

# GeoCollector

Describiendo colecciones de recursos para la Web de Datos

**Abargues, Carlos; Granell, Carlos; Huerta, Joaquín**

## Resumen

El descubrimiento de información en una IDE se ha convertido en una tarea crítica ante la cantidad de datos generados y que pasan en la mayoría de los casos por la elaboración de metadatos que describan dicha información. Habitualmente la descripción de estos recursos no es óptima llegando a duplicar los metadatos de recursos similares y aislándolos entre ellos así como entre otros recursos relacionados.

La Web de Datos se visiona como un nuevo paradigma para la Web en la que en lugar de conectar documentos se conecte información o datos siguiendo los principios establecidos por el Linked Data, que definen un estilo de publicación de datos enlazados en la Web. Siguiendo esta filosofía, el uso de colecciones de recursos u objetos compuestos podría optimizar la descripción de recursos similares y su interconexión con otros facilitando el descubrimiento de estos.

En este trabajo se presenta *GeoCollector*, una herramienta multiplataforma que permite la creación de colecciones de recursos heterogéneos de forma sencilla para el usuario con una clara componente geográfica. *GeoCollector* se basa en el modelo de datos abstracto definido por la especificación Object Reuse & Exchange del Open Archive Initiative (OAI-ORE), un estándar para la agregación de recursos Web, con el objetivo de estructurar y relacionar los distintos recursos agregados, y optimizar la descripción de éstos a nivel de colección. De esta forma, el uso de metadatos a distintos niveles (colección y recurso) permite la adición de nuevas propiedades y relaciones con otros recursos y su geolocalización mediante el uso del conjunto de geometrías simples definidos por *GeoRSS*. Esta herramienta no sólo permite definir la estructura de la relación sino que ofrece un cliente de búsqueda basado en *OpenSearch-Geo* gracias al cual es posible encontrar nuevos recursos. Por último, la aplicación permite la serialización de las colecciones creadas en distintos formatos como por ejemplo *RDF/XML*, *RDFa*, *Atom*, *N3*, *Turtle* y *KML*.

La agrupación de recursos similares o relacionados mediante colecciones ofrece varias ventajas. En principio, el uso de colecciones evita la duplicación de los metadatos de los recursos agregados, facilita el descubrimiento de recursos similares que conforman la agregación y, permite su enlace con otros datos y recursos ya existentes. Al usar un modelo basado en tecnologías como la arquitectura Web o los principios del Linked Data asegura la interoperabilidad de las colecciones con otros sistemas y su inserción de forma transparente en la futura Web de Datos. Este mismo aspecto es potenciado gracias a las distintas serializaciones que se ofrecen y que permiten su uso por otros sistemas como los motores de búsqueda.

## PALABRAS CLAVE

Colecciones, Web de Datos, Linked Data, Descubrimiento, Metadatos

## 1. INTRODUCCIÓN

En determinados escenarios el disponer de datos no sólo bien documentados sino visibles y que faciliten la búsqueda de otros recursos interconectados resulta clave para la realización de distintas tareas. En el campo SIG existen distintas estrategias para la publicación de datos geográficos pero sin embargo y en parte debido a las necesidades en cuanto a interoperabilidad, especialmente en el entorno de las IDEs, estos métodos de publicación se reducen en muchos casos a la publicación de datos y metadatos en sistemas que implementen un servicio de catálogo OGC CSW [1] como *GeoNetwork*<sup>1</sup> o similares. Para la correcta publicación de los datos en estos sistemas de catálogo el usuario necesita introducir un conjunto de metadatos gracias a los cuales otros usuarios podrán encontrar los datos que describen. El proceso de descripción de los recursos puede resultar una tarea ardua al igual que la actualización de sus metadatos y que en muchos casos representa una tarea incluso duplicada al describir recursos relacionados o muy similares que comparten parcialmente la información que los describe. Esto representa un claro problema en cuanto a efectividad que puede ralentizar el proceso de publicación y actualización de datos geográficos. Emplear métodos para la publicación y descripción de recursos que eviten el duplicar información mediante la reutilización de información ya publicada aligeraría este proceso derivando en una tarea más sencilla, rápida y eficaz.

Otro aspecto importante de las actuales herramientas para la publicación de datos geoespaciales es la homogeneidad de los datos que permiten publicar. En la actualidad, para realizar estudios SIG no siempre son suficientes datos de naturaleza geoespacial sino que otros tipos de recursos (multimedia, documentos, etc.)

---

<sup>1</sup> <http://geonetwork-opensource.org>

deberían poder ser distribuidos con éstos para que el profesional pueda llevar a cabo su labor. Este concepto de elemento compuesto o colección de recursos puede no ser explotado convenientemente como sí lo está siendo en otros campos como las librerías digitales [2, 3]. Así pues, puede resultar beneficioso indagar en métodos para la descripción y distribución conjunta de recursos heterogéneos integrando en el dominio geoespacial aquellos recursos que hasta la fecha no habían sido incorporados y explotados eficazmente en este ámbito.

La Web Semántica [4] el Linked Data [5] y movimientos derivados como el Linked Open Data<sup>2</sup> (LOD) están promoviendo la publicación de información o datos como tripletas basada en el formato RDF <cita> para constituir lo que se conoce como la Web de Datos. Esta nueva capa que se solapa con la tradicional Web documental alberga a día de hoy millones de tripletas proyectándose como una inmensa base de conocimiento abierto, distribuido e interconectado. Un punto fundamental en esta nueva capa es que mucha de esta información está enlazada entre ella de manera que este conocimiento es representado como una estructura reutilizable y navegable mediante estos enlaces de la que agentes humanos y automáticos pueden extraer información y navegar para encontrar otra relacionada. Toda esta información, disponible públicamente, ofrece una oportunidad sin igual no sólo para enriquecer semánticamente la descripción de los recursos sino para también enlazar recursos de distinta índole entre sí. En este contexto el usuario puede beneficiarse del trabajo de otros reutilizando información ya existente y al mismo tiempo colaborar publicando sus datos de forma pública y abierta. Esta filosofía ha llegado incluso a ser adoptada por gobiernos como los de EEUU<sup>3</sup> o Reino Unido<sup>4</sup>, que están publicando datos de dominio público de forma que puedan ser consultados y reutilizados por los ciudadanos. Estas iniciativas comparten ciertas características con el concepto de IDE como es la función de dar acceso a datos públicos. Pese a sus muchas ventajas los actuales sistemas de publicación no explotan totalmente esta información y no facilitan la tarea de enlazar los recursos publicados con otros. Así pues, el desarrollar herramientas o sistemas que permitiesen enlazar datos o recursos de las IDE con información publicada como Linked Data mejoraría la descripción de ambas partes permitiendo el descubrimiento de unas a través de otras. En definitiva un mayor número de enlaces entre recursos o datos puede representar una mayor cantidad de información relacionada y accesible y por tanto un mayor espacio de conocimiento que facilite el descubrimiento de una mayor cantidad de información. En la actualidad también existen iniciativas en España<sup>5</sup> centradas en el enriquecimiento de esta Web de Datos con datos geoespaciales del territorio nacional español. Disponiendo de esta información públicamente accesible sólo queda enlazar y de esta forma interconectar otros recursos o datos con ella.

