

26/10/2017

## Linaje geoespacial en entornos distribuidos



En el ámbito de las Tecnologías de Información Geoespacial (TIG) se trabaja con grandes cantidades de datos procedentes de diferentes orígenes y que se procesan en entornos distribuidos, en la nube. Investigadores de la UAB proponen nuevos mecanismos para describir y retener, en este tipo de entornos, el linaje de los datos, es decir, los metadatos que describen sus fuentes y los procesos que dan lugar a cada capa de información.

La irrupción de la web, supuso una revolución en el acceso y el intercambio de datos y servicios en la mayoría de campos científicos. En el ámbito de las Tecnologías de Información Geoespacial (TIG), este hecho ha comportado la posibilidad de usar y trabajar con datos y herramientas de geoprocесamiento en entornos distribuidos ubicados en la nube. Este cambio de paradigma ha contribuido a aumentar considerablemente los estudios que combinan datos procedentes de diferentes disciplinas y orígenes, posibilitando la generación de nuevos productos o servicios cartográficos, como por ejemplo describir un aspecto de los posibles impactos del cambio climático en la sociedad. A pesar de los beneficios, también se ha incrementado la dificultad de uso y de interpretación de los datos debido a la falta de información auxiliar sobre el contexto de generación de los propios datos (lo que llamamos metadatos).

Así pues, necesitamos un medio para transportar los metadatos y disponer de información sobre los autores, la escala, las unidades, el sistema de proyección, el formato y la calidad de los datos y de los servicios geoespaciales. Además, en un contexto científico caracterizado por el rigor, determinar la calidad y la veracidad de los datos es capital. En este sentido, disponer de información de linaje en los metadatos geoespaciales puede ayudar a discernir la idoneidad de uso de determinados productos geográficos en estudios científicos concretos. El linaje es definido como aquella parte de los metadatos que describe las fuentes y los procesos utilizados en la elaboración de un producto geoespacial. A pesar de la reconocida importancia y relevancia del linaje geoespacial, su presencia en los metadatos geoespaciales es en general todavía escasa, y en muchos casos fragmentada, en parte debido a la falta de herramientas para retener esta información.

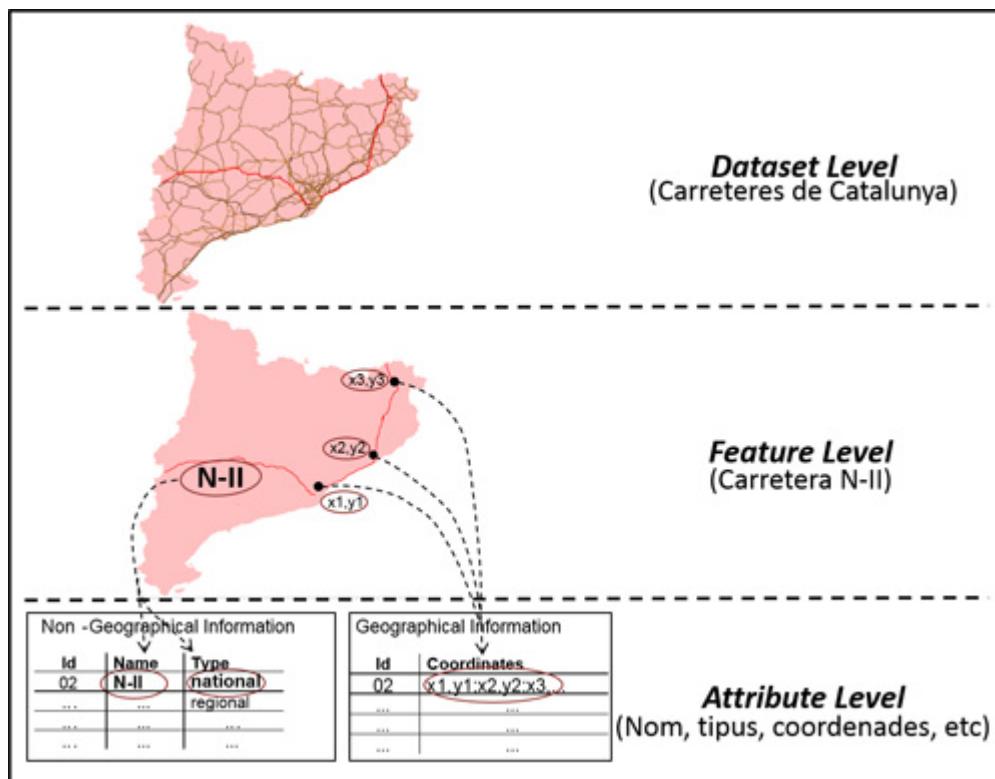


Figura 1. Esquema conceptual representando los tres niveles de granularidad de la información geoespacial.

El objetivo del trabajo presentado es el de proponer nuevos mecanismos para describir y retener el linaje geoespacial en un paradigma distribuido. Para ello, se ha usado el modelo propuesto por uno de los organismos que velan por el buen funcionamiento en la web: el W3C, pero se ha adaptado para que sea aplicable a las particularidades de la información geoespacial. La estructura modular del W3C PROV y la capacidad de definición de las relaciones entre diferentes elementos, permite describir el apellido geoespacial a los diferentes niveles de granularidad de la información geoespacial (Figura 1). Así pues, podemos describir las fuentes de datos y los procesos que han dado lugar a una capa de información de manera general (por ejemplo, todas las carreteras de Cataluña), pero también, elemento a elemento (y así saber el origen y el día de actualización de la descripción de la carretera N-II). Esta descripción puede ser exportada para poder ser usada en la Infraestructura de Datos Espaciales (IDE) Europea

(INSPIRE) pero también se puede integrar en el modelo semántico global de los datos enlazados de la web (linked data), incrementando la interoperabilidad e interconexión entre diferentes modelos científicos.

**Guillem Closa**

[Guillem.closa@uab.cat](mailto:Guillem.closa@uab.cat)

Grumets Research Group, Departamento de Geografía

Universitat Autònoma de Barcelona

**Joan Masó**

[Joan.maso@uab.cat](mailto:Joan.maso@uab.cat)

Grumets Research Group, CREAF

Universitat Autònoma de Barcelona

**Referencias**

Closa, G., Masó, J., Proß, B., Pons, X. **W3C PROV to describe provenance at the dataset, feature and attribute levels in a distributed environment.** *Computers, Environment and Urban Systems Volume 64*, July 2017, Pages 103-117.  
<https://doi.org/10.1016/j.compenvurbsys.2017.01.008>

[View low-bandwidth version](#)