

IDEZar: Procesos, herramientas y modelos urbanos aplicados a la integración de datos municipales procedentes de fuentes heterogéneas

López Pellicer, F.J., Álvarez, P., Muro-Medrano, P.R..

Departamento de Informática e Ingeniería de Sistemas Informáticos.
Universidad de Zaragoza
Edificio Ada Byron, C/ María de Luna 1, 50018 Zaragoza
{fjlopez,alvaper,prmuro}@unizar.es

Resumen

Gran parte de los esfuerzos de la Infraestructura de Datos Espaciales de Zaragoza han sido aplicados al diseño e implementación de procesos de integración de datos de referencia de baja calidad. Tomando como base la propuesta de directiva INSPIRE, se ha desarrollado un conjunto de procesos, herramientas y modelos que ayudan a integrar de forma homogénea datos de referencia producidos en la administración local.

Palabras clave: Datos de referencia, administración local, geodatabase, INSPIRE, integración

1 Introducción

La iniciativa IDEZar (Infraestructura de Datos Espaciales de Zaragoza) es el resultado de la colaboración iniciada en el año 2004 entre el Ayuntamiento y la Universidad de Zaragoza para la implantación de una Infraestructura de Datos Espaciales (IDE) a nivel local. Dos elementos relevantes en esta IDE son la política de integración de datos y la arquitectura basada en servicios. Su importancia deriva de la existencia en el ámbito municipal de datos espaciales caracterizados por su gran volumen y heterogeneidad. Hay una amplia variedad de temáticas (identificadores de propiedad, figuras de protección medioambiental, restricciones urbanísticas, etc.) como plataformas (GIS, CAD, aplicaciones de oficina) amen de diferentes requerimientos de calidad, disponibilidad, validez, extensión y formato.

Además presentan problemas como lagunas de información, duplicidades e inconsistencias entre fuentes. La IDE, vista como herramienta que facilita el acceso y la explotación de información espacial clave [1], debe ser capaz de acceder a fuentes heterogéneas de datos y ofrecer una visión homogénea de la información como valor añadido. Adelantándose a la futura directiva INSPIRE, la política de integración de datos ha establecido mecanismos para que se “reduzcan las duplicaciones en la recopilación de datos, y (se) promuevan y respalden la armonización, la difusión y la utilización de los datos de una forma lo más amplia posible” [2] dentro del ayuntamiento. La arquitectura basada en servicios, que sigue los criterios de la *Web Services Architecture* [3] impulsada por Open Geospatial Consortium (OGC), permite que los Servicios municipales exploten el valor añadido al facilitar la construcción de aplicaciones y adaptadores sobre los servicios de la IDE.

Este artículo presenta una solución para la implementación de la política de integración de datos. Esta pasa por establecer un modelo que describe los datos espaciales usados por la administración local en la línea establecida por INSPIRE y que denominamos *modelo urbano* por la temática de la mayoría de sus datos. En este modelo los datos que INSPIRE denomina *datos de referencia*, datos que son utilizados para caracterizar espacialmente a otros datos, tienen un papel diferenciado pues son los responsables de homogenizar la información. Este modelo se implementa en una *geodatabase* poblada a partir de la *integración de datos* procedentes de *fuentes heterogénea* mediante la aplicación de un conjunto de *procesos y herramientas*. Además, este modelo ha simplificado el diseño de nuevos servicios y aplicaciones.

Este artículo se organiza como sigue, el apartado 2 incide en la relación entre INSPIRE, los datos de referencia y la administración local, el apartado 3 caracteriza los problemas detectados desde 2004 hasta la fecha, el apartado 4 muestra nuestra propuesta y el apartado 5 su aplicación. Finalizamos con unas breves conclusiones.

2 La directiva INSPIRE, los datos de referencia y la Administración local

La directiva INSPIRE señala que temas considera como datos de referencia pero no quién es responsable de su creación. Este punto analiza que temas son idealmente del ámbito local y qué implicaciones tiene para su gestión.

Los anexos I y II de la directiva INSPIRE concretan qué temas son considerados de referencia. En la Tabla 1 se recogen aquellos temas y subtemas en los que la administración local española es potencialmente responsable de su gestión y/o mantenimiento. Esta selección se basa en las competencias municipales y en el uso de dichos datos en la gestión municipal.