En la Web los recursos geográficos cobran cada día más importancia y su cantidad y uso aumenta a la par que la disparidad entre tipos de usuarios que hacen uso de ellos. En la actualidad, para localizar recursos pertenecientes a una IDE, el usuario suele realizar búsquedas a través de un geoportal u otra interfaz que posiblemente realiza peticiones a algún servicio de catálogo. Aunque eficaz, este método no tiene por qué ser necesariamente el único para descubrir recursos si lo que se intenta es potenciar su visibilidad al publicarlos. De todos es conocida la importancia de los motores de búsqueda como Google<sup>6</sup>, Yahoo<sup>7</sup> o Bing<sup>8</sup> para la inmensa mayoría de usuarios de la Red. El poder publicar recursos y que éstos sean visibles para los *crawlers* o robots de estos motores incrementaría de manera considerable la visibilidad de dichos recursos haciéndolos accesibles a través de sus servicios y por ende accesibles a un volumen de usuarios de los que posiblemente y hasta la fecha no gozan los geoportales. Además, el concepto de motor de búsqueda también se ha adaptado a la Web Semántica existiendo ya herramientas específicas que permiten buscar<sup>9</sup> y navegar<sup>10</sup> por contenido hecho público siguiendo la filosofía del Linked Data. Una ventaja de estos motores de búsqueda que no trabajan sobre el contenido de documentos como los actuales sino sobre datos o información expresada en RDF es la de que ofrecen al usuario enlaces desde el propio recurso a otros que tienen relación con él permitiendo al usuario descubrir mucha más información relacionada. Esta aproximación tiene claras ventajas respecto a los métodos de búsqueda que, por ejemplo, se basan en la coincidencia o aparición de ciertos valores o texto en distintos campos de por ejemplo un archivo de metadatos. Mediante estos motores de búsqueda sobre datos basados en RDF no sólo la información que contiene cierto texto es devuelta como resultado sino que otros recursos que aun no conteniendo esa información pueden ser de utilidad para el usuario también lo son. En cualquiera de los casos la publicación de información geoespacial de forma acorde a como trabajan los actuales motores de búsqueda incrementaría su visibilidad y por tanto su accesibilidad.

Por los motivos anteriormente expuestos parece razonable investigar nuevos métodos para la publicación de datos más acordes con la evolución que está siguiendo la Web y que faciliten la tarea al usuario al mismo tiempo que represente un método efectivo para la localización de la información publicada por otros usuarios potenciales. Así pues parece necesario el ofrecer métodos que permitan describir como un único elemento

---

<sup>2</sup> <http://linkeddata.org>

<sup>3</sup> <http://www.data.gov>

<sup>4</sup> <http://www.data.gov.uk>

<sup>5</sup> <http://geo.linkeddata.es>

<sup>6</sup> <http://www.google.es>

<sup>7</sup> <http://es.yahoo.com>

<sup>8</sup> <http://www.bing.com>

<sup>9</sup> [http://semanticweb.org/wiki/Category:Semantic\\_search\\_tool](http://semanticweb.org/wiki/Category:Semantic_search_tool)

<sup>10</sup> [http://semanticweb.org/wiki/Category:Semantic\\_browser](http://semanticweb.org/wiki/Category:Semantic_browser)

compuesto a distintos recursos heterogéneos que aporten toda la información necesario al usuario independientemente de su naturaleza. Al mismo tiempo el reducir esfuerzos en la descripción de los recursos deriva inevitablemente en un proceso de publicación y descripción de los recursos más ágil y sencillo. Esta simplificación podría implementarse, por ejemplo, describiendo propiedades comunes a todos los elementos a nivel de la colección o elemento compuesto que los agrupa evitando replicar la misma información en varios recursos. Considerando lo anteriormente expuesto un primer paso para desarrollar nuevos sistemas de publicación de recursos de forma que se agilice o incluso se llegue a automatizar el proceso evitando tareas innecesarias por parte del usuario pasa por la adaptación de la información ya existente a nuevos contextos como el de la Web de Datos. De esta forma no sólo se publica esta información en un espacio más abierto y accesible sino que puede ser reutilizado por otros usuarios del mismo modo que información creada por estos usuarios puede ser aprovechada para enriquecer su descripción.

Este documento presenta el trabajo realizado en la creación del prototipo de aplicación GeoCollector. Esta aplicación puede definirse como una aplicación para la gestión de colecciones de recursos heterogéneos y que persigue facilitar la realización de los siguientes objetivos:

- Crear y describir elementos complejos basados en colecciones de elementos heterogéneos.
- Agilizar y hacer más eficiente la descripción de recursos evitando la réplica de información.
- Relacionar recursos no geoespaciales con otros de naturaleza geoespacial para su introducción y aprovechamiento en el campo SIG.
- Reutilizar recursos o datos ya existentes en el ámbito geoespacial y su inclusión en la Web de Datos.
- Incrementar la visibilidad de recursos o datos ya existentes en el ámbito geoespacial mediante su representación en formatos procesables por agentes automáticos como los motores de búsqueda.

El trabajo se expone de la siguiente forma. En la sección 2 se presenta el concepto de colecciones de recursos heterogéneos y el modelo de datos abstractos sobre el que se basa la herramienta. La sección 3 presenta un análisis de las partes más importantes de la herramienta GeoCollector como son el modelo interno de las colecciones, la interfaz de usuario, la herramienta para búsqueda de recursos o las distintas serializaciones en las que permite exportar las colecciones creadas por el usuario. Finalmente la sección 4 expondrá las conclusiones de este trabajo.

## 2. COLECCIONES DE RECURSOS HETEROGÉNEOS

El *Object Reuse and Exchange* (ORE) es un estándar para la descripción e intercambio de agregaciones o colecciones de recursos Web propuesto por la *Open Archive Initiative* (OAI)<sup>11</sup> y cuya primera versión vio la luz en 2008. ORE define un modelo de datos abstractos para la representación de agregaciones de recursos Web identificados mediante su URI. Uno de los objetivos de ORE, además de representar un estándar para la identificación de agregaciones, es describir los constituyentes y límites de estas. Este estándar se fundamenta en 3 pilares básicos:

- Arquitectura de la World Wide Web [6] cuyas bases son:
  - o Un recurso es cualquier ítem de interés.
  - o URI es usado como identificador global para recursos.
  - o La representación es el conjunto de datos correspondiente al estado del recurso en el momento en el que su URI es accedida mediante algún protocolo como HTTP.
  - o Un enlace representa la conexión entre dos recursos.
- Web Semántica, *Linked Data* y *Cool URIs* para la Web Semántica [7]. Habitualmente en la Web las URI describen documentos Web y cuando estas son accedidas ponen a disposición del usuario una representación de dicho documento. En la Web Semántica las URI pueden describir no sólo documentos Web sino entidades del mundo real y entidades abstractas. A diferencia de los documentos Web, dichas entidades no disponen de una representación que indique que significan como recurso. La iniciativa del *Linked Data* describe una aproximación para obtener información de recursos sin representación basada en las siguientes normas:
  - Uso de URIs HTTP para la descripción de recursos.
  - Publicación de un documento que ofrezca información sobre el recurso en una URI HTTP distinta a la del propio recurso.
  - Uso de *Cool URIs* para la Web Semántica para el descubrimiento de la URI HTTP que ofrece información sobre el recurso a partir de su propia URI.
- *Resource Description Framework* (RDF). Los documentos propuestos por el *Linked Data* están habitualmente expresados en RDF/XML, serialización del RDF basada en XML y que representa el

---

<sup>11</sup> <http://www.openarchive.org/ore>

modelo de datos para la Web Semántica. Este modelo se basa en declaraciones del tipo sujeto-predicado-objeto llamadas triplas. Estas triplas expresan relaciones de un recurso (sujeto) identificado mediante su URI. El predicado, también identificado por su URI indica el tipo o naturaleza de la relación. Por último, el objeto representa el valor actual de la relación mediante un valor literal (cadena, entero, etc.) o mediante una URI denotando posiblemente otro recurso. En muchas ocasiones los sujetos tienen varias relaciones o predicados y pueden llegar a compartir objetos de una relación formando estructuras similares a los conocidas en el ámbito de las matemáticas como grafos.

