

Reutilización de los conjuntos de datos geoespaciales: una cuestión de linaje

Uso del WPS para capturar el linaje geoespacial

Closa, Guillem; Masó, Joan; Zabala, Alaitz; Pesquer, Lluís;

El linaje es la parte de los metadatos que proporciona información sobre las fuentes y los procesos utilizados para generar un producto geográfico determinado. El interés en la información de linaje por parte de todos los niveles administrativos radica en su utilidad para determinar la calidad de un conjunto de datos, su usabilidad y su trazabilidad hacia las fuentes originales. Además, en un entorno distribuido, donde los datos geoespaciales y herramientas de geoprocésamiento están disponibles como servicios, la información de linaje es aún más valiosa para verificar la fiabilidad de los datos y puede tener un papel importante en la determinación de su validez legal.

Aunque la importancia de la información del linaje está más que documentada, los actuales estándares de información geográfica no proporcionan una completa descripción del linaje que asegure la reproducibilidad de los conjuntos de datos geoespaciales. Tradicionalmente, la comunidad geoespacial ha usado las normas ISO 19115 e ISO 19115-2 para describir el linaje. Sin embargo, existen algunas lagunas en el modelo, como la incompleta descripción de los parámetros, que impiden una reproducibilidad de los conjuntos de datos. Además, la escasez de herramientas que capturen automáticamente la información de linaje es otro de los obstáculos que impiden la generalización y normalización de su uso.

En este sentido, hemos desarrollado una herramienta que captura y representa la información del linaje basada en el uso combinado del estándar *Web Processing Service (WPS)* con el modelo ISO de linaje. La herramienta, desarrollada en el marco del programa de SIG y Teledetección MiraMon, presenta una visualización gráfica del linaje y permite a los usuarios editar información agregando o eliminando los procesos, parámetros de entrada y de salida, o fuentes en cadenas de procesos geoespaciales. Además, permite la reproducción de los procesos anteriores y extender su aplicación a situaciones similares.

La tarea principal ha consistido en la descripción de todas las aplicaciones del MiraMon (MSA), sus entradas, salidas y descripción del proceso, con documentos de *WPS DescribeProcess response*. Los documentos DescribeProcess response utilizan su propio lenguaje XML para describir los parámetros exactos y valores involucrados en una ejecución (pueden ser números, conjuntos de datos y referencias a servicios de datos o el envolvente). Esto permite al sistema capturar la información del linaje cuando se ejecuta una MSA y almacenarla en el archivo de metadatos como una parte de la información de calidad y del linaje. Finalmente, el linaje se puede visualizar y editar en la interfaz del GeMM (Gestor de Metadatos y Relaciones de Base de Datos del MiraMon). La herramienta mantiene las dependencias hacia los conjuntos de datos de origen y a sus respectivos archivos de metadatos.

La descripción de los módulos de MiraMon como documentos de *WPS DescribeProcess* nos ha permitido también generar automáticamente los documentos de ayuda del programa, y significa el punto de partida para desarrollar un servicio WPS de todos los módulos del MiraMon en un futuro próximo.

PALABRAS CLAVE

IDE, LINAJE, TRAZABILIDAD, METADATOS, WPS

AUTORES

Guillem Closa

g.closa@creaf.uab.cat

Grumets Research Group,
CREAF. Edifici C. Universitat
Autònoma de Barcelona.
08193 Bellaterra, Catalonia,
Spain

Joan Masó

joan.maso@uab.cat

Grumets Research Group,
CREAF. Edifici C. Universitat
Autònoma de Barcelona.
08193 Bellaterra, Catalonia,
Spain

Alaitz Zabala

alaitz.zabala@uab.cat

Grumets Research Group, Dep
Geografia. Edifici B.
Universitat Autònoma de
Barcelona.
08193 Bellaterra, Catalonia,
Spain

Lluís Pesquer

l.pesquer@creaf.uab.cat

Grumets Research Group,
CREAF, Edifici C, Universitat
Autònoma de Barcelona
08193 Bellaterra, Catalonia,
Spain