

Perfil de metadatos para direcciones desarrollado dentro de la *Best Practice* “EURADIN” y su implementación en la herramienta CatMDEdit

A. González¹, A. S. Maganto¹, M.A. Manso², C. Rodríguez², M.J. Bravo²,
J. Barrera³, F. Gracia³, J. Noguera Iso⁴, M. Cabello⁵

¹Instituto Geográfico Nacional
(agjimenez@fomento.es, asmaganto@fomento.es)

²Laboratorio de Tecnologías de la Información Geográfica (Latingeo).
Universidad Politécnica de Madrid (UPM)
(m.manso@upm.es, carlos.metadatos@gmail.com, mjbravo@gmail.com)

³GeoSpatiumLab
(jesusb@geoslab.com, fgcrespo@geoslab.es)

⁴Dpto. de Informática e Ingeniería de Sistemas, Universidad de Zaragoza
(jnog@unizar.es)

⁵TRACASA (Trabajos Catastrales S.A.)
(mcabello@tracasa.es)

Resumen

Este artículo describe el trabajo realizado para definir un perfil de metadatos, y las herramientas asociadas, que permitan documentar datos geográficos conteniendo información relativa a direcciones. Este trabajo se enmarca dentro del proyecto europeo EURADIN, cuyo objetivo es constituir una red de buenas prácticas para promover la armonización de direcciones europeas, garantizando la interoperabilidad, reutilización y explotación del contenido de los datos.

Palabras clave: EURADIN, INSPIRE, perfil de metadatos, CatMDEdit, ISO19115, direcciones, especificaciones de datos, GML.

1. Introducción

Las direcciones constituyen uno de los temas recogidos en el anexo I de la Directiva INSPIRE¹ y por tanto formarán parte de los conjuntos de datos que componen la Infraestructura de Datos Espaciales de la Comunidad Europea. Con la finalidad de constituir una red de buenas prácticas (*Best Practices*) para promover la armonización de direcciones europeas garantizando la interoperabilidad, reutilización y explotación del contenido de los datos surge, en el año 2008, el proyecto EURADIN² (EURadin ADdress INfrastructure), dentro de la convocatoria eContentplus de contenidos digitales para la Sociedad de la Información que realiza anualmente la Comisión Europea.

En este proyecto, liderado por el Gobierno de Navarra, participan 30 socios de 16 diferentes países europeos y tiene una duración de dos años (junio 2008-junio 2010). El plan de trabajo se organiza en 9 paquetes de trabajo (*Work Packages*) que parten del análisis del estado del arte de direcciones en Europa y finalizan con la puesta en práctica del modelo de datos y de metadatos definidos en el proyecto a través de una solución *gazetteer* interoperable con los servicios existentes(*figura 1*)



Figura 1: Relación de los paquetes de trabajo del proyecto EURADIN

¹ <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:ES:PDF>

² <https://www.euradin.eu/>

El paquete 4 (WP4) tiene como objetivo llevar a cabo la armonización de los metadatos para esta temática, definiendo para ello un perfil de metadatos (*EURADIN profile*) aplicado a la descripción de las direcciones conforme a las normas de ejecución de metadatos definidas por INSPIRE, apoyándose en los perfiles implementados en cada uno de los sistemas de direcciones implicados en el desarrollo de esta *Best Practice*, y cumpliendo con la norma Internacional ISO 19115:2003 de metadatos. Además debe desarrollarse dentro del marco establecido por la Especificación de Datos de INSPIRE correspondiente al grupo temático de Direcciones (DS-Addresses) [1], y basarse en las conclusiones y resultados obtenidos en otros paquetes de trabajo (principalmente el WP2: Valoración inicial y requisitos de usuarios y el WP3: Modelo de datos).

Otro objetivo específico del WP4 ha sido evaluar la viabilidad de este perfil con su puesta en práctica tomando como referencia modelos de datos de direcciones de diferentes países europeos. Para ello se ha desarrollado una herramienta *Open Source* que facilita la generación semi-automática de metadatos acorde a dicho perfil. Dicha herramienta es una adaptación de la herramienta CatMDEdit [2], llamada CatMDEdit-EURADIN, en la que se ha implementado el perfil EURADIN.

Este artículo pretende describir las tareas desarrolladas por el WP4 y detallar el perfil de metadatos adoptado que permite la descripción detallada de los datos de direcciones tanto a nivel de base de datos (dataset) como a nivel de entidad (feature type). A continuación se detalla su implementación en la herramienta CatMDEdit-EURADIN para la generación de metadatos según dicho perfil. Esta descripción se completa con un conjunto de recomendaciones que los sistemas de direcciones implicados deben seguir para la producción de metadatos según el perfil EURADIN a partir de los metadatos generados previamente según sus perfiles locales. Por último se finaliza el artículo con un apartado de conclusiones.

2. Proceso de generación del perfil de metadatos para EURADIN

Para llevar a cabo la definición del perfil EURADIN se estableció una metodología constituida por cuatro fases. La primera fase consistió en la realización de una encuesta entre los socios participantes en el proyecto para conocer el estado de desarrollo de cada una de estas organizaciones en materia de

metadatos; una vez recopilada la información, en una segunda fase se realizó una comparación entre los distintos perfiles de metadatos “locales” de direcciones facilitados; seguidamente, en la tercera fase se procedió a la elección de los elementos de metadatos a incluir en el perfil y finalmente, en la cuarta fase se generó la descripción detallada de cada uno de los elementos de metadatos seleccionados.

A continuación se describen cada una de estas fases.

2.1 Estado del arte de los metadatos en los sistemas de los socios participantes

Para conocer la situación inicial en materia de metadatos, se analizaron las conclusiones obtenidas en el paquete de trabajo 2 “Valoración inicial y requisitos de usuario” (WP2) [3], obtenidas en base a una encuesta sobre metadatos realizada entre los socios del proyecto.