El modelo de datos propuesto por ORE que muestra la **Figura 1** se puede resumir en los siguientes puntos:

- Para referenciar y localizar de forma inequívoca a la agregación de recursos Web, se crea un nuevo recurso llamado *Aggregation* (agregación). Como cualquier otro recurso, la *Aggregation* dispone de una URI que la identifica, sin embargo, dado que se trata de una construcción conceptual no dispone de representación.
- Siguiendo las recomendaciones del *Linked Data*, un nuevo recurso es introducido para ofrecer información acerca del elemento *Aggregation*. Este recurso, llamado *Resource Map* (mapa de recursos) puede ser considerado como la representación de una *Aggregation* y ofrece información acerca de la misma. Un *Resource Map* expresa la agregación que describe y los recursos que forman parte de esta, pero también puede representar relaciones y propiedades de los recursos que constituyen la agregación así como sus metadatos. Estos mapas de recursos pueden ser serializados en distintos formatos entre los que destacan Atom [8], RDF/XML, RDFa, N3 o Turtle.
- Para poder trabajar mediante HTTP tanto con las agregaciones como con sus mapas de recursos ambas poseen una URI HTTP y los principios de las *Cool URIs* para la Web Semántica son aplicados para facilitar el descubrimiento de la URI de un mapa de recursos a través de la URI de la agregación a la que describe. ORE propone dos aproximaciones. La primera comprende el uso de redirecciones HTTP 303 que a partir de la URI de la agregación dirige al cliente a la URI del mapa de recursos. La segunda aproximación se basa en el uso de *Hash URIs* o identificadores de fragmentos.

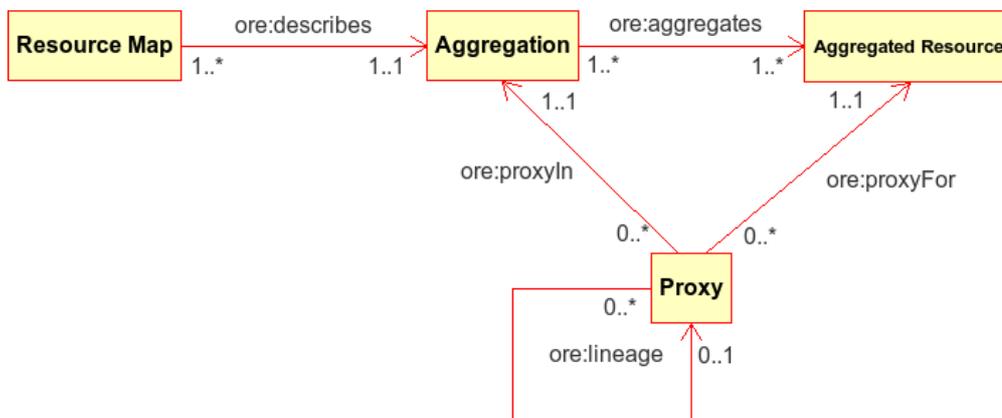


Figura 1: Modelo de datos definido por OAI-ORE.

Así pues las entidades básicas del modelo son la agregación, el recurso agregado y el mapa de recursos. Las agregaciones consisten en uno o más recursos que pueden presentar relaciones, en muchos casos de tipo semántico, entre ellos. Uno o más de los recursos de la agregación puede tener tipos que pueden ser definidos por varios vocabularios u ontologías. Además pueden existir relaciones entre la agregación y/o recursos de la agregación y unidades de información externas a la misma que pueden ser a su vez otros recursos o agregaciones. Cada mapa de recursos describe una única agregación, sin embargo una agregación puede ser descrita por uno o más mapas de recursos. Estos mapas de recursos son los que enumeran los recursos agregados que conforman a nivel lógico la agregación. Por último, los mapas de recursos pueden incluir propiedades adicionales acerca de la agregación y los recursos de esta. Este modelo de datos permite la anidación, pudiendo ser una agregación un recurso agregado de otra y así sucesivamente. Por último, ORE introduce el concepto de *Proxy*. Un *Proxy* se puede describir como un recurso agregado en el contexto específico de la agregación de la que forma parte. Dicho *Proxy* dispone de una URI que lo identifica dentro del contexto de la agregación a la que pertenece. Este sistema permite la creación de relaciones entre recursos que sólo tienen sentido en el contexto propio de la agregación. Un ejemplo del uso de este tipo de elemento se puede encontrar en la descripción de un orden secuencial/temporal entre los recursos agregados.

En cuanto al uso de metadatos, estos pueden y deben ser usados tanto para describir la agregación como sus mapas de recursos. La especificación OAI-ORE recomienda el uso elementos del *Dublin Core* [9] y de Foaf<sup>12</sup> para añadir metainformación referente a la autoría y última modificación del mapa de recursos, que representan el conjunto mínimo de metadatos necesarios en un agregación descrita en ORE. Sin embargo, los metadatos que se pueden añadir no están limitados al uso de estos esquemas sino que pueden ser ampliados dependiendo de la necesidad o finalidad de la agregación. Por ejemplo, en el ámbito SIG, se podrían añadir metadatos usando la estructura definida por el estándar ISO 19115 [10]. De esta forma se puede hacer uso de cualquier vocabulario para expresar metadatos y relaciones, y así describir el tipo de los recursos agregados o relaciones entre ellos u otros recursos que no forman parte de la agregación.

ORE permite la serialización de los mapas de recursos en RDF/XML, RDFa y Atom XML pero debido a que está basado en RDF otras serializaciones de este formato como Turtle o n3 pueden ser usadas.

El descubrimiento de mapas de recursos es una precondition de uso en ORE. El estándar sugiere una variedad de métodos agrupados en tres categorías: *Batch Discovery*, *Resource Embedding* y *Proxy URI*. En la primera categoría se presentan métodos para el descubrimiento en masa de mapas de recursos mediante el uso de *Atom Feeds*, archivos sitemap con enlaces a los archivos que contengan los mapas de recursos o el uso del estándar OAI-PMH (*Protocol for Metadata Harvesting*)<sup>13</sup>. El siguiente grupo de métodos propone la inserción de enlaces a los mapas de recursos embebidos dentro de los recursos que pertenecen a las colecciones (links en HTML o metadatos en Adobe XMP<sup>14</sup> dentro de archivos PDF). Los Proxy URI presentan una aproximación parecida pero esta vez apuntando a los elementos *Proxy* dentro del archivo ORE que representa el mismo recurso desde el que se enlaza.

### 3. GEOCOLLECTOR

GeoCollector es una aplicación diseñada para la creación de colecciones o agregaciones de recursos heterogéneos. Esta herramienta se ejecuta en el equipo del usuario y está basada en Java y desarrollado como aplicación Eclipse RCP (*Rich Client Platform*)<sup>15</sup> por lo que puede ser portada a distintas plataformas como Windows, Linux o Mac OS X sin dificultad. Las colecciones creadas mediante esta aplicación siguen un modelo basado en el especificado por el OAI-ORE y permite no sólo agrupar recursos heterogéneos en colecciones sino describir tanto los recursos como la agregación en si mediante elementos *Dublin Core*, dotar de geometría y localización geográfica a elementos que carezcan de ella así como relacionar los distintos recursos entre si y con otros ya existentes. GeoCollector también permite acceder a fuentes de datos que expongan su funcionalidad mediante *OpenSearch*<sup>16</sup>. Por último GeoCollector permite la exportación de las colecciones creadas mediante su editor a formatos que facilitan la publicación de dichas colecciones como *Linked Data* (RDF/XML, N3, Turtle), que facilitan su indexación por motores de búsqueda tradicional conservando toda la información estructurada (RDFa), que permiten su utilización y visualización en herramientas y sistemas SIG (KML, KML+MIMEXT) y que ofrecen métodos alternativos que permitan al usuario detectar fácilmente posibles actualizaciones en dichas colecciones (Atom).

#### 3.1. MODELO DE DATOS DE GEOCOLLECTOR

GeoCollector se basa en el modelo OAI-ORE para definir y describir las colecciones de recursos. En el desarrollo de esta herramienta se ha implementado un modelo similar, aunque no exacto, al definido por OAI-ORE, que se define a continuación y que ha sido desarrollado mediante Eclipse EMF (*Eclipse Modelling Framework*)<sup>17</sup>. EMF es un *framework* para el modelado y generación de código que facilita la creación de herramientas y otras aplicaciones basadas en modelos de datos estructurados. Mediante la especificación de este modelo, EMF es capaz de generar las distintas clases que implementan dicho modelo así como sus *adapter* que permiten su visualización y edición. Así pues, mediante EMF se ha creado un modelo estructurado representado en la Figura 2.

