizados con estas herramientas facilitan la toma de decisiones al gestor de la información. Algunas de estas herramientas se detallan a continuación (algunos libres y de código abierto distribuidos bajo una licencia GNU, y otros de carácter comercial, distribuidos con licencia de pago).

#### 5.1.1 QGIS

Es un software libre y gratuito. QGIS pretende ser una herramienta de fácil uso de los SIG, proporcionando funciones y características comunes a otras herramientas de este ámbito. El objetivo inicial fue proporcionar un visor de datos SIG. QGIS ha alcanzado el punto en su evolución en el que está siendo utilizado por muchos usuarios en sus necesidades de visualización en sus datos diarios SIG.

Una de las grandes fortalezas de QGIS es que trabaja en cualquiera de los sistemas operativos: GNU/Linux, BSD, Unix, Mac OSX, Windows y Android, funcionando de manera similar en todos ellos. Está desarrollado utilizando el conjunto de herramientas Qt (<http://qt.nokia.com>) y C++.

Además, una de las grandes versatilidades de QGIS es su facilidad de interconexión con muchas bases de datos geoespaciales como PostGIS, SpatialLite, ORACLE Spatial y otras. QGIS tiene una conexión directa

con bases de datos SpatialLite y existen varios plug-ins (complementos) que facilitan el manejo de esta base de datos.

Con esta herramienta se puede visualizar, administrar, editar, realizar análisis espaciales y componer mapas. Funciona con un gran número de formatos vectoriales y ráster. Entre sus funciones se encuentran: soporte wms, búsqueda de atributos, transparencia de capas vectoriales y mejora antialiasing, una potente funcionalidad de análisis mediante la integración con grass, mejoras en la edición de capas vectoriales y posibilidad de editar shapefiles y capas ogr.

Para conocer información sobre el proyecto QGIS y descargar la aplicación se puede consultar la URL:

<http://www.qgis.org/es/site/>

### 5.1.2 ARCGIS

ArcGIS Desktop consta de un conjunto de aplicaciones: ArcMap, ArcCatalog, ArcGlobe, ArcScene, ArcToolbox y ModelBuilder. Con estas aplicaciones se puede realizar cualquier tarea de SIG, simple o avanzada.

Bajo el nombre ArcGIS Desktop se comercializan tres licencias: ArcInfo, ArcEditor y ArcView, que comparten un mismo núcleo y un número de funciones que varía de la versión más completa (ArcInfo) hasta la más simple (ArcView). Cada una de ellas está compuesta por dos aplicaciones diferentes:

- ArcMap (semejante a ArcView 3.x) y ArcCatalog (semejante al Explorador de Windows).

Por otra parte, integrado en éstos dos se encuentra ArcToolbox (conjunto de herramientas de conversión y análisis de datos). Empleando estas tres aplicaciones juntas se puede realizar cualquier tarea SIG: creación, edición, análisis y representación de información geográfica.

De manera resumida, la familia ArcGIS se compone de los productos que se describen a continuación:

- ArcReader: es un visualizador de mapas gratuito con el que se puede visualizar, consultar e imprimir mapas de forma sencilla.
- ArcGIS ArcView: es la siguiente generación de la versión ArcView 3.x. Comprende las aplicaciones ArcCatalog y ArcMap, ambas con una versión reducida de ArcToolBox.
- ArcGIS ArcEditor: está diseñado para crear y editar geodatabases. Abarca una versión completa de ArcMap, incluyendo todo el editor de objetos y ArcCatalog. Con el ArcEditor se pueden crear y modificar bases de datos y esquemas de bases de datos para ficheros shapefiles, coverages, geodatabases personales y geodatabases corporativas.

- ArcGIS ArcInfo: es el producto más completo de la familia ArcGIS. Incluye todas las aplicaciones de ArcView y ArcEditor, con todas las herramientas disponibles de ArcToolbox y una versión completa de ArcInfo Workstation (Arc, ArcEdit, ArcPlot, Arc language y todas las extensiones). Permite crear datos, actualizarlos, consultarlos, hacer mapas y realizar análisis.
- ArcSDE: permite trabajar con datos geográficos almacenados en un sistema de gestión de bases de datos relacionales como Oracle, Microsoft SQL Server, IBM MD2 o Informix. ArcSDE gestiona el intercambio de información entre ArcGIS y la base de datos relacional, permitiendo a varios usuarios compartir y editar de forma simultánea los datos geográficos almacenados en un servidor.
- ArcIMS: es una herramienta que permite publicar a través de internet mapas y datos incluyendo herramientas de visualización, consultas a la base de datos, simbolización y análisis sin que el usuario tenga necesidad de tener instalado ningún software SIG en su ordenador.
- ArcPad: herramienta para el mapeo móvil, orientada a proyectos de captura de datos y recogida de la información geográfica mediante posicionamiento GPS en campo. ArcPad añade a los dispositivos móviles (Pocket PC, Tablet PC, etc.) funcionalidades GIS y GPS, proporcionando la integración de acceso a bases de datos y sistemas de posicionamiento global en cualquier parte en tiempo real.

### 5.1.3 GVSIG

Es un Sistema de Información Geográfica (SIG) diseñado para capturar, almacenar y manipular la información geográfica. Con un formato sencillo y compatible con la mayor parte de los soportes manejados por los SIG (maneja datos vectoriales y ráster), se convierte en otra potente herramienta para efectuar análisis alfanuméricos con una componente espacial. Es software libre, gratuito, creado bajo la arquitectura Java, con licencia GNU/GPL, lo que permite su uso y mejora por los usuarios, para plataformas Linux, Windows y Mac OS X. Su gran número de extensiones ofrece un gran número de funcionalidades. Es un software libre y gratuito.

Las principales características como cliente de sobremesa son:

- Proyecto Open Source.
- Acceso a los formatos más usuales tanto vectoriales como ráster.
- Importante implementación de servicios OGC: WMS, WFS, WCS, WFS-T y WFS-G.
- Elevado grado de aceptación e implantación: implantación en Administraciones públicas (tanto en Andalucía, España como a nivel mundial).
- Interfaz amigable e intuitiva: disponible en varios idiomas.



#### 5.1.4 KOSMO

El proyecto Kosmo es la primera plataforma SIG libre corporativa, distribuida bajo licencia GNU/GPL. Kosmo es una herramienta desarrollada a partir de la plataforma JUMP (Java Unified Mapping Platform, <http://www.vividsolutions.com/jump>) y de la librería de componentes SIG Geotools (<http://www.geotools.org>). Existen versiones estables tanto para Windows como Linux. Es un software libre y gratuito.

En él se hace uso intensivo y se desarrollan y/o integran las herramientas necesarias para satisfacer las necesidades de la mayoría de los usuarios (desde muy básicos hasta muy complejos).

Su diseño y arquitectura están basados en la gestión y análisis de la información territorial a través de Bases de Datos Espaciales, dotándolo así de carácter corporativo.

El proyecto está en pleno desarrollo y el primero de sus componentes Kosmo-Desktop sigue en continua evolución, estando ya disponible para aquellos que requieran de una funcionalidad avanzada en un SIG de escritorio potente.

Se debe destacar que a día de hoy ya se encuentra implementado en numerosos sistemas de producción altamente exigentes en sus requisitos tanto de estabilidad como de funcionalidad.

