

## Crowd data, linked data y open data y cómo seguir usando INSPIRE

M. Arias de Reyna Domínguez<sup>(1)</sup>

<sup>(1)</sup> Ingeniera Informática, GeoCat bv, maria.arias@geocat.net

### RESUMEN

*Aunque INSPIRE lleva ya con nosotros casi una década, es poco conocido y usado fuera del ámbito institucional. Esto ha hecho que se desarrollen paralelamente otros formatos para almacenar e intercambiar datos colaborativos y abiertos. ¿Es posible seguir usando INSPIRE mientras la gente demanda otro tipo de formatos? ¿Se puede tener un IDE que compagine INSPIRE y RDF y no morir en el intento?*

**Palabras clave:** crowd-data, open-data, linked-data, GeoNetwork, CKAN.

## INTRODUCCIÓN

Como en muchos otros ámbitos de las nuevas tecnologías, en el ámbito de los catálogos de datos hay diversas corrientes de pensamiento y actuación que no siempre son compatibles entre sí. En el caso de los datos geográficos, este distanciamiento entre las corrientes de actuación es bastante marcado, lo que nos lleva a considerar que existen dos universos completamente diferentes, y en muchos casos incompatibles, sobre la manera de tratar los datos.

### El universo INSPIRE

Hace ya casi una década que INSPIRE entró en nuestras vidas. La idea era muy sencilla: encontrar un formato y unas reglas de definición de datos que permitieran un intercambio fácil y rápido de datos geográficos entre países europeos. Para ello se utilizó como base el lenguaje de marcado XML, muy de moda en aquel momento.

INSPIRE resultó ser muy útil a la hora de unificar datos geográficos europeos y su implantación en años posteriores facilitó mucho la tarea de desarrollar herramientas de forma conjunta entre diferentes países. Donde antes cada país definía sus propias normas y especificaciones, ahora todos se esforzaban en ser interoperables. Lo cual fue a su vez muy útil a la hora de compartir datos entre países fronterizos.

Una de las grandes novedades de INSPIRE fue la incorporación de tesauros multiidioma: Un tesoro es un listado de palabras clave que se comparten entre diferentes catálogos, de forma que, de alguna manera, datos que están en diferentes catálogos pueden estar relacionados entre sí al utilizar una misma palabra clave. Si el tesoro además tiene en cuenta las traducciones de la palabra clave (como agua, water y Wasser), permite que catálogos que ni siquiera están en el mismo idioma puedan relacionar entre sí datos mediante sus palabras clave, cada una definida en el idioma del catálogo.

INSPIRE ha seguido evolucionando desde su primera versión, adaptándose a lo que sus usuarios iban pidiendo. Aunque algunas voces discordantes protestaban porque era un formato demasiado estricto y muchas veces añadía un montón de texto inútil para datos simples, funcionó (y sigue funcionando) de forma razonablemente bien durante años. Tan buena aceptación tuvo, que incluso hubo quien habló de “tropicalizar” INSPIRE no hace mucho en estas mismas jornadas, exportándolo a países al otro lado del océano.

### El universo Open-Data

Con la irrupción de los movimientos de apertura de datos (open-data), el panorama cambió. Ahora lo importante no eran los estándares. Ni siquiera la interoperabilidad. Lo importante era poner a disposición del mayor número de personas en el menor tiempo posible la mayor cantidad de datos que tuviéramos disponibles. El responsable del dato no se hacía responsable de su formato. Era el usuario del dato, normalmente el desarrollador de una aplicación, el que se preocuparía de darle sentido a esos datos.

Esto hizo que irrumpieran muchos formatos para datos abiertos basados en los diferentes casos de uso que los desarrolladores de aplicaciones iban encontrando. Con el tiempo, algunos de esos formatos se convirtieron en estándares de facto, pero seguían evolucionando de forma errática y sin seguir una guía común. A esto también contribuyó el hecho de que el movimiento open data de liberación de datos tuviera casos de uso muy diferentes incluso dentro de un mismo país o temática.