Este modelo describe parcialmente el modelo definido mediante la especificación OAI-ORE además de añadir otros elementos específicos para el caso de uso de la herramienta. A partir del diagrama del modelo de datos podemos observar que los elementos y relaciones del modelo de datos abstracto definido por OAI-ORE se corresponden con los elementos *Aggregation*, *ResourceMap*, *AggregatedResource* y *Proxy* y las distintas relaciones entre ellos. Estos cuatro elementos y sus relaciones permiten la descripción de colecciones o agregaciones (representadas por el elemento *Aggregation*) de recursos representados por el elemento *AggregatedResource*.

---

<sup>12</sup> <http://xmlns.com/foaf/spec>

<sup>13</sup> <http://www.openarchives.org/pmh>

<sup>14</sup> <http://www.adobe.com/products/xmp/overview.html>

<sup>15</sup> <http://www.eclipse.org/rcp>

<sup>16</sup> <http://www.opensearch.org>

<sup>17</sup> <http://www.eclipse.org/emf>

Además, como se define en OAI-ORE, las agregaciones necesitan de un elemento *ResourceMap* que simboliza su serialización puesto que las agregaciones son elementos abstractos. Además OAI-ORE define el elemento *Proxy* asociado a un determinado recurso agregado y que representa el rol específico de este en el contexto de la agregación lo que dota a este elemento de una gran versatilidad para la descripción de los recursos que componen dicha agregación. Como es especificado en OAI-ORE todos los recursos se identifican por su URI. Dado que todos estos comparten esta y otras características se ha definido el elemento *Resource* del que derivan todos ellos.

Además de estos elementos básicos se ha extendido el modelo con otros específicos del contexto de la aplicación. El primer grupo de estos elementos es el formado por los elementos *Relationship* y *RelationshipAttr* que simbolizan relaciones y atributos para estas. Estas relaciones se basan en el concepto de tripleta definido en el contexto del *Linked Data* por el que se definen relaciones de tipo sujeto-predicado-objeto. Así el elemento *Relationship* implementaría el predicado de esta relación que podría ser entre elementos o entre un elemento y un determinado valor y que se identifica por su URI al igual que otros elementos del modelo. Además de su URI y el valor en caso de que la relación asigne un valor a un elemento en lugar de definir una relación entre elementos, cada relación puede tener un número indefinido de atributos. Otro grupo de elementos añadido al modelo es el representado por los elementos referentes a los metadatos de los elementos de la agregación, representados por *Metadata* y *DCMetadata*. El motivo de definir una interfaz y una clase que la implementa es la futura ampliación o extensión del modelo. Así pues, hasta la fecha se ha definido la clase *DCMetadata* que simboliza un elemento donde almacenar todos aquellos metadatos de un determinado recurso definidos mediante el uso de *Dublin Core*. Como se puede observar en la Figura 2, esta clase define todos aquellos elementos definidos como *Dublin Core Elements*. En futuras versiones de esta herramienta se espera poder ampliar la parte del modelo correspondientes a metadatos con otros modelos más específicos y complejos como los basados en ISO. Por último se han incluido en el modelo las clases necesarias para dotar de geometría y de geolocalización a todos aquellos componentes de la colección. Como en el caso de los metadatos, se ha definido la interfaz *Geometry* para permitir futuras extensiones del modelo con nuevos tipos de geometrías. Debido a la necesidad de encontrar un modelo de geometrías simples para este primer prototipo de la aplicación se ha optado por basar estas geometrías en el modelo definido por la versión *Simple* de GeoRSS<sup>18</sup>. Esta versión define las geometrías *Point*, *Box*, *Polygon*, *Circle* y *Line*. Mediante estas geometrías simples podemos georreferenciar fácilmente todos aquellos recursos que formen parte de la colección e incluso la colección en sí misma (esta podría representarse mediante una geometría *Box* que comprendiera todas las geometrías asociadas a los recursos agregados a modo de *Bounding Box*). En futuras versiones estas geometrías podrían extenderse con otras como las definidas en el estándar OGC GML [11] u OGC KML [12].

---

<sup>18</sup> <http://www.georss.org/simple>

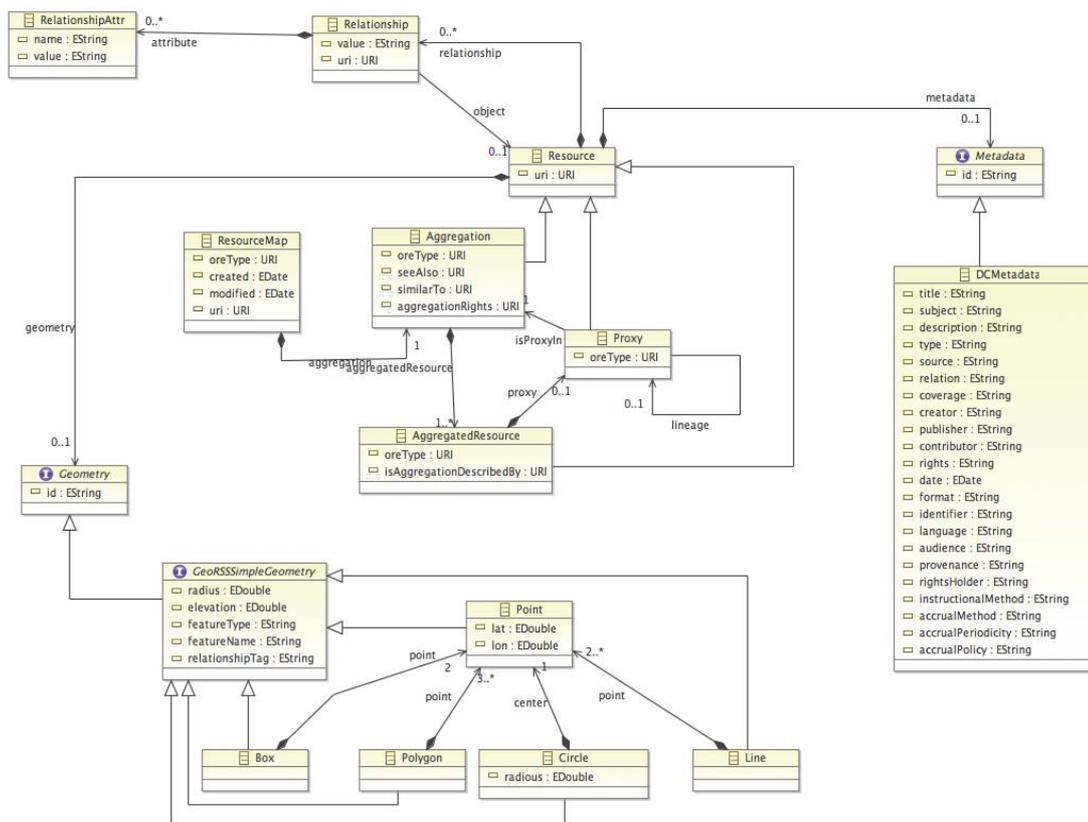


Figura 2: Diagrama ecocore del modelo de datos usado por GeoCollector

Con este modelo de datos y mediante el *framework* Eclipse EMF se han generado todas las clases y *adapters* necesarios para implementar y trabajar con el modelo así como un editor básico que permite al usuario trabajar de forma amigable con instancias del modelo. Otra de las ventajas de EMF es que genera automáticamente todo el código necesario para la serialización del modelo (por defecto en formato XMI) así como la validación del mismo. Gracias a esto la herramienta permite la validación de cualquier colección para asegurar su exactitud y conformidad con el modelo antes de ser reutilizada por terceros.

### 3.2. INTERFAZ DE USUARIO

La interfaz de GeoCollector está dividida en vistas como cualquier aplicación basada en Eclipse RCP y actualmente la aplicación dispone de 3 vistas principales:

- Vista de editor: El editor creado mediante EMF se incluye dentro de la aplicación Eclipse RCP y permite la visualización de las colecciones creadas por el usuario en forma de árbol como muestra la **Figura 3**. El usuario, mediante menús contextuales puede crear nuevos elementos dependiendo del elemento seleccionado y siempre siguiendo el modelo establecido. Este editor también permite la validación de un determinado elemento o conjunto de elementos del modelo evitando el tener que validar todo el modelo por completo. Por defecto, para cada elemento, el editor muestra su tipo e identificador usando la vista de propiedades para visualizar y editar el resto de atributos.

