

Data processing using Web Processing Service orchestration within a Spatial Data Infrastructure

T. Wehrmann^{a,*}, S. Gebhardt^b, V. Klinger^a, C. Künzer^a

^a DLR, German Remote Sensing Data Center, Oberpfaffenhofen, Germany – (thilo.wehrmann, steffen.gebhardt, verena.klinger, claudia.kuenzer)@dlr.de

^b University of Würzburg, Department of Geography, Würzburg, Germany.

Abstract – Standardised web services are the backbone of the ELVIS system, the *Environmental Visualisation and Information System* supporting Vietnamese stake holders in the fields of water related management processes based on open source technologies. They support distributed and loosely coupled, component based architecture of the system. In cooperating OGC compliant web services for data access, visualisation and data processing the system is extendible to external data resources and other proprietary software solutions. The base idea behind the architecture is the orchestration of decoupled web resources representing data sets and functionality to model more complex business processes. The system covers most aspects of administrative business processes including spatial and non-spatial data ingestion and dissemination, necessary data processing and visualisation techniques. In combination with a semantics enabled data management ELVIS is capable to produce value added information products to water management related tasks.

1. INTRODUCTION

The current developments in geographic information systems (GIS) direct towards resources over the web. Though the demand of transforming distributed data into necessary information products acts as a major driver (Alameh, 2003). Spatial Data Infrastructures (SDI) as successor of desktop based GIS environments are widely accepted to process and distribute large collections of spatially distributed geo data repositories (Beaumont, 2005, Bernard, 2005). The application of distributed spatially enabled data is one of the key elements of such infrastructures leading to new requirements for standards, interfaces and functionality. Accepted standards for web driven data management with processing components were demanded and standard bodies like the Open GIS Consortium (OGC) or the International Organization for Standardization (ISO) reacted in providing simple but powerful service definitions. Independence of file formats and proprietary protocols

1.1 Spatial business processes supported by ELVIS

Several recent projects (German–Vietnamese WISDOM project and Central Asia Project CAWA) require an accurate system to manage, to process and to distribute project results for the targeted stake holders in the participating countries. E.g. the WISDOM system enables end users to perform analyses on very specific questions, such as regional planning activities, and facilitates integration of the dispersed datasets to support decision making processes. As new data is acquired, data is updated or existing data is processed according to the users' needs, the stored data and new data increase in value by

providing new information, as well as transforming the presentation of this information into a suitable manner. The aim of the information system is to provide necessary and additional information for Vietnamese stake holder organisations. Thus existing business processes should be supported directly or indirectly as best as possible to ensure acceptability of the technology. Requirements analysis identified those possible workflows. A workflow is an automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules (Keens, 2007). The analysis mapped several actions to services which can be provided by the system and enables the composition of service chains.

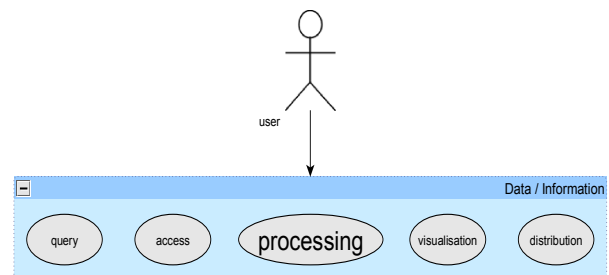


Figure 1. ELVIS use cases for data / information retrieval

The analysed business cases of surveyed stake holder organisations consist of one or more of the following functional groups which affect available data or information (figure 1):

- search for data
- access relevant data
- combine / process data to information
- generate output media (map, report)
- distribute information product

Most of these services can be directly mapped to functions, which in case of an atomic representation of a single process, can be reusable composed to more complex processing chains for different purposes.

The article is structured in the following way. First OGC compliant web services with data centric and processing centric focus are characterised. Following different orchestration types are introduced focusing on the implemented orchestration engine. The article ends with a conclusion based on experiences gained within the project and a short outlook.

* Corresponding author. This is useful to know for communication with the appropriate person in cases with more than one author.

2. WEB SERVICES

Starting point of any SDI is data which are provided by services over the web (Kiehle, 2006). According to W3C the definition of a web service is a software system designed to support interoperable machine-to-machine interactions over a network (Booth, 2004). Service oriented architecture (SOA) and Service Oriented Architecture Protocol (SOAP) address these specifications and provide technical solutions for distributed systems.

Beside technical implementations standards agreed by standard-setting bodies such as ISO TC211, W3C or OGC foster the base for web services. In principle they ensure interoperability of services, exchange of data and policies and ensure reusability of provided services.