Uno de los ejemplos más destacados de este movimiento de datos abiertos en el ámbito geográfico es OpenStreetMap. Diseñado como una base de datos geográfica, la idea era poder añadir de forma sencilla todo tipo de datos localizados, a los que se les iba asociando una serie de etiquetas, predefinidas o no.

Dentro de un contexto más general, el movimiento de datos abiertos fue poco a poco confluyendo en un formato común conocido como linked-data:

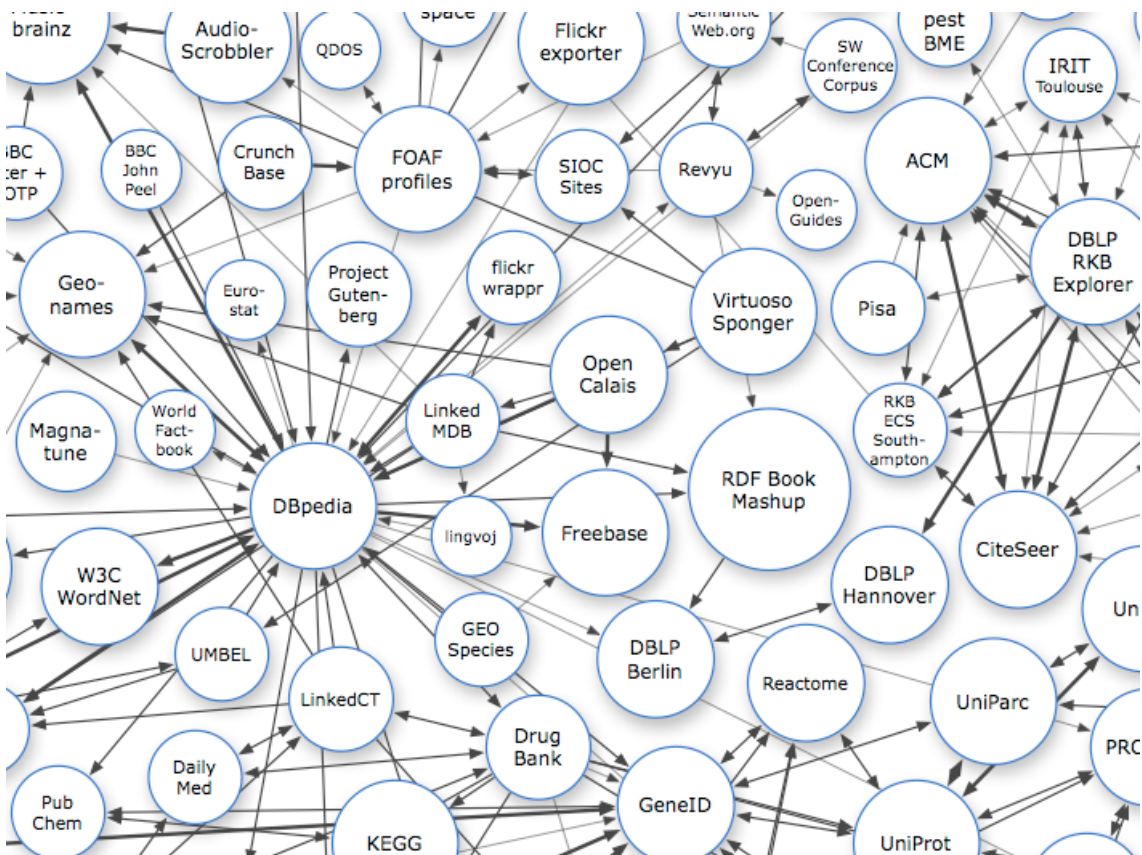


Figura 1: Representación del formato linked-data.

Este formato se basa en una premisa sencilla: casi todos los datos referencian a su vez a otros datos. Como una gran base de datos distribuida a lo largo de todo internet, es posible que al definir un dato alguno de sus atributos no sean más que un enlace que apunte a otro objeto previamente definido en la misma red.

