

Unidades administrativas, una perspectiva ontológica

Javier Lacasta Miguel
Francisco Javier López Pellicer
Juanjo Floristán Jusué
Javier Nogueras Iso
Francisco Javier Zarazaga Soria

Departamento de Informática e Ingeniería de Sistemas.
Universidad de Zaragoza
e-mail: {jlacasta,fjlopez,juanf, jnog ,javy }@unizar.es

Resumen

El conjunto de datos formado por las unidades administrativas que conforman la organización territorial de un país tiene numerosas aplicaciones en una Infraestructura de Datos Espaciales. Para explotarlo de forma adecuada es necesario poseer un modelo de información capaz de afrontar su diversidad. Este trabajo propone la representación de las unidades administrativas mediante una ontología de dominio, describiendo su estructura y sus usos. También se muestra su aplicación al caso de España.

Palabras clave: Unidades Administrativas, Ontologías, Gazetteer, Nomenclator, Identificadores Geográficos.

1 Introducción

Bajo el nombre genérico de *unidades administrativas* como resultado de cambios políticos, económicos, sociales y culturales, y para centralizar o descentralizar la gestión administrativa, los estados han creado a lo largo de los siglos diferentes tipos de entidades de diversa naturaleza legal para cubrir ámbitos territoriales específicos. Dado su carácter oficial, son conocidas y usadas a todos los niveles (público, privado, personal) para identificar recursos explícitamente (dónde se encuentra, a dónde se aplica) o como parte de otros identificadores de referencia (direcciones, límites administrativos, listas de entidades).

Dado su papel como datos de referencia para la construcción de otros conjuntos de datos geográficos más elaborados, las unidades administrativas se han convertido

en uno de los recursos más relevantes a publicar a través de distintas iniciativas Infraestructuras de Datos Espaciales (IDE) que han surgido a nivel local, regional o nacional. De hecho, la propuesta de directiva INSPIRE para el establecimiento de una Infraestructura de Información Espacial en Europa incluye en su anexo I a las unidades administrativas como uno de los temas prioritarios de publicación [1]. Además exige que sean publicadas cumpliendo unos determinados requisitos de calidad para facilitar la interoperabilidad de datos caracterizados geográficamente en base a ellos. Un buen ejemplo de aplicación lo encontramos en datos fronterizos. Su localización podría estar definida solo mediante sus coordenadas, pero si se utiliza adicionalmente una unidad administrativa como referencia se evitan claramente ambigüedades.

Las unidades administrativas distan de ser una colección estable y uniforme de tipos e instancias. Su diversidad, sus peculiaridades y su carácter evolutivo, hacen necesario un modelo coherente que facilite su gestión y simplifique su uso. Este trabajo, propone la representación de las unidades administrativas mediante un modelo general basado en una ontología de dominio reutilizable, que defina los conceptos básicos de las unidades y sus relaciones, y una serie de ontologías de aplicación, cada una describiendo en detalle los tipos de unidades administrativas específicos de un país y sus instancias.

El artículo se organiza como sigue. La sección 2 describe el estado del arte en la caracterización de las unidades administrativas. La sección 3 describe la ontología creada y sus características. La sección 4 muestra los usos que dicha ontología puede tener en una IDE. Termina con unas conclusiones y trabajo futuro.

2 Estado del arte en la caracterización de las unidades administrativas

Un estado para organizar su territorio define mediante leyes que tipos de unidades administrativas pueden existir dentro de un área determinada y qué atribuciones legales tienen. Entre ellas puede estar la potestad de organizarse a su vez de forma autónoma creando nuevas unidades de orden inferior. Las unidades administrativas creadas pueden con el tiempo cambiar de forma, fusionarse con otras e incluso, si hay un cambio en la norma legal, evolucionar a otro tipo de entidad. Este conjunto de características hace que no solo diferentes países tengan distintos tipos de unidades administrativas, sino que distintas zonas de un mismo país tengan tipos distintos, no necesariamente relacionados.