Las conclusiones de este análisis fueron:

- De un total de 31 participantes, el 54,8% de los encuestados utilizan metadatos para la descripción de los conjuntos de datos de direcciones.
- La mayoría de los que generan metadatos (58,8%) lo hacen conforme a la norma ISO 19115; tan solo el 5,9% utiliza Dublin Core y el 11,8% se basa en ambas normativas.
- A la pregunta relativa al formato de distribución de los metadatos, el 58,8% usa como formato de intercambio XML.
- El 64,7% de los participantes que generan metadatos tienen un perfil de metadatos propio.
- El 52,9% utilizan una herramienta para la generación de sus metadatos.

Como resumen de las conclusiones se puede afirmar que más de la mitad de los socios implicados generan metadatos desde los últimos 4 años, utilizan como base la Normativa Internacional ISO 19115 y definen perfiles compuestos por un número mínimo de elementos a describir, aunque en ningún caso estos perfiles son específicos de direcciones; además, el formato de intercambio más común es el XML y suele utilizarse alguna herramienta que facilite su generación.

2.2 Comparativa entre perfiles/ejemplos de metadatos

El análisis de la encuesta realizada en el WP2, aunque permitió tener una visión general sobre el estado de los metadatos, fue insuficiente para analizar los perfiles

aplicados en cada sistema de direcciones. Por este motivo se realizó una segunda encuesta con los siguientes objetivos:

- Identificar qué información (elementos) deberían componer un perfil de metadatos para la descripción de direcciones.
- Conocer, de forma detallada, las normas utilizadas en los diferentes países y la información descrita en los metadatos creados.

A partir de toda la información recopilada, se inició un análisis comparativo basado en los perfiles proporcionados y en ejemplos de metadatos de aquellos que no disponían de un perfil. El análisis requirió un laborioso proceso de homogenización de los perfiles/ejemplos, debido a las siguientes diferencias:

- Algunos de los perfiles/ejemplos de metadatos proporcionados no se basaban en ISO 19115 (p.e. CSI-PIEMONTE) por lo que fue necesario analizar cada uno de los elementos descritos en los perfiles/ejemplos y localizar el correspondiente de la Normativa ISO 19115..
- El idioma utilizado en algunos casos era distinto del inglés o español (portugués checo, etc.) por lo que fue necesario traducirlos a un idioma común.
- En la mayoría de los casos el nombre de los elementos variaba en relación con el asignado por la Normativa Internacional ISO 19115, incluso aunque estuviesen basados en dicha especificación, por lo que fue necesario identificar cada elemento.

La identificación de los elementos de cada perfil/ejemplo con los ítems de ISO 19115, se realizó en una matriz comparativa en la que se relacionaron los elementos de cada uno de los perfiles con sus correspondientes en: el núcleo de la norma ISO 19115 y el resto de la norma en general, las Normas de Ejecución de metadatos de la directiva INSPIRE y el perfil de metadatos definido en la versión disponible (v 2.0) de la Especificación de Datos de INSPIRE correspondiente al grupo temático de Direcciones.

Finalmente, a partir del estudio de dicha matriz se concluyó la existencia de un total de 15 perfiles/ejemplos de metadatos (Tabla 1) y se detectaron los elementos de metadatos más utilizados.

2.3 Elección de los elementos de metadatos del perfil de EURADIN

El proceso de elección de los ítems a incluir en el perfil EURADIN se llevó a cabo considerando dos factores importantes. En primer lugar era obligatorio cumplir con las normativas internacionales y reglamentos de metadatos europeos existentes y por ello se incluyeron como elementos fijos los pertenecientes a:

- Normas de Ejecución de metadatos de la Directiva INSPIRE (R.I.INSPIRE) [4].
- Especificación de Datos de INSPIRE correspondiente al grupo temático de Direcciones
- Núcleo de la Normativa internacional ISO 19115 [5].

El segundo lugar se tuvo en cuenta la recurrencia de los elementos en la matriz comparativa. Por ello, además de los ítems pertenecientes a las normativas citadas, se seleccionaron aquellos elementos considerados en más del 45% de los perfiles analizados.

Una vez establecido el primer conjunto mínimo de elementos para constituir el perfil se realizó un proceso de consenso (*Tabla 2*) entre los miembros del grupo de trabajo para determinar la relevancia de los elementos descartados en esta primera selección y en su caso, su posible incorporación al perfil para la descripción de la información de direcciones.

2.4 Descripción de las particularidades de cada elemento

Otro aspecto que ha de tenerse en cuenta a la hora de definir un perfil, además de la selección de sus elementos, es la descripción de cada una de sus particularidades, es decir, su definición, origen, ruta del esquema ISO/TS 19139, obligatoriedad/condicionalidad, multiplicidad, tipo de dato, dominio de los valores, instrucciones de implementación, un ejemplo, etc.

En el caso del perfil EURADIN se determinó que, inicialmente, todas estas particularidades se heredarían de las características descritas en las Normas de Ejecución para metadatos de la Directiva INSPIRE y en la Norma ISO 19115, y se acordó la prevalencia de la primera en caso de discrepancia entre ambas pues se trata de un documento legal con rango de Directiva y por tanto más restrictivo en el

ámbito europeo. En muchas ocasiones se particularizaron las características de los elementos a fin de personalizarlos para el tema de direcciones.

