

NORMALIZACIÓN Y GEOCODIFICACIÓN DE FUENTES DE INFORMACIÓN ADMINISTRATIVA

VI JORNADAS IBÉRICAS DE INFRAESTRUCTURAS DE DATOS ESPACIALES

Sevilla, 6 de Noviembre de 2015

Fernando Martínez, Gema Galera, Javier Marzo y Elisa Caballero
INSTITUTO DE ESTADÍSTICA Y CARTOGRAFÍA DE ANDALUCÍA

Índice

1. INTRODUCCIÓN
2. NORMALIZACIÓN DE LAS DIRECCIONES POSTALES
3. GEOCODIFICACIÓN CON aLink
4. RESULTADOS Y CONCLUSIONES

Introducción



ISO/TC 2011 → Incrementar IG

La mayoría de fuentes administrativas presentan un gran potencial cartográfico

FUENTES ADMINISTRATIVAS

Direcciones postales

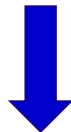
FUENTES DE REFERENCIA

Direcciones postales
geocodificadas

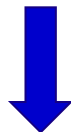
ENLACE

GEOCODIFICACIÓN

Fuente de Referencia en Andalucía



Callejero Digital de Andalucía Unificado (CDAU)



Callejero Digital de Andalucía Unificado

Proyecto Avances Datos Recursos Participar

Estás en: Inicio > Recursos > Productos > **Información alfanumérica**

Recursos

Productos

- **Información alfanumérica**
- Información geográfica

Herramientas

Formación técnica

Breves CDAU

Enlaces

Información alfanumérica

Una de las utilidades básicas de la información de CDAU es servir como base cartográfica para la geocodificación de registros administrativos.

Entre las aplicaciones que permiten automatizar este proceso de asignación de coordenadas geográficas a direcciones postales se encuentra aLink: Herramienta de Fusión de Ficheros.

Se trata de una aplicación libre y gratuita de fusión de ficheros, desarrollada por el Instituto de Estadística y Cartografía de Andalucía, que permite geocodificar registros enlazando las direcciones postales coincidentes de dos ficheros de datos y asignando las coordenadas geográficas X e Y (del fichero que las contiene), a las direcciones postales del fichero que se desea geocodificar.

Para llevar a cabo un proceso de geocodificación con aLink: Herramienta de Fusión de Ficheros, es necesario disponer de la aplicación, y descargar los siguientes ficheros de datos provincializados a partir de la información que contiene CDAU:

Ficheros de viales. Contiene la información alfanumérica de las vías de CDAU por provincias con las coordenadas puntuales al centro de la vía.

Ficheros de portales. Contiene la información alfanumérica de los portales de CDAU por provincia con las coordenadas exactas a cada portal.