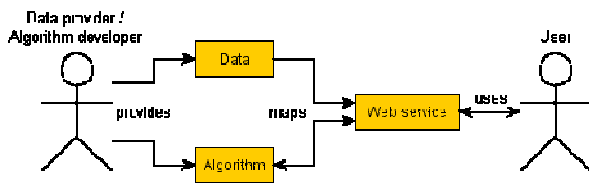


Figure 2. Application of web services for data and process abstraction

Hence, web services encapsulate discrete software based functions providing access using a prescribed interface (OASIS). Fig. 2 shows how services add a level of abstraction to data and algorithms (Kiehle, 2006) and support a transparent interconnection between data producers, algorithm developers and analysts and hence, provide workflow interoperability.

2.1 Data centric services

The data management system within ELVIS employs the PostgreSQL/POSTGIS database for storing and registering spatial data (Gebhardt *et al.*). This object-relational database was chosen over a relational database to tag spatial objects to tabular data, improving the retrieval of census and observational data at regional, provincial, and local areas. The data model also incorporates additional styling aspects of the spatial datasets through Styled Layer Descriptions (SLD) and technical metadata like Web Mapping Service (WMS) layer specifications, allowing retrieval of rendered maps. Metadata elements of the spatial data are based on the ISO19115 standard. XML structured information of the SLD and metadata are stored in an open source XML database.

Data ingestion, respectively registration is performed using Data Entry Portal (DEP) application. It is the interface for an automatic incorporation of spatial data into the data management system. This application comprises analysing data for standard compliance, automatic metadata generation and disseminating the dataset to the data management system.

Interoperability is ensured due to common used standards, e.g. by OGC. Service specification defines the implemented interfaces for all OGC compliant web services using accepted operations and standardised metadata. It enables the automated integration of the used web services und their application.

Selection of necessary standards used within OGC compliant services:

- The Geographic Markup Language (GML) is used to encode spatially enabled features (Lake *et al.* 2004)
- The Catalogue Service for Web (CSW) supports data and service queries (Nebert, 2004).
- The Web Mapping Service (WMS) provides cartographic representations of the data using existing WWW technologies (de La Beaujardiere, 2004).
- Web Coverage Service (WCS) enables the direct access to raster based data (Evans, 2003)
- The Web Feature Service (WFS) distributes vector based data over the internet.

Beside the mentioned standards there exists of course additional service specifications handling transformation, filtering, styling and many more covering all aspects of spatial data application.

2.2 Processing centric services

Since 2007 Web Processing Service (WPS) with version 1.0.0 is proclaimed by Open GIS Consortium. The web service provides a standardised access to spatially distributed GIS functionality over the internet. The interface handles process identification and control. Moreover it defines input and output parameters of the process execution. Web requests are allowed via HTTP GET and POST method, and Simple Object Access Protocol (SOAP) with Web Services Description Language (WSDL) as specification. Similar to other existing OGC compliant web services service communication is based on HTTP and XML dialects.

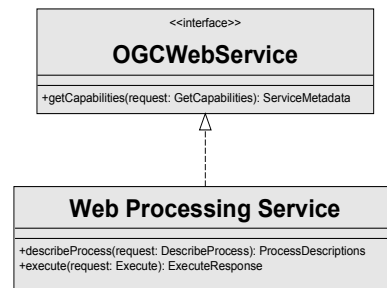


Figure 3. WPS interface UML diagram

According to the OGC reference model (ORM) in fig. 3 it provides three mandatory service operations, including *getCapabilities()*, *describeProcess()* and *execute()*. The *getCapabilities()* operation is inherited by the common OGC web service interface class and describes the service itself, its supported service metadata, operations and processes in an XML document and allows machine based service discovery. The *describeProcess()* interface returns the full description of a single process with its mandatory and optional input and output parameters. The *execute()* operation starts the processing task. Each request can be controlled using this method. The execution status of the running process is accessible in asynchronous mode through additional client polls. The application of the provided interfaces enables easy and machine understandable service chaining and makes (geospatial) computational tasks ubiquitous available for the network.

The specification of the WPS interface is standardised according to OGC specifications. However the communication between the OGC interface and the implemented function is language- and platform dependent tributary to the adopted interface implementation. There are a wide range of open source implementations providing WPS capabilities including 52North WPS (<http://52north.org/wps>), degree WPS from latlon (<http://www.degree.org>) and PyWPS (<http://pywps.wald.intevation.org>) from J. Cepicky. The first two solutions are Java based server applications; the latter one uses Python.