Por último, con referencia al nivel de aplicabilidad del perfil y con el fin de alcanzar una solución capaz de aportar el grado de descripción de los datos definido en el modelo de datos de direcciones de la especificación de INSPIRE, fue necesario definir el nivel jerárquico de aplicación de cada elemento componente del perfil:

- **Conjunto de datos o *Dataset***: que aporta una descripción general del sistema de direcciones (es decir, las direcciones de una base de datos, etc.).
- **Fenómeno o *Feature Type***, que se corresponde con el archivo de metadatos referente a cada fenómeno o componente de la dirección, según la definición proporcionada en Especificación de Datos de Direcciones de la Directiva INSPIRE. Según dicha especificación una dirección postal se compone de los fenómenos: nombre de la unidad administrativa, nombre del área de la dirección (superficies inferior al término municipal necesaria para especificar una dirección inequívocamente (ej. entidad singular de población), nombre de la vía, y descriptor postal.

Este hecho es muy significativo ya que se ha definido un perfil con particularidades propias según el nivel jerárquico (conjunto de datos o fenómenos) al que estuviese asociado. Como consecuencia de esta apreciación se incluyó una nueva característica a describir dentro de los elementos de metadatos: la granularidad.

3. Perfil EURADIN: solución adoptada

Una vez definidos los ítems que constituyen el perfil de metadatos de EURADIN y las características propias de cada uno de ellos, se obtuvo como resultado un **perfil** compuesto por **44 elementos de metadatos** (*figura 1*). de tal forma que:

- Todos los ítems se corresponden con elementos de la ISO 19115.
- Incluye todos los elementos del Core de la ISO 19115.
- Están representados todos los elementos de las Normas de Ejecución de metadatos de la Directiva Europea INSPIRE (R.I.INSPIRE) [4].

- Están representados todos los elementos de la Especificación de Datos para Direcciones de la Directiva INSPIRE (D.S.Addresses) [1].
- Incluye 6 elementos adicionales necesarios para la descripción de datos relativos a direcciones.

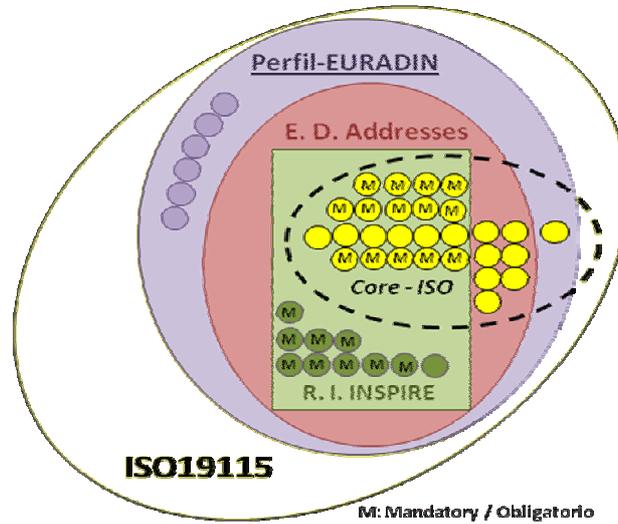


Figura 1; Gráfico esquemático de los ítems del perfil de EURADIN

El perfil de metadatos de EURADIN se encuentra documentado en sus “Guidelines” que incluye un Anexo con la descripción detallada de cada elemento.

| Nº | Nombre del socio | Información analizada |
|-----------|---|------------------------------|
| 1/2 | Gobierno de Navarra (GN)/ TRACASA | Perfil de metadatos |
| 3/4 | Instituto Geográfico Nacional de España (CNIG/IGN) / Institut Cartographic Valencia (ICV) | Perfil de metadatos |
| 6 | Danish Enterprise & Construction Authority | Encuesta y ejemplo metadatos |
| 8/24 | Regione PIEMONTE / CSI-PIEMONTE | Encuesta y ejemplo metadatos |
| 9 | Regione LOMBARDIA | Encuesta y ejemplo metadatos |
| 11 | Lantmäteriet (NLS) (Sweden) | Encuesta |
| 13 | 'VÚGTK, v.v.i. (Cze) | Perfil de metadatos |
| 14/28 | Instituto Cartografico Portugues (ICP) / FCT-UNL | Perfil de metadatos |
| 15 | Kadaster (Netherlands) | Perfil de metadatos |
| 17 | Intelligent Addressing Limited (UK) | Perfil de metadatos |
| 25 | State Enterprise Centre of Registers (Lithuania) | Perfil de metadatos |
| 26 | Institute of Geodesy, Cartography and Remote Sensing (FÖMI)(Hungary) | Perfil de metadatos |
| 27 | GeoX TérinformatikaiKf (Hungary) | Ejemplo de metadatos |
| 30 | Regione Toscana | Encuesta |
| -- | Flemish profile (Belgium) | Perfil de metadatos |

Tabla 1. Miembros participantes y tipo de información analizada

A continuación se muestran todos los elementos de metadatos que componen el perfil EURADIN, especificándose su grado de obligatoriedad (M: Mandatory, C: Conditional, O: Optional), su granularidad a través de colores (nivel de dataset, nivel de feature type o ambos niveles) así como el grupo al que pertenecen según la clasificación de las Normas de Ejecución de Metadatos de INSPIRE (B.1 – B.10).

| | Name | Definition |
|--------------------|--------------------------------|---|
| B.1 Identification | Resource title (M) | Name by which the cited resource is known. |
| | Alternative Title (O) | Acronym or name in another language by which the cited information is known. |
| | Resource abstract (M) | Brief narrative summary of the content of the resource. |
| | Purpose (O) | Summary of the intentions with which the resource(s) was developed |
| | Resource Type (M) | Scope to which metadata applies |
| | Hierarchy LevelName (C) | Name of the hierarchy levels for which the metadata is provided |
| | Resource Locator (C) | Location (address) for on-line access using a Uniform Resource Locator address or imilar addressing scheme. |
| | Unique Resource Identifier (M) | A value uniquely identifying an object within a namespace |
| | ~ code (M) | Alphanumeric value identifying an instance in the namespace |