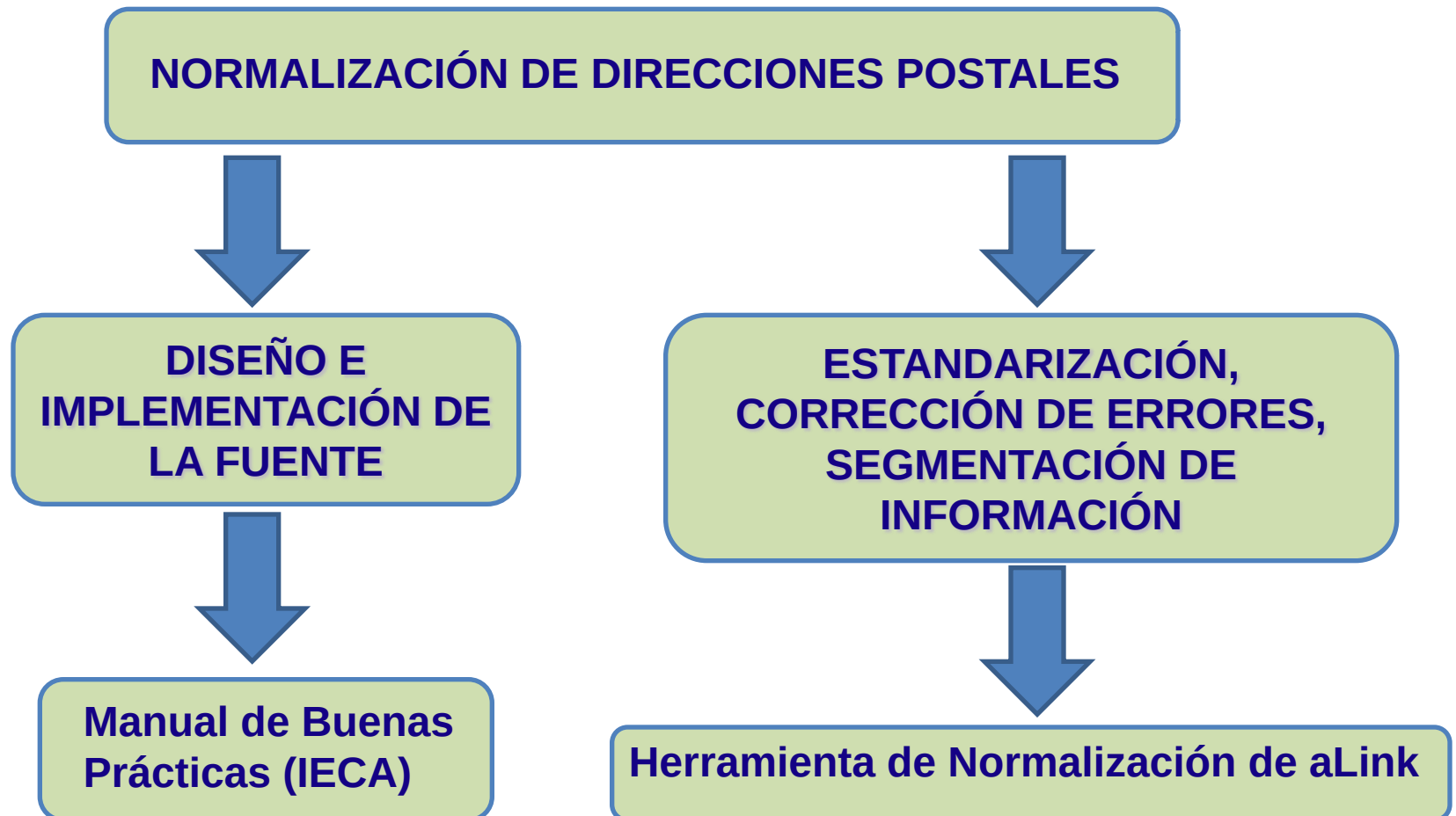
La fecha de referencia de los datos de cada uno de los ficheros es el 8 de junio de 2015.

Especificaciones del estándar ISO 19100 IG e INSPIRE Data Specification on Adresses

Índice

1. INTRODUCCIÓN
2. NORMALIZACIÓN DE LAS DIRECCIONES POSTALES
3. GEOCODIFICACIÓN CON aLink
4. RESULTADOS Y CONCLUSIONES

Establecer enlaces entre ficheros requiere la normalización previa de los campos comunes



Propuesta de recogida de la dirección postal en los formularios e implementación en los Sistemas de Información

DESCRIPCIÓN	FORMULARIO	SISTEMA DE INFORMACIÓN	
		Etiqueta	Tipo dato
DESCRIPCIÓN A NIVEL DE VÍA			
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)
DESCRIPCIÓN A NIVEL DE NÚMERO			
Código de Portal	-	id_por_pk	Entero (long. 9)
Número de portal/Km	NUMERO/Km	-	-
Calificador del número (letra)	CALIF. NUM	-	-
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	-	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)
DESCRIPCIÓN A NIVEL DE INMUEBLE			
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	cos_hue_dgc	Texto (long. 21)
INFORMACIÓN COMPLEMENTARIA			
Tipo de agrupación	COMPLEMENTO DOMICILIO	nom_tipo_agrupacion	Texto (long. 100)
Nombre de agrupación		nom_agrup	Texto(long. 100)
Otros datos de ubicación (ODUB)		txt_app	Texto (long. 200)

aLink: Herramienta de Fusión de Ficheros

INTERFAZ PRINCIPAL

Aplicación libre y gratuita, desarrollada por el IECA, a partir de FEBRL. Combina técnicas en distintas etapas para llevar a cabo un proceso de fusión de ficheros de grandes volúmenes de datos



