Using WPS to describe geospatial provenance

1 Guillem Closa, 1 Joan Masó, 2Alaitz Zabala, 1 Lluís Pesquer
1 Grumets Research Group, CREAT
2 Grumets Research Group, Dep Geografia
GEOSPATIAL PROVENANCE

- Is part of the metadata that provides information about SOURCES and PROCESSES used to generate a specific geographical product.
**PROCESSES**

Such elements can be referenced using:

- The name of the operation, or a full description of the operation.
- Valid URI of the operation.
- Referencies to an engine used.
- Referencies to an algorithm used.
GEOSPATIAL PROVENANCE

Source

fileIdentifier:MDE_CONGOST_2

ProcessStep:DENSRAS

Source:MDE_CONGOST_5

Source:PARAM=2

ProcessStep:RETALLA

Source:MDE_CAT

Source:POL_CONGOST

SOURCES

Such elements can be referenced using:

- A descriptive citation.
- An element id and a metadata id.
- An element URI or a metadata URI.
- Constant or variable element that modified the behaviour of the algorithm.
PROVENANCE UTILITY:

✓ Data quality assessment.
✓ Data attribution and geospatial data trustworthiness.
✓ Help reproducibility information and replication processes.
✓ Data usability assessment.

OBSTACLES FOR USING PROVENANCE

✗ The lack of provenance standards which captures the complete workflow.
✗ The lack of automated tools for capturing provenance information.
Developed in the context of the MiraMon GIS and RS software.

Based on the combined use of **Web Processing Service (WPS)** with the **LI_Lineage** of **ISO 19115-1** and **LE_ProcesStep** of **ISO 19115-2**.

Use the WPS **DescribeProcess** documents to capture Lineage.

- The DescribeProcess documents can be used in a local execution (No web).

Captures and represents automatically provenance information.

Allows to edit provenance information manually using the **GE MM** (MiraMon Metadata and Database manager) interface.

Provenance information is stored as a part of quality metadata in a .rel format (MiraMon Metadata format).
ISO + WPS COMBINATION
➢ WPS to describe Provenance

- To capture provenance information all the MiraMon modules (MSA) have been documented with WPS DescribeProcess documents.

<table>
<thead>
<tr>
<th>Help documents of the MiraMon modules</th>
<th>TRANSFORMED IN:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WPS DescribeProcess documents</td>
</tr>
<tr>
<td></td>
<td>ProcessDescription</td>
</tr>
<tr>
<td></td>
<td>Inputs</td>
</tr>
<tr>
<td></td>
<td>Outputs</td>
</tr>
<tr>
<td></td>
<td>Module syntax</td>
</tr>
</tbody>
</table>
WPSTO DESCRIBE PROVENANCE

fileldentifier:MDE_CONGOST_2

ProcessStep:DENSRAS

Source:MDE_CONGOST_5 Source:PARAM=2

ProcessStep:RETTALLA

Source:MDE_CAT Source:POL_CONGOST
WPS TO DESCRIBE PROVENANCE

fileIdentifier:MDE_CONGOST_2

ProcessStep:DENSRAS

Source:MDE_CONGOST_5  Source:PARAM=2

ProcessStep:RETALLA

Source:MDE_CAT  Source:POL_CONGOST
PROVENANCE CAPTURATION

WPS DescribeProcess documents

Uses

Is Generated by

Provenance Engine

Part of

Lineage Information

Can be visualized with

GeMM Interface

Can be edited and modified with

Metadata Archives (.rel format)

Part of:

Dependencies

Part of:

Outputs (raster, vector, etc)

Part of:

MiraMon Architecture

MiraMon Module Engine

GeMM Metadata Engine
1. Name, and metadata file.
2. Main menu.
3. Lineage tab.
4. Lineage tree.
5. Description and editing process chain, individual processes, inputs and outputs.
1. Identifier and process description.
2. Organizations.
3. Dates.
1. Add and remove sources, change order
2. Identifier and purpose
3. Type of value, units, value.
4. Location of the source
5. Input, Output, Input-Output
PROVENANCE REPRESENTATION
CONCLUSIONS

- Geospatial provenance facilitates geospatial data evaluation for reuse, and brings us closer to the replication of process chains.

- The combination of WPS DescribeProcess documents with ISO model provides a more complete provenance description.

- Lineage tool allows to capture provenance information automatically and edit it manually.

- This process is also used to generate automatically parts of the modules help and develop a WPS service prototype of the MiraMon modules (MSA).
Thanks!

g.closa@creaf.uab.cat