Existen numerosos trabajos que describen diferentes modelos generales para representar cualquier tipo de entidad geográfica [2][3], e incluso estándares

internacionales como ISO19109 [4]. En el ámbito más restringido de las unidades administrativas existen recopilaciones de los modelos administrativos de los distintos países usando distintos modelos de organización del conocimiento, en forma de listas [5][6], tesauros [7] u ontologías [8]. El estándar internacional ISO19112 [9] describe el modelo lógico de un diccionario de nombres geográficos autorizados (nomenclátor), permitiendo describir las unidades administrativas como una jerarquía, pero no considera problemas como las especializaciones de tipos, el multilingüismo o la duración temporal (entendida como extensión temporal en el mundo legal).

El principal problema de los modelos y recopilaciones citados es la falta de una semántica adecuada en la representación de los tipos de unidades administrativas y las relaciones espaciales, temporales y de dependencia que presentan. Además, los datos que contienen o no son exhaustivos, presentando solo partes del modelo de unidades administrativas de uno o varios países o no se garantiza que los nombres usados para identificar las unidades sean los oficialmente reconocidos.

3 Ontología de Unidades Administrativas

Para solventar los problemas de representación señalados en el punto anterior y siguiendo el esquema descrito por Guarino [10], se ha creado un modelo de ontologías (Figura 1) en tres niveles. (1) Una ontología de alto nivel que define tipos de datos y relaciones generales e independientes del contexto. (2) Una ontología de dominio donde se definen los conceptos y relaciones reutilizables en el contexto de los modelos administrativos de diferentes países. (3) Y una ontología de aplicación por país, donde se representan los tipos específicos de unidades administrativas de cada país, junto con las instancias concretas de las unidades existentes.

Como ontología de alto nivel, se ha seleccionado la descrita en [11] por contener todos los conceptos y relaciones básicas necesarias para construir la ontología de dominio. Esta ontología mezcla los conceptos de alto nivel de Wordnet [12], PENMAN [13] y CYC [14] entre otros, como una extensión de la ontología propuesta por Sowa [15]. El uso de esta ontología de alto nivel en otros entornos, simplifica la combinación de nuestra ontología de dominio con otras ya existentes.