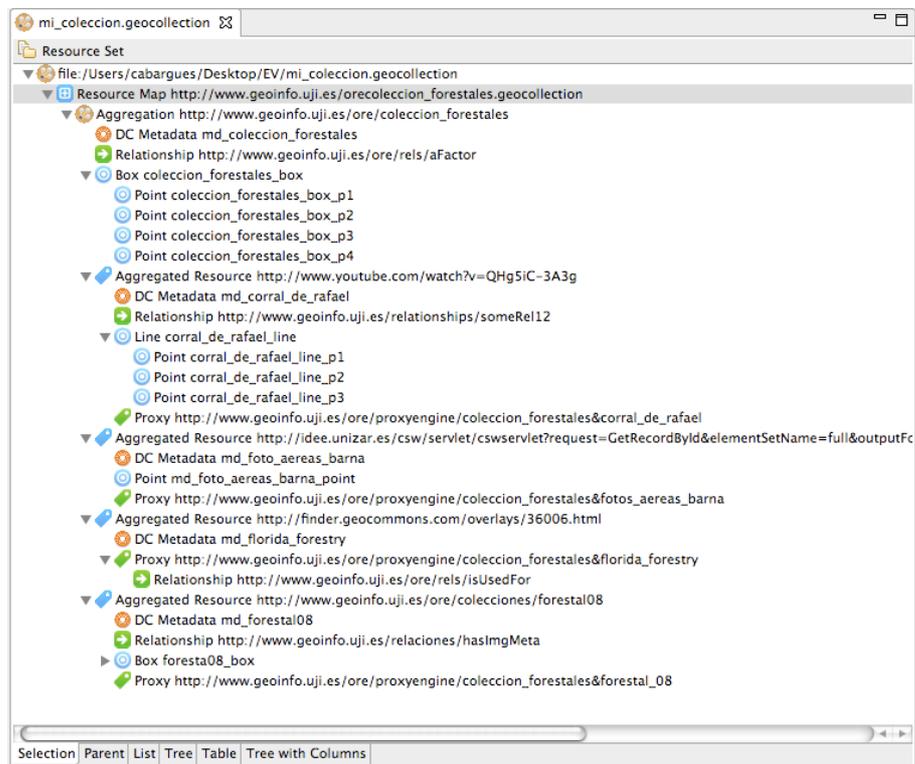


Figura 3: Vista de editor de GeoCollector.

- Vista de propiedades: Esta vista permite la visualización y edición de todos los atributos y relaciones del elemento seleccionado en la vista de editor. Estos atributos pueden ser detallados mediante simples campos de texto, listas u otros elementos que ofrece la plataforma Eclipse RCP dependiendo del tipo de atributo y su cardinalidad. La **Figura 4** muestra un ejemplo de esta vista con los atributos referentes a los metadatos (elemento *DCMetadata*) de un determinado elemento del modelo.
- Vista de búsqueda: Esta vista ofrece una interfaz para la búsqueda de recursos que puedan ser reutilizados como recursos agregados de la aplicación y cuyo funcionamiento se explica en mayor detalle en la siguiente sección.

### 3.3. BÚSQUEDA Y ACCESO A RECURSOS

La creación de colecciones de recursos implica el conocer la URI de cada uno de los recursos que forman o formarán parte de dicha colección. A la hora de crear una de estas colecciones puede darse el caso en el que el usuario no conozca con exactitud la URI del recurso a agregar o incluso que quiera descubrir nuevos recursos para relacionarlos con otros ya conocidos. Para facilitar al usuario la tarea de encontrar recursos se ha incorporado a la aplicación un nuevo módulo de búsqueda que en esta versión del prototipo se basa en *OpenSearch*. Este conjunto de tecnologías representan un método para la publicación de resultados de búsqueda de forma conveniente para su agregación a la vez que permite publicar resultados de búsqueda de un modo estándar y en un formato accesible. *OpenSearch* define un modo estándar de definir y describir motores de búsqueda, sintaxis de búsqueda y formato del resultado de las búsquedas. En la actualidad, *OpenSearch* está incrementando su popularidad en el mundo de las tecnologías geoespaciales gracias al desarrollo y de su extensión geoespacial (*OpenSearch-geo*) y a las posibilidades que puede ofrecer como interfaz unificada de búsqueda para distintos servicios [13, 14].

Property	Value
Accrual Method	
Accrual Periodicity	
Accrual Policy	
Audience	
Contributor	Universitat Jaume I de Castellón
Coverage	España
Creator	Carlos Abargues
Date	
Description	Colección de recursos para estudio sobre incendios forestales.
Format	
Id	md_coleccion_forestales
Identifier	md_coleccion_forestales
Instructional Meth	
Language	es-ES
Provenance	
Publisher	Universitat Jaume I de Castellón
Relation	
Rights	<a href="http://creativecommons.org/licenses/by-sa/2.5/">http://creativecommons.org/licenses/by-sa/2.5/</a>
Rights Holder	Universitat Jaume I de Castellón
Source	
Subject	Incendios forestales
Title	Colección de recursos forestales e incendios.
Type	<a href="http://purl.org/dc/dcmitype/Collection">http://purl.org/dc/dcmitype/Collection</a>

Figura 4: Vista de propiedades de GeoCollector.

GeoCollector hace uso de un pequeño cliente *OpenSearch* basado en Java que le permite interactuar con servicios y motores de búsqueda que ofrezcan esta interfaz. Así pues el cliente permite, a partir de la URL de un documento de descripción de servicio *OpenSearch*, parsear dicho documento, recibir los parámetros de la búsqueda y devolver de forma conveniente los resultados de la misma. En la actualidad, GeoCollector sólo ofrece un campo para los *Search Terms* o términos de la búsqueda aunque se está trabajando para generar automáticamente y en tiempo de ejecución la interfaz del módulo de búsqueda en base a los parámetros que permita el servicio con el que trabajar y que están definidos en las sintaxis de las búsquedas. *OpenSearch* permite definir varios tipos de formatos de salida distintos para los resultados de la búsqueda siendo los más comunes los RSS, Atom, HTML o JSON.

La extensión geoespacial de *OpenSearch* amplía su uso para poder especificar parámetros que permitan la búsqueda en base a la localización de los elementos que ofrece un determinado servicio o motor de búsqueda. Además y dada la naturaleza de los resultados también se ha ampliado el formato en el que estos son presentados popularizándose el uso de *feeds* RSS o Atom junto con GeoRSS para definir la localización geográfica y geometría de los resultados de la búsqueda así como otros formatos como el estándar OGC KML. Este tipo de extensión permite ofrecer una interfaz unificada para la búsqueda de información, sea esta de carácter geoespacial o no. Por estos motivos *OpenSearch* y su extensión *geo* representan un método de búsqueda y acceso a información idóneo.

Gracias a su flexibilidad y sencillez, *OpenSearch* puede usarse para describir como interrogar casi cualquier servicio de búsqueda. De esta forma es posible crear descriptores mediante *OpenSearch* que definan por ejemplo como interrogar a una instancia de *GeoNetwork*, a un motor de búsqueda como Google o a un servicio de vídeos como *YouTube*<sup>19</sup>. Gracias a esto el usuario de GeoCollector es capaz de interrogar motores de búsqueda, catálogos y demás sistemas que ofrezcan su información mediante *OpenSearch* y recuperar los recursos o resultados de búsqueda para ser incorporados como recursos agregados de la colección. La **Figura 5** muestra la interfaz de la sección de búsqueda en el prototipo.

