

Aspectos de modelos e infraestructura de servicios para el soporte de un servicio nacional estándar de nomenclátor en Web

F.J .López-Pellicer, R. Bejar,
F. J. Zarazaga-Soria, P. R. Muro-Medrano

Departamento de Informática e Ingeniería de Sistemas
Universidad de Zaragoza, María de Luna 1, 50018 Zaragoza
Tlf: 976 762 134 Fax: 976 761 914. e-mail: iaaa@unizar.es

Resumen

Este artículo describe cómo se concibe un nomenclátor para los *Sistemas de Información Geográfica*, las *Bibliotecas Digitales*, los *Procesos de Estandarización Toponímica* y la *Cartotoponimia* y cómo influye en su publicación en una *Infraestructura de Datos Espaciales*.

Palabras clave: IDE, SIG, Biblioteca Digital, Procesos Estandarización, Nomenclátor.

1 Introducción

Un *nombre geográfico* es un nombre propio usado consistentemente en un lenguaje para referirse o identificar a un lugar, fenómeno o área que tenga una identidad reconocible en la superficie de la Tierra [1] cuya “principal función (...) es servir como etiqueta, y como tal, su significado semántico, aun cuando sea evidente, es una consecuencia de su rol como etiqueta” [2]. Una colección de nombres geográficos organizada de una determinada forma es un *nomenclátor*. Una de esas formas es el *nomenclátor nacional estándar* cuyo propósito final es normalizar el uso de nombres geográficos en las administraciones públicas. En este nomenclátor encontramos fenómenos naturales, poblaciones humanos, divisiones políticas, áreas administrativas, rutas de transporte y estructuras creadas por el hombre -millones de nombres a

nivel nacional -. Solo la microtoponimia no utilizada en la administración o sin relevancia cultural suele quedar fuera de esta colección.

En una *Infraestructura de Datos Espaciales* (IDE), entendida como la colección de tecnologías, políticas y acuerdos institucionales que facilitan la disponibilidad y acceso a datos espaciales [3], *nomenclátor* es el nombre genérico del *conjunto de acuerdos, estándares e infraestructuras* que dan soporte al acceso a *colecciones de nombres geográficos* y a servicios basados en dichas colecciones. El *nomenclátor nacional estándar* puede publicarse a través del punto de acceso a la IDE correspondiente acompañado de los correspondientes estándares e infraestructuras. Por ejemplo, en la IDE de España se está elaborando una propuesta de Modelo de Nomenclátor de España [4] con el fin de facilitar el intercambio de nomenclátors nacionales estándar entre las administraciones y la implementación de servicios de nomenclátor.

Cómo sean esos acuerdos, estándares e infraestructuras es el resultado del equilibrio entre los distintos tipos de usuario del *nomenclátor nacional estándar*: los *Sistemas de Información Geográfica*, las *Bibliotecas Digitales*, los *Procesos de Estandarización Toponímica* y la *Cartotoponimia*. Para cada uno el nomenclátor tiene un rol distinto: un *Sistema de Referencia*, un *Diccionario Geográfico*, un *Nomenclátor Estándar* o un *Índice de Topónimos*. El problema que se afronta en una IDE es organizar de tal manera la publicación del nomenclátor nacional estándar de tal forma que satisfaga de forma adecuada cada uno de estos roles.

Este artículo tiene dos objetivos. El primero es describir dichos modelos de nomenclátor, señalar las similitudes y diferencias, y las implicaciones sobre las infraestructuras. El segundo es proponer una infraestructura que facilite la publicación de un nomenclátor teniendo en cuenta las consecuencias de los requerimientos detectados. Ambos objetivos se analizan en el contexto de la publicación del servicio nacional estándar de nomenclátor mediante una IDE.