| | Name | Definition |
|---------------------------|-----------------------------------|--|
| B.1. Identification | Resource Language (C) | The language(s) used within the datasets |
| | Dataset characterSet (C) | Full name of the character coding standard used for the dataset. |
| | Spatial representation type (O) | Method used to spatially represent geographic information |
| | referenceSystemIdentifier (M) | Description of the spatial and temporal reference systems used in the dataset. |
| | ~ authority (O) | person or party responsible for maintenance of the namespace |
| | + authority > title (M) | Name by which the cited resource is known. |
| | + authority > Alternate title (O) | short name or other language name by which the cited information is known. |
| B.2 Classif. partial data | ~ code (M) | alphanumeric value identifying an instance in the namespace |
| | Topic Category (C) | Main theme(s) of the dataset |

| | Name | Definition |
|-------------------------|---|--|
| B.3 Keyword | ~ keyword value (M) | Commonly used word(s) or formalised word(s) or phrase(s) used to describe the subject |
| | ~ Originating controlled vocabulary (C) | Name of the formally registered thesaurus or a similar authoritative source of keywords. |
| | + Originating controlled vocabulary > title (M) | Name by which the cited resource is known. |
| | + Originating controlled vocabulary > date > (M) | Reference date for the cited resource |
| | + Originating controlled vocabulary > date > date (M) | Reference date for the cited resource |
| | + Originating controlled vocabulary > date > dateType (M) | Event used for reference date |
| B.4 Geographic Location | Geographic Bounding Box (M) | |
| | ~ westBound Longitude (M) | Western-most coordinate of the limit of the dataset extent, expressed in longitude in decimal degrees (positive east). |
| | ~ eastBound Longitude (M) | Eastern-most coordinate of the limit of the dataset extent, expressed in longitude in decimal degrees (positive east). |

| | | | | | | | | |
|--|-----------------------|--|-------------|--------------------|--|--|------------------------------|---|
| | ~ codeSpace (O) | Name or identifier of the person or organisation responsible for namespace | B.3 Keyword | Keyword (M) | | | ~ southBound Latitude (M) | Southern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degrees (positive north). |
|--|-----------------------|--|-------------|--------------------|--|--|------------------------------|---|

| | Name | Definition | | Name | Definition | | Name | Definition |
|--------------------------|--|---|--|--|---|---------------------------------------|---------------------------------|--|
| B.4 Geographic Location | ~ northBound Latitude (M) | Northern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degrees (positive north). | B.7 Conformity | + value (M) | quantitative value or values, content determined by the evaluation procedure used. | B.7 Conformity | + value (M) | Quantitative value or values, content determined by the evaluation procedure used. |
| | Temporal Reference (M) | | | ~ DQ_Completeness Omission (O) | Data absent from the dataset, as described by the scope | | ~ DQ_Conceptual Consistency (O) | adherence to rules of the conceptual schema |
| B.5 Temporal Reference | ~ Temporal Extent (O) | Time period covered by the content of the dataset. | | + value unit (M) | value unit for reporting a data quality result | | + value unit (M) | value unit for reporting a data quality result |
| | ~ Date of publication (C) | Reference date for the cited resource – publication | | + value (M) | Quantitative value or values, content determined by the evaluation procedure used. | | + value (M) | Quantitative value or values, content determined by the evaluation procedure used. |
| | ~ Date of last revision (C) | Reference date for the cited resource – revision | | ~ DQ_Absolute ExternalPositional Accuracy (O) | Closeness of reported coordinate values to values accepted as or being true | | ~ DQ_Domain Consistency (O) | adherence of values to the value domains |
| | ~ Date of creation (C) | Reference date for the cited resource – creation | | + value unit (M) | value unit for reporting a data quality result | | + value unit (M) | value unit for reporting a data quality result |
| | Lineage (M) | General explanation of the data producer's knowledge about the lineage of a dataset | | + value (M) | Quantitative value or values, content determined by the evaluation procedure used. | | + value (M) | Quantitative value or values, content determined by the evaluation procedure used. |
| B.6 Quality and validity | Spatial Resolution (C) ~ equivalent scale | Level of detail expressed as the scale denominator of a comparable hardcopy map or chart | | ~ DQ_Thematic Classification Correctness (O) | Comparison of the classes assigned to features or their attributes to a universe of discourse | | ~ DQ_Temporal Consistency (O) | accuracy of the temporal attributes and temporal relationships of features |
| | ~ Distance (C) | Ground sample distance | | + value unit (M) | value unit for reporting a data quality result | | + value unit (M) | value unit for reporting a data quality result |
| | Quality measurements | | | + value (M) | Quantitative value or values, content determined by the evaluation procedure used. | | + value (M) | Quantitative value or values, content determined by the evaluation procedure used. |
| B.7 Conformity | ~ DQ_Completeness Commission (O) | Excess data present in the dataset, as described by the scope | ~ DQ_Non Quantitative Attribute Accuracy (O) | Accuracy of non-quantitative attributes | Conformity (C) | | | |
| | + value unit (M) | value unit for reporting a data quality result | + value unit (M) | value unit for reporting a data quality result | ~ Degree (M) | Indication of the conformance result. | | |