<sup>19</sup> <http://www.youtube.com>

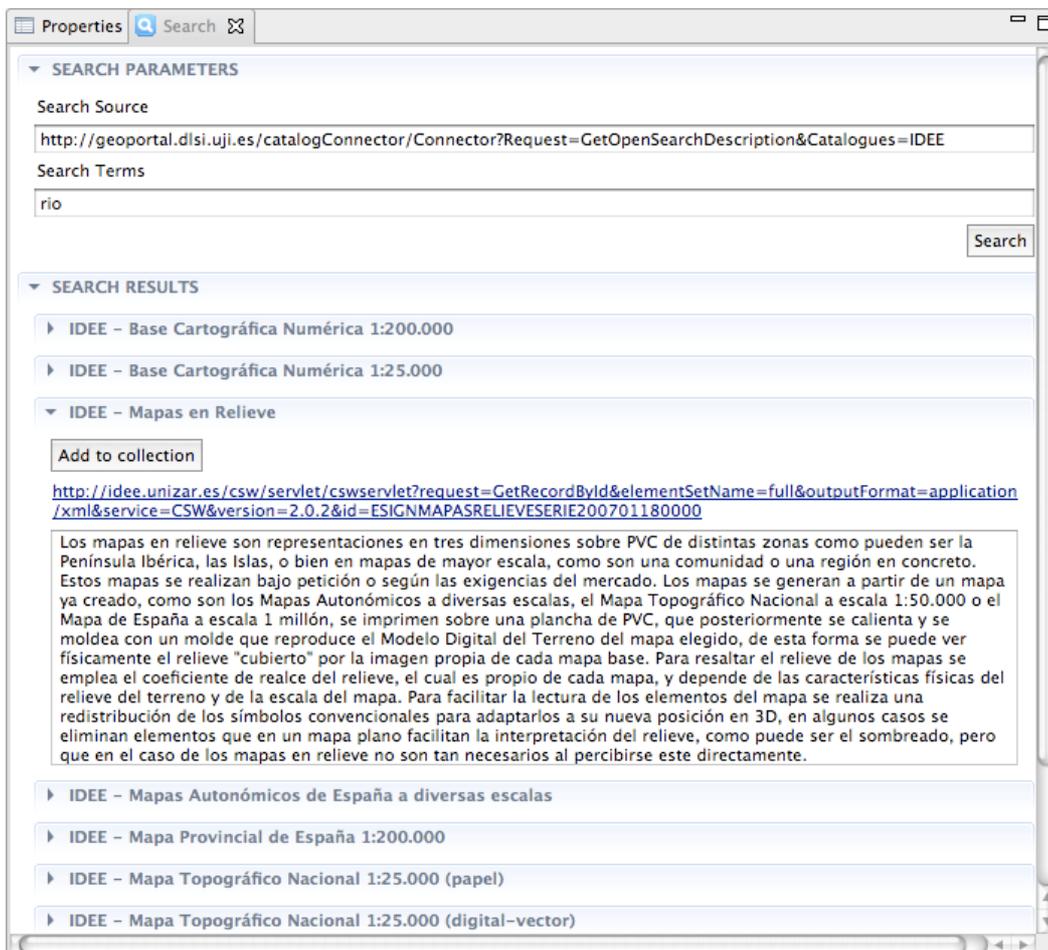


Figura 5: Vista de búsqueda y resultados de una búsqueda mediante OpenSearch en GeoCollector.

Aunque actualmente es necesario que el usuario introduzca la URL del descriptor *OpenSearch* se está trabajando en un gestor para los servicios de búsqueda que permita al usuario gestionarlos de forma amigable mediante un panel de preferencias y que permita registrar o eliminar servicios y seleccionar mediante una lista aquellos previamente registrados. Como muestra la **Figura 5** el usuario además de la URL del descriptor puede introducir una serie de términos de búsqueda (en las sucesivas versiones se generarán automáticamente todos los campos para los parámetros definidos en el descriptor). Los resultados de la búsqueda son creados tras procesar y parsear el resultado enviado por el servicio y mostrado como una lista de secciones desplegadas que muestran el título del resultado, y su URL junto a la descripción al desplegar la sección que lo representa. Como se puede observar también se incluye un botón que permite añadir directamente el resultado consultado como elemento de la colección. Al añadir dicho elemento se crea automáticamente un recurso agregado para la colección cuya URI corresponde con la URL del resultado de la búsqueda. Además se crea otro elemento *DCMetadata* asociado a este recurso agregado y que simboliza sus metadatos. Tanto el título como la descripción son usados para rellenar los campos del elemento *DCMetadata* que se corresponden con los elementos *Description* y *Title* definidos por la especificación *Dublin Core*. En la versión actual del prototipo estos dos campos son los únicos importados directamente del resultado de la búsqueda puesto que representan el mínimo conjunto de información que es devuelto habitualmente por los servicios de búsqueda. No obstante esta información puede ser ampliada, dependiendo siempre del servidor o motor de búsqueda sobre el que se trabaje. En aquellos casos en los que la cantidad de metadatos devueltos sea mayor, estos podrían ser reutilizados para describir aquellos recursos agregados a la colección. Por ejemplo, un caso en el que se podría obtener más información sería aquel en el que se estuviese interrogando a una instancia de *GeoNetwork*. Habitualmente las entradas del catálogo ofrecen mucha más información que únicamente el nombre y descripción del contenido al que hacen referencia y por tanto esta información podría ser reutilizada, adaptándola al modelo de metadatos usado siempre que fuese necesario.

### 3.4. SERIALIZACIÓN

Al crear una colección mediante GeoCollector esta se guarda en disco y serializa de forma transparente al usuario. Al crear un modelo de datos mediante Eclipse EMF esta herramienta también genera automáticamente todo el código necesario para serializar y parsear convenientemente las instancias basadas en este modelo. Además, EMF permite la serialización del modelo en varios formatos siendo el más común el *XML Metadata Interchange* (XMI). El formato XMI como su nombre bien indica está basado en XML y representa una especificación comúnmente usada para el intercambio de diagramas. En el caso GeoCollector este formato permite la representación de toda la información relacionada con la colección y sus componentes en un lenguaje basado en XML .

Pese a su utilidad XMI no es un lenguaje usado fuera de su contexto y mucho menos en el ámbito de la información geoespacial. Además este formato no se incluye entre aquellos recomendados por la OAI para expresar colecciones en su formato ORE enfocados principalmente para su uso en la Web Semántica o Web de Datos. Es por esto que el poder representar las colecciones en otros formatos facilitaría el uso de las mismas por otros usuarios o sistemas. GeoCollector permite la exportación de colecciones creadas por usuarios a los formatos RDF/XML, N3, N-Triples, Turtle, Atom, RDFa, KML y KML junto con la extensión MIMEXT [15].

OAI-ORE está basado en los principios de la arquitectura Web al igual que otras tecnologías actualmente en auge como son los servicios REST y que subyace detrás de *Linked Data*. Este modelo de datos para la descripción de colecciones encaja perfectamente dentro de este contexto. El describir colecciones permite no sólo agrupar en una única entidad lógica distintos conceptos que comparten ciertas características que los hacen susceptibles de ser agrupados sino que también permite el enlazar y de esta forma conectar datos que hasta la fecha podrían no estar. Además también se pueden crear nuevas colecciones entre recursos ya enlazados, enlazar datos a distintos niveles (nivel de recurso o nivel de colección, colecciones anidadas) y crear enlaces o conexiones que sólo tienen sentido cuando un dato o recurso en concreto forma parte de una colección determinada mediante el uso de *Proxy*. Considerando esto, parece pues lógico que aquellas colecciones que sigan el modelo OAI-ORE puedan y deban ser representadas mediante formatos reutilizables en la Web de datos. En el contexto de los sistemas de información geográfica la creación de estas colecciones de recursos heterogéneos y su publicación facilitaría la actual tendencia de publicar información o datos siguiendo los principios del *Linked Data* además de colaborar en su interconexión con otros datos. Así la serialización de estas colecciones en formatos basados en RDF cumplen con los objetivos establecidos permitiendo la incorporación de recursos geoespaciales al espacio definido por la Web de Datos permitiendo al mismo tiempo su interconexión con información ya existente. Estas conexiones permitirían la reutilización de información o datos ya existente evitando la tarea de replicarlos al mismo tiempo que permitiría descubrir más información o hacer visibles los recursos que forman parte de la colección mediante estos enlaces.