Este artículo se organiza de la siguiente manera. En el apartado 2 se define qué rol tiene un nomenclátor para cada una de estas áreas. En el apartado 3 se describe las necesidades que dan origen a estos roles. En el apartado 4, se muestra que cada uno de los modelos tiene características y peculiaridades propias que complican que solo a partir del nomenclátor nacional se cumpla una su adecuada publicación en la IDE. En el apartado 5 se afronta cómo una IDE debe publicar un nomenclátor nacional estándar y cómo cada una de estas publicaciones responde a cada una de dichas áreas. En el apartado 6 se describe una solución que se está utilizando para publicar nomenclátors. Finalmente, en el apartado 7 se establecen algunas conclusiones.

2 Roles de un nomenclátor

Tal como se ha indicado, hay cuatro áreas que requieren vistas diferentes del *nomenclátor nacional estándar*: los *Sistemas de Información Geográfica* (SIG), las *Bibliotecas Digitales* (BD), los *Procesos de Estandarización Toponímica* (PET) y la *Cartotoponimia* (CT).

Para los SIG un nomenclátor es el *conjunto de estándares e infraestructuras* que dan soporte al uso de *nombres geográficos como identificadores espaciales con tipo* dando lugar a *Sistemas de Referencia por Identificadores Geográficos*. Un ejemplo es el nomenclátor definido en el estándar ISO 19112 [5]. En estas condiciones, por ejemplo, la geometría de una finca puede ser sustituida por una representación compacta, comprensible y multiuso: “parcela catastral Coso 6, Zaragoza”.

En una DB un nomenclátor es un *diccionario geoespacial de nombres geográficos*. Sus componentes esenciales son *nombres geográficos, huellas espaciales y tipos*. Dentro de esta aproximación el nomenclátor puede tomar la forma de *tesauro de nombres geográficos*. Un ejemplo del primer tipo es el nomenclátor de *Alexandria Digital Library* (ADL Gazetteer) [6] y de los segundos es el *Tesauro Getty de Nombres Geográficos* (TGN) [7]. Los nombres contenidos se utilizan para *georreferenciar indirectamente recursos electrónicos* [8]. Por ejemplo, el área aplicación de una norma del ayuntamiento de Zaragoza publicada en su página Web puede ser descrita como “World; Europe; España; Aragón; Zaragoza; Zaragoza” según TGN o “Zaragoza; Spain” indicando que es de tipo “populated places” según ADL Gazetteer.

Como resultado de un PET el *nomenclátor* es el medio por el cual se disemina los *nombres geográficos oficiales* establecidos. El modelo que sirve de base a estos nomenclátos ha sido asentado en a lo largo de las Conferencias de las Naciones Unidas en la Estandarización de Nombres Geográficos que han tenido lugar desde 1967 [1]. Así un nomenclátor es una *lista en un orden lógico de los nombres geográficos organizados por divisiones administrativas*. Si el nombre geográfico es oficial debe adjuntar información adicional como el *tipo* de entidad geográfica identificada, su *localización* y *extensión*, así como *nombres alternativos* y *variaciones en la ortografía*, y para cada nombre su *lenguaje*. También se considera aceptable incluir información como la *elevación*, el *número de habitantes*, su *localización en hojas de series cartográficas oficiales*, *información gramatical*, *transliteraciones*, etc. Dada su complejidad, en 1977 se propuso la elaboración de *nomenclátos concisos* como paso intermedio al nomenclátor nacional. En el caso de España [4], este es una lista que contiene el nombre geográfico preferente aplicado cada una de las entidades culturales o

naturales con nombre. Cada entrada incluye, además, otros nombres de uso restringido y aquellas denominaciones oficiales anteriores de menos de 100 años, cada una con su idioma respectivo y su fuente de procedencia. Para identificar la entidad a la que se aplica se indica su tipo, las divisiones administrativas hasta el segundo orden que la incluyen, un punto representativo así como la hoja de una serie cartográfica donde este punto se sitúa.