La siguiente tabla muestra las características generales de los SIG de escritorio presentados:

| CARACTERÍSTICAS                              | CLIENTES PESADOS  |   |   |   |
|--|---|---|---|---|
|  | QGIS  | ArcGIS  | gvSIG   | Kosmo   |
| Estándares OGC                               | WMS/WMTS, WCS y WFS.                                    | WMS, WMC, WFS, GML Y CSW  | WMS,WFS/WFS-T, WCS, WPS                                   | WMS, WFS  |
| Formatos Ráster Soportados                   | GRASS GIS, GeoTIFF, TIFF, JPG, etc.                     | ECW, JPG, MrSID, TIFF, geoTIFF, ArcIMS, IMG(Erdas), RAW, etc.                         | BMP, GIF, TIF, TIFF, JPG, JPEG, PNG, VRT, ADF, NOS, etc.  | ECW, MrSID, TIFF, ECW, PNG, BMP, GIF, etc.                  |
| Formatos Vectoriales Soportados              | ESRI SHP, ArcInfo coverages, Mapinfo, GRASS GIS, etc.   | ESRI SHP, DXF, DWG(2000), ArcIMS, DGN(v7).  | ESRI SHP, GML, KML,DXF, DWG, DGN                          | ESRI SHP, DXF   |
| Conectividad a Bases de datos Relacionales   | PostgreSQL/PostGIS, SpatialLite, Oracle Spatial, MySQL  | Oracle Spatial PostgreSQL/PostGIS, DB2, Informix y SQL Server.                        | Oracle Spatial PostgreSQL/PostGIS, MySQL, JDBC            | Oracle Spatial, PostgreSQL/PostGIS y MySQL.                 |
| Orientación del SIG                          | Vectorial/ Ráster                                       | Vectorial / Ráster  | Vectorial / Ráster  | Vectorial / Ráster  |
| Lenguaje Documentación de Soporte en Español | Si  | Si  | Si  | Si  |
| Última versión                               | 2.8.1   | 10.3  | 2.1   | 3.0   |
| Desarrolladores                              | OSGeo   | ESRI  | Asociación gvSIG  | SAIG  |
| Lenguaje de desarrollo                       | C++, Python, Qt   | Python  | Java  | Java  |
| Página oficial                               | <a href="http://www.qgis.org/">http://www.qgis.org/</a> | <a href="http://www.esri.com/software/arcgis">http://www.esri.com/software/arcgis</a> | <a href="http://www.gvsig.com/">http://www.gvsig.com/</a> | <a href="http://www.opensig.es/">http://www.opensig.es/</a> |

Cuadro 6.- Sistemas de Información Geográfica

## 5.2 BASES DE DATOS ESPACIALES

Tradicionalmente las operaciones de análisis espacial tanto ráster como vectorial se han efectuado de forma local usando un software SIG de escritorio. En la actualidad, las **bases de datos espaciales** se han convertido en una alternativa a los softwares SIG tradicionales y SQL es el lenguaje que utilizan los Sistemas de Gestión de Bases de Datos (SGBD) para gestionar sus bases de datos tanto alfanuméricas como espaciales.

Entre las ventajas de utilizar una base de datos espacial se tienen:

- Permiten definir las relaciones espaciales existentes entre los objetos geográficos (topología).
- Mejoran la seguridad de los datos, ya que permiten la creación de permisos a usuarios y grupos de usuarios.
- Proveen facilidades para la manipulación de grandes volúmenes de datos.
- Organizan los datos de manera íntegra evitando redundancias.

Entre los sistemas gestores de bases de datos que existen actualmente en el mercado se pueden destacar dos, uno libre y otro comercial.

Dentro del mundo del software libre, **PostgreSQL** y su extensión espacial **PostGIS** se han convertido en uno de los sistemas gestores de bases de datos más utilizados, empleándose en múltiples proyectos y aplicaciones SIG tanto de escritorio como web. PostGIS se publica bajo la Licencia Pública General de GNU y actualmente es un producto consolidado que ha demostrado su eficiencia. A juicio de muchos, es muy similar a la versión geográfica del sistema gestor de bases de datos comercial Oracle, **Spatial Oracle**. Además, otro aspecto a tener en cuenta es que ha sido certificado en 2006 por el Open Geospatial Consortium (OGC), lo que garantiza la interoperabilidad con otros sistemas.

### 5.3 POSIBILIDADES ANALÍTICAS QUE OFRECE LA INFORMACIÓN GEOCODIFICADA

La geocodificación de los registros de un fichero posibilita la generación de su correspondiente archivo en formato SHAPE (.shp), el más utilizado cuando se trabaja con información geográfica vectorial. Ello permite su inmediata integración en cualquier SIG de escritorio, software que facilita el almacenamiento, la gestión y el análisis de la información espacial georreferenciada.

Una de las principales características de los SIG es que, aunque almacenan y representan la información espacial mediante distintas capas temáticas, éstas pueden ser relacionadas entre sí no sólo a través de posibles campos en común sino también mediante sus características topológicas o geométricas (contigüidad, inclusión, proximidad, etc.). Consecuentemente, la geocodificación de nueva información temática supone siempre un importante valor añadido porque, además de que es posible analizar sus propias características espaciales, permite ponerla en relación con otras capas temáticas que ya se encuentran definidas con el fin de poder realizar análisis espaciales entre ellas (ver figura 14).

A través de estos análisis se pueden establecer relaciones de proximidad entre distintas capas (áreas de influencia, cálculo de distancias, etc.), análisis de redes (cálculo de rutas, conectividad, etc.), entre otras.

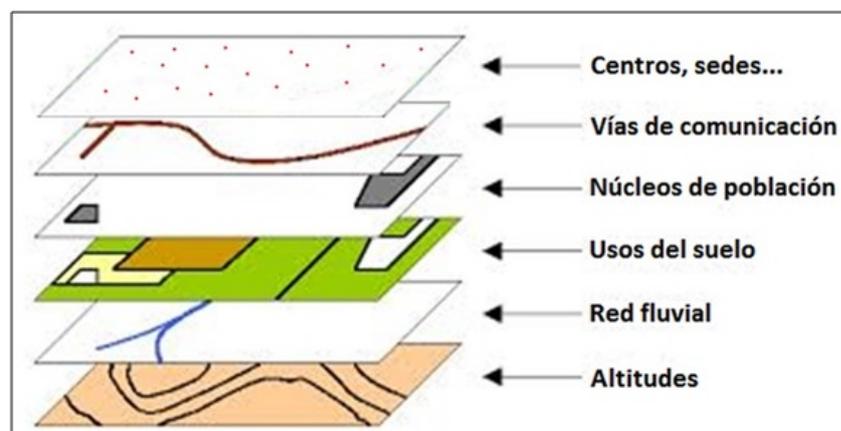


Figura 14.- Ejemplo de posibles capas temáticas

Debido a todas estas potencialidades, los SIG tienen su aplicación en campos disciplinares muy dispares: urbanismo, marketing, desarrollo rural, gestión de equipamientos sociales, medio ambiente, obras públicas, etc.

### 5.3.1 Tabla de atributos y selección de elementos

Los SIG de escritorio, que administran y combinan de manera integrada la información geográfica, proporcionan un conjunto amplio de herramientas analíticas que rebasan las funcionalidades básicas de cualquier visualizador. Los archivos SHAPE, que son los que almacenan la información geométrica de los elementos de las capas, tienen siempre asociadas una tabla de atributos (con extensión .dbf) que puede incluir diversos campos con información alfanumérica. Ello permite tanto conocer los atributos de un elemento seleccionándolo en el mapa (figura 15, a), como ubicarlo geográficamente seleccionándolo en la tabla de atributos (figura 15, b).