| | | Name | Definition | | | Name | Definition | |
|--|---|------|---|--|---|---|---|---|
| B.7 Conformity | ~ Specification (M) | | Citation of product specification or user requirement against which data is being evaluated. | B.9 Organisations responsible for the establishment, management, maintenance and distribution of spatial data sets | Responsible Organisation (M) | | ~ contactInfo >address >electronicMailAddresses (M) | Address of the electronic mailbox of the responsible organisation or individual |
| | + Specification > title (M) | | Name by which the cited resource is known. | | Responsible party (M) | | maintenanceAndUpdateFrequency (M) | Frequency with which changes and additions are made to the resource after the initial resource is completed |
| | + Specification > date (M) | | Reference date for the cited resource | | ~ individualName (O) | | Distribution Format (O) | Provides a description of the format of the data to be distributed. |
| | + Specification > date > date (M) | | Reference date for the cited resource | | ~ organisationName (M) | | ~ Name (M) | Name of the data transfer format(s) |
| | + Specification > date > dateType (M) | | Event used for reference date | | ~ positionName (O) | | ~ Version (M) | Version of the format (date, number, etc.) |
| B.8 Constraint related to access and use | Condition applying to access and use (M) | | | | ~ contactInfo >address >electronicMailAddress (M) | | Metadata file identifier (O) | Unique identifier for this metadata file |
| | ~ use limitation (M) | | Restrictions on the access and use of a resource or metadata | | Responsible party role (M) | | parent identifier (C) | Metadata file identifier of the which this metadata are a subset (child) |
| | Limitations on public access (M) | | | | DistirbutorContact (M) | | Metadata point of contact (M) | |
| | ~ accessConstraints (O) | | Access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the resource. | | ~ individualName (O) | | ~ organisationName (M) | Name of the responsible organisation |
| | ~ otherConstraints (C) | | Other restrictions and legal prerequisites for accessing and using the resource or metadata. | | ~ organisationName (M) | | ~ electronicMailAddress (M) | Address of the electronic mailbox of the responsible organisation or individual |
| B.9 Organisations responsible for the establishment, management, maintenance and distribution of spatial data sets | Responsible Organisation (M) | | | | B.10 Metadata on Metadata | ~ contactInfo >address >electronicMailAddress (M) | | |
| | Responsible party (M) | | Identification of, and mean of communication with, person(s) and organisation(s) associated with the resource(s) | parent identifier (C) | | | | |
| | ~ individualName (O) | | Name of the responsible person- surname given name, title separated by a delimiter | Metadata point of contact (M) | | | | |
| | ~ organisationName (M) | | Name of the responsible organisation | ~ organisationName (M) | | | | |
| | ~ positionName (O) | | Role or position of the responsible person | ~ electronicMailAddress (M) | | | | |
| | ~ contactInfo >address >electronicMailAddress (M) | | Address of the electronic mailbox of the responsible organisation or individual | | | | | |
| | Responsible party role (M) | | Function performed by the responsible party | | | | | |
| | DistirbutorContact (M) | | Party from whom the resource may be obtained. This list need not be exhaustive | | | | | |
| | ~ individualName (O) | | Name of the responsible person- surname given name, title separated by a delimiter | | | | | |
| | ~ organisationName (M) | | Name of the responsible organisation | | | | | |
| | ~ contactInfo >address >electronicMailAddress (M) | | Address of the electronic mailbox of the responsible organisation or individual | | | | | |

| | ~ classification (M) | Name of the handling restrictions on the resource. | | ~ positionName (O) | Role or position of the responsible person | | ~ role (codelist) (M) | function performed by the responsible party |
|---------------------------|----------------------|--|---------------------------|----------------------------|---|---------------------------|-------------------------------|---|
| | Name | | Name | | Name | | Name | |
| | Definition | | Definition | | Definition | | Definition | |
| B.10 Metadata on Metadata | ~ individualName (O) | name of the responsible person- surname given name, title separated by a delimiter | B.10 Metadata on Metadata | Metadata Language (M) | Language used for documenting metadata | B.10 Metadata on Metadata | Metadata standard name (O) | Name of the metadata standard (including profile name) used |
| | ~ positionName (O) | name of the responsible organisation | | Metadata character set (C) | Full name of the character coding standard used for the metadata set. | | Metadata standard version (O) | Version (profile) of the metadata standard used |
| | Metadata Date (M) | Data that the metadata was created | | | | | | |

 Nivel Dataset
  Nivel Feature type
  Nivel Dataset & feature

Tabla 2. Elementos de metadatos del perfil EURADIN

4. Personalización de la herramienta CatMDEdit para la generación de metadatos según el perfil EURADIN

4.1 La herramienta CatMDEdit

CatMDEdit³ es una aplicación de escritorio multiplataforma, distribuida bajo licencia de Software Libre (*GNU Lesser General Public License*) e internacionalizada a seis idiomas diferentes, cuyo objetivo consiste en promover la creación de metadatos como un mecanismo para facilitar la recuperación y procesamiento posterior de los datos de una manera efectiva. Por ello, la herramienta proporciona diversos mecanismos para la creación semi-automática de metadatos con el fin de minimizar la interacción del usuario y ocultar la

³ <http://catmdedit.sourceforge.net>

complejidad inherente de los propios estándares de metadatos. Otras funcionalidades relevantes que ofrece esta aplicación y que merece la pena destacar son las siguientes:

Definición y gestión de repositorios de metadatos, facilitando la selección y filtrado de los registros de metadatos de acuerdo a distintos criterios.

- Edición de metadatos de acuerdo a las normas internacionales ISO 19115 (*Geographic Information – Metadata*) y Dublin Core (ISO 15836). En el caso de ISO 19115 se proporcionan además cuatro interfaces de edición especializadas para esta norma de acuerdo a distintos perfiles: modelo general de ISO19115, "Núcleo Español de Metadatos" (NEM), las normas de ejecución de metadatos establecidas por INSPIRE para su implementación sobre ISO 19115/19119, y una propuesta de perfil de metadatos de recursos relacionados con la Directiva Marco del Agua (esta propuesta fue establecida en el contexto del proyecto SDIGER).
- Posibilidad de modificar la herramienta para soportar nuevos estándares y perfiles de metadatos que se adapten a las necesidades del usuario.
- Intercambio de registros de metadatos de acuerdo a distintos estándares (ISO19115, Dublin Core y CSDGM) y formatos (diferentes codificaciones sobre XML y RDF).
- Diferentes estilos de presentación de registros de metadatos en HTML y Excel.
- Herramientas adicionales para facilitar la edición de metadatos: validación de metadatos, agenda de contactos, gestión de tesauros (y vocabularios controlados), integración con herramientas SIG para la definición de la extensión geográfica y la visualización de los datos.