Finalmente, para la CT un nomenclátor es un *índice de topónimos*, la lista de *nombres geográficos normalizados* de una *unidad administrativa* en la que junto a cada *nombre oficial* hay al menos la suficiente información adicional que permite localizarlo en forma escrita en un mapa e identificar a qué entidad representada gráficamente se refiere. En mapas digitales, esta lista de topónimos tiene prácticamente los mismos atributos que la información representada gráficamente: una etiqueta, un atributo de tipo y un punto, sin información adicional. Hay que tener en cuenta que la CT es un proceso orientado a mostrar gráficamente en un mapa las formas escritas de una entidad geográfica acompañando a su representación gráfica mediante un punto, una línea o un área [1].

3 Necesidades cubiertas por un nomenclátor

Hemos visto en el apartado anterior los cuatro roles de un nomenclátor: un *Sistema de Referencia*, un *Diccionario Geográfico*, un *Nomenclátor Estándar* o un *Índice de Topónimos*. Estos roles están relacionados con necesidades como poder referenciar espacialmente en forma humana – la traducción a coordenadas es relevante –, describir espacialmente en forma humana – la traducción a coordenadas no es relevante –, minimizar los costes por errores de interpretación y reforzar la autoridad nacional.

Los nomenclátors SIG, creados por expertos en información geográfica y en geomática, son el resultado de la necesidad de *normalizar la descripción indirecta de propiedades espaciales mediante nombres geográficos*. Los nomenclátors BD, creados principalmente por biblioteconomistas e historiadores, surgen de la necesidad de *catalogar documentos que pueden tener referencias espaciales futuras, contemporáneas e históricas*. Es decir, los nomenclátors SIG contienen información cuyo uso es referencial mientras que los nomenclátors BD contienen información cuyo uso es descriptivo.

La información contenida en un nomenclátor PET no sirve para un uso práctico descriptivo o referencial. En realidad tiene objetivos que dan satisfacción a necesidades sociales en lo económico y en lo político. El económico es derivado de la reducción de costes derivados de errores en el uso los nombres geográficos

por administraciones, empresas y ciudadanos. El político es resultado del aumento de la autoestima nacional tanto mediante la protección y/o recuperación de la herencia lingüística como por el refuerzo de la autoridad sobre el territorio al establecer que nombres son de uso oficial. Entre otros, este resultado se ve en forma de mapas en los que aparecen las formas oficiales y/o estandarizadas de una entidad.

Los nomenclátors CT solo tienen como objetivo ser reflejo de las formas escritas presentadas en un mapa. En ese sentido cubren las mismas necesidades que los PET pero circunscritos al ámbito de la comunicación visual de localizaciones.

Pero estas necesidades se pueden satisfacer parcialmente con nomenclátors producidos para otros usos. En este sentido un nomenclátor es un artefacto fácilmente reutilizable (Figura 1).



Figura 1 Un mismo nomenclátor tiene muchos usos

4 Comparación de los modelos de nomenclátor

En los dos apartados anteriores se han revisado los roles y las necesidades que cubren los nomenclátors en diferentes áreas. En este apartado resumimos las principales características que comparten los modelos de contenidos de nomenclátor y proponemos qué contenidos debería tener el nomenclátor que se publicara mediante una IDE para dar satisfacción a las diferentes vistas.

Tomemos como modelos prototípicos el *ADL Gazetteer Content* para las BD,

el *Modelo de Nomenclátor de España* para los PET, el estándar *ISO 19112* para los SIG y la capa de topónimos de la mayoría de las *Bases Topográficas* para la CT. La Figura 2 muestra dichos modelos de una forma simplificada.