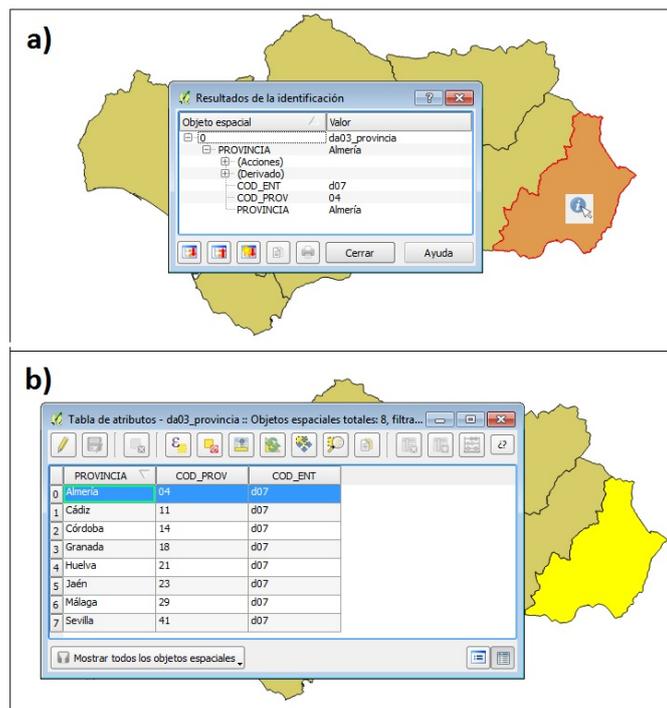


Figura 15.- a) Identificación de elementos sobre el mapa; b) Selección de elementos en la tabla de atributos

Una de las grandes potencialidades que ofrecen los SIG de escritorio es la posibilidad de seleccionar elementos según sus atributos o según la localización de sus elementos con respecto de otras capas temáticas. Por ejemplo, disponiendo de una capa puntual de minas, en la que uno de los atributos fuera el volumen de producción anual, podrían seleccionarse aquellas que superaran un determinado umbral de esa variable. Además, mediante la selección por localización podríamos añadir la condición de que tales

minas estuvieran situadas en un parque natural (para lo que, obviamente, sería necesario disponer de una capa de parques naturales).

Cualquier análisis (espacial o estadístico) podrá realizarse de acuerdo a los elementos seleccionados. Además, también es posible generar nuevas capas de información a partir de una selección efectuada.

### **5.3.2 Obtención de información a partir de relaciones espaciales entre capas (Análisis espacial)**

La existencia de relaciones espaciales supone una ventaja fundamental a la hora de asignar información alfanumérica procedente de otros ficheros, especialmente en aquellos casos en los que no existen campos en común. Por ejemplo, sería posible asignar códigos de secciones censales a una capa puntual de monumentos de una región. En este caso concreto podría realizarse una operación de enlace espacial entre la capa de monumentos y la capa de secciones censales, cuyo resultado sería una capa puntual donde la tabla de atributos contendría la información original de la capa de monumentos y la procedente de la capa de secciones censales. De esta forma se sabría a qué sección censal pertenece cada monumento (ver figura 16).

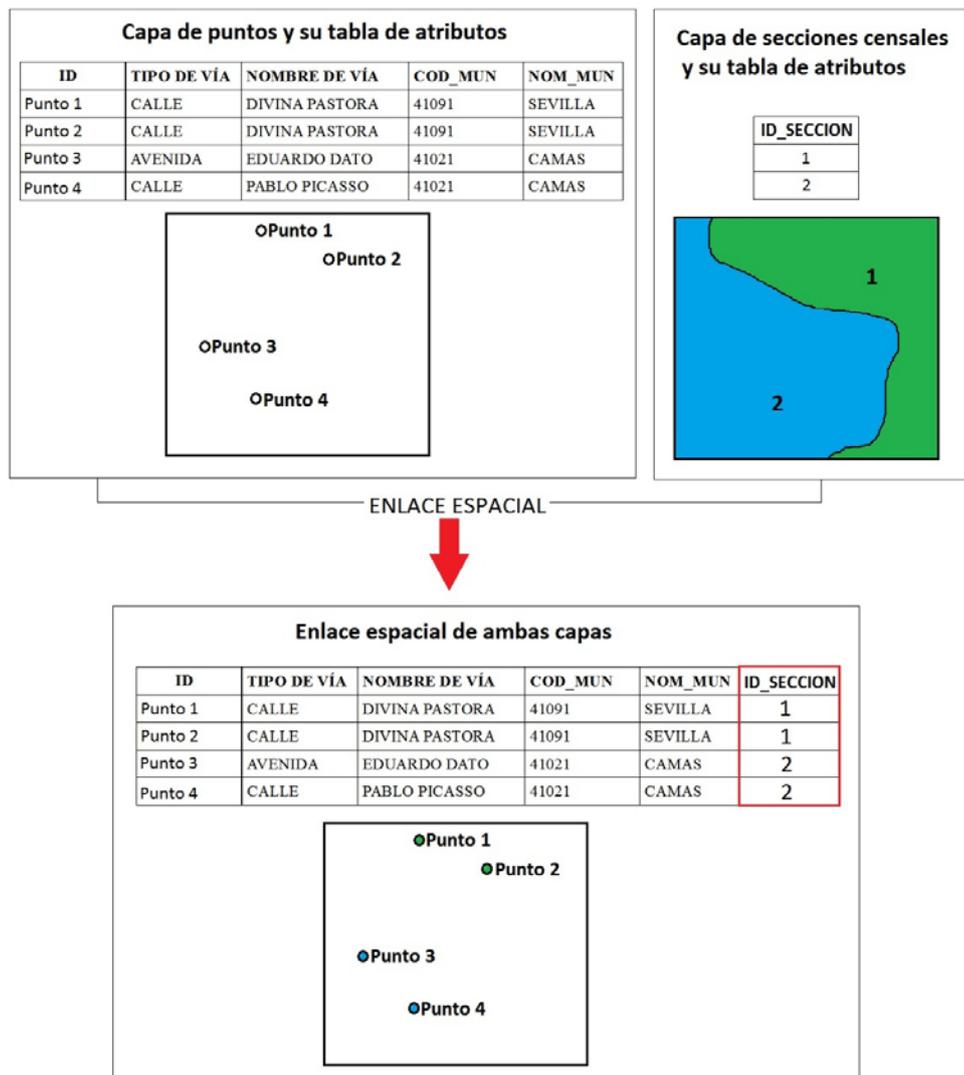


Figura 16.- Ejemplo de asignación de información alfanumérica a partir de relaciones espaciales entre capas

El abanico de posibilidades de análisis espacial es realmente amplio. A continuación se muestran algunas de las operaciones espaciales más habituales en el ámbito de los SIG, especialmente en relación a la utilización de capas de información puntuales.

### Análisis de superposición

Los análisis de superposición solventan uno de los problemas más habituales en cartografía: el conocimiento de las relaciones *de contacto* entre distintas variables o fenómenos que se dan sobre un mismo ámbito geográfico. Por ejemplo, sería posible saber si existen núcleos urbanos sobre zonas inundables, identificar los centros de observación ornitológica que se encuentran situados dentro de

parques naturales, conocer los usos del suelo en los que se ubican los parques eólicos (ver figura 17), etc.