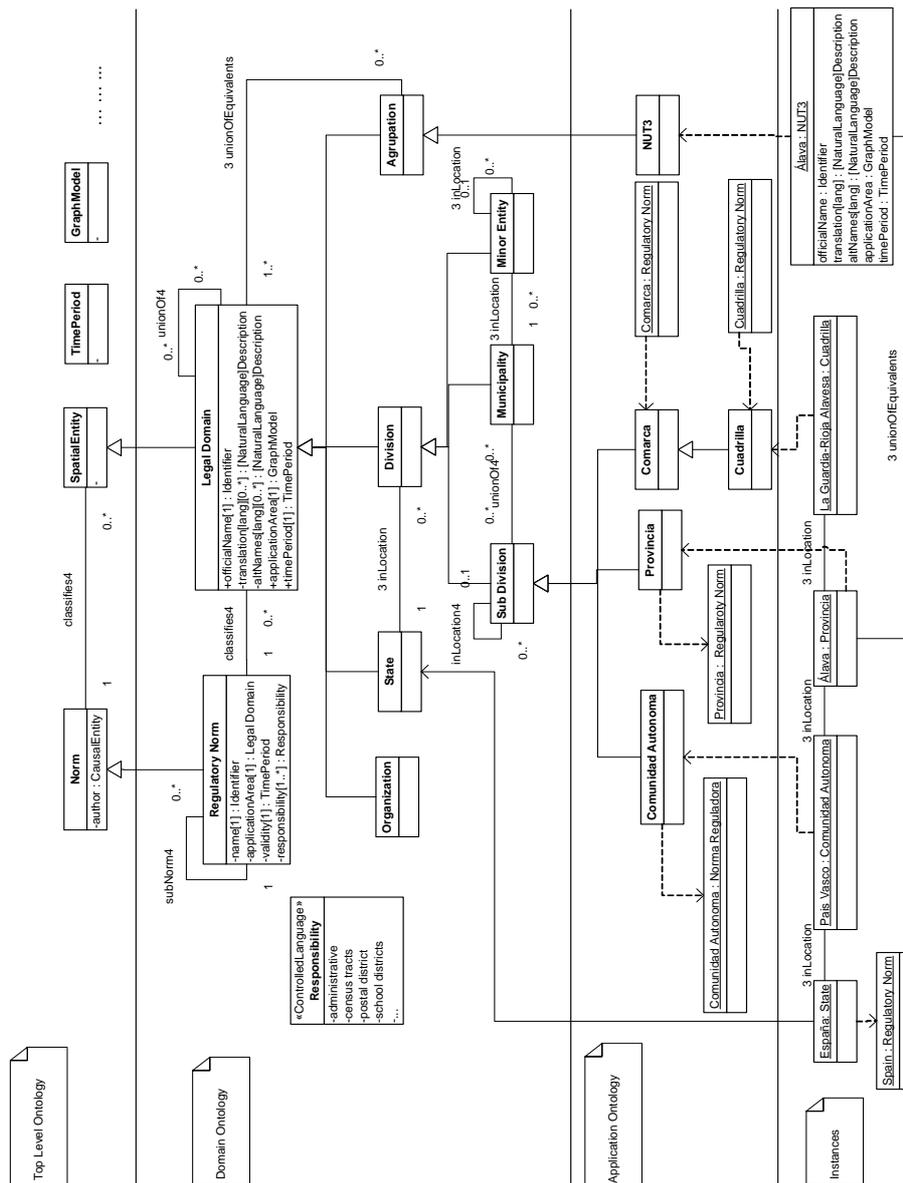


Figura 1: Ontología de Unidades Administrativas

Dado que las unidades administrativas representan dominios legales cuya existencia se basa en una norma oficial con carácter de ley que los regula. La ontología de dominio esta basada en los conceptos generales de “Norma

Reguladora” y “Dominio Legal”. Además, dada la organización jerárquica de los países, es habitual que una norma de validez a otras para regular administrativamente zonas específicas. En la ontología, las unidades administrativas son especializaciones de un “Dominio Legal” y distinguimos cuatro tipos, “Estado”, “División”, “Organización” y “Agrupación”. Los tres primeros son equivalentes a los dominios jurisdiccionales básicos definidos en ISO 15944-5 [16]. Así, un “Estado” es una entidad que tiene naturaleza legal propia y con la cualidad de reconocimiento entre pares. Las divisiones son las distintas particiones del territorio de un estado, y de entre ellas consideramos el “Municipio” como la entidad básica de división, agrupándose entre ellos para formar subdivisiones mayores o dividiéndose para formar entidades menores. Las organizaciones son agrupaciones de dominios jurisdiccionales, creados con una finalidad determinada (ONU, Euro-regiones de la UE). A esos tres tipos básicos se añade el concepto “Agrupación”. Se utiliza para tratar jerarquías funcionales con propósitos distintos al organizativo. Un claro ejemplo son las NUTS [17], divisiones de la Unión europea establecidas con fines estadísticos y que crean divisiones artificiales. Por ejemplo, establece la división “Noroeste” de España que cubre las comunidades autónomas de Galicia, Principado de Asturias y Cantabria.

Para modelar los tipos de unidades específicos de un país, se ha creado una ontología de aplicación que extiende los tipos descritos en la ontología de dominio y que ha sido enriquecida con las instancias existentes para los tipos definidos. En este primer paso nos hemos centrado en el caso de España, sin perder de vista la futura generación de las de los países limítrofes, creando la ontología de aplicación de las unidades administrativas por encima del municipio. Los datos utilizados son los proporcionados por el Registro de Entidades Locales de Ministerio de Administraciones Públicas, entidad encargada de recopilar los nombres oficiales de las unidades administrativas de España (asignados por las administraciones competentes), y por el Instituto Nacional de Estadística, que aporta información sobre los cambios año por año.