La presente sección analiza el papel de la herramienta CatMDEdit como medio para facilitar la creación de metadatos que documenten los datos de direcciones creados conforme al modelo definido en el proyecto EURADIN. Para ello, la aplicación ha sido adaptada para permitir la edición de metadatos conforme a los perfiles específicos de la norma Internacional ISO 19115:2003 definidos en el proyecto, así como para facilitar la creación automática de metadatos de datos de direcciones a partir de ficheros GML [6] que sigan el modelo de datos para direcciones propuesto por INSPIRE.

4.2 Edición de metadatos de acuerdo al perfil de metadatos EURADIN

El perfil de metadatos EURADIN [7] distingue elementos de metadatos diferentes para los conjuntos de datos (*dataset*) y para los tipos de fenómenos (*feature type*), de tal manera que cada tipo de recurso tiene su propio conjunto de elementos de metadatos y reglas para completarlos. Por este motivo, y con el fin de facilitar al usuario la creación de metadatos para los diferentes recursos considerados en el ámbito del proyecto, se han desarrollado en CatMDEdit dos interfaces diferentes de metadatos: uno para los conjuntos de datos (*dataset*) y otro para los tipos de fenómenos (*feature type*).

En la Figura 3 se muestran las dos interfaces de edición incluidas en la herramienta para el proyecto EURADIN. Además, CatMDEdit proporciona una interfaz adicional de acuerdo al modelo completo de metadatos ISO 19115:2003 [5] en el que se pueden editar todos los elementos de la norma, así como una vista HTML que muestra un resumen de los elementos de metadatos completados.

Por último, hay que destacar que las interfaces de edición de metadatos añadidas a CatMDEdit para el proyecto EURADIN proporcionan al usuario todas las utilidades y ayudas presentes en la versión *Open Source* de la herramienta, proporcionando de esta forma información sobre los diferentes elementos de metadatos que constituyen cada perfil junto con pautas concretas para su edición. Así, por ejemplo, se muestra al usuario la definición de cada elemento conforme al perfil de metadatos EURADIN, las condiciones de obligación del elemento de acuerdo a dicho perfil, el número del elemento de metadatos en el diccionario de datos de ISO 19115:2003 incluyendo su ruta completa, un ejemplo del elemento de acuerdo al perfil de metadatos EURADIN, etc.

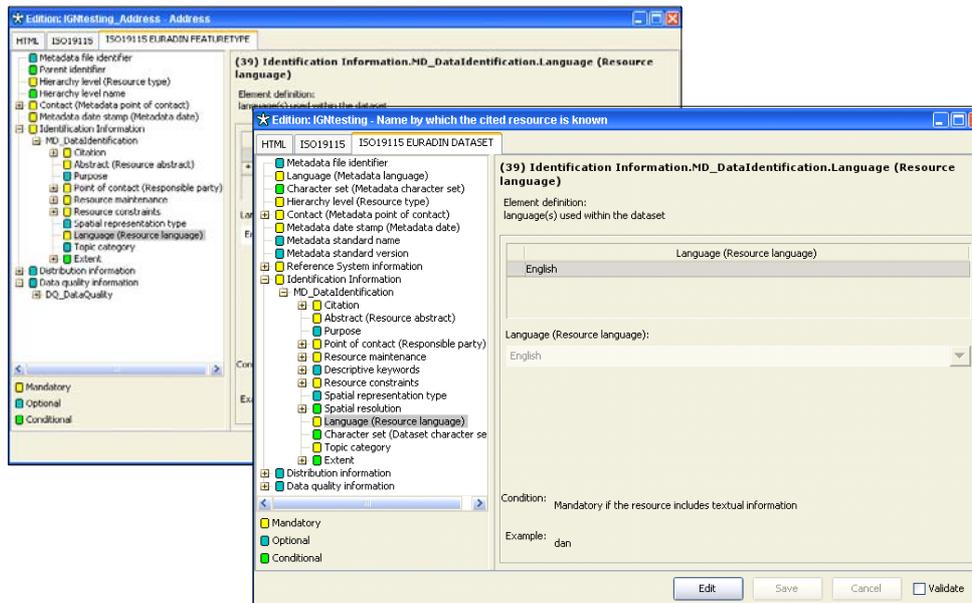


Figura 3: Interfaces de edición de metadatos

4.3 Generación automática de metadatos para datos de direcciones

Dentro del modelo de datos para direcciones propuesto en INSPIRE se han definido las direcciones como un conjunto suficientemente flexible de componentes (unidad administrativa, nombre de calle, descriptor postal, etc.). El propósito de este modelo de datos es permitir que distintas comunidades de usuarios que hayan acordado la utilización del mismo puedan intercambiar datos fácilmente. Para ello, el formato de intercambio de datos elegido es GML (*Geography Markup Language*), un lenguaje basado en XML definido por OGC (*Open Geospatial Consortium*) para expresar fenómenos geográficos.

Con el objetivo de facilitar la creación de metadatos para datos de direcciones del proyecto EURADIN, se ha llevado a cabo la integración en la herramienta CatMDEdit de un mecanismo que automatiza la generación de metadatos de ficheros GML conforme a los esquemas de aplicación de XML publicados por el *Joint Research Centre* [8] de acuerdo a los modelos de datos definidos por INSPIRE.

Como en el perfil de metadatos definido en el proyecto EURADIN se consideran metadatos tanto a nivel de *dataset* como a nivel de *feature types*, CatMDEdit considera también esta dualidad creando: un metadato para el conjunto de datos (*dataset*), y varios metadatos adicionales para describir de forma separada cada conjunto de instancias de los tipos de fenómenos (*feature type*) encontrados.