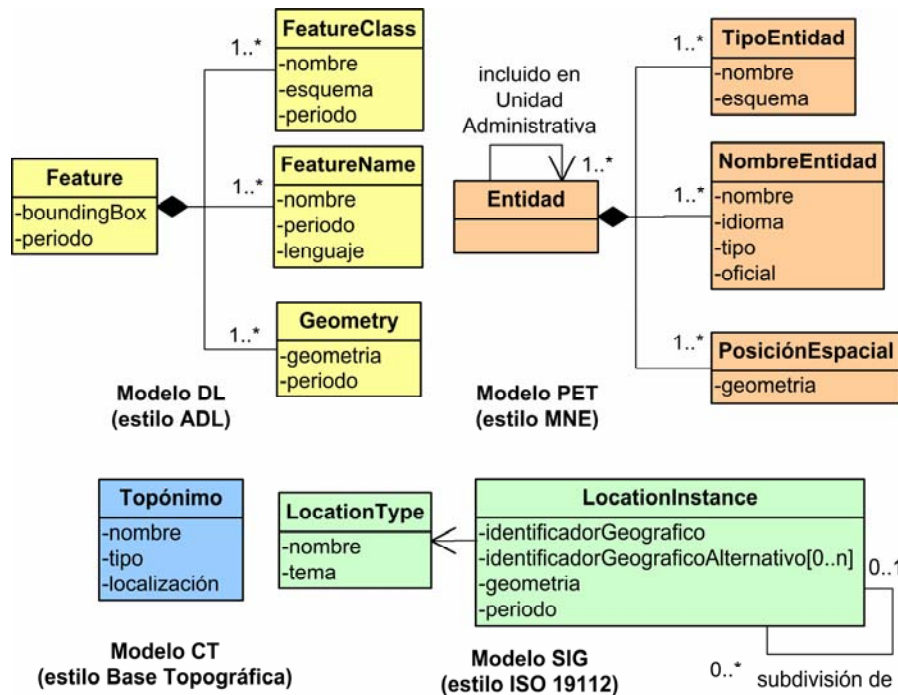


Figura 2 Diversos modelos de nomenclátor

Para estos ejemplos el nombre y su aplicación a entidades naturales y culturales es un elemento inseparable. De hecho se habla de forma indistinta de catálogos de nombres y catálogos de instancias. Por ejemplo, en el estándar ISO 19112 se define un nomenclátor como “*catálogo de instancias (de localización) de una o más clases de fenómenos que contienen cierta información sobre posición*” y un poco más adelante lo define también como un “*catálogo de identificadores geográficos que describen instancias de localización*” o abstracciones del mundo.

El tratamiento de los nombres es muy variado. El diccionario geográfico y el nomenclátor estándar tienen la información más amplia, destacando en particular el idioma. En ambos casos se incluyen nombres alternativos. No se da el caso en los sistemas de referencia y en los índices de topónimos.

Solo para describir e identificar es necesario un tesoro de tipos. En cuanto la tipología, solo el sistema de referencia define simultáneamente un sistema de tipos. En los diccionarios geográficos en forma de tesoro, el tesoro de tipos forma parte de él. En el resto de los casos los tipos vienen dados.

La localización espacial se describe de diferentes formas. El diccionario geográfico y el sistema de referencia requieren de la extensión y en su ausencia de la posición, mientras que los nomenclátors estándar y los índices de topónimos se conforman con una posición aproximada. Estos utilizan para completar la información espacial referencias a hojas de series cartográficas y a unidades administrativas.

El aspecto temporal o es sobredimensionado o es implícito. El aspecto temporal en el diccionario geográfico tiene como objetivo la descripción histórica: conocer para cada atributo su valor a lo largo del tiempo. En el sistema de referencia no hay aspecto temporal tal cual, solo es una marca de validez. El aspecto temporal en los nomenclátors estándar y los índices de topónimos esta implícito en su condición de “foto fija” asociada a la fecha de su publicación.

Las relaciones jerárquicas de contención son relevantes para los sistemas de referencia y los diccionarios geográficos si se presentan como tesoros.