Anexo I y II	Definición INSPIRE	Subtema local
Nombres geográficos	Nombres de zonas, regiones, localidades, ciudades, periferias, poblaciones o asentamientos, o cualquier rasgo geográfico o topográfico de interés público o histórico.	Nombres geográficos locales.
Unidades Administrativas	División del territorio nacional en unidades de administración a nivel local, regional y nacional. Las unidades administrativas estarán separadas por límites administrativos. También se incluirán las fronteras del territorio nacional y las costas.	Unidades administrativas locales descentralizadas.
Redes de transporte	Redes de carreteras, ferrocarril, transporte aéreo y por vía navegable e infraestructuras correspondientes. Se incluirán las conexiones entre redes diferentes.	Red local de vías urbanas, caminos y vía rurales.
Identificadores de propiedad	Localización geográfica de las propiedades, efectuada sobre la base de la denominación de las direcciones, por ejemplo el nombre de la vía pública, el número de la finca, el código postal.	Tipo y nombre de vía pública y número de la finca.
Parcelas Catastrales	Áreas determinadas por límites catastrales y caracterizados por una situación jurídica específica de propiedad.	Propuestas de alteraciones catastrales.

Fuentes: Ley 7/1985 reguladora de Bases de Régimen Local, formatos de intercambio del Instituto Nacional de Estadística (INE), formatos de intercambio Dirección General del Catastro (DGC), elaboración propia

Tabla 1 Temas de referencia local

Salvo en el caso de las Parcelas Catastrales, el subtema local es tan amplio que la administración local en su conjunto es responsable de la gestión de la mayoría de los datos. Cuando se aplique la directiva INSPIRE los sistemas de información locales deberán adaptarse para que estos temas soporten una identificación no ambigua de estos datos (sea o no con un sistema común de identificadores únicos

[4]), tengan sus modelos establecidos y, por ende, existan mecanismos que eviten inconsistencias y errores.

3 Problemas en la manipulación de datos de referencia

Hay una serie de problemas en los datos de referencia que tienen que ser resueltos a nivel de datos para llegar a la situación señalada en el apartado anterior. Estos problemas se han detectado según se iban cumpliendo las metas marcadas a lo largo de estos dos años para IDEZar [5] con vistas al desarrollo de un geoportal orientado a los ciudadanos (<http://www.zaragoza.es/idezar>). Los problemas detectados son:

- **Baja calidad espacial.** Es frecuente que no se disponga de información espacial adecuada o falten relaciones con otros datos espaciales importantes para su explotación
- **Duplicidad de fuentes.** Algunos de los temas tienen más de una lista de referencia que se solapan.
- **Un uso no sistemático.** No todos los elementos de un tema están espacialmente caracterizados de la misma manera. Por ejemplo, al usar las direcciones en zonas rurales para identificar centros municipales discrecionalmente se indica o no la junta vecinal en la que se encuentra.
- **Identificadores alternativos ad-hoc.** En ausencia de reglas o acceso a listas controladas se identifican los datos temáticos mediante identificadores alternativos ad-hoc. Por ejemplo se pueden encontrar varias formas de escribir el nombre de la calle.
- **Sesgo estadístico.** El diseño de los nombres geográficos, las unidades administrativas e los identificadores de propiedad está orientado a identificar población. Gran parte de esta información se intercambia con el Instituto Nacional de Estadística (INE) y son sus requisitos los que guían cómo deben ser los datos [6].
- **Describen mal zonas despobladas.** No son relevantes los nombres geográficos que describen el territorio pero que no tienen población o cuya población es diseminada en sentido estadístico.
- **Falta de representación espacial adecuada.** Excepción a este hecho son los temas relacionados con el urbanismo.
- **Direcciones como texto libre.** Consecuencia indirecta de los problemas anteriores.

4 Propuesta de solución

Para afrontar los problemas indicados en el punto anterior y encaminar la IDE en la línea del apartado 2 nuestra propuesta plantea establecer como elemento central un modelo de datos de referencia urbanos acompañado de procesos y herramientas de integración de información y de servicios de red para su explotación.

