

RESTFUL IMPLEMENTATION OF CATALOGUE SERVICE FOR GEOSPATIAL DATA PROVENANCE

Liangcun Jiang, Peng Yue, Xianchang Lu

State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129
Luoyu Road, Wuhan, 430079, China -
Email: pyue@whu.edu.cn

KEY WORDS: REST, Catalogue Service, Data Provenance, Geospatial Service, RESTful Web Service

ABSTRACT:

Provenance, also known as lineage, is important in understanding the derivation history of data products. Geospatial data provenance helps data consumers to evaluate the quality and reliability of geospatial data. In a service-oriented environment, where data are often consumed or produced by distributed services, provenance could be managed by following the same service-oriented paradigm. The Open Geospatial Consortium (OGC) Catalogue Service for the Web (CSW) is used for the registration and query of geospatial data provenance by extending ebXML Registry Information Model (eBRIM). Recent advance of the REpresentational State Transfer (REST) paradigm has shown great promise for the easy integration of distributed resources. RESTful Web Service aims to provide a standard way for Web clients to communicate with servers based on REST principles. The existing approach for provenance catalogue service could be improved by adopting the RESTful design. This paper presents the design and implementation of a catalogue service for geospatial data provenance following RESTful architecture style. A middleware named REST Converter is added on the top of the legacy catalogue service to support a RESTful style interface. The REST Converter is composed of a resource request dispatcher and six resource handlers. A prototype service is developed to demonstrate the applicability of the approach.

1. INTRODUCTION

With the advancement of Web Service technologies, Geographic Information System (GIS) has evolved into Geographic Information Service (GIService). In a GIService environment, where data are often consumed or produced by distributed services, geospatial data provenance becomes important since it helps data providers and consumers to evaluate the quality and reliability of geospatial data. Consequently, new challenges have emerged, including sharing and integrated analysis of geospatial data provenance in a service-oriented environment. The Open Geospatial Consortium (OGC) Catalogue Service for the Web (CSW) specifies the interface to query and register descriptive information (metadata) about geospatial data, services and related resources. It can be used for the registration and query of geospatial data provenance by extending ebXML Registry Information Model (eBRIM) (Yue et al., 2011).

Recent advance of the REpresentational State Transfer (REST) paradigm has shown great promise for the easy integration of distributed resources. REST is a set of principles and constraints imposed on the communication between clients and servers in designing distributed systems and applications (Fielding, 2000). RESTful Web Service aims to provide a standard way for Web clients to communicate with servers based on REST principles. The existing approach for provenance catalogue service could be improved by adopting the RESTful design.

This paper presents the design and implementation of a catalogue service for geospatial data provenance following RESTful architecture style. In the study, the PROV-DM model developed by World Wide Web Consortium (W3C) is extended to represent geospatial data provenance. It is combined with a catalogue service to provide the storage, management, and share of geospatial data provenance. According to the extended model, five kinds of resource entities are defined, namely, Data resource, Service resource, Provenance resource, DataList resource, and ServiceList resource. According to REST principles, every resource is assigned with a unique Uniform Resource Identifier (URI). These identified resources can be retrieved via HTTP GET on their URIs. A middleware named REST Converter is added on the top of the legacy catalogue service to support a RESTful style interface. The REST Converter is composed of a resource request dispatcher and six resource handlers. A prototype service is developed to demonstrate the applicability of the approach.

2. RELATED WORK

2.1 Geospatial Data Provenance

Some of the earliest research on geospatial data provenance can be traced back to early 1990s, when lineage of derived products in GIS is characterized as information that describes materials and transformations applied to derive the data (Lanter, 1991). As data provenance becomes increasingly important in

geospatial domain, some scholars give an overview of concepts, technologies, and methodologies related to geospatial data provenance, and refer it as the derivation history (lineage) of a geospatial data product (Yue & He, 2009; Yue et al., 2010; Di & Yue, 2013).

To achieve the efficient management and share of geospatial provenance information, an appropriate and standard-based information model to present provenance is needed. Existing mainstream provenance models include the Open Provenance Model (OPM) (Moreau et al., 2008), W3C PROV Data Model (PROV-DM) (Moreau & Missier, 2013), and ISO 19115 lineage model (2003). The OPM, a generic provenance model, is mainly used in scientific workflow domain. The PROV-DM, developed on the basis of OPM, is more universal and extensible, and has become a W3C recommendation. Although the ISO 19115 is suitable for most cases of geospatial data provenance, it does not readily support the automatic processing of provenance information. In this study, we adopt PROV-DM as our base model for broad interoperability. Then we extend the base model by mapping ISO 19115 model to PROV-DM and make some adjustments to fit the geospatial service context. Finally, extensions of the eBRIM model are made to enable the registration of geospatial data provenance.