Considerando lo anterior, si se creara un único repositorio del que se sirvieran todos los servicios del nomenclátor estándar este debería ser una al estilo de las BD pero incluyendo: un tesoro de tipos y la huella de la entidad siguiendo los requisitos de los sistemas de referencia, los puntos de localización con la idea de ser utilizados en mapas y un tratamiento diferenciado a las relaciones de jerarquía según sean de división o de contención.

5 Papel del nomenclátor estándar en una IDE

En este apartado se analiza las implicaciones de publicar un nomenclátor nacional estándar mediante el punto de acceso de una IDE. Desde nuestro punto de vista las implicaciones están condicionadas principalmente por las necesidades y objetivos de la visión PET del nomenclátor. El resto de las visiones condicionan el diseño de servicios específicos.

Hay que tener en cuenta que dado que en el contexto europeo el concepto de IDE es el resultado acumulado en el campo de los datos geográficos de una continuada serie de Directivas de la Comunidad Europea, aquellas que traten de nombres geográficos condicionarán cómo se publica este nomenclátor. La más importante de todas ellas es la propuesta de directiva INSPIRE. En ella, los nombres geográficos definidos como cualquier rasgo geográfico o topográfico

de interés público o histórico son un conjunto de datos espaciales que las autoridades públicas deben de mantener y difundir. INSPIRE todavía no ha sido aprobada por lo que las reglas de implementación no han sido definidas. Es previsible que las reglas de implementación para los nombres geográficos siga las recomendaciones de los PET por tres motivos: los nomenclátors estándares PET se suelen publicar en los puntos de acceso de las IDE nacionales para su difusión, suelen coincidir la autoridad la que administra la IDE y con la que elabora el PET, y ambas participan en el proceso de decisión de INSPIRE.

Según los PET el éxito de un nomenclátor nacional estándar está en la capacidad de dar a autoridades, empresas y público en general información útil sobre los nombres oficiales de una forma rápida y sencilla. Por ello, para una óptima difusión los nombres normalizados, la IDE debe proporcionar servicios de catálogo que usen nombres geográficos estándar para describir, servicios de proceso que los utilicen referenciar y servicios de datos que produzcan mapas donde aparezcan rotulados.

En consecuencia, el punto público de descubrimiento, acceso y distribución de datos y servicios espaciales de la IDE nacional debe, en relación con los nomenclátors estándar:

- Describir los *tipos* de nomenclátors que hace disponibles, qué *servicios* los proveen y qué *modelo de distribución* se aplica. Por ser datos de naturaleza oficial se debe considerar en la política de distribución la *política de notificación y difusión* de las novedades, las correcciones y las actualizaciones.
- Difundir el *modelo de contenido* de cada uno de ellos así como su implementación en diferentes *formatos de intercambio estandarizados*. Estos formatos de intercambio deben recoger el contenido en función del servicio y del las necesidades usuario. No tienen las mismas necesidades de información las administraciones, las empresas o los ciudadanos.
- Facilitar su descubrimiento publicando su existencia en los *servicios de catálogo* y su difusión haciéndolo accesible mediante *servicios de datos* (colecciones de datos, elaboración de mapas digitales) y *servicios de proceso*, en particular el *servicio de nomenclátor*. Estos servicios deberían poderse integrar en aplicaciones desarrolladas por terceros. Notar que lo relacionado con servicios de búsqueda y catálogo está influenciado por la visión BD mientras que la elaboración de mapas digitales está influenciado por la visión CT.
- Si se ha descentralizado la autoridad o la gestión toponímica el nodo de

la IDE y en función de las políticas de descentralización y actualización, se establecerán *procesos de harvesting y búsquedas en cascada*.