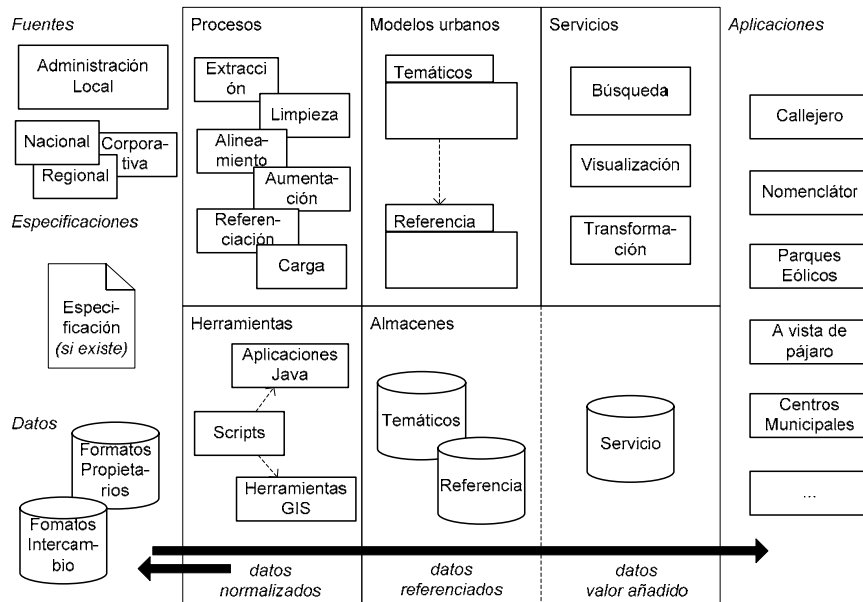


Figura 1 Propuesta de Solución

4.1 Modelo urbano

El elemento central es un modelo de datos de referencia urbana. Alrededor de él se establecen una serie de modelos de datos temáticos que utilizan el modelo de datos de referencia para caracterizarse espacialmente. Forman parte del modelo de referencia las *unidades administrativas* (juntas municipales y vecinales), los *códigos postales*, las *vías urbanas* y los *números de finca o de policía*. Ejemplos de modelos temáticos son el *transporte público urbano*, las diferentes *regulaciones fiscales*, las *zonas saturadas*, las *divisiones censales* y los *centros municipales*. Este modelo se implementa en una geodatabase. El único requisito que se exige es la

compatibilidad de los almacenes de los datos de referencia y temáticos con la especificación SFS [7] de OGC.

4.2 Procesos y herramientas

Los procesos de incorporación y homogenización de información transforman los datos en formato propietario o legado obtenidos de las diversas fuentes en datos correctamente referenciados. Estos procesos se implementan mediante scripts que encadenan aplicaciones Java y aplicaciones ETL-GIS. El primer paso del proceso de integración es la puesta bajo control tanto de las fuentes de datos, los datos que estas produce, su especificación (si existe) e información de proceso: cuándo se producen, se producen en sistemas legados, son relevantes para el proceso administrativo, cómo se intercambian y en que formatos. Tras analizar los datos bajo control se diseñan los *procesos de extracción, transformación y carga* de la información en la geodatabase. Los procesos de transformación pueden clasificarse en:

- **Procesos de limpieza.** Aplican heurísticas de limpieza de datos. Comprueban los dominios de los campos y su contenido.
- **Procesos de alineamiento.** Relacionan dos o más datos entre si. Dependiendo de la fuente, además de detectar duplicados, se detectan inserciones, modificaciones y borrados.
- **Procesos de aumentación.** Aplican heurísticas, procesos y herramientas para aumentar la información espacial asociada a los datos.
- **Procesos de referenciación.** Herramientas y procesos que permiten asociar datos temáticos de forma sistemática con datos de referencia almacenados en la geodatabase.

Tras la transformación disponemos de datos normalizados y enriquecidos que actualizaran mediante procesos de carga la geodatabase. Adicionalmente estos datos pueden ser devueltos a la fuente como datos normalizados.

4.3 Servicios de Red

En nuestra propuesta las aplicaciones Web acceden a la información mediante *servicios Web de búsqueda y visualización*. Estos servicios se caracterizan por no acceder directamente a los almacenes de datos temáticos y de referencia. Existen almacenes con datos con valor añadido adaptados a los requerimientos de las aplicaciones. Estos almacenes se han construido utilizando *servicios de Web de*

transformación que acceden directamente a los almacenes datos temáticos y de referencia.