3. ORCHESTRATION OF WEB SERVICES

3.1 Web service orchestration (WSO)

Single atomic web processes which support reusability and flexibility are limited in providing complex functionality. On the other hand complex web services are not flexible enough for multiple applications. To solve this problem web service orchestration seems to be the logical conclusion. Web service orchestration (WSO) is the loosely coupled composition of these provided processes (Veerawarana, 2005). There are several possibilities to arrange web processes, including Business Process Execution Language (BPEL) engines, compositing WPS processes and other workflow mechanisms (Pratt, 2010). The goal of web service orchestration is the specification of collaborative business processes which are based on existing web services.

BPEL is an OASIS standard describing necessary interactions to describe and manage workflows and control web services. Moreover it can be used to specify abstract processes from real world business cases. It is mainly used to composite provided web services which are described by the Web Service Definition Language (WSDL) specification. Both are XML based standards to define process flows, resp. describing underlying web services. There exists a wide range of BPEL supporting orchestration engines for different applications and requirements. Due to compatibility of WPS with SOAP and WSDL, WPS processes can be orchestrated using BPEL (Chen, 2010). In Kiehle, 2006 and Meng, 2010 two examples of BPEL applications for the orchestration of OGC compliant services are presented. The advantage of using BPEL like techniques comprises existing standards and applications to describe composed process chains in both human and machine readable ways.

On the other hand compositing WPS services can follow a centralised service chaining concept or a cascading service chaining concept (Weiser, 2007). The centralised composition process invokes all containing processes and thus controls the entire work flow. The alternative would be a coupling of the services itself, that means the processing services iteratively invoke themselves (nested requests). Both concepts are supported by the WPS interface. Of course a single big WPS process can always consist of multiple sub processes with the drawback of lacking reusability and flexibility.

3.2 Implementation of a compositing WPS orchestration

The proprietary implementation applies the concept of a centralised composition process. The orchestration uses a single process to control the underlying processes, executes the processes in the particular order and assigns the necessary input

parameters for the WPS processes. A rule based workflow definition in a XML derivate describes the processing order of the workflow chain and the necessary chaining of each following processes' input and output parameters. To describe the current order of a workflow iterative and parallel execution modes are supported. A single processing step includes all necessary information to execute the underlying WPS process. In the implementation predefined workflows and dynamic workflows which are specified during invocation request are supported. Hence, the implementation is not fully compliant to the WPS specification because the *describeProcess()* method cannot response properly for a dynamically given process chain.

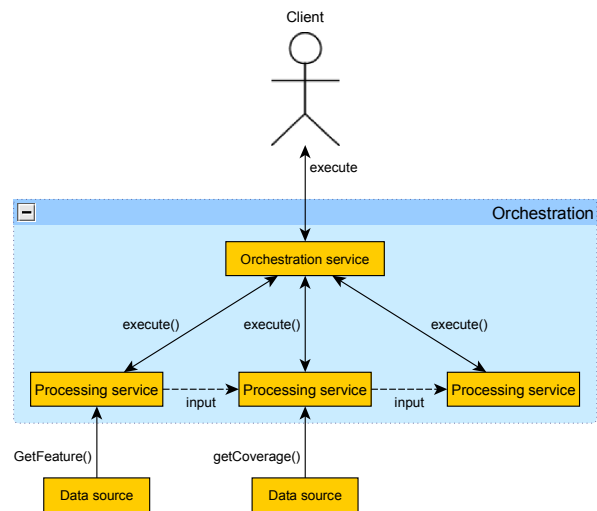


Figure 4. Schematic illustration of service orchestration using a compositing WPS

Fig. 4 refers to the modus operandi of a centralised composition. In contrast to a cascading orchestration chaining concept each addressed WPS process does not have any information about other processes. The orchestration service has this intrinsic information, ensures the valid specification of each service and controls the necessary processes with the required input parameters. Results of certain processes are input variables of following processes. The orchestration service needs to observe current running processes to provide an iterative workflow configuration. Due to missing notification capabilities of the WPS specification this is done using status request polling in certain intervals. The application of a Web Notification Service (WNS) would overcome this problematic solution. The last process should provide archiving or an ingestion of the result because otherwise it would be in responsibility of the requesting client to persistently store the final result. The access to auxiliary data and a cleaning process are tasks of a single service, respectively of the orchestration service.

4. CONCLUSIONS AND OUTLOOK

This paper has described the capabilities and characteristics of the implemented WPS orchestration inside the ELVIS architecture. The application of this OGC introduced specification shows the reusability of spatially enabled processes over web. The easy integration of new functionality increases overall practicability and primary, prepares the base of a Spatial Data Infrastructure's processing capability. With the implementation of atomic processes and the reconfiguration of

those components it is possible to build complex processing chains, even dynamically. Even when WPS specifications are still in progress (like the discussion about profiles according to different viewpoints and the extension with semantically enabled metadata) the potential they provide right now are enormous and their impact in spatially distributed services will change geospatial infrastructure in the future. One of the next steps in processing components within a SDI will be the integration of semantic aspects to enable self-organising services, resp. fully autonomous processing composition.