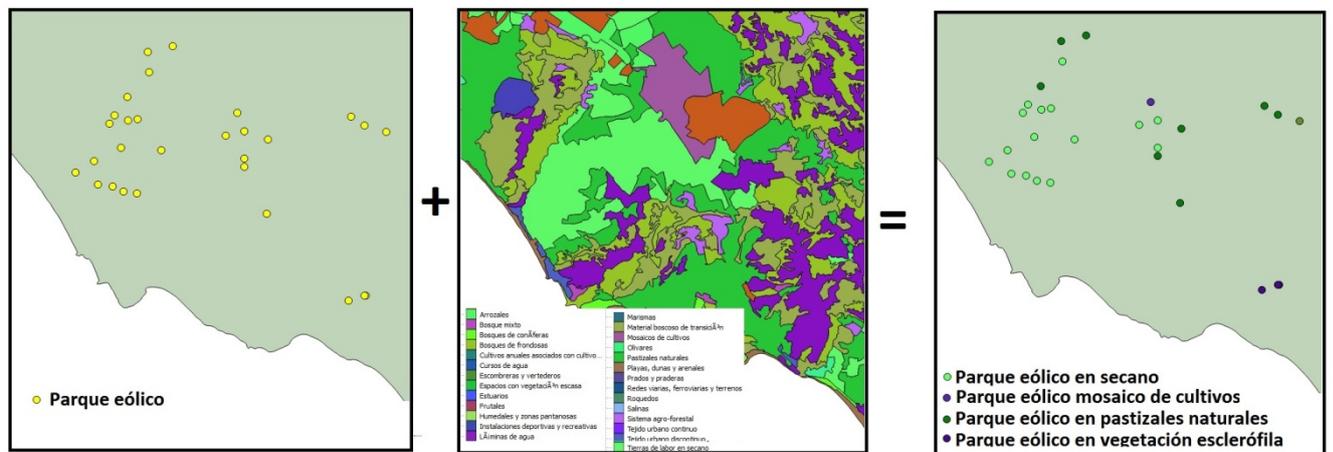


Figura 17.- Intersección de capas

### **Análisis de proximidad**

Además de las relaciones de contacto entre capas, el otro conjunto fundamental de operaciones espaciales es el que permite realizar análisis de proximidad. Así, por ejemplo, a partir de un elemento seleccionado se pueden identificar todos aquellos que se encuentran a una distancia dada de éste. En definitiva, los análisis de relaciones de proximidad permiten resolver cuestiones del tipo “¿qué está cerca de...?”.

Existen diferentes herramientas que se basan en la proximidad de los elementos para realizar estos análisis, entre los que pueden destacarse las áreas de influencia y el análisis de redes.

### **Áreas de influencia**

Un área de influencia se define como el espacio geográfico circundante en torno a un elemento dado.

Éstas se pueden construir a partir de elementos puntuales, lineales y poligonales y se suelen utilizar para analizar las entidades que se encuentran dentro de una distancia crítica de algún elemento o ámbito. Por ejemplo, podrían construirse las áreas de influencia de estaciones de tratamiento de residuos para analizar su posible impacto sobre parques naturales (ver figura 18).

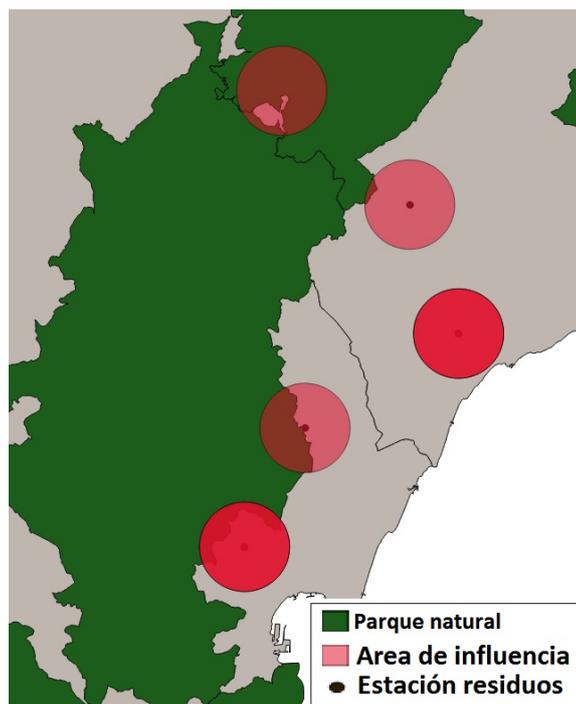


Figura 18.- Zonas de influencia

Además, se pueden construir zonas de influencia en anillos múltiples para clasificar las áreas de alrededor de una entidad en clases de distancia cercana, distancia moderada y distancia larga para un análisis, es decir, se pueden crear entidades en varias distancias específicas alrededor de una entrada (ver figura 19).

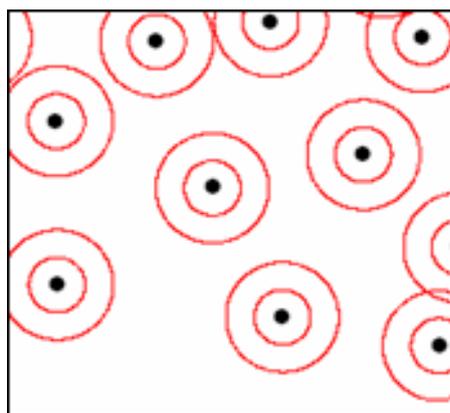


Figura 19.- Zonas de influencia de anillos múltiples

### Análisis de redes

Algunos análisis de proximidad requieren que las mediciones se realicen en términos de coste de desplazamiento, y no únicamente a partir de la distancia. El movimiento de personas, el transporte de bienes y servicios, las comunicaciones y los flujos de energía pueden ser estudiados mediante el análisis de redes de una forma más fidedigna, permitiendo la identificación del camino óptimo entre dos localizaciones teniendo en cuenta las vías de comunicación (ver figura 20), la asignación de servicios de emergencia según su accesibilidad, etc.



Figura 20.- Cálculo del camino óptimo mediante análisis de redes

Estos análisis necesitan de elementos interconectados mediante líneas, como las carreteras, las vías, los ríos, etc. Estas conexiones se pueden considerar redes.

Con estos análisis también es posible localizar los elementos más cercanos a uno o más servicios, ordenándolos en función del coste empleado en alcanzarlos y caracterizando la relación tanto por ese coste como por la longitud de la red que los separa.

Otra particularidad es la conectividad entre tramos de una línea, que permite identificar los tramos de red que conectan a través de un punto dado (ver figura 21). De este modo puede localizarse una zona de avería, aislarla, reconfigurar la red, seleccionar los usuarios afectados, calcular el consumo previsto, etc.