Las unidades administrativas tienen un “área de aplicación” la cual gestionan. Para representarla, los distintos modelos citados anteriormente han usado aproximaciones diferentes: un punto, un *bounding box* o mediante una geometría más o menos detallada. En los dos primeros casos se produce una gran pérdida de información, y aunque existen técnicas para regenerar las extensiones espaciales de una colección de datos usando como base el centroide o el *bounding box* [18] solo producen resultados aproximados, que aunque pueden ser suficiente para ciertos usos, no bastan si se quiere identificar de forma precisa la unidad administrativa a la que pertenece cierta entidad espacial. Por ello, en nuestra ontología de unidades de España, incluimos la información de geometrías proporcionadas por el Instituto

Geográfico Nacional, organismo encargado en España de mantener las líneas límite entre jurisdicciones.

La experiencia nos ha mostrado que es muy frecuente que si algún servicio de una IDE necesita las unidades administrativas solo necesita un pequeño número de ellas y de sus relaciones, no siendo necesario que el sistema contenga toda la ontología. Por ejemplo en un servicio Web que proporcione el inventario del patrimonio industrial de Aragón, solo es necesario que tenga un subconjunto de la organización jerárquica de la comunidad (municipios y provincia a la que pertenecen), siendo irrelevante el resto del contenido de la ontología. En estos casos es interesante poder generar vistas, adaptadas a las necesidades de cada servicio, que solo contengan los elementos y relaciones que se necesiten. Dichas vistas, cuando sean usadas para los usos descritos en la siguiente sección, es conveniente que sean presentadas en forma de estructuras jerárquicas tales como tesauros con el objetivo de simplificar su uso. Como formato de almacenamiento e intercambio de las vistas se ha decidido usar SKOS [19], ya que se está estableciendo como estándar de facto en la Web semántica para la representación de modelos simples de conocimiento. Dichas vistas pueden ser proporcionadas a los distintos servicios de la IDE para tareas de creación de metadatos o búsqueda, tanto clásica (buscador con un selector de unidades) como exploratoria (creación de un mapa de tópicos de unidades ajustado a la colección de metadatos).

4 Aplicaciones de la Ontología de Unidades Administrativas

En [20] se detallan los usos que las ontologías léxicas pueden proporcionar a una IDE y que se pueden resumir en, facilitar la creación del contenido de los metadatos, facilitar las consultas a servicios y facilitar la armonización del contenido. La ontología de unidades administrativas puede tener estos mismos usos en ciertos contextos, pero la geometría asociada posibilita otros nuevos usos descritos aquí.

4.1. Facilitar la creación del contenido de los metadatos

En una IDE, los recursos de información almacenados están descritos mediante metadatos [21],[22] con el objetivo de mejorar los procesos de recuperación de la información; pero esta mejora depende principalmente de la calidad de su contenido. Una forma de mejorar el contenido es el uso para ciertos campos de un vocabulario previamente determinado, de forma que se aumente su homogeneidad y se simplifiquen los procesos de recuperación de información. Aquí, el uso de una ontología para modelar la terminología permite además razonar sobre ella.

En este contexto, las herramientas de creación de metadatos pueden usar la ontología creada para proporcionar a los usuarios toda o parte de la jerarquía de unidades administrativas de un país y sus vecinos. Además, pueden hacer uso de la información espacial asociada para generar de forma automática el subconjunto de unidades administrativas que intersectan o contienen la posición espacial extraída del dato asociado o del bounding box ya creado en el metadato. Por ejemplo, el metadato del catálogo de IDEZar que describe el campo eólico de “Los Labrados” ha sido elaborado por una herramienta que no utiliza esta ontología siendo asignado por error únicamente al municipio de Zaragoza a pesar de parte de él está en los municipios de Cuarte de Huerva y Cadrete.

4.2. Facilitar consultas a servicios

Dentro del proceso de consulta a los servicios de una IDE, el uso de restricciones de carácter espacial es habitual, tanto en catálogos de datos, como en nomenclátors, servidores de mapas o servidores de coberturas. Gracias a que en nuestra ontología almacenamos la geometría de las unidades, y dado que cubren todo el territorio de un país a distintos niveles, los servicios de búsqueda que la usen pueden proporcionar los nombres de las unidades para fijar restricciones espaciales a las consultas, y luego ejecutar las búsquedas en base a la geometría espacial asociada.