Normalización de las direcciones postales con aLink

HERRAMIENTA DE NORMALIZACIÓN

Instituto de Estadística y Cartografía de Andalucía

Andalucía se mueve con Europa



Instituto de Estadística y Cartografía de Andalucía
CONSEJERÍA DE ECONOMÍA, INNOVACIÓN, CIENCIA Y EMPLEO

Unión Europea
Fondo Europeo de Desarrollo Regional

Herramienta de Normalización

Archivo Herramientas Ayuda

Nuevo Salir Ejecutar

Fichero a normalizar

(Ninguno)

Tipo de normalización

☐ Nombres propios ☐ Direcciones postales ☐ NIF / DNI / NIE

Requerimientos del Sistema

Nombres propios **Direcciones postales** **NIF / DNI / NIE**

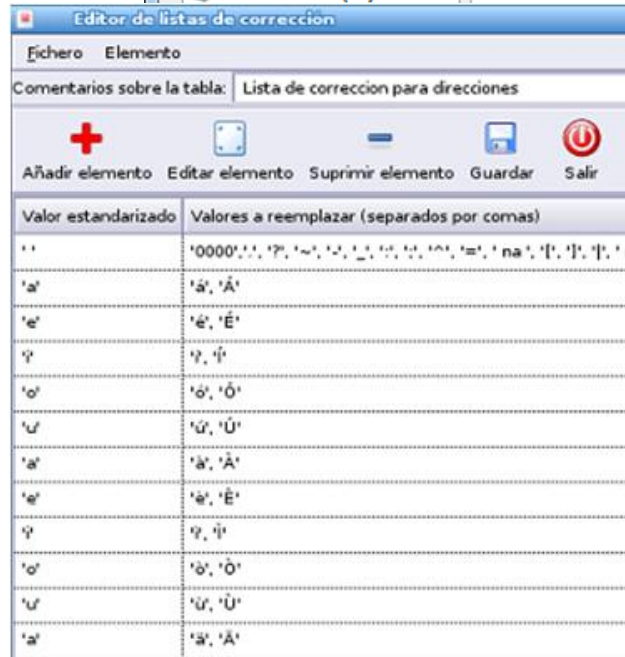
Campo a normalizar:

Lista de corrección: (Ninguno)

Tablas de búsqueda: app

Modelo Oculto de Markov: (Ninguno)

Campos de salida: Seleccionar...



Modelos Ocultos de Markov

- Se utilizan para segmentar campos de texto.
- Reconocen patrones dentro de un fichero, describiendo el proceso que con más probabilidad ha generado los datos.

DIRECCIÓN
C/ SOL, NUMERO 3, CARMONA
AVDA SAN FRANCISCO JAVIER, 156, SEVILLA
PLAZA DE ESPAÑA, S/N, DOS HERMANAS
LEONARDO DA VINCI, 22, SEVILLA



Tipo de vía	Nombre de vía	Identificador de numeración	Ein (Número)	Municipio
calle	sol	numero	3	carmona
avenida	san francisco javier		156	sevilla
plaza	de españa	sin_numero		dos hermanas
	leonardo da vinci	22		sevilla

Desagregación de las direcciones postales según el CDAU

Direcciones postales

Determinar el tipo de desagregación:

Selección de campos de salida:

☒ Tipo de vía	☒ Escalera	☐ Id. de sector
☒ Nombre de vía	☐ Id. de planta	☐ Sector
☒ Id. de numeración	☒ Planta	☐ Id. de manzana
☒ Entidad inferior de numeración	☐ Id. de puerta	☐ Manzana
☒ Calificador ent. inf. de numeración	☒ Puerta	☐ Id. de parcela
☒ Entidad superior de numeración	☐ Id. de letra	☐ Parcela
☒ Calificador ent. sup. de numeración	☐ Letra	☐ Id. de nave
☐ Id. de bloque	☒ Entidad singular	☐ Nave
☒ Bloque	☒ Municipio	☐ Tipo de zona
☐ Tipo de edificio	☒ Provincia	☐ Zona
☐ Edificio	☐ Id. de código postal	☒ Otros datos de ubicación (ODUB)
☐ Id. de portal	☒ Código postal	
☒ Portal	☒ Tipo de agrupación	
☐ Id. de escalera	☒ Agrupación	

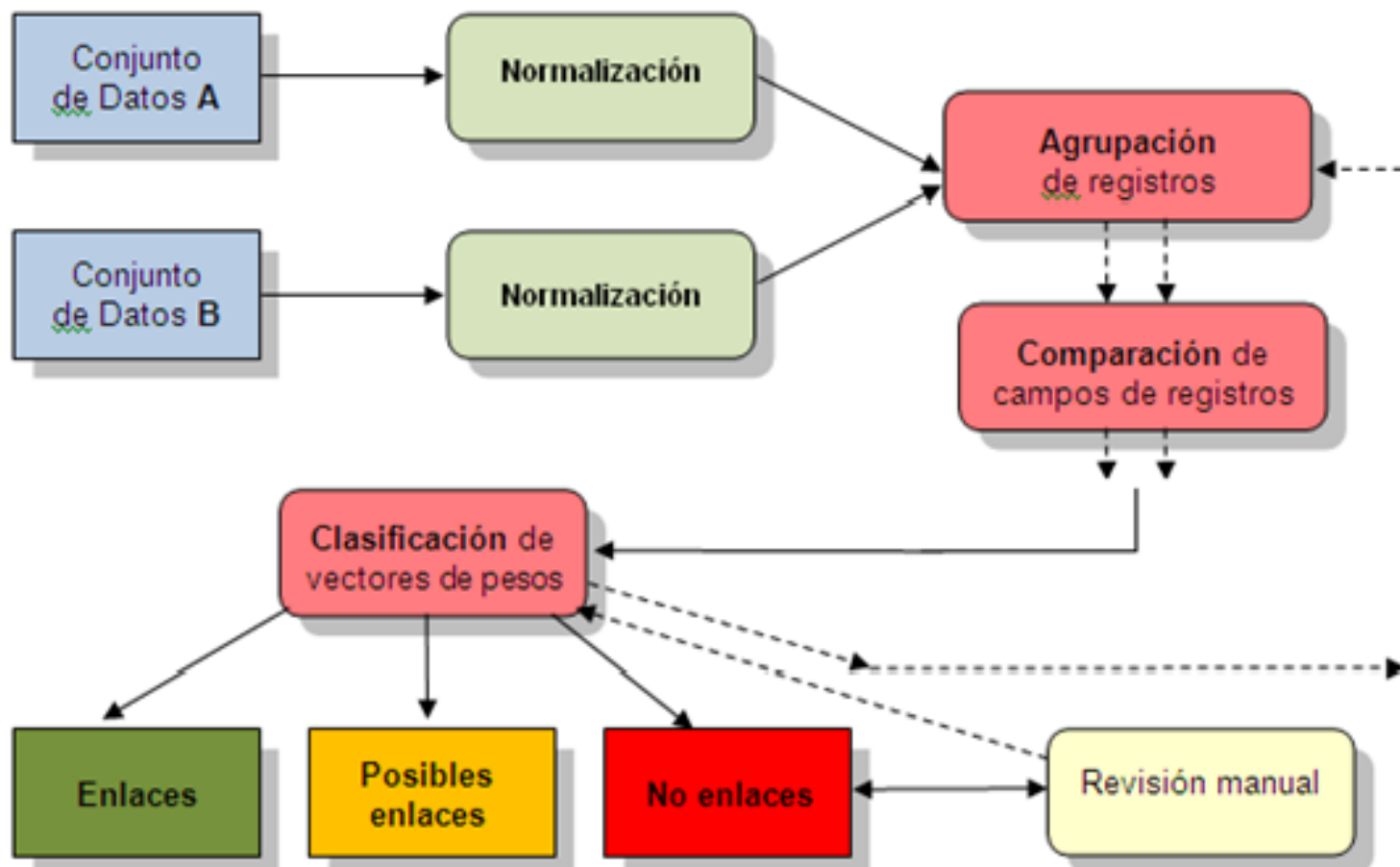
Fichero normalizado

[illegible]