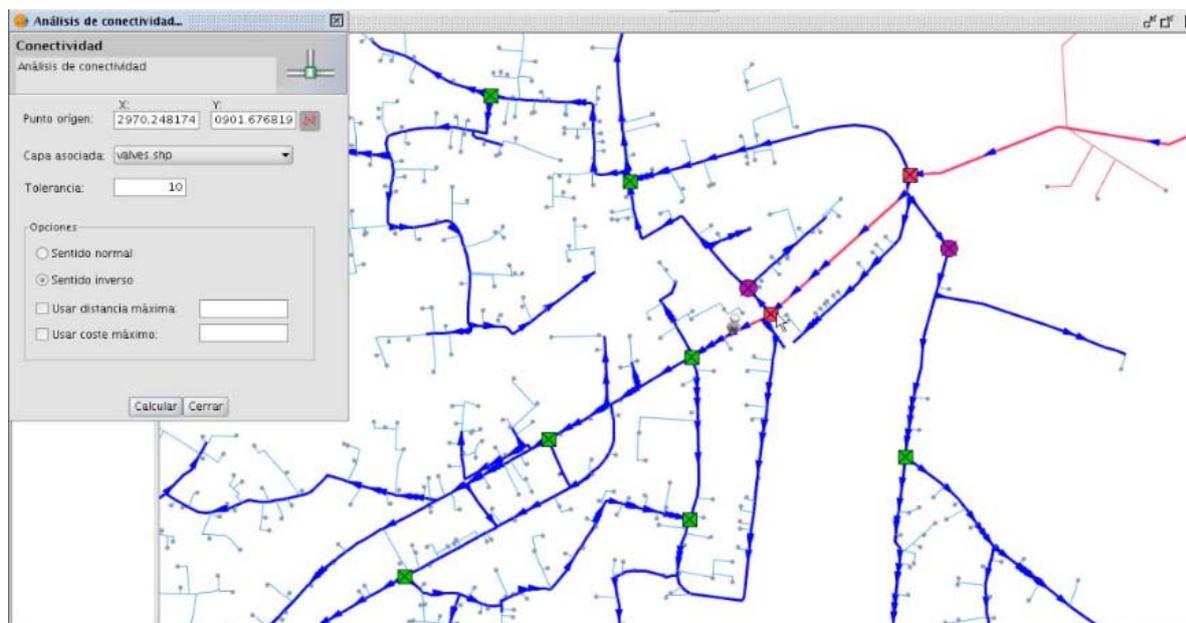


Figura 21.- Análisis de conectividad

### Creación de superficies

Existen muchas variables que varían espacialmente de forma continua. Es el caso de las variables climáticas, el relieve, el tipo de suelo, etc. Los SIG disponen de herramientas que permiten crear superficies continuas georreferenciadas a partir de información puntual. Por ejemplo, a partir de los datos de temperatura media anual recogidos en observatorios meteorológicos podría construirse una superficie continua que expresara esa variable en cualquier punto del territorio, en base a técnicas de interpolación espacial (ver figura 22).

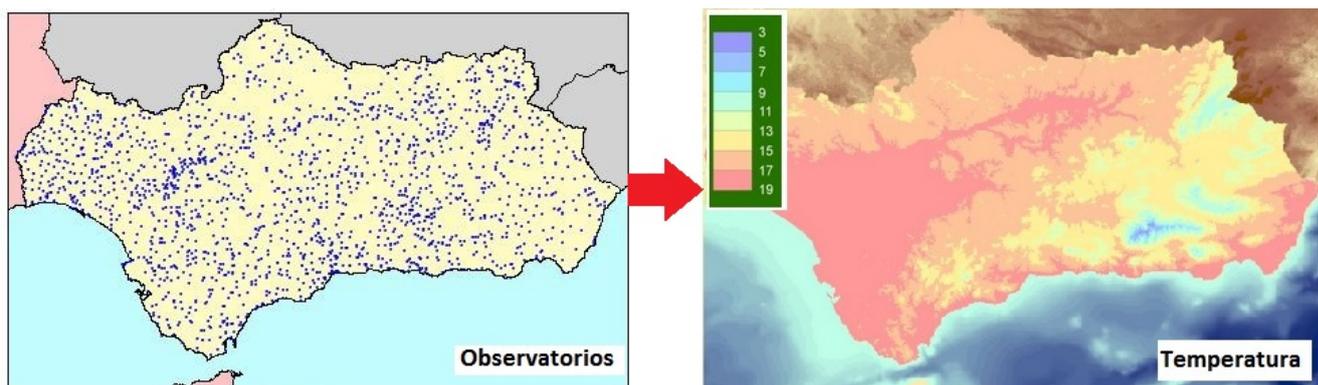


Figura 22.- Espacialización de variables puntuales

## 6. DIFUSIÓN DE LA INFORMACIÓN

### 6.1 PUBLICACIÓN DE DATOS ESPACIALES

La información espacial obtenida en el proceso de geocodificación no sólo puede usarse en equipos locales y distribirse como ficheros de un determinado formato. La información geocodificada también puede ser difundida mediante servicios web estándar, permitiendo cubrir las siguientes necesidades:

- Disponer de manera fácil, cómoda y eficaz de los datos geográficos existentes y poder usarlos en diferentes plataformas (PC, Tablet, Smartphone, ...) con la única necesidad de un navegador
- Oportunidad de reutilizar la información geográfica generada en un proyecto para otras finalidades diferentes, dado el alto coste de su producción, de manera que se encuentre accesible directamente para toda la ciudadanía
- Poder visualizarla en cualquiera de los visores cartográficos disponibles a tal efecto

De esta forma se accede a los datos de forma fluida y el usuario accede a la información estándar de manera que puede visualizarla, consultarla y en función de las características de los servicios prestados y los servidores, descargarla o realizar análisis espaciales.

Un **servicio web estándar** se basa en un sistema distribuido con una arquitectura cliente/servidor. Esta arquitectura consiste básicamente en un cliente (navegador web) que realiza peticiones a otro programa (servidor) que le da respuesta.

Para servir información espacial por la red o a través de internet, es necesario:

- Un **Servidor Web** que se encarga de mantenerse a la espera de las peticiones "http" llevadas a cabo por un cliente y resolverlas.
- Un **Servidor de Mapas** que se encarga de la representación de la información geográfica en forma de un archivo de imagen digital conveniente para la exhibición en una pantalla de ordenador.

Para poder acceder a los servicios de mapas publicados en Internet es necesario conocer los protocolos que permitan que los clientes se entiendan con los servidores. En este sentido, la existencia de un gran número de protocolos hace necesario establecer un idioma común que permita la transmisión de información en formato digital a través de Internet. Esta estandarización y normalización de un protocolo único permiten que la información pueda ser interoperable.

Las técnicas de interoperabilidad permiten la consulta, descarga y actualización de información de forma independiente a la tecnología que se usa para la visualización, resultando necesario publicar el conjunto de datos obtenidos con servicios web estándar.

Las normas internacionales para la interoperabilidad son desarrolladas por Open Geospatial Consortium (OGC) y éstas son asimiladas por el Centro de Normalización Europea CEN TC/287, y a su vez incorporadas por los organismos nacionales. En el caso de España es AENOR CTN/148.

## 6.2 SERVIDORES DE MAPAS

Los servidores de mapas permiten la interacción con la información espacial almacenada en servidores de datos espaciales accesibles vía web. De forma genérica, los servidores de mapas son aplicaciones informáticas que realizan la publicación de información espacial en la web. Estos servidores de mapas facilitan la accesibilidad de los datos espaciales a través de la web.

El usuario se conecta a los servicios prestados por éstos a través de clientes, tanto con aplicaciones web que permiten la consulta de estos servidores de mapas desde el navegador (clientes ligeros), como con aplicaciones SIG de escritorio con módulos que permiten la conexión a servidores de mapas, como por ejemplo QGIS (clientes pesados).