Los elementos de metadatos generados se han clasificado en cuatro grandes grupos de acuerdo al método empleado para la extracción de su valor. Estos cuatro tipos de generación son los siguientes:

- Método 1: el valor de un elemento se extrae directamente de un fichero GML de forma que el valor usado en el metadato será el mismo valor que el de la propiedad en dicho GML. Éste es el caso de elementos como el título, la descripción, o la codificación de caracteres, donde el propio GML presenta campos que recogen dichos valores.
- Método 2: el valor de un elemento se calcula a partir del valor de una propiedad en el GML o los valores de un conjunto de propiedades en el GML. Tal es el caso de la extensión geográfica o temporal (calculadas ambas como combinación de los valores presentes en cada una de las entidades de direcciones presentes en el GML), o los informes de calidad (calculados en base a una serie de evaluaciones realizadas sobre el propio GML).
- Método 3: el valor de un elemento se corresponde con un término de un tesoro o lista controlada, como ocurre en el caso de las palabras clave, o el tipo de recurso.
- Método 4: algunos elementos de metadatos tienen un valor fijo o se completan con un valor por defecto si no es posible extraer dicho valor aplicando los métodos anteriores. De esta forma, hay algunos campos, como el punto de contacto o el linaje, que, según el perfil de metadatos, es obligatorio completar pero que difícilmente pueden extraerse del fichero GML porque éste no cuenta con ningún campo para recoger tales valores. Por ello, se les asigna un valor por defecto que luego el usuario podrá matizar en caso necesario.

El proceso seguido para generar los archivos de metadatos se puede observar en la Figura 4 y consta de los siguientes pasos:

1. En primer lugar, se analiza sintácticamente el fichero GML y se genera un metadato parcial para el conjunto de datos de acuerdo a los tres primeros métodos anteriormente expuestos.

2. Este metadato parcial se completa con los valores fijos y por defecto para rellenar todos los elementos obligatorios definidos por el perfil de metadatos de INSPIRE para conjuntos de datos. En este momento se consigue un metadato válido.
3. Además existe la posibilidad de permitir al usuario definir plantillas con valores que se añadirán al metadato general para que sea más detallado.
4. El último paso en este proceso es almacenar el fichero de metadatos en el repositorio.

En paralelo a estas operaciones se genera un metadato por cada tipo de fenómeno encontrado en el GML, es decir, se genera un metadato para describir conjuntamente las instancias de cada tipo de fenómeno y se almacena en el mismo repositorio que el metadato del conjunto de datos.

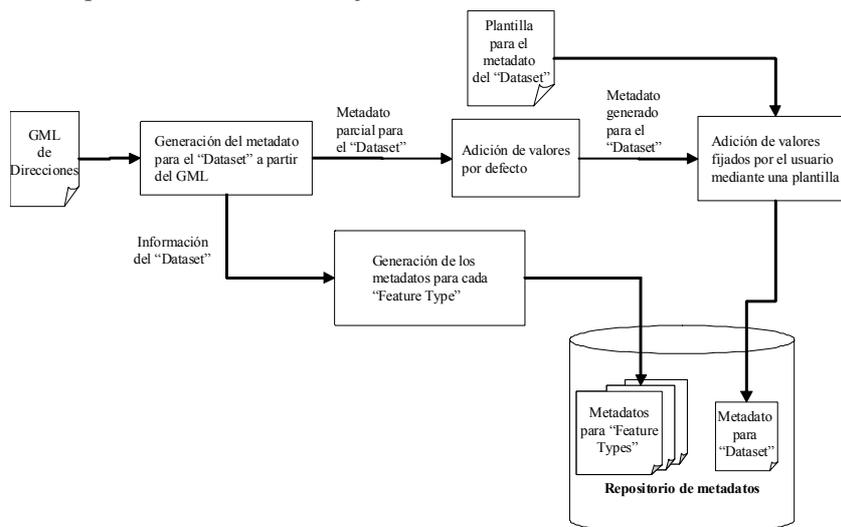


Figura 4 Proceso de generación de un metadato a partir de un fichero GML

4.4 Adaptación de metadatos creados con otras herramientas

A partir de la encuestas llevadas a cabo para conocer el “estado del arte” de los metadatos se comprobó que algunos de los socios ya habían creado metadatos o estaban desarrollando sus perfiles, bien para datos de direcciones bien para otros

tipos de datos, antes de la puesta en marcha del proyecto. Estos ficheros de metadatos no siguen el perfil de EURADIN pero contienen información que puede ser interesante en el contexto del proyecto. Por ello, como parte del proyecto EURADIN, se realizó un estudio de cada uno de dichos perfiles de metadatos, analizando su compatibilidad con la versión de CatMDEdit desarrollada para EURADIN y proporcionando una serie de recomendaciones (incluidas dentro de las “Guidelines” de la herramienta [9]) a tener en cuenta si se quieren importar en dicha herramienta metadatos que no se han creado previamente con la misma.

Las conclusiones generales que se obtuvieron del estudio realizado son las siguientes:

- Algunos participantes ya habían creado sus ficheros de metadatos con versiones anteriores de la herramienta CatMDEdit por lo que la importación de los mismos en la nueva versión de la aplicación es inmediata gracias a la compatibilidad hacia atrás ofrecida por ésta.
- Ciertos perfiles de metadatos no siguen la norma ISO 19115:2003, por lo que los ficheros creados conforme a dichos perfiles no podrán ser importados en CatMDEdit-EURADIN.
- Existen participantes que han definido listas controladas propias, para algunos elementos, diferentes de las especificadas en ISO 19115:2003. Dichas listas controladas no son soportadas por CatMDEdit por lo que será necesario convertir los valores a los recomendados por ISO 19115:2003 antes de importar los ficheros de metadatos en CatMDEdit-EURADIN.
- Ciertos perfiles de metadatos se han implementado mediante un esquema XML propio, diferente del recomendado por la norma ISO 19139:2007 [10]. Estos esquemas no son soportados por CatMDEdit, por lo que será necesario adaptar los ficheros de metadatos creados conforme a los mismos al esquema general de la norma.
- En el caso de los metadatos de series, algunos participantes han optado por documentar en un mismo fichero toda la información relativa a la serie y a todas las unidades que la componen. Para permitir la correcta importación de estos metadatos en CatMDEdit, se deberá crear un fichero de metadatos independiente para la serie y otro por cada una de las unidades que pertenecen a ella.