2.2 Representational State Transfer

REST is a hybrid style derived from several network-based architectural styles and defines a uniform connector interface (Fielding, 2000). As Roy Fielding is one of the principal contributors of the Web protocols, REST can be considered as a post-hoc description of the features of the Web that made the Web successful. It should be understood that REST is not a standard, but an architectural style. The REST describes six constraints applied to the architecture, namely client-server, stateless, cacheable, layered system, code on demand (optional), and uniform interface. Any applications conforming to these constraints can be characterized as “RESTful”. Fundamental principles in the design of REST interface can be concluded as follows:

- Identification of resources.
- Manipulation of resources through representations.
- Self-descriptive messages.
- Hypermedia as the engine of application state.

Resource, referenced with a unique URI, is the central idea of REST. Only limited operations are used to manipulate the resources: HTTP GET to retrieve resources; HTTP PUT to create resources; HTTP POST to update resources; HTTP DELETE to remove resources. Resources are manipulated through their representations which can be more than one format (HTML, XML, etc.). Relationships between resources are expressed by hyperlinks, through which state of the representation is transferred.

2.3 RESTful Web Service

A RESTful Web Service is a Web API developed based on HTTP and REST principles. Because of its scalability and simplicity, OGC forms a RESTful Service Policy Standards Working Group (SWG) in September 2011. This SWG aims to develop a policy standard for the structure and content of RESTful approach to the implementation of geospatial services. Candidate OGC Services to be considered includes WMS, WFS, WCS, SOS, SPS, WPS, CSW, WMTS, and profiles thereof (OGC, 2011).

Since REST principles have values in simplifying and unifying access to services, some related work has been done. Existing geospatial services re-designed in RESTful style include WCS (Mazzetti et al., 2009), WMTS (Mas ó et al., 2010), SOS (Page, 2009; Janowicz et al., 2013), and WPS (Foerster, 2011). Besides, some resources on RESTful geospatial services are available, including ESRI’s ArcGIS Server REST API, PostGIS RESTful Web Service, GeoREST (2010), and FeatureServer (2012).

3. APPROACH

In order to build a RESTful Catalogue Service for geospatial data provenance, a middleware called REST Converter is added on top of the legacy catalogue service. By this means, the legacy interface of catalogue service is reserved to interoperate with existing Web Services. Meanwhile, the rest principle is supported. The design of a RESTful Catalogue Service for geospatial data provenance is based on a 3-tier Web architecture (Figure 1).

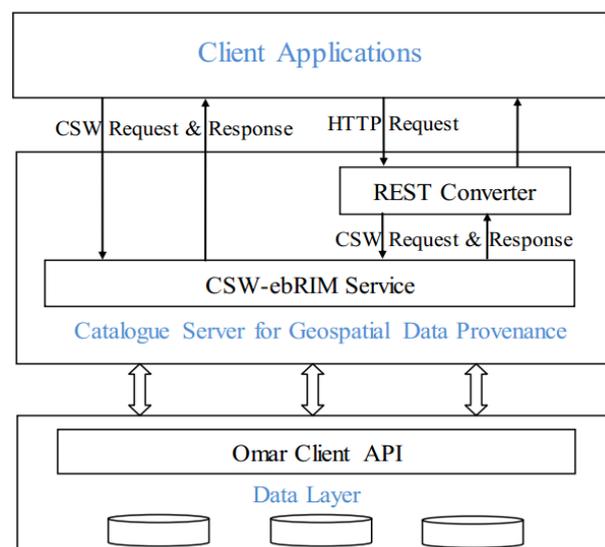


Figure 1. Architecture of RESTful geospatial provenance service

Clients include both thin clients (e.g. Browsers) and thick clients (e.g. Web Services Client Applications). The client layer can interact with the server layer using HTTP GET method or CSW requests. The server layer mainly contains a REST Converter and a CSW module. The REST converter converts RESTful requests into CSW requests and submits them to the CSW

module that can work as a single application to provide catalogue service. In the data layer, Omar, an open source software program, is used to implement the ebXML registration. It provides a client API that interacts with the database.

In the development of the RESTful Catalogue Service for geospatial data provenance, two key issues need to be addressed. Since resource is the most essential element in RESTful Web Services, one key issue is how to identify all resources with global URIs. The other issue is to comply with REST principles while at the same time reusing legacy catalogue services.