La principal característica de los clientes ligeros es que son accesibles a través de un navegador Web. No necesitan ninguna instalación en el puesto cliente y muestran mapas en formatos JPEG o PNG. El tipo de página que generan es independiente de la tecnología usada en el servidor.

El proceso de publicación (ver figura 23), petición del cliente y respuesta que siguen los servidores de mapas vienen dados por los siguientes pasos:

1. En el servidor de mapas se configuran las fuentes de datos donde se almacenan las capas.
2. Se diseña la representación de las capas. En función de las características y atributos de los datos es posible la agrupación de valores, el filtrado etiquetado, etc.
3. Se configuran las escalas de visualización y capas y se publica en el servidor de mapas.
4. El servidor de mapas espera las peticiones clientes mediante el protocolo específico para cada tipo de servicio.
5. El cliente realiza la petición mediante su conexión a la web (http) solicitando las capas, estilos o resolución deseados.
6. El servidor procesa dicha petición y compone el mapa con las capas solicitadas.
7. El servidor de mapas devuelve como resultado de la solicitud del cliente el mapa resultante.

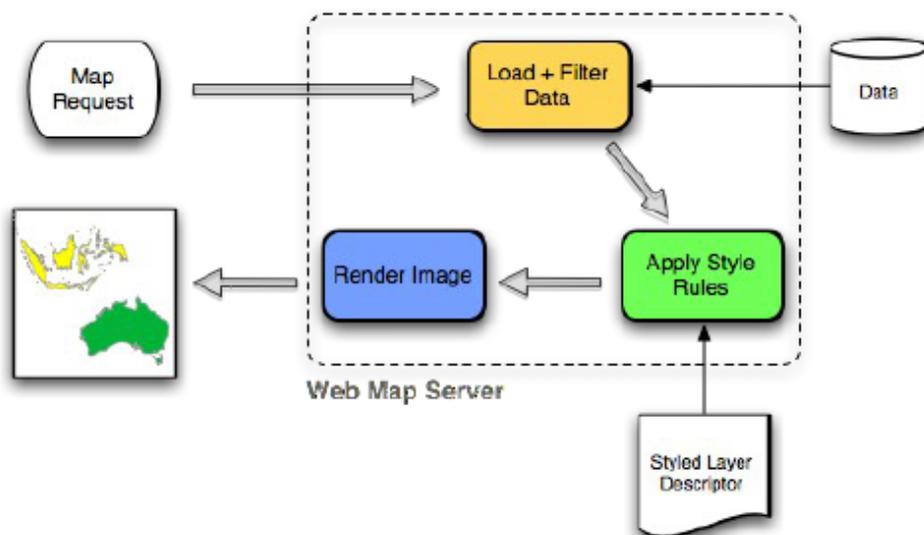


Figura 23.- Servidores de mapas

Los servidores de mapas deben ser interoperables, es decir, deben de ser consultables por medio de unas especificaciones estandarizadas independientes del servidor o cliente concreto que se use. Tal y como se ha explicado anteriormente, ésta es la condición mediante la cual los sistemas heterogéneos pueden intercambiar procesos o datos. Es difícil imaginar soluciones interoperables sin el uso de estándares abiertos, es decir, que cumplen unas reglas generales.

Los servidores de mapas de código abierto más utilizados que cumplen uno o varios de los principales estándares en relación con el acceso a datos espaciales son:

- **Geoserver**

Geoserver está desarrollado en Java, lo que le hace ser multiplataforma, permitiendo a los usuarios compartir y editar datos geoespaciales. Diseñado para la interoperabilidad, publica datos de cualquier fuente de datos espaciales con estándares abiertos. Permite generar servicios Web Map Service (WMS), Web Feature Service (WFS) y Web Coverage Service (WCS).

También incluye un cliente integrado OpenLayers para previsualizar capas de datos, y soporta la publicación de datos geográficos para Google Earth a través del uso de enlaces de red utilizando KML para ello.

El enlace a la página web del proyecto Geoserver es:

<http://geoserver.org/>



Las tareas básicas que deben realizarse para la publicación de información geográfica con Geoserver son:

- Configuración de un espacio de trabajo, donde agrupar datos publicados.
  - Configuración de un almacén de trabajo con aquellos datos que se van a mostrar a través del servicio.
  - Configuración de los datos que permite configurar cómo se puede publicar la información alfanumérica, qué campos se quieren utilizar y los estilos.
  - Publicación de los datos.
- 
- **Mapserver**

MapServer es una plataforma para la publicación de datos espaciales y aplicaciones de cartografía interactiva para la web.

El enlace a la página web del proyecto Mapserver es:

<http://mapserver.org/>



Además de estos dos servidores de mapas, existen numerosos servidores utilizados en mayor o menor medida y que cumplen con los estándares OGC:



## 6.3 SERVICIOS

Los **servicios** son las funcionalidades accesibles mediante tecnologías telemáticas interoperables que se ofrecen al usuario para interactuar con los datos geográficos. Los estándares de servicios geográficos más usuales son:

- **WMS** (Servicio de mapas): La información se muestra como imágenes compuestas por una o varias capas y generadas a partir de información espacial vectorial o ráster. El servicio Web Map Service (WMS) produce, dinámicamente, mapas de datos referenciados espacialmente a partir de información geográfica. Estos mapas se generan generalmente en un formato de imagen como PNG, TIFF, GIF o JPEG.
- **WFS** (Servicios de objetos): Ofrece la información en capas de datos compuesta por objetos vectoriales. El servicio Web Feature Service (WFS) ofrece acceso de lectura a datos vectoriales utilizando GML (Geography Markup Language) como protocolo subyacente para realizar consultas espaciales, recuperar los datos y manipular la geometría.

Otros estándares que también se utilizan son el Web Coverage Service (WCS) que es un servicio análogo al WFS para datos ráster y el Catalogue Service (CSW), que es un servicio de catálogo que permite la publicación y búsqueda de metadatos. Por otra parte, los servicios WMTS permiten la visualización de la información de manera más rápida mediante la definición de una teselación estándar, independiente del ámbito que se solicite en cada momento, que permite almacenar las teselas que componen las respuestas en una memoria intermedia (caché).