Índice

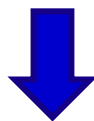
1. INTRODUCCIÓN
2. NORMALIZACIÓN DE LAS DIRECCIONES POSTALES
3. GEOCODIFICACIÓN CON aLink
4. RESULTADOS Y CONCLUSIONES

Etapas de un proceso completo de geocodificación con aLink



Prioridades para geocodificar según fuente de referencia

ORDEN DE PRIORIDAD	TIPO DE GEOCODIFICACIÓN	FUENTE DE REFERENCIA	INFORMACIÓN UTILIZABLE
1	Portal exacto	CDAU de portales	Dirección postal (con número de vía) y referencia catastral
2	Centroide de la parcela	Catastro	Referencia catastral
3	Portal cercano	CDAU de portales	Dirección postal (con número de vía)
4	Centro de la vía	CDAU de vías	Dirección postal (sin número de vía)



Guía de Geocodificación de Fuentes de Información Administrativa (IECA)

Resultado de la geocodificación con aLink

FICHERO DIRECCIONES NORMALIZADAS

R	S	T
TVIA_OF	NVIA_OF	EIN
CALLE	REDONDA	16
CALLE	SEÑOR DE LA EXPIRACION	35
PLAZA	GARCIA LORCA	3
CALLE	REAL ALTA	5
CALLE	ERAS BAJAS DE PINOS PUENTE	50
CALLE	ERMITA	9
RBLA	ERAS	23
CALLE	ENMEDIO	2
CALLE	RAFAEL ALBERTI	10
AVDA	ANDALUCIA	9
CALLE	SAN JOSE	63
CALLE	REAL	130
CALLE	SAN MARTIN	11
CALLE	JARDINES DE PINOS PUENTE	6

ENLACE MEDIANTE LOS CAMPOS:

- TIPO DE VÍA
- NOMBRE DE VÍA
- NÚMERO DE LA VÍA

CDAU

D_VIAL	INE_VIA	NOM_T	NOM_VIA	NUM_POR_DES	REFCATPARC	NOM_TIPO_A	NOM_AG	INE_NUCL	NOM_NUCLEO	INE_M	NOM_MUNICI	COD_P	X_CDAU	Y_CDAU
347000141	1816100630	CALLE	REDONDA	16				18158000701	PINOS PUENTE	18158	PINOS PUENTE	18240	433035.45673	4123000.21611
314000006	1811800054	CALLE	SEÑOR DE LA EXPIRACION	35				18116000101	LANJARON	18116	LANJARON	18420	457447.74834	4085967.75428
281000047	1807300040	PLAZA	GARCIA LORCA	3				18071000201	DURCAL	18071	DURCAL	18650	449697.24223	4093680.37642
336000208	1814800012	CALLE	REAL ALTA	5				18145000101	OGIJARES	18145	OGIJARES	18151	446044.51054	4108453.80026
347000161	1816100060	CALLE	ERAS BAJAS (PINOS PUENTE)	50	3732506VG3233A			18158000701	PINOS PUENTE	18158	PINOS PUENTE	18240	433611.26285	4122976.26521
360000386		CALLE	ERMITA	9				18175000301	SANTA FE	18175	SANTA FE	18320	436454.64079	4116231.63126
383000025	1805900239	RAMBLA	ERAS	23				18907000101	ALCUDIA DE GUADIX	18907	VALLE DEL ZALAB	18511	491317.65243	4123544.16596
383000050	1805900018	CALLE	ENMEDIO	2				18907000201	CHARCHES	18907	VALLE DEL ZALAB	18511	503953.33166	4127458.17256
336000017	1814800040	CALLE	RAFAEL ALBERTI	10				18145000101	OGIJARES	18145	OGIJARES	18151	446240.6095	4108272.25095
314000063	1811800003	AVENIDA	ANDALUCIA	9				18116000101	LANJARON	18116	LANJARON	18420	456801.48389	4085891.386
281000257	1807300113	CALLE	SAN JOSE	63	9442004VF4994A			18071000201	DURCAL	18071	DURCAL	18650	449290.82861	4094111.14999
314000009	181800044	CALLE	REAL	130	7061737VF5876S			18116000101	LANJARON	18116	LANJARON	18420	456941.33696	4085871.98162
276000014	1806600050	CALLE	SAN MARTIN	11				18066000101	DEIFONTES	18066	DEIFONTES	18570	447402.59678	4131137.03724
347000169	1816100083	CALLE	JARDINES (PINOS PUENTE)	6				18158000701	PINOS PUENTE	18158	PINOS PUENTE	18240	433183.57315	4122976.96663