3.1 Identifying resources

In the implementation of RESTful Web Services, it is a crucial step to identify all required resources. With a particular focus on provenance entities including data and services used in deriving geospatial data products, we define five types of resources, namely Data, Service, Provenance, DataList, and ServiceList. In the provenance, a data product could have descendants or ancestors, which are represented using Inputs and Outputs resources. Then URIs are assigned to these resources as follows:

Data resource: `/geoprov/rest/datalist/{dataId}`

Service resource: `/geoprov/rest /servicelist/{serviceId}`

Provenance resource:

`/geoprov/rest/datalist/{dataId}/geoprov?outputschema={output schema}`

Inputs: `/geoprov/rest /datalist/{dataId}/inputs`

Outputs: `/geoprov/rest /datalist/{dataId}/outputs`

DataList:

`/geoprov/rest/datalist?name={ name }&key word={ key word }&bbox={ bbox }`

ServiceList:

`/geoprov/rest/servicelist?name={ name }&key word={ key word }&type={ type }`

For Provenance resource, parameter is provided for users to choose the encoding mode. The default parameter is “w3c”, and it also can be “ebrim” or “iso”. For DataList and ServiceList resource, parameters such as resource name, key word information for describing resource and bounding box can be appended to a URI to act as filters for querying.

As resources are always returned through a representation in a RESTful Web Service, their encodings have to be defined. Resources can be encoded in XML, JSON, or other formats. Since the extended PROV-DM model is represented in XML in this study, we provide resources with XML representations. Figure 2 shows an example of XML representation of geospatial data resource. The XML representation contains links to related resources following which state transfer can be achieved.

```
<?xml version="1.0" encoding="UTF-8"?>
<rest:document xmlns:rest="http://localhost:8081/ns/rest"
  xmlns="http://localhost:8081/ns/rest">
  <rest:GeoData id="urn:uuid:c96b4e97-0494-4d6e-a70e-cdecf40a02c4">
    <rest:type>prov:Entity</rest:type>
    <rest:name>output data</rest:name>
    <rest:description>this is a test data that used to verify...</rest:description>
    <rest:keywordList>
      <rest:keyword>test</rest:keyword>
      <rest:keyword>conflation</rest:keyword>
      <rest:keyword>reference layer</rest:keyword>
    </rest:keywordList>
    <rest:Envelope>
      <rest:lowerCorner>180 -90</rest:lowerCorner>
      <rest:upperCorner>180 90</rest:upperCorner>
    </rest:Envelope>
    <rest:publisher>Wuhan University</rest:publisher>
    <rest:format>gml</rest:format>
    <rest:datauri>http://localhost:8081/geoprov/data/testdata3.gml</rest:datauri>
    <rest:geoprov>http://localhost:8081/rest/geoprov/urn:uuid:c96b4e97-0494-4d6e-a70e-cdecf40a02c4/geoprov</rest:geoprov>
  </rest:GeoData>
</rest:document>
```

Figure 2. Example of a Data resource represented in XML

3.2 Designing a REST Converter

The core module of a RESTful Catalogue Service for geospatial provenance is the REST Converter. The REST Converter extracts the user’s query from the URI, encodes it as a valid CSW request, sends the request to CSW interface, and converts the response into an XML representation for resources.

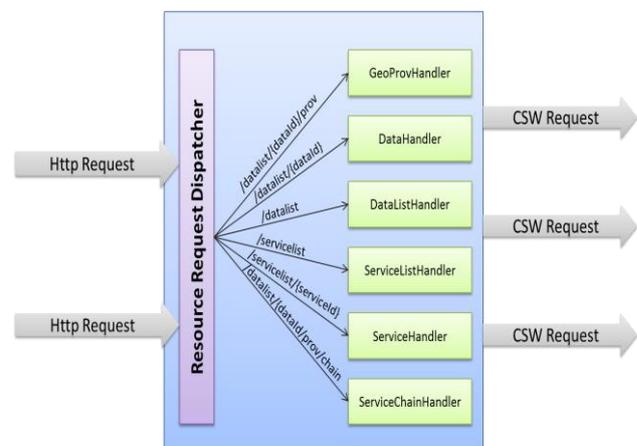


Figure 3. Mechanism of the REST Converter

As illustrated in Figure 3, the REST Converter consists of a resource request dispatcher and six resource handlers. The dispatcher is responsible for parsing requested URI and forwarding HTTP requests to corresponding handlers. The handler is responsible for parsing a HTTP request, and converting it into CSW requests. Besides, the responses from Catalogue Service are converted into XML representations by the handlers.