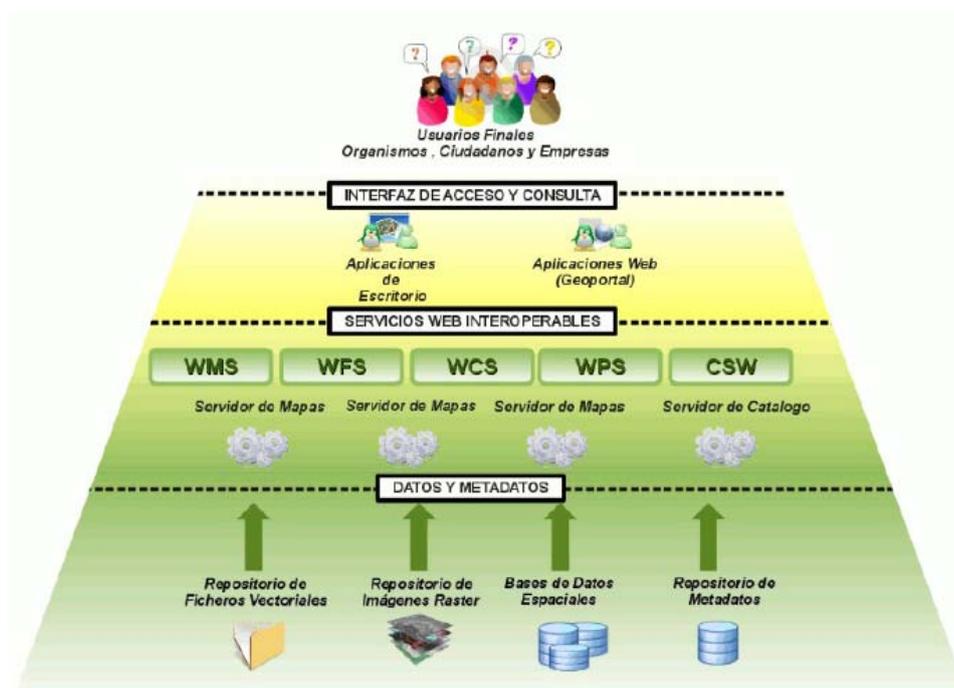


Figura 24.- Interoperabilidad con datos geográficos

Tal y como muestra la figura 24, los usuarios acceden a los Servicios Interoperables mediante aplicaciones web (clientes ligeros) o de escritorio (clientes pesados), que conectarán con los Servidores de Mapas que cumplen con las especificaciones de OGC, y donde se encuentran los datos geográficos almacenados.

### 6.3.1 Visores de Servicios

La información geográfica que genera un servicio web viene expresada mediante una URL, aunque ésta no es un ejecutable. Así, por ejemplo, si se usa el servicio web del Nodo del Instituto de Estadística y Cartografía de Andalucía (IECA) que viene expresado como [http://www.ideandalucia.es/wms/mta400v\\_2008?](http://www.ideandalucia.es/wms/mta400v_2008?), no se obtendría nada pinchando sobre él.

Para poder visualizar la información espacial que generan los servicios web existentes y hacerlos visibles al exterior es necesario la utilización de visores. Estos deben ser capaces de consultar servicios estándares de mapas cumpliendo las especificaciones de la OGC y las normativas ISO.

Estas aplicaciones pueden ser:

- a) Clientes pesados que son aplicaciones SIG escritorio y necesitan instalarse en local (QGIS, ARCGIS, Google Earth, etc.)
- b) Clientes ligeros con acceso desde un navegador web y sin instalación. Existen diversos visores de servicios. Entre otros, el IECA pone a disposición de los usuarios dos herramientas a través de su web
  - **IDEAvisor** <http://www.ideandalucia.es/IDEAvisor/>
  - **Visor3D**: es un visor cartográfico 2D/3D de fácil manejo y que permite cargar servicios wms al igual que capas locales para su posterior visualización en 2D, 2,5D (perspectiva) y 3D (vuelo virtual).  
<http://www.ideandalucia.es/portal/web/ideandalucia/herramientas/visor3d>

Por otro lado, el SIG Corporativo de la Junta de Andalucía ha desarrollado **Mapea**, que es una herramienta para la inserción de visores cartográficos en páginas web (Mapshup) pero que a su vez permite generar directamente una dirección o URL e insertarla en cualquier navegador. El configurador de mapas permite diseñar de manera fácil e intuitiva una URL de Mapea. Es un visor para usuarios finales no especializados.

Mediante el uso del configurador de mapas el usuario puede seleccionar las capas que desea visualizar de las existentes en un grupo predefinido. Se oferta, además, la posibilidad de especificar URLs de servicios

externos (WMS, WMC, KML). El usuario puede guardar también archivos KML en el servidor, quedando disponibles para su uso en el mapshup de mapas.

En la interfaz el usuario puede ver todas las funcionalidades ofertadas seleccionando las que considere oportunas. Una funcionalidad interesante es la posibilidad de centrar la vista del mapa de acuerdo a una búsqueda efectuada sobre el servicio web del Callejero, permitiendo buscar calles o sedes de la Junta de Andalucía.

Para publicar un mapa en la web con la información geocodificada se accede a Mapea a través de la siguiente dirección: <http://www.ideandalucia.es/IDEAmapshup/componente/genmapshup.jsp>

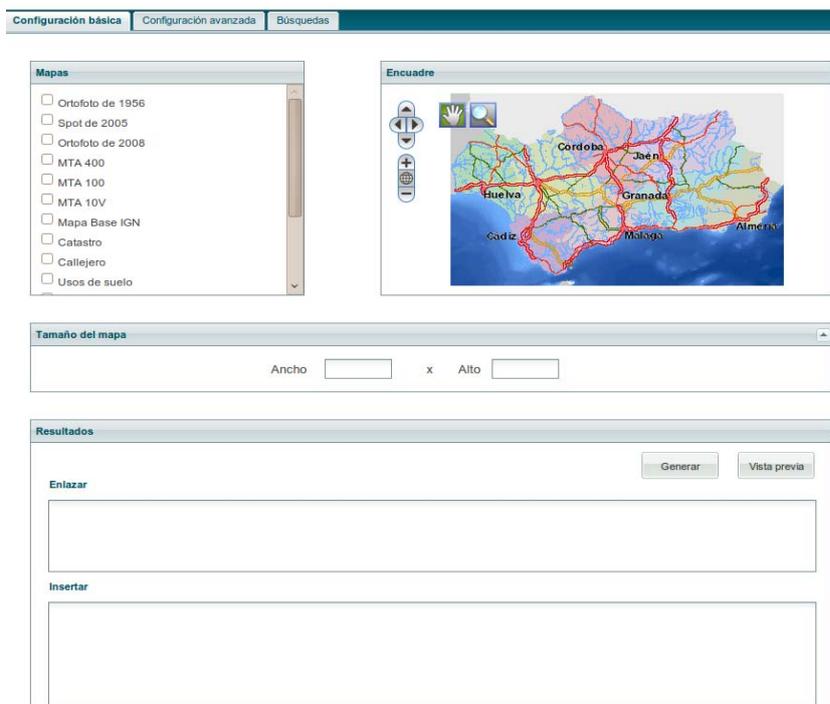


Figura 25.- Pantalla inicial de la herramienta de generación de visores de mapas (mapshup)

Esta herramienta (figura 25) consta de tres pestañas que permiten configurar el visor web de mapas que se va a generar:

- a) Configuración básica, pantalla que se muestra por defecto.

En este caso el usuario puede seleccionar una serie de capas precargadas en el sistema para que se muestren como cartografía de base o "fondo cartográfico" de los datos que previamente haya geocodificado. Para ello se debe seleccionar del conjunto etiquetado como "Mapas" aquellos que se deseen añadir.

En la opción “Encuadre”, situada a la derecha, se puede establecer el área de visualización que usa el visor generado. Si no se hace uso de ella se establece un área por defecto, correspondiente a la región andaluza completa. Para establecer el mencionado campo solo se debe modificar el mismo interactuando con el mapa de la derecha, haciendo zoom y desplazando el centro del mapa con el icono con forma de mano.

En esta pantalla también se da la posibilidad de introducir el “Tamaño del mapa” a incluir en la web, expresado en píxeles (los valores habituales oscilan entre: 600x400 para un mapa incrustado o 800x600 y superiores para mapas a pantalla completa).

El Botón “Generar” sirve para crear la URL para ser distribuida a los ciudadanos y al público objetivo del mapa creado o permite que el mapa generado sea embebido en cualquier página web mediante el componente html IFRAME.