FICHERO GEOCODIFICADO

Información adicional											Coordenadas		
TVIA	NVIA_OF	EIN	INE_VIA	NOM_TI	NOM_VIA	REFCATPARC	INE_NUCLEO	NOM_NUCLEO	INE_MU	NOM_MUNICI	COD_P	X_CDAU	Y_CDAU
CALLE	REDONDA	16	1816100630	CALLE	REDONDA		18158000701	PINOS PUENTE	18158	PINOS PUENTE	18240	433035.45673	4123000.21611
CALLE	SEÑOR DE LA EXPIRACION	35	1811800054	CALLE	SEÑOR DE LA EXPIRACION		18116000101	LANJARON	18116	LANJARON	18420	457447.74834	4085967.75428
PLAZA	GARCIA LORCA	3	1807300040	PLAZA	GARCIA LORCA		18071000201	DURCAL	18071	DURCAL	18650	449697.24223	4093680.37642
CALLE	REAL ALTA	5	1814800012	CALLE	REAL ALTA		18145000101	OGUARES	18145	OGUARES	18151	446044.51054	4108453.80026
CALLE	ERAS BAJAS DE PINOS PUENTE	50	1816100060	CALLE	ERAS BAJAS (PINOS PUENTE)	3732506VG3233A	18158000701	PINOS PUENTE	18158	PINOS PUENTE	18240	433611.26285	4122976.26521
CALLE	ERMITA	9		CALLE	ERMITA		18175000301	SANTA FE	18175	SANTA FE	18320	436454.64079	4116231.63126
RBLA	ERAS	23	1805900239	RAMBLA	ERAS		18907000101	ALCUDIA DE GUADIX	18907	VALLE DEL ZALABI	18511	491317.65243	4123544.16596
CALLE	ENMEDIO	2	1805900018	CALLE	ENMEDIO		18907000201	CHARCHES	18907	VALLE DEL ZALABI	18511	503953.33166	4127458.17256
CALLE	RAFAEL ALBERTI	10	1814800040	CALLE	RAFAEL ALBERTI		18145000101	OGUARES	18145	OGUARES	18151	446240.6095	4108272.25095
AVDA	ANDALUCIA	9	1811800003	AVENIDA	ANDALUCIA		18116000101	LANJARON	18116	LANJARON	18420	456801.48389	4085891.386
CALLE	SAN JOSE	63	1807300113	CALLE	SAN JOSE	9442004VF4994A	18071000201	DURCAL	18071	DURCAL	18650	449290.82861	4094111.14999
CALLE	REAL	130	181800044	CALLE	REAL	7061737VF5876S	18116000101	LANJARON	18116	LANJARON	18420	456941.33696	4085871.98162
CALLE	SAN MARTIN	11	1806600050	CALLE	SAN MARTIN		18066000101	DEFONTES	18066	DEFONTES	18570	447402.59678	4131137.03724
CALLE	JARDINES DE PINOS PUENTE	6	1816100083	CALLE	JARDINES (PINOS PUENTE)		18158000701	PINOS PUENTE	18158	PINOS PUENTE	18240	433183.57315	4122976.96663