Por ejemplo: si se definen los atributos de un libro, muchos de estos atributos como su autor o su editorial no son más que punteros que apuntan a los objetos autor y editorial correspondientes. De esta forma, si se definen varios libros de un autor, no es necesario describir al autor en cada uno de los libros, bastará con

enlazar al objeto que define al autor. Así mismo, si el autor cambia alguna de sus características, como por ejemplo su país de residencia, no será necesario actualizar los objetos libro. Es un concepto similar al del tesoro INSPIRE, pero en vez de referirse únicamente a un conjunto de palabras clave ya conocidas, se puede enlazar con cualquier tipo de objeto.

Este formato linked-data es ya un estándar de facto muy reconocido en el mundo de los datos abiertos y raro es encontrar un catálogo de datos que no lo implemente.

Cuando hablamos de linked-data en metadatos, se suelen utilizar las siglas RDF. RDF es un estándar del W3C para la definición de metadatos y que soporta el lincado necesario para linked-data.

#### **DCAT: El formato para unificarlos a todos (los metadatos)**

Reunidos por una parte desarrolladores y usuarios de INSPIRE y de la otra usuarios y desarrolladores del mundo open data, se hizo patente la necesidad de tener un formato común a ambos mundos. O, al menos, la forma de convertir de uno a otro formato de forma más o menos unívoca para que se pudieran reutilizar los datos open-data en los catálogos INSPIRE y viceversa.

Así es como surge el formato DCAT (Data Catalog Vocabulary), una recomendación del W3C con poco más de un año de vida que define cómo utilizar RDF para compaginarlo con INSPIRE. La idea es bien sencilla: INSPIRE define una serie de tipos de objetos como catálogo, recurso, autor,... etc. Lo único que se necesitaba entonces era una forma de definir en RDF estos mismos objetos. Que es lo que DCAT define.

Este formato DCAT es el que permite que catálogos clásicamente INSPIRE como GeoNetwork hayan incorporado una interfaz RDF que permite hacer búsquedas en los datos (INSPIRE) del catálogo y devolverlos en formato linked-data (DCAT).

La transformación de los datos en sí mismos, sin embargo, se hace de forma diferente. Datos en formato KML, geoJSON, Shapefile o SQLite ya se están pudiendo convertir de forma más o menos aceptable a formatos OGC compatibles con INSPIRE. Al no tener tantas restricciones el dato INSPIRE como el metadato INSPIRE, estas conversiones son mucho más sencillas, dando lugar en muchos casos a que ni siquiera haga falta el proceso de conversión.

Antes hemos mencionado las interfaces RDF incluidas en GeoNetwork como ejemplo de cómo poder convertir del formato más estricto INSPIRE al menos estricto linked-data. Sin embargo, hacer la conversión inversa (de menos a más estricto) no es un ejercicio tan trivial.

En este aspecto, cabe destacar un proyecto fundado por la Comisión Europea que pretende desarrollar una extensión a CKAN, el catálogo de datos abiertos por excelencia, de forma que permita un tratamiento especial de los datos geográficos para poder realizar una conversión a DCAT más directa. Esta extensión en desarrollo llamada Publicamundi y disponible en repositorios abiertos, añade validaciones INSPIRE y simplifica el paso del formato linked-data al formato

INSPIRE. No cubre, sin embargo, el caso de los datos actualmente existentes en un formato laxo y diferente del DCAT, al menos actualmente.

### **CONCLUSIONES**

Es ciertamente posible acercar ambos universos y utilizar tanto datos en los nuevos formatos abiertos como usar INSPIRE. Sin embargo, ambas partes deberán colaborar más estrechamente en el futuro si no queremos que esta brecha se haga cada vez más grande hasta hacerse insalvable.

En cualquier caso, a día de hoy, es ciertamente difícil (por no decir imposible) conseguir una plataforma software que cumpla todos los requisitos de ambos universos: la flexibilidad de un dato que sólo cumpla linked-data (y no DCAT) y la rigurosidad de validar dicho dato en INSPIRE.

### **AGRADECIMIENTOS**

Me gustaría agradecer a mis compañeros de desarrollo de GeoNetwork, a los impulsores y desarrolladores de CKAN y Publicamundi y tantos otros que hacen que la compartición de datos geográficos sea cada día más fácil.

Mención especial para los compañeros de pyCSW, que recientemente han superado el proceso de incubación y ya forman parte de la familia OSGEO.