- En función de la política de distribución, establecer *servicios de harvester* para mantener sincronizadas las copias remotas de los usuarios de forma incremental.
- En función de la política de notificación, establecer *servicios de suscripción-notificación* para notificar los cambios en el contenido de los nomenclátors.
- Finalmente, como parte de la política de difusión, proporcionar *aplicaciones de nomenclátor*, tanto de demostración como para usuarios avanzados, que permitan realizar preguntas basadas en nombres geográficos utilizando mapas como medio para transmitir el contexto espacial de las respuestas.

La visión SIG del nomenclátor es determinante en el diseño de los *Servicios de Nomenclátor* entendidos como parte de los *Servicios de Proceso* (SP) de la IDE. La funcionalidad que deberían ofrecer en base al nomenclátor estándar estaría formada por:

- La traducción de nombres geográficos a sus coordenadas geoespaciales; es un papel similar a los *Servicios de Transformación de Coordenadas*, salvo que uno de los sistemas de referencia que intervienen en la transformación es un *Sistema de Referencia Basado en Identificadores Geográficos*.
- El soporte, e incluso la sustitución en su lugar, a otros SP como los *Servicios de Geocodificación* y los *Servicios de Tesauro Geográficos*.
- El soporte a *Servicios de Datos* basados en localización como los *Servicios de Directorio*.
- El soporte a *Servicios de Catálogo* que realicen búsquedas basadas en descripciones espaciales textuales.

Un ejemplo de la arquitectura resultante se puede ver en la Figura 3.

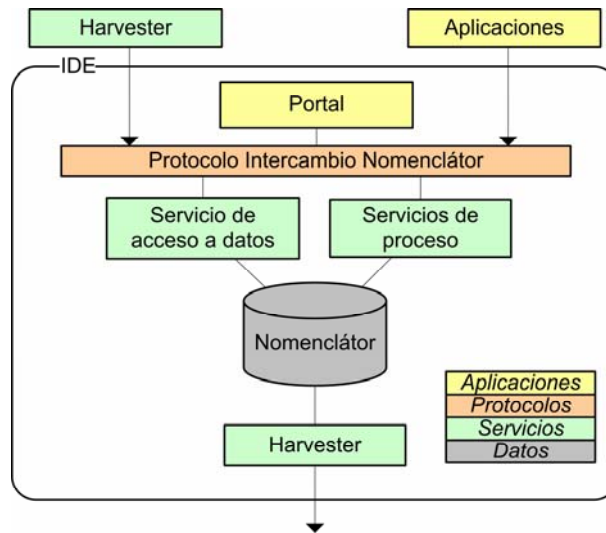


Figura 3 Arquitectura básica con harvesting

6 Nuestra solución

En los apartados anteriores hemos analizado los modelos y servicios relacionados con la publicación de un nomenclátor estándar en una IDE. En este apartado proponemos una infraestructura orientada a facilitar la publicación de un nomenclátor teniendo en cuenta las consecuencias de los requerimientos detectados.

Nuestra propuesta es un sistema libremente basado en los diccionarios geográficos en el que hay al menos tres tipos de elementos de primera clase o recursos (Figura 4): el nombre, la entidad y el tipo. Este sistema utiliza el vocabulario *Dublín Core* [9] como base de una *ontología* para definir las propiedades de cada uno de estos recursos. Además, utilizando esta misma ontología se describen tanto las fuentes de datos como Con esta ontología se definen adicionalmente fuentes de colecciones de nombres y pasarelas entre el modelo del repositorio y los modelos de cada una las fuentes. Utilizando esta información se diseñan y construyen las transformaciones que a partir de elementos de la colección fuente se crean entidades (*features*) con su tipo (*type*) y sus topónimos (*names*). De igual manera se han establecido pasarelas y transformaciones estables entre el modelo del repositorio y los modelos de vistas optimizadas para cada uno de los formatos y servicios de la IDE.