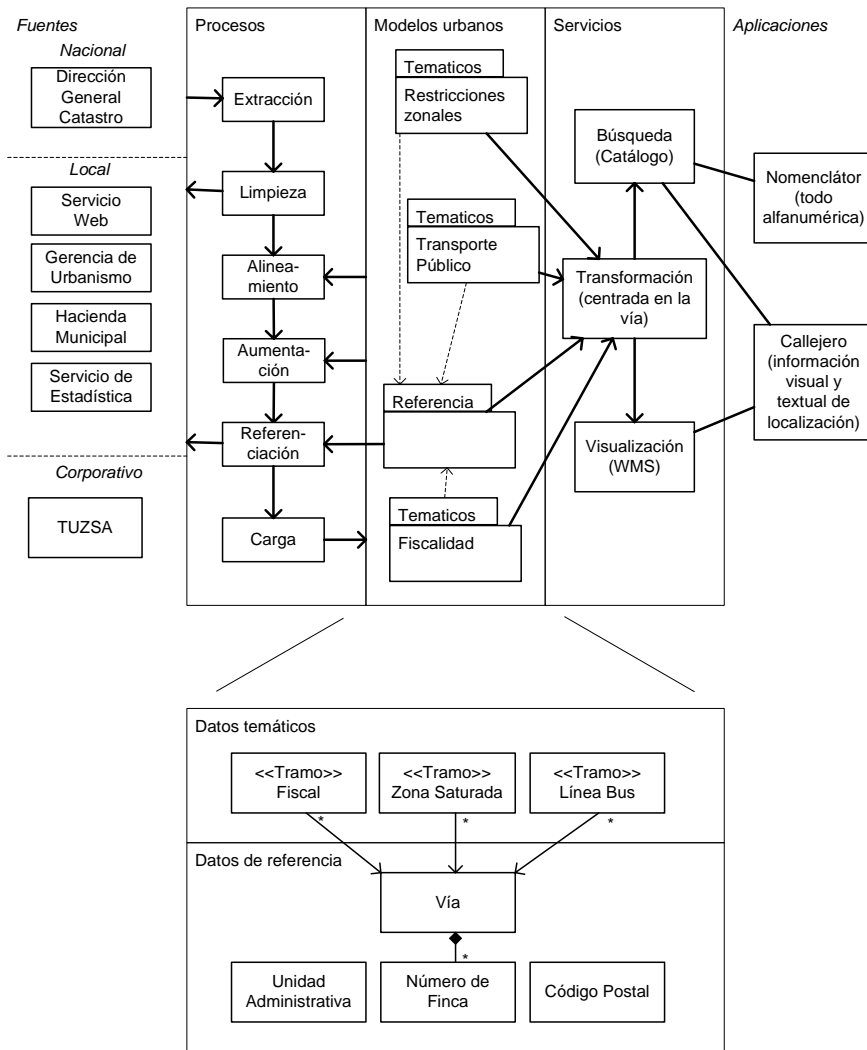


Figura 2 Caso de aplicación del Callejero y Nomenclátor

5 Caso de aplicación

Esta propuesta se ha aplicado en la construcción de las aplicaciones callejero y nomenclátor de IDEZar (ver Figura 2). El callejero es una de las aplicaciones más empleadas en la IDE local. El nomenclátor de calles es una aplicación diseñada para presentar toda la información textual asociada a las calles y que inicialmente estaba desperdigada en diferentes bases de datos. Los datos de referencia relevantes para la elaboración de un callejero son la vía, los números de finca a ella asociados así como la unidad administrativa (junta municipal, junta vecinal) y el código postal. Para la elaboración del nomenclátor se requiere además de información temática sobre fiscalidad, sobre restricciones zonales y sobre transporte público referenciada a rangos de portales o tramos de vías. En ambas aplicaciones casos el elemento clave es la vía.

Para cada tema los datos servidos proceden de más de una fuente como la Dirección General de Catastro, el Servicio de Información Geográfica y el Servicio Web del ayuntamiento.