El servicio de acceso al catálogo de datos es un caso especial dado que los metadatos contienen palabras claves de lugar con los nombres de las unidades administrativas que contienen el dato referenciado. Por tanto, se puede hacer directamente una búsqueda temática (con restricciones de tipo “*contenido jurídicamente en*”), pero pudiendo aprovechar la existencia de la geometría de las unidades administrativas para los casos en que las palabras clave de lugar no estén introducidas en los metadatos y para permitir cierto grado de restricciones topológicas (“*adyacente a*” y “*contenido espacialmente en*”). Así pues, una consulta que contenga como restricción la unidad “Zaragoza” (Municipio) devolvería todos los metadatos que en sus palabras clave de lugar contienen el municipio de Zaragoza o cuyo *bounding box*, según el tipo de restricción indicada, esta contenido, intersecta o cubre el polígono de Zaragoza.

La existencia de una jerarquía detallada de tipos de unidades existentes y la asignación de estos tipos a las instancias, simplifica la localización de recursos en áreas fronterizas en las que los tipos de unidades administrativas existentes son diferentes. Por ejemplo si estamos buscando los “*Municipios*” con instalaciones de esquí existentes en una zona fronteriza de los pirineos entre España y Francia, la ontología permite descubrir que las “*Municipalités*” francesas son equivalentes a

los “*Municipios*” españoles, y así poder mostrar la “*Municipalité*” de Luz-Saint Sauveur que contiene la estación de esquí de Luz-Ardiden. También la existencia de tipos permite recuperar las unidades administrativas de un área geográfica que tengan una funcionalidad concreta, tales como distritos o secciones censales.

4.3. Facilitar la armonización del contenido

La ontología de unidades administrativas también facilita la armonización de los metadatos procedentes de otras organizaciones u otros dominios de aplicación con respecto de los metadatos ya existentes en el sistema y todos ellos respecto a las búsquedas realizadas por los usuarios.

En [20] se describe un proceso de armonización de ontologías léxicas basado en el uso de Wordnet [11], pero este proceso no es aplicable para las unidades administrativas, ya que Wordnet no es exhaustivo (solo incluye países, regiones y capitales de regiones). A pesar de esto, ciertas tareas de armonización pueden realizarse. La comparación de las etiquetas alternativas a los nombres oficiales y la geometría existente en la ontología, con el texto existente en las palabras clave de lugar y el *bounding box* del metadato ayuda a detectar errores en los metadatos y corregirlos. Además, la ontología también puede ser usada como base de conocimiento para la identificación de unidades administrativas en consultas realizadas por un usuario. Por ejemplo, si se pregunta por “Parques eólicos de Zaragoza”, se obtendrían, la provincia y el municipio en Aragón y la población de Lugo dejando al usuario la selección de la adecuada. Para facilitar esta tarea, se le puede proporcionar toda la información disponible en la ontología (nombres oficiales y alternativos, jerarquía organizativa y posición espacial).

Conclusiones

En este artículo se ha presentado una ontología de unidades administrativas para modelar las divisiones administrativas de un país y se han descrito las ventajas que aportaría la inclusión de dicha ontología dentro de una IDE. El modelo presentado está organizado en tres capas, una ontología de alto nivel que define tipos de datos y relaciones generales e independientes del contexto, una ontología de dominio donde se definen los conceptos y relaciones reutilizables en el contexto de los modelos administrativos de diferentes países y una ontología de aplicación donde se representan los tipos específicos de unidades administrativas de un país, junto con sus instancias. Respecto a los usos en una IDE, la existencia de una geometría asociada a las unidades permite facilitar la creación del contenido de los metadatos, las consultas a servicios y la armonización del contenido.

Los siguientes pasos de este trabajo serán desarrollar un proceso automático de generación de las instancias a partir de la información origen para los casos de España y Francia y su integración real dentro de una IDE para mostrar su potencial real de aplicación. Además, se analizará el problema del mantenimiento de la información referenciada, ya que la alta frecuencia de cambios existente en la organización administrativa de un país (forma, estructura, nombre o potestades administrativas) hace que un gran número de metadatos se queden desfasados respecto de la situación administrativa real.