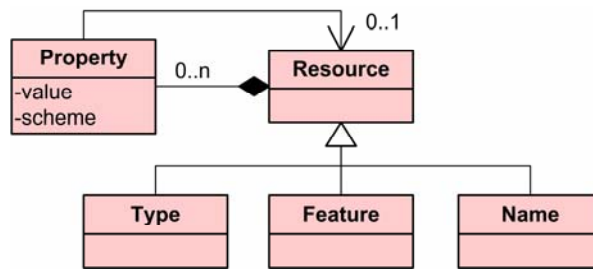


Figura 4 Modelo básico

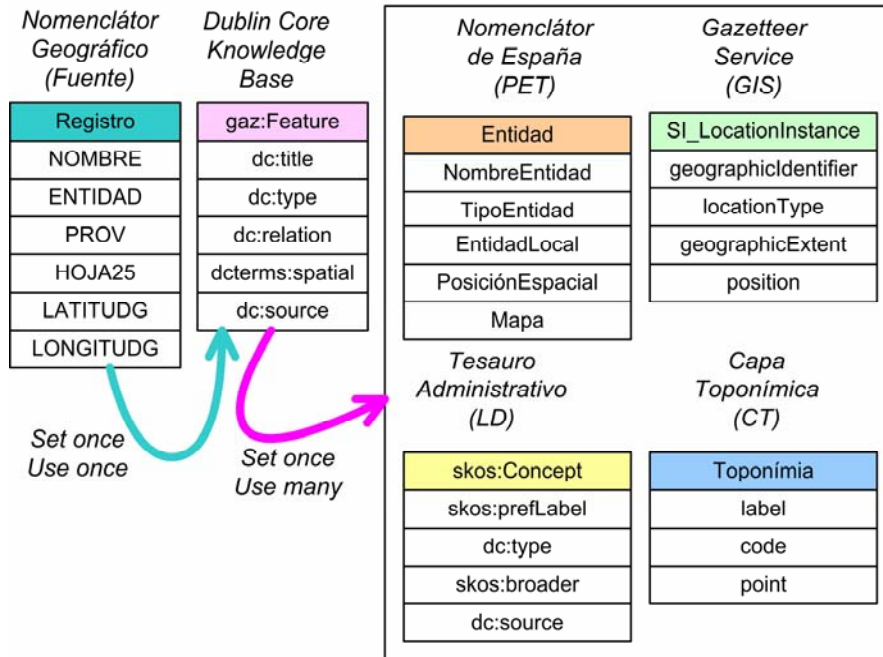


Figura 5 Correspondencias

Por ejemplo (ver Figura 5), utilizando dichas pasarelas a partir de los datos del Nomenclátor Geográfico del IGN se puebla la base de conocimientos. Esta labor es dependiente de la fuente, y es posible que en sucesivas recolecciones se tenga que realizar ajustes (*set once, use once*). La ventaja aparece al transformar de la base de conocimientos a los modelos de cada uno de los servicios (*set once, use many*), que es estable mientras que la base de conocimientos no sufra un cambio sustancial.

De forma resumida, el sistema esta formado por procesos e adaptación y carga, uno por cada tipo de fuente, un framework de transformación al modelo de

Dublín Core, procesos de transformación a modelos optimizados para diferentes servicios y una serie de tareas de limpieza, integración de fuentes, deduplicación y aumentación de información orientados a mejorar y enriquecer la colección de nombres (Figura 6)