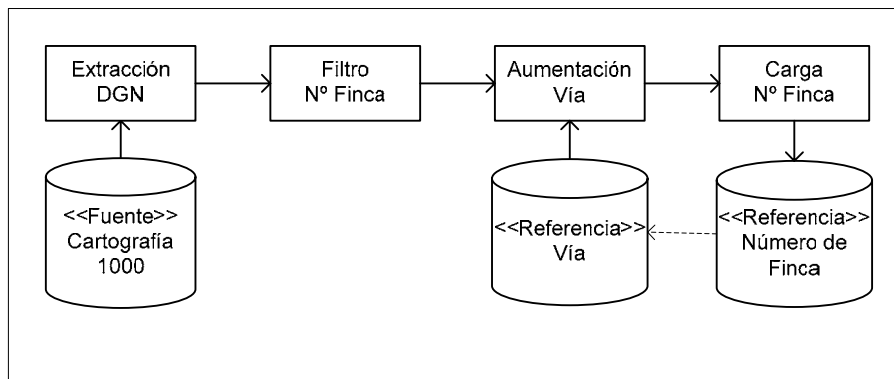


Figura 3 Procesado de números de finca

Por ejemplo, para obtener los números de finca (ver **Figura 3**) que se encontraban en la base cartográfica municipal en formato Microstation Design se utilizó una herramienta ETL – GIS. Se aplicaron diversas heurísticas de filtro y limpieza adecuadas a los usos del ayuntamiento para extraer las etiquetas que representaban números de finca. Al faltar un atributo que informara de la calle a la que el número de finca se refiere, se combinó aplicando una serie de heurísticas con la información ya almacenada de vías, las cuales procedían de una fuente distinta. Los

números de finca, ya correctamente referenciados a vías, fueron entonces almacenados.

Las dos aplicaciones, callejero y nomenclátor, comparten el mismo servicio de búsqueda, basado en tecnología de catálogo. Solo el callejero utiliza un WMS para situar visualmente la calle o la dirección buscada. Aprovechando tecnologías que transforman consultas SQL en XML se han desarrollado herramientas de extracción, transformación y carga que producen un almacén de información optimizado para responder a consultas sobre direcciones. La compatibilidad de la geodatabase con SFS hace trivial la conversión de los datos de direcciones en una capa servida por el WMS.

6 Conclusiones

Es evidente que la administración local es la autoridad que mejor conoce las parcelas donde residimos y trabajamos, cómo las identificamos, cuales son las rutas en el primer y último kilómetro de cualquier viaje y con qué nombres identificamos las comunidades donde residimos. Es por tanto un actor básico en cualquier infraestructura espacial y se verá obligado, tanto por necesidades de modernización de la administración como por los requerimientos de directivas europeas, a dar los pasos necesarios para el uso de procesos, herramientas y modelos adecuados con el fin de mantener adecuadamente los datos de referencia. IDEZar es un ejemplo de cómo se está realizando este cambio.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el proyecto TIC2003-09365-C02-01 del Plan Nacional de Investigación Científica y Desarrollo Tecnológico del Ministerio de Educación y Ciencia de España. El trabajo de J. López (ref. B136/2006) ha sido parcialmente financiado por una beca del Gobierno de Aragón.

Referencias

- [1] Muro-Medrano, P.R, Gould, M., Bernabé, M.A. “IDE-E: Technological advances for a Webbased National Spatial Data Infrastructure, convergence with the European initiative INSPIRE TIC2003-09365-C02”, Jornada de Seguimiento de Proyectos, Programa Nacional de Tecnologías Informáticas, 2005

- [2] Comisión de las Comunidades Europeas, “Propuesta de DIRECTIVA DEL PARLAMENTO EUROPEO Y DEL CONSEJO por la que se establece una infraestructura de información espacial en la Comunidad (INSPIRE)”, Julio 2004 , COM(2004) 516 final.2004/0175 (COD)
- [3] Open Geospatial Consortium: “OpenGIS Web Services Architecture”, Reference number OGC 03-025, 2003.
- [4] Recommendation for Second Reading on the Council common position for adopting a directive of the European Parliament and of the Council establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) (12064/2/2005 – C6-0054/2006 – 2004/0175(COD)), 13 de junio de 2006
- [5] López Pellicer, F.J., Álvarez, P. Muro-Medrano, “IDEZar: un ejemplo de implantación de una IDE local” Jornadas Técnicas de las Infraestructuras de Datos Espaciales de España – JIDEE 05, Madrid, 2005
- [6] Ministerio de la Presidencia. “Orden de 11 de Julio de 1997 sobre Comunicaciones electronicas entre las Administraciones publicas referentes a la Informacion de los Padrones municipales”. BOE de 16 de Julio de 1997, nº 0169
- [7] Open Geospatial Consortium: OpenGIS® Simple Features Implementation Specification for SQL (SFS), 99-049, 1999