Una vez generada la URL se puede obtener una vista previa del resultado pulsando el botón “Vista Previa”. Dicha vista previa solo estará disponible después de generar la URL.

El resultado mostrado en el hueco etiquetado como “Enlazar” puede ser copiado directamente en un navegador para empezar a utilizar el visor.

El usuario puede optar también por incluir un Mapea personalizado en su propia web. Para ello, una vez generada la URL en la ventana etiquetada como “Insertar” aparece el código que el usuario debe copiar y pegar en el lugar deseado dentro de su web. El tamaño que tiene Mapea dentro de la web también puede ser definido por el usuario especificando un valor a las etiquetas “Ancho” y “Alto” incluidas en la interfaz. Si no se especifica ningún valor se tomará el valor por defecto, esto es, una ventana de 600x400 píxeles.

En resumen, con la configuración básica de la herramienta de mapshup el usuario puede generar un mapa con las capas precargadas en la herramienta y con los niveles de zoom y encuadre seleccionados. Para incorporar información de la geocodificación de los datos alfanuméricos realizada es necesario acceder a la pestaña de configuración avanzada.

#### b) Configuración avanzada.

En esta vista los usuarios de la herramienta pueden añadir a los mapas que generen, además de las capas que vienen preconfiguradas en el sistema, otros servicios de mapas definidos según los estándares WMC (Web Map Context) o WMS (Web Map Service). Además, se permite la inclusión

de ficheros KML publicados en Internet o resultantes de la ejecución del proceso de geocodificación. Para estos últimos solo será necesario indicar la dirección del archivo, mientras que para los anteriores, además de la URL, habrá que proporcionar el nombre del contexto y de la capa, respectivamente.

Para subir un fichero KML residente en el equipo local tan solo hay que pulsar el botón "Examinar" y buscar el fichero con los datos geocodificados en el sistema local de archivos. Una vez seleccionadas las capas a mostrar, establecidos los niveles de visualización y las herramientas que se le van a mostrar al usuario, hay que ir a la pestaña de configuración básica y pulsar el botón "Generar". En este momento se generará una URL y un código HTML, que contienen la información del mapa generado y que pueden ser compartidos para su visualización desde cualquier navegador web de uso común. En el apartado "Ficheros KML" se posibilita que el usuario guarde un archivo KML en el servidor. Este archivo está disponible para su explotación en Mapea.

Para WMC y WMS el usuario también tiene la posibilidad de seleccionar un "EPSG" para el mapa. Esto es posible eligiendo su valor en el desplegable de la derecha. La proyección puede estar expresada en grados o metros.

En la "Configuración de otras acciones" se ofrece la posibilidad de definir los parámetros "center" y "zoom" permitidos en Mapea. Para hacer uso de esta opción el usuario podrá seleccionar un valor de "Zoom" de entre los ofrecidos en el combo dispuesto para tal fin. Para establecer el valor de las coordenadas X e Y basta con hacer click sobre el punto del mapa sobre el que se deberá centrar la nueva vista.

Con el fin de ser más precisos en la elección de este centro se dota al usuario de la posibilidad de hacer zoom sobre el mapa de referencia, así como desplazarse sobre el mismo.

Se pueden añadir tantos servicios o capas como se deseen, pero debe tenerse en cuenta que la aplicación puede llegar a sobrecargarse o ralentizar su funcionamiento si son demasiadas.

- c) Búsqueda. Configuración de las utilidades de búsqueda que se van a ofrecer a la persona que visite el mapa en la web.



En el manual de usuario de Mapea se describen otras funcionalidades, así como una amplia explicación del uso de la herramienta. Este manual se encuentra en la web de Mapea:

<http://www.ideandalucia.es/portal/web/ideandalucia/herramientas/mapea>

El SIG Corporativo ha creado también la herramienta de **MAPEA Edita**, que integra la posibilidad estándar WFS-T, de tal manera que junto a los controles de navegación de Mapea incorpora también los controles de edición, que además de permitir cargar y dibujar los datos, permite editarlos. Mapea Edita cuenta con botones para guardar cambios, editar atributos, mover elementos..., y además, descargar los datos modificados en un fichero con formato CSV.

## ÍNDICES DE ESQUEMAS, CUADROS Y FIGURAS

### ESQUEMAS

|  |    |
|--|----|
| Esquema 1.- Proceso de geocodificación de fuentes de información: fases, fuentes y herramientas .....                | 6  |
| Esquema 2.- Normalización de fuentes de información: fases, posibilidades de actuación, agentes y herramientas ..... | 10 |
| Esquema 3.- Proceso de geocodificación de fuentes y registros administrativos.....                                   | 44 |
| Esquema 4.- Etapas del proceso de enlace de registros.....   | 47 |

### CUADROS

|   |    |
|---|----|
| Cuadro 1.- Recogida de información de variables territoriales en formularios y sistemas de información..... | 15 |
| Cuadro 2.- Recogida en el formulario .....  | 16 |
| Cuadro 3.- Implementación de los datos recogidos en el sistema de información.....                          | 16 |
| Cuadro 4.- Recogida y normalización de la dirección postal.....   | 17 |
| Cuadro 5.- Sistemas de referencia, sistemas de coordenadas y códigos EPSG.....                              | 34 |
| Cuadro 6.- Sistemas de Información Geográfica .....   | 62 |

### FIGURAS

|   |    |
|---|----|
| Figura 1.- Propuesta de recogida del campo domicilio en un formulario.....  | 18 |
| Figura 2.- Resultado del campo domicilio tras su normalización .....  | 22 |
| Figura 3.- Interfaz principal de aLink: Herramienta de Fusión de Ficheros.....                                    | 23 |
| Figura 4.- aLink: Herramienta de Normalización .....  | 24 |
| Figura 5.- Desagregación de campos de salida con aLink: a medida o desagregación según los campos del CDAU.....   | 26 |
| Figura 6.- Estructura y orden de los componentes de una dirección postal con Nordir .....                         | 28 |
| Figura 7.- NorDir.....  | 29 |
| Figura 8.- Latitud y longitud .....   | 32 |
| Figura 9.- Husos en España .....  | 33 |
| Figura 10.- Interfaz de Calar .....   | 35 |
| Figura 11.- Visualizador del Callejero Digital de Andalucía Unificado (CDAU) .....                                | 41 |
| Figura 12.- Geodir.....   | 48 |
| Figura 13.- TELEGEO.....  | 53 |
| Figura 14.- Ejemplo de posibles capas temáticas .....   | 63 |
| Figura 15.- a) Identificación de elementos sobre el mapa; b) Selección de elementos en la tabla de atributos..... | 64 |
| Figura 16.- Ejemplo de asignación de información alfanumérica a partir de relaciones espaciales entre capas.....  | 66 |
| Figura 17.- Intersección de capas .....   | 67 |
| Figura 18.- Zonas de influencia.....  | 68 |
| Figura 19.- Zonas de influencia de anillos múltiples.....   | 68 |
| Figura 20.- Cálculo del camino óptimo mediante análisis de redes .....  | 69 |
| Figura 21.- Análisis de conectividad.....   | 70 |
| Figura 22.- Espacialización de variables puntuales .....  | 70 |



|   |    |
|---|----|
| Figura 23.- Servidores de mapas.....  | 73 |
| Figura 24.- Interoperabilidad con datos geográficos .....                                       | 75 |
| Figura 25.- Pantalla inicial de la herramienta de generación de visores de mapas (mapshup)..... | 77 |