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. Lacasta (ref. B139/2003) y J. López (ref. B136/2006) ha estado parcialmente financiado por una beca del Gobierno de Aragón.

Referencias

- [1] Commission of the European Communities (CEC), 2004. Proposal for a Directive of the European Parliament and of the Council establishing an infrastructure for spatial information in the Community (INSPIRE). COM(2004) 516 final, 2004/0175 (COD).
- [2] D. Manov, A. Kiryakov, B. Popov, D. Ognyanoff, M. Goranov. "Experiments with Geographic Knowledge for Information Extraction". Workshop of Analysis of Geographic References. "005. Edmonton, Canada.
- [3] R. Irie B. Sundheim. "Resources for Place Analysis". Fourth International Conference on Language Resources and Evaluation- LREC2004. 317-320. 2004. Lisbon, Portugal.
- [4] ISO 19109. "Geographic information -- Rules for application schema" International Organization for Standardization (ISO), 2005.
- [5] Registro de entidades locales del GDAL. <http://www.dgal.map.es/>
- [6] Relación de municipios del Instituto Nacional de Estadística (INE). <http://www.ine.es/>
- [7] Tesouro EUROVOC extendido por el senado para incluir los municipios de España <http://www.senado.es/basesdedatos/index.html>
- [8] M. S. Chaves, M. J. Silva, and B. Martins. A Geographic Knowledge Base for Semantic Web Applications. In Proc. of the 20th Brazilian Symposium on Databases, Minas Gerais, Brazil, pages 40–54, October, 3–7 2005.

- [9] ISO 19109. "Geographic information -- Spatial referencing by geographic identifiers" International Organization for Standardization (ISO), 2003.
- [10] Guarino, N. Formal Ontologies and Information Systems. Proceedings of FOIS'98, 3-15. 1998.Trento, Italy
- [11] P. Martin, "Using the WordNet Concept Catalog and a Relation Hierarchy for Knowledge Acquisition", Proceedings of Peirce'95, 4th International Workshop on Peirce, University of California, Santa Cruz, August 18, 1995.
- [12] C. Fellbaum (Ed). Wordnet, An Electronic Lexical Database. MIT Press. 1998.
- [13] J. Bateman. "Upper modeling: Organizing knowledge for natural language processing". In Proc. of Fifth International Workshop on Natural Language Generation, 1990. Pittsburgh, PA.
- [14] D. Lenat, R.Guha. "Building large knowledge-based systems: representation and inference in the Cyc project". Addison-Wesley, 1990.
- [15] J. Sowa. "Conceptual Graphs Summary. In Conceptual Structures: current research and practice", England , Ellis. Horwood Workshops, 1992.
- [16] ISO 15944-5 Information technology — Business Agreement Semantic Descriptive Techniques. Part 5: Identification and Mapping of Various Categories of Jurisdictional Domains. International Organization for Standardization (ISO), 2005.
- [17] Regulation (EC) No 1059/2003 of the European Parliament and of the Council of 26 May 2003 on the establishment of a common classification of territorial units for statistics (NUTS) (Official Journal L 154, 21/06/2003)
- [18] H. Alani, C. Jones, D. Tudhope, "Voronoi-based region approximation for geographical information retrieval with gazetteers ", International Journal of Geographical Information Science, 2005, vol. 15, no. 4, pag. 287 — 306.
- [19] A. Miles, B. Matthews, M. Wilson. SKOS Core: Simple Knowledge organization for the WEB. Proceedings of the International Conference on Dublin Core and Metadata Applications, 2005.
- [20] J. Lacasta, P. Muro, J. Noguera, J. Zarazaga, "Web Ontology Service, a key component of a Spatial Data Infrastructure", Proceedings of the 11th EC GI & GIS Workshop, ESDI Setting the Framework, 2005, 19-22
- [21] ISO 19115. Geographic information – Metadata. International Organization for Standardization (ISO), 19115:2003
- [22] ISO 15836. Information and documentation - The Dublin Core metadata element set. International Organization for Standardization (ISO). 15836:2003.