5. Conclusiones

Dentro del ámbito europeo de la información geográfica se ha definido un perfil de metadatos “EURADIN profile”, como parte de la Best Practice EURADIN para documentar datos geográficos que contienen información relativa a direcciones.

Para su definición se ha llevado a cabo un trabajo estructurado en varias fases de las cuales se han extraído una serie de conclusiones, entre las que cabe destacar:

- Este perfil es el resultado de un conjunto de actividades de consenso llevadas a cabo por los miembros del proyecto EURADIN, en especial por aquellos participantes más avanzados en temas de metadatos.
- Se ha desarrollado a partir del análisis exhaustivo de los perfiles/ejemplos de metadatos empleados y facilitados por los sistemas de direcciones implicados en el proyecto y, bajo la premisa de ser desarrollado dentro del marco normativo definido por las especificaciones más relevantes en el ámbito de los metadatos (ISO 19115:2003, Normas de Ejecución de metadatos de INSPIRE) y de basarse en la especificación de datos de INSPIRE sobre direcciones.
- La premisa de cumplir el modelo de datos de direcciones ha condicionado la definición de un perfil aplicable a dos niveles de granularidad (dataset, feature type), siendo esta característica algo muy particular en la temática de los metadatos.

En cuanto a la herramienta de metadatos, aparte de la edición de metadatos de acuerdo al perfil definido, cabe destacar el esfuerzo realizado para automatizar lo máximo posible la generación automática de metadatos a partir de los ficheros de datos de direcciones en formato GML. Este mecanismo de autogeneración permite facilitar el trabajo del usuario eliminando parte de la complejidad inherente a la definición de un perfil de metadatos derivado de la norma ISO 19115 (una norma que define más de 400 elementos estructurados de forma jerárquica). A partir de los metadatos generados, el usuario solo necesita revisar y modificar aquellos elementos (Ej.: *palabras clave, resumen*) que tengan un carácter más subjetivo. Asimismo, este mecanismo de autogeneración se podría extender para otras especificaciones de datos donde también se establezca un formato de intercambio de datos estandarizado sobre la base de GML.

El perfil de EURADIN es un buen ejemplo de que es posible alcanzar la interoperabilidad entre los metadatos para las direcciones de los países europeos y es un primer paso en la puesta en práctica de la Directiva INSPIRE en Europa.

En cuanto a las futuras líneas de trabajo se espera revisar en los próximos meses el perfil creado para adaptarlo a la versión final de la Especificación de Datos de INSPIRE sobre Direcciones cuya su publicación ha sido posterior (septiembre 2009) a los trabajos desarrollados para la creación de este perfil. Así mismo, dicha revisión implicará también la actualización de la herramienta CatMDEdit-EURADIN para adaptarla a los cambios que pueda sufrir la nueva versión del perfil.

Referencias

[1] European Commission, Joint Research Centre, INSPIRE Thematic Working Group Addresses. “D2.8.I.5 Data Specification on Addresses –Guidelines, v3.0”

http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_AD_v3.0.pdf, Septiembre 2009

[2] Nogueras-Iso J, Barrera J, Gracia-Crespo F, Laiglesia S, Muro-Medrano PR (2008). “Integrating catalog and GIS tools: access to resources from CatMDEdit thanks to gvSIG”. In *4as Jornadas Internacionales gvSIG, 3-5 December 2008*. Valencia, Spain

[3] Final WP 2 report on INITIAL ASSESSMENT, USER REQUIREMENTS, and BEST PRACTICES - D2%201_WP2FinalReport. (<https://www.euradin.eu/Lists/Results%20List/DispForm.aspx?ID=1&Source=https%3A%2F%2Fwww%2Eeuradin%2Eeu%2FPages%2FResults%2Easpx>). 1, Enero, 2009.

[4] DT Metadata. Draft Implementing Rules for Metadata (D1.3). (http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/draftINSPIREMetadata1Rv2_20070202.pdf). 02, Febrero, 2007.

[5] International Organization for Standardization (ISO). ISO 19115:2003, Geographic information – Metadata. (http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=26020).

[6] Cox, S., Daisey, P., Lake, R., Portele, C., Whiteside, A., 2003. OpenGIS Geography Markup Language (GML) Implementation Specification, Version 3.0 : OpenGIS Project Document OGC 02-023r4. Open Geospatial Consortium Inc.

[7] EURADIN Metadata Profile: Technical Guidelines. EURADIN Consortium. <https://www.euradin.eu/Pages/Results.aspx>, Mayo 2009

[8] European Commission (EC), Joint Research Centre. “GML Application Schemas. INSPIRE Data Specifications”. http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/DataSpecifications/D2.8-I_GML-Application-Schemas_v2.0.zip, 2008.

[9] Metadata tool: CatMDEdit-EURADIN. Recommendations to facilitate the transformation into EURADIN Profile. EURADIN Consortium. <https://www.euradin.eu/Pages/Results.aspx>, Mayo 2009

[10] International Organization for Standardization (ISO). ISO 19139:2007 Geographic information - Metadata - XML schema implementation.