In the traditional CSW approach, users need to be familiar with the CSW-ebRIM model, and construct corresponding CSW queries. For RESTful requests, all users need to know is the resource's URI and corresponding HTTP method.

5. CONCLUSIONS AND FUTURE WORK

This paper presents design and implementation of a catalogue service for geospatial data provenance following RESTful architecture style. It illustrates how REST principles can be applied into a catalogue service to facilitate the management and share of geospatial data provenance. RESTful approach to the design and standardization of geospatial services will lead to both improved interoperability and greater accessibility for these services. The results demonstrate the applicability of the RESTful Catalogue Service for managing geospatial data provenance. It also paves the way for the investigation of how REST architecture style could be used in geospatial service development.

Future work will include implementing more CSW optional operations in the REST Converter, and improving the integration and interoperation of geospatial data provenance with other geospatial service.

ACKNOWLEDGEMENTS

The work is supported by National Basic Research Program of China (2011CB707105), National Natural Science Foundation of China (41271397), and LIESMARS Special Research Funding.

REFERENCES

Di, L., Yue, P., Ramapriyan, H. K., & King, R. L., 2013. Geoscience Data Provenance: An Overview. *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 99, pp. 1-8.

FeatureServer, 2012. <http://featureserver.org/> (Accessed 18 Aug. 2013).

Fielding, R. T., 2000. Architectural styles and the design of network-based software architectures. UNIVERSITY OF CALIFORNIA, IRVINE. <http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm> (Accessed 18 Aug. 2013)

Foerster, T., Brühl, A., & Schäffer, B., 2011. RESTful web processing service. In *Proceedings 14th AGILE International Conference on Geographic Information Science, Utrecht, Netherlands* (p. 8).

GeoREST, 2010. <http://code.google.com/p/georest/> (Accessed 18 Aug. 2013).

ISO 19115:2003. Geographic Information – Metadata[EB/OL]. http://www.iso.org/iso/catalogue_detail?csnumber=26020 (Accessed 18 Aug. 2013).

Janowicz, K., Bröring, A., Stasch, C., Schade, S., Everding, T., & Llaves, A., 2013. A restful proxy and data model for linked sensor data. *International Journal of Digital Earth*, 6(3), pp. 233-254.

Lanter, D. P., 1991. Design of a lineage-based meta-data base for GIS. *Cartography and Geographic Information Systems*, 18(4), pp. 255-261.

Masó J., Pomakis, K., & Julià N. OpenGIS Web Map Tile Service Implementation Standard| OGC®. *Documento online*. Disponible en: <http://www.opengeospatial.org/standards/wmts> (Accessed 18 Aug. 2013).

Mazzetti, P., Nativi, S., & Caron, J., 2009. RESTful implementation of geospatial services for Earth and Space Science applications. *International Journal of Digital Earth*, 2(S1), pp. 40-61.

Moreau, L., Freire, J., Futrelle, J., McGrath, R. E., Myers, J., & Paulson, P., 2008. The open provenance model: An overview. In *Provenance and Annotation of Data and Processes*, Springer Berlin Heidelberg, pp. 323-326.

Moreau, L. & Missier P., 2013. "PROV-DM: The PROV Data Model". <http://www.w3.org/TR/prov-dm/> (Accessed 18 Aug. 2013).

OGC, 2011. RESTful Services Policy SWG. <http://www.opengeospatial.org/projects/groups/restfulswg> (Accessed 18 Aug. 2013)

Page, K. R., De Roure, D. C., Martinez, K., Sadler, J. D., & Kit, O. Y., 2009. Linked sensor data: Restfully serving rdf and gml. *Proc. Semantic Sensor Networks*, 49.

Yue, P., Gong, J., & Di, L., 2010. Augmenting geospatial data provenance through metadata tracking in geospatial service chaining. *Computers & Geosciences*. 36 (3), pp. 270-281.

Yue, P., & He, L., 2009. Geospatial data provenance in cyberinfrastructure. In: *Proceedings of the 17th international conference on geoinformatics (Geoinformatics 2009)*, 12–14 August 2009 (4pp). Fairfax, USA: IEEE Publication.

Yue, P., Wei, Y., Di, L., He, L., Gong, J., & Zhang, L., 2011. Sharing geospatial provenance in a service-oriented environment. *Computers, Environment and Urban Systems*, 35 (2), pp. 333-343.