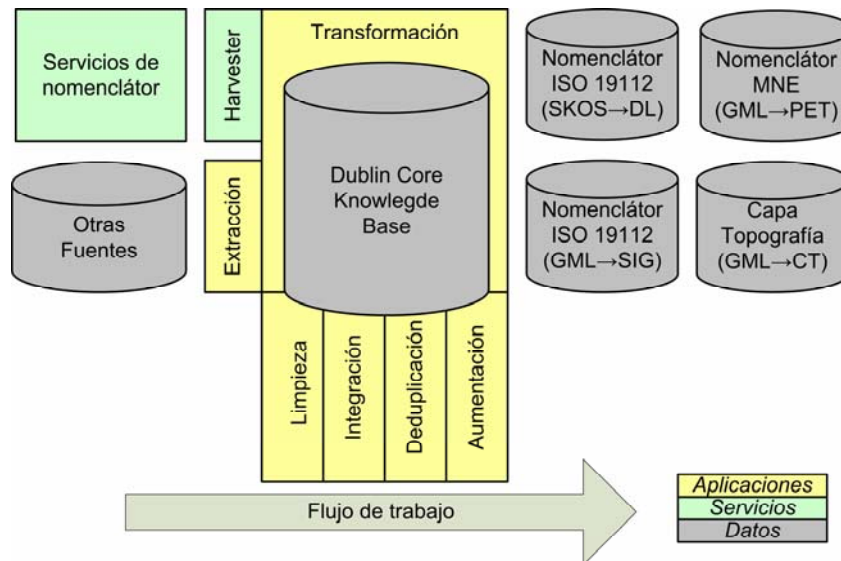


Figura 6 Explotación del nomenclátor

7 Conclusiones

Se pueden señalar las siguientes conclusiones:

- El diseño del nomenclátor en una IDE es complejo por tener que responder a necesidades de descripción, referenciación, normalización y dibujo, por señalar algunas. El modelo resultante es sobrecargado.
- Por la particular preponderancia del nomenclátor PET, la mayor parte del contenido publicado va a ser consistente con sus requisitos. El modelo PET no es adecuado ni para describir, ni para referenciar, aunque si vale para la producción de mapas digitales. En consecuencia las visiones BD y SIG van a tener problemas de adecuación de los datos, pero no así la visión CT.
- Las necesidades del PET implica servicios adicionales en una IDE como servicios de harvesting y de búsqueda en cascada sobre

colecciones de datos (que no de metadatos), y servicios de suscripción-notificación.

- Debido a la complejidad del problema se necesita una infraestructura especializada e integrar colecciones de nombres y generar cada una de las vistas. Además este sistema debe servir para, partiendo de los nomenclátors PET, integrar información de otras fuentes y poder aumentar la calidad de las visiones BD y SIG.

Agradecimientos

El trabajo de F. J. López-Pellicer (ref. B136/2006) ha estado parcialmente financiado por una beca del Gobierno de Aragón.

Referencias

- [1] United Nations Group of Experts on Geographical Names, UNGEGN (2006), *Manual for the national standardization of geographical names*, United Nations Publication, New York.
- [2] Kadmon, N. (2000), *Toponymy: the lore, laws and language of geographical names*, Vantage Press, New York.
- [3] Douglas D. Nebert, G., ed. (2004), *Developing Spatial Data Infrastructures: The SDI Cookbook*, Global Spatial Data Infrastructure Association.
- [4] Alcázar, A., Azcárate M. (2006), *Situación actual del Nomenclátor Geográfico Conciso de España*, en *Technical Papers of the 23rd Session of the UNGEGN*, UNGEGN.
- [5] AENOR (2005), *UNE-EN ISO 19112:2005 Información geográfica. Sistemas de referencia espaciales por identificadores geográficos*.
- [6] Alexandria Digital Library Gazetteer, (1999-) , Santa Barbara CA: Map and Imagery Lab, Davidson Library, University of California, Santa Barbara. Copyright UC regents. <http://www.alexandria.ucsb.edu/gazetteer/>
- [7] Getty Thesaurus of Geographic Names. The J. Paul Getty Trust. http://www.getty.edu/research/conducting_research/vocabularies/tgn/
- [8] Hill, L.L. (2000), Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints, en *ECDL '00: Proceedings of the 4th European Conference on Research and Advanced Technology for Digital Libraries*, Springer-Verlag, London, pp. 280—290.
- [9] Dublin Core Metadata Initiative (2005), *DCMI Abstract Model*. <http://es.dublincore.org/documents/abstract-model/>