Índice

1. INTRODUCCIÓN
2. NORMALIZACIÓN DE LAS DIRECCIONES POSTALES
3. GEOCODIFICACIÓN CON aLink
4. RESULTADOS Y CONCLUSIONES

aLink: Resultados

- **Trabajar con grandes volúmenes de datos**
- **Versatilidad (procesos iterativos)**
- **Normalización total de la dirección postal**
- **Geocodificación del 90% de los registros.
Depende de la información que aparece en la
dirección**

Directorio de empresas y establecimientos con actividad económica en Andalucía

	Nº registros	Geocodificados				No geocodificados
Provincia		Exacto	Cercano	Centro vía	Zona cercana	
Almería	45.166	21.950 (48.60%)	6.595 (14.60%)	4.243 (9.39%)	575 (1.27%)	11.803 (26.13%)
Cádiz	68.948	42.174 (61.17%)	7.830 (11.36%)	11.083 (16.07%)	895 (1.30%)	6.966 (10.10%)
Córdoba	54.383	35.093 (64.53%)	6.843 (12.58%)	7.699 (14.16%)	85 (0.16%)	4.663 (8.57%)
Granada	62.200	41.194 (66.23%)	7.218 (11.6%)	9.197 (14.79%)	372 (0.6%)	4.219 (6.78%)
Huelva	29.328	17.845 (60.85%)	3.553 (12.11%)	3.900 (13.30%)	790 (2.69%)	3.240 (11.05%)
Jaén	39.126	26.207 (66.98%)	4.546 (11.62%)	4.010 (10.25%)	772 (1.97%)	3.591 (9.18%)
Málaga	120.913	75.699 (62.61%)	10.389 (8.59%)	22.420 (18.54%)	611 (0.51%)	11.794 (9.75%)
Sevilla	126.054	87.603 (69.50%)	18.877 (14.98%)	10.696 (8.49%)	180 (0.14%)	8.698 (6.90%)
Total	546.118	347.765 (63.68%)	65.851 (12.06%)	73.248 (13.14%)	4.280 (0.78%)	54.974 (10.07%)

Descarga de aplicación aLink

The screenshot shows the IDEAndalucía website interface. At the top left is the logo of the Junta de Andalucía. The main navigation bar includes 'IDEANDALUCIA', 'DATOS Y SERVICIOS', 'HERRAMIENTAS', and 'IDES'. The 'HERRAMIENTAS' menu is open, displaying a list of tools: Nomenclator, Mapea, Visor 3D, Telegeo, Calar, Nordir y Geodir, Descargas, Linea, and aLink. A large blue arrow points to the 'aLink' option. Below the navigation bar, there are several sections: 'Nodos IDEAndalucía' with a description of accessing public administration nodes; 'Catalogo' with a description of interoperable cartographic services; and a 'Visualizador de información geográfica interoperable' section. Each section has a corresponding image and a button to 'Accede' (Access) to the respective service.

JUNTA DE ANDALUCÍA

ideAndalucía

CONTACTO AYUDA

IDEANDALUCIA DATOS Y SERVICIOS HERRAMIENTAS IDES

NOMENCLATOR ANDALUCIA 3D MAPEA

CALAR DESCARGAS GEOCODIFICAR

Instituto de Estadística y Cartografía de Andalucía
CONSEJERÍA DE ECONOMÍA, INNOVACIÓN, CIENCIA Y EMPLEO

Fotografías aéreas de Andalucía

Nodos IDEAndalucía

Acceso a los Nodos de la IDEAndalucía de las distintas administraciones públicas.

Accede a los Nodos

Catalogo

Acceso a los servicios cartográficos interoperables que ofrecen en Internet las distintas administraciones públicas.

Accede al catálogo

Visualizador de información geográfica interoperable. Permite visualizar y superponer cartografía y ortomágenes de diversas fuentes.

Accede al Visor

Fin

Gracias por vuestra atención