Toda esta información expresada en formatos basados en RDF como RDF/XML, N3 o Turtle no sólo permite ofrecer una visión estructurada de la información sino que también permite su indexación mediante motores de búsqueda específicos para la Web de Datos e incluso navegar entre los distintos elementos enlazados usando navegadores creados a tal efecto. Pese a ello, y de cara a potenciar la visibilidad de los datos hay que considerar otros formatos que son reconocidos por otros motores de búsqueda como Google<sup>20</sup>, Yahoo<sup>21</sup> o Bing<sup>22</sup> que son realmente los usados por la inmensa mayoría de usuarios de la Web. Entre otros formatos estos tipos de motores trabajan e indexan contenido expresado mediante formatos como HTML o XHTML. Es por ello que GeoCollector también permite serializar las colecciones de recursos haciendo uso del formato RDFa. Esta recomendación de W3C añade una extensión a nivel de atributo al lenguaje XHTML para poder embeber de esta forma metadatos dentro de este tipo de documentos. En otras palabras RDFa permite describir tripletas RDF dentro de documentos XHTML. RDFa ofrece características muy importantes que hacen de su uso una opción a considerar. Entre estas características destacan principalmente dos:

- Los formatos basados en RDF son conocidos por ser fácilmente procesables por agentes automáticos gracias a como se estructura la información, sin embargo estos formatos resultan complicados de interpretar por agentes humanos. RDFa al ser aplicado sobre documentos XHTML ofrece información estructurada fácilmente procesable por agentes automáticos que al estar representada sobre XHTML también resulta legible para agentes humanos. De esta forma RDFa permite la correcta interpretación de los datos por cualquier tipo de agentes.
- Los motores de búsqueda más populares y con mayor visibilidad en la Web trabajan principalmente con documentos basados en HTML y XHTML lo que por tanto excluye en muchos casos aquellos documentos en formatos como RDF/XML, N3, Turtle. Al embeber la información mediante el formato RDFa dentro de documentos XHTML estos son procesados convenientemente por los motores de búsqueda, que pese a desconocer o no interpretar convenientemente la estructura de la información sí indexan el contenido de estos documentos. Mediante la publicación de la descripción de colecciones en este formato estas son susceptibles de aumentar su visibilidad al poder ser correctamente procesadas por la mayoría de motores de búsqueda y por tanto accesibles a un mayor público.

---

<sup>20</sup> <http://www.google.es>

<sup>21</sup> <http://es.yahoo.com>

<sup>22</sup> <http://www.bing.com>

El uso de RDFa permite no sólo el distribuir y reutilizar la información gracias a que se basa en RDF sino que también permite dotar a esa información de mayor visibilidad al no estar acotada al nicho de la Web Semántica o los sistemas de información geográfica. Es de destacar, que algunos motores de búsqueda como Google también indexan otros formatos como documentos Adobe PDF o Microsoft Word. Entre todos estos destaca el formato *Keyhole Markup Language*, estándar OGC desde 2008. Desde hace algunos años aquellos documentos especificados siguiendo este formato también son analizados y parte de su contenido indexado por el motor de búsqueda de Google [16] y por tanto accesible desde su servicio de búsqueda. Es por esto que una hipotética serialización de la colección de recursos en este formato también incrementaría su visibilidad.

GeoCollector hace uso de la librería *Foresite*<sup>23</sup> para serializar las colecciones de recursos heterogéneos en los formatos RDF/XML, N3, N-Triples, Turtle y Atom. Foresite es una librería con versión para Java y Python para la construcción, parseo, manipulación y serialización de colecciones de recursos basadas en OAI-ORE. Esta librería también ofrece al usuario varias opciones para serializar sus colecciones entre las que destacan las anteriormente descritas. Así, se han implementado métodos para traducir aquellas colecciones creadas mediante GeoCollector a instancias definidas en *Foresite* para facilitar su serialización. Actualmente tanto las serializaciones en RDF/XML, N3, N-Triples, Turtle, RDFa y Atom producen una representación correcta de estas colecciones. Estos formatos permiten tanto la inserción de metadatos mediante elementos *Dublin Core* como las geometrías asociadas a los recursos mediante el uso de GeorSS. El uso de ambos en formatos derivados de RDF no es nuevo < cita de ambos (webs)> y permiten crear nuevos enlaces o añadir información a datos existentes.

Pese a que actualmente la serialización en Atom no es óptima su funcionamiento es importante y conlleva una serie de ventajas. Posiblemente la más importante de estas ventajas es la posibilidad de que cualquier usuario pueda suscribirse al *feed* que representa dicha colección. Gracias al funcionamiento de los distintos clientes de *feeds* el usuario puede estar continuamente monitoreando dicho hilo y recibir notificación ante cualquier modificación. Este modelo de comunicación de actualizaciones presenta ciertos inconvenientes como la necesidad de estar continuamente interrogando al *feed* para detectar posibles actualizaciones sin embargo nuevos modelos, como el definido por el reciente *Pubsubhubub*<sup>24</sup> ofrecen mejoras en cuanto a la eficiencia de las comunicaciones. De esta forma el usuario podría no sólo acceder a la información que necesita sino que, mediante otras herramientas como los clientes de noticias, también podría recibir automáticamente notificación de las posibles actualizaciones sobre dicha información permitiéndole estar siempre al día ante posibles actualizaciones.

Como se ha indicado, GeoCollector también permite la representación de las colecciones en formato KML y en KML usando además su extensión para la anotación de recursos multimedia MIMEXT. Mediante estos dos tipos de representaciones podemos no sólo expresar la colección en si sino también utilizarla en otras aplicaciones o sistemas propios del contexto de la información geográfica distribuyendo de esta forma la información referente a esta colección en un formato procesable por este tipo de sistemas. Un ejemplo de estas aplicaciones son por ejemplo los globos virtuales mediante los cuales se podría ofrecer una visualización tridimensional del contenido de la agregación. La serialización en KML representa la colección mediante el elemento KML *Folder* y sus recursos agregados mediante el elemento *Placemark*. Las URL que identifican a cada uno de los elementos de la colección son usados como *id* o identificadores de cada uno de los elementos KML (*Folder* y *Placemark*). Debido a la naturaleza heterogénea de los recursos agregados la visualización de los mismos sobre estas aplicaciones parece complicado sino imposible. Es por ello que también se ofrece la posibilidad de exportar la colección al formato KML usando la extensión MIMEXT. Esta extensión del lenguaje KML permite no sólo georreferenciar y anotar recursos mediante el uso de un nuevo elemento *Resource* dentro de archivos KML sino que ofrece información acerca de la localización del recurso (URL) y su tipo (información del tipo MIME del recurso en forma de tipo-subtipo). Esta información adicional sobre el recurso ofrece un mayor control sobre la información descrita en un archivo KML permitiendo mejorar la visualización de los mismos de forma que esta sea lo más acorde posible al tipo de contenido que cada elemento *Resource* representa. La información referente al tipo de contenido de cada recurso es extraída durante el proceso de conversión de la colección a este formato. OAI-ORE permite la anidación de colecciones creando mapas de colecciones. Dentro de una colección se distingue a un recurso agregado que también es una colección mediante la relación *ore:isDescribedBy* que relaciona e indica que ese recurso agregado es descrito por otro recurso que en este caso es el *ResourceMap* o serialización de dicha colección. En el modelo usado por GeoCollector esta relación se ha definido explícitamente por que puede ser creada directamente sin usar ningún otro elemento de tipo *Relationship*. En el caso de las serializaciones basadas en RDF esta relación se representa mediante una tripleta más sin embargo en el caso de KML se puede ampliar la funcionalidad de esta relación para por ejemplo poder visualizar el contenido de esa colección anidada junto con el contenido de la colección que la contiene. Esta funcionalidad es fácilmente implementable en KML mediante el uso del elemento *NetworkLink*. Este elemento del lenguaje permite especificar un enlace que apunta normalmente a otro archivo KML cuyo contenido será cargado al acceder a dicho elemento en el archivo KML que lo contiene. De esta forma podemos usar el elemento *NetworkLink* para representar ese anidamiento de colecciones y poder navegar y acceder al contenido de estas. Esta característica también es implementada durante el proceso de traducción en el cual, si un recurso agregado presenta una relación de tipo *ore:isDescribedBy* a un recurso de tipo KML, este recurso agregado es serializado como un elemento *NetworkLink*