One the other hand there are still some open issues and drawbacks of the WPS specification like the incapacity of streaming binary data in a BPEL engine controlled environment and the missing interface description between the WPS interface and the algorithm implementation. The simple mechanism of the loosely coupled control of services and the underlying processes lack of standardised methods to halt or stop processes, ensure persistent processing capabilities or provide more sophisticated process flow controls (e.g. conditions, exceptions etc.)

“Processing as a service” derived from the “Everything as a Service” paradigm allows separation between algorithm development by the scientific domain and service application from the service domain and allow the collaboration of scientific work in large scale projects and the introduction of new business models. These features will become more and more important for upcoming demands from project donors and decision makers.

5. ACKNOWLEDGEMENT

This research was mainly undertaken in the framework of the German-Vietnamese research project “WISDOM”, funded by the German Ministry of Education and Research BMBF (FKZ: 0330777) and the Vietnamese Ministry of Science and Technology (MOST).

6. REFERENCES

- Alameh, N. 2003. Chaining geographic information Web services. *IEEE Internet Computation*. 7(5). Pp 22-29.
- Bernard, L., Kanellopoulos, I., Annoni, A., Smits, P. 2005. The European geoportal-one step towards the establishment of a European Spatial Data Infrastructure, *Computer, Environment and Urban Systems*. 29(1). pp 15-31.
- Beaumont, P., Longley, P.A., Maguire, D.J. 2005. Geographic information portals – a UK perspective. *Computers, Environment and Urban Systems*. 29. pp 49-69.
- Booth, D., Hass, H., McCabe, F., Newcomer, E., Champion, M., Ferris, C., Orchard, D. 2004. Web services architecture. W3C working group note. <http://www.w3.org/TR/ws-arch/> (accessed 2.9.2010)
- Chen, N., Di, L., Yu, G., Gong, J. 2010. Geo-processing workflow driver wildfire hot pixel detection under sensor web environment. *Computers & Geosciences*. 36. pp 362-372.
- Foerster, T., Schaeffer, B., Brauner, J., Jirka, S. 2222. Integrating OGC Web Processing Services into Geospatial Mass-Market Applications.
- Foerster, T., Letho, L., Sarjakoski, T., Sarjakoski, L.T., Stoter, J. 2010. Map generalization and schema transformation of geospatial data combined in a Web service context. *Computers, Environment and Urban Systems*, 34(1). Pp. 79-88.
- Gebhardt, S., Wehrmann, T., Klinger, V., Schettler, I., Huth, J., Künzer, C., Dech, S., 2010. Improving data management and dissemination in web based information systems by semantic enrichment of descriptive data aspects. *Computers & Geosciences*. In Press, Corrected Proof.
- Keens, S. (ed.). 2007. OWS-4 WPS IPR. Discussions, findings, and use of WPS in OWS-4.
- Kiehle, C. 2006. Business logic for geoprocessing of distributed geodata. *Computers & Geosciences*. 32. pp 1746-1757.
- Kiehle, C. (ed.), Foerster, T. 2010. OGC OWS-7 Web Processing Service Profiling Engineering Report. OGC Public engineering report.
- Meng, X., Xie, Y., Bian, F. 2010. Distributed geospatial analysis through web processing service: a case study of earthquake disaster assessment. *Journal of Software*, 5(6), pp. 671-679.
- Pebesma, E., Cornford, D., Dubois, G., Heuvelink, G.B.M., Hristopoulos, D., Pilz, J., Stöhlker, U., Morin, G., Skoien, J.O. 2010. INTAMAP: The design and implementation of an interoperable automated interpolation web service. *Computers & Geosciences*. In press.
- Pratt, A., Peters, C., Siddeswara, G., Lee, B., Terhorst, A. 2010. Exposing the Kepler Scientific Workflow System as an OGC Web Processing Service. iEMSs. Ottawa, Canada.
- Schut, P., 2007). OpenGIS Web Processing Service: Open Geospatial Consortium. Wayland, MA, USA.
- Veerawarana, S., Curbera, F., Leymann, F., Storey, T., Ferguason, D.F. 2005. Web services platform architecture: soap, wsdl, ws-policy, ws-addressing, ws-pbel, ws-reliable messaging and more. Prentice Hall.
- Weiser, A., Zipf, A. 2007. Web service orchestration of OGC web services for disaster management. Geomatics solutions for disaster management, Toronto. Canada. pp. 239-254.