<sup>23</sup> <http://foresite.cheshire3.org/>

<sup>24</sup> <http://code.google.com/p/pubsubhubub/>

que apunta a dicho recurso KML. En caso de que se especifique dicha relación pero a un recurso que no es de tipo KML el uso de *NetworkLink* carece de sentido y por tanto dicha relación será especificada del mismo modo que el resto de relaciones y metadatos dentro del archivo KML. Los metadatos y relaciones especificadas en GeoCollector para cada recurso son también añadidos a las representaciones generadas mediante cualquiera de los dos tipos de salida en KML. En primer lugar tanto los metadatos *Dublin Core* añadidos a cada recurso mediante el elemento *DCMetadata* como las distintas relaciones especificadas mediante el elemento *Relationship* son añadidas a la representación KML de cada recurso mediante el uso del elemento KML *ExtendedData*. Este elemento permite añadir código XML dentro del propio archivo KML. Así, mediante *ExtendedData* sólo es necesario especificar el espacio de nombres del código XML que en nuestro caso serán el espacio de nombres en el que se definen los elementos *Dublin Core* y aquellos en los que se definen las distintas relaciones de las que se hace uso. De esta forma, aunque el contenido dentro del elemento *ExtendedData* no es comúnmente visualizado por herramientas como los globos virtuales esta información es igualmente contenida en el archivo y por tanto puede ser reutilizada y procesada por otros agentes. Además de insertar todos los metadatos y relaciones en elementos *ExtendedData* algunos metadatos son también utilizados para describir los recursos y la agregación mediante los elementos más comúnmente usados en KML. Así el elemento *Title* de los metadatos en *Dublin Core* se usa para especificar el valor del elemento *name* tanto en los elementos *Folder* (colección) como *Placemark*, *Resource* o *NetworkLink* (recurso agregado). Del mismo modo el elemento *Dublin Core Description* se utiliza para dar valor al campo KML con el mismo nombre. La especificación KML también contempla la especificación de la autoría los documentos KML mediante el uso de elementos definidos en Atom como *name*. Estos campos de autoría también pueden ser rellenados extrayendo información de los metadatos *Dublin Core* como por ejemplo del elemento *creator*. Finalmente el proceso de conversión a KML implica que aquellas geometrías asignadas a cada uno de los recursos y especificadas por GeoRSS Simple deben ser también transformadas a aquellas geometrías especificadas en el estándar KML.

Por último hay que destacar la complejidad o más bien, imposibilidad de adaptar el concepto de *Proxy* a cualquier elemento KML. No parece lógico pues el representar un mismo recurso por duplicado en KML pese a que ambos tengan distintas relaciones. Es por esto que toda la información referente a un *Proxy* de un determinado elemento es incluida en KML como metadatos en elementos *ExtendedData*.

#### 4. CONCLUSIONES

En este trabajo se han revisado ciertos puntos cuya solución representarían mejoras en cuanto a la publicación de información geográfica como son la duplicidad de sus metadatos. Además se han presentado las ventajas de incluir y enlazar la información geoespacial dentro del espacio de conocimiento que representa la Web de Datos. GeoCollector intenta ofrecer mejoras en este y otros aspectos permitiendo al usuario la creación de colecciones de recursos heterogéneos basándose en tecnologías como en las que se fundamenta la propia Web de Datos. Además GeoCollector incluye tecnologías que facilitan al usuario la búsqueda de nuevos recursos y las distintas serializaciones permiten entre otras cosas facilitar el descubrimiento de esta información por agentes automáticos como son los motores de búsqueda.

#### 5. AGRADECIMIENTOS

Este trabajo ha sido financiado parcialmente por el proyecto CENIT España Virtual, subvencionado por en CDTI dentro del programa Ingenio 2010 a través del Centro Nacional de Información Geográfica (CNIG).

#### 6. REFERENCIAS

- [1] Open Geospatial Consortium. OpenGIS Catalogue Service (CSW) Implementation Specification, Version 2.0.2. Open Geospatial Consortium Inc. (Open GIS Consortium Inc.) <http://www.opengeospatial.org/standards/cat> (2007)
- [2] S. Choudhury, T. DiLauro, A. Szalay, E. Vishniac, R. Hanish, J. Steffen, R. Milkey, T. Ehling and R. Plante. Digital data preservation for scholarly Publications in astronomy. *International Journal of Digital Curation* , 2(2) (2007)
- [3] K. Cheung, J. Hunter, A. Lashtabeg, and D. J. SCOPE - a scientific compound object publishing and editing system. In 3rd International Digital Curation Conference (2007)
- [4] Berners-Lee T, Hendler J, Lassila, O. The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*,

284(5): 34-43 (2001)

- [5] Bizer, C., Cyganiak, R., & Heath, T. How to Publish Linked Data on the Web. Technical report. <http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial> (2007)
- [6] Jacobs, I. & Walsh, N. Architecture of the World Wide Web. Technical report. <http://www.w3.org/TR/webarch> (2004)
- [7] Sauer mann, L., Cyganiak, R., & Völkel, M. Cool URIs for the Semantic Web. Technical report. <http://www.dfki.uni-kl.de/sauer mann/2006/11/cooluris> (2007)
- [8] M. Nottingham and R. Sayre, The Atom Syndication Format, Internet Engineering Task Force (IETF) Memo RFC 4287. [Online]. <http://www.ietf.org/rfc/rfc4287.txt> (2005)
- [9] NISO. The Dublin Core Metadata Element Set: ANSI/NISO Z39.85-2007. American National Standards Organization. (2007)
- [10] ISO. Geographic information - Metadata. ISO/TS 19115:2003, International Organization for Standardization (ISO) (2003)
- [11] Open Geospatial Consortium. OpenGIS Geography Markup Language (GML) Encoding Standard, Version 3.2.1. Open Geospatial Consortium Inc (Open GIS Consortium Inc). <http://www.opengeospatial.org/standards/gml> (2007)
- [12] Open Geospatial Consortium. OpenGIS Keyhole Markup Language (KML) Implementation Specification, Version 2.2.0. Open Geospatial Consortium Inc (Open GIS Consortium Inc). <http://www.opengeospatial.org/standards/kml> (2008)
- [13] Óscar Fonts. Adding custom search engines to OpenLayers with OpenSearch 'geo'. In Proceedings of the 2010 Free and Open Source Software for Geospatial Conference (FOSS4G 2010). Barcelona, Spain (2010)
- [14] Oscar Fonts, Joaquín Huerta, Laura Díaz, Carlos Granell. OpenSearch-geo: El estándar simple para buscadores de información geográfica. Actas de las IV Jornadas de SIG Libre 2010 (SIG Libre 2010). Girona, Spain (Mar 2010)
- [15] Carlos Abargues, Arturo Beltrán, Carlos Granell. MIMEXT: a KML extension for georeferencing and easy share MIME type resources. In M. Painho, M.Y. Santos, H. Pundt (Eds): Geospatial Thinking (Proceedings of the 13th AGILE International Conference on Geographic Information Science , AGILE 2010). Guimarães, Portugal, May 2010, pp. 315-334, ISBN 978-3-642-12325-2. (2010)
- [16] Carlos Abargues. Discovery and retrieval of Geographic data using Google. PhD thesis, Univesitat Jaume I de Castellón, Castellón. <http://hdl.handle.net/10362/2536> (2009)

## 7. CONTACTOS

**Carlos ABARGUES**  
abargues@uji.es  
Universitat Jaume I de Castelló  
(UJI)  
Institute of New Imaging  
Technologies (INIT)

**Carlos GRANELL**  
carlos.granell@uji.es  
Universitat Jaume I de Castelló  
(UJI)  
Institute of New Imaging  
Technologies (INIT)

**Joaquin HUERTA**  
huerta@uji.es  
Universitat Jaume I de Castelló  
(UJI)  
Institute of New Imaging  
Technologies (INIT)