Full Meta Objects for flexible geoprocessing workflows: profiling WPS or BPMN?

Julian Rosser
Nottingham Geospatial Institute
University of Nottingham
Nottingham, U.K.
julian.rosser@nottingham.ac.uk

Amir Pourabdollah
School of Computer Science
University of Nottingham
Nottingham, U.K.
amir.pourabdollah@nottingham.ac.uk

Roger Brackin
School of Geography
University of Nottingham
Nottingham, U.K.
lgxrcb@nottingham.ac.uk

Mike Jackson
Nottingham Geospatial Institute
University of Nottingham
Nottingham, U.K.
mike.jackson@nottingham.ac.uk

Didier G. Leibovici
Nottingham Geospatial Institute
University of Nottingham
Nottingham, U.K.
didier.leibovici@nottingham.ac.uk

Abstract

The design and execution of scientific workflows is an important and necessary function within many disciplines. In the geospatial community, workflows are used to generate and quality assure new spatial datasets based on sequences of processing steps or complex modelling. Using BPMN for representing these workflows allows stakeholders to discuss the scientific conceptual approach behind this modelling whilst also being able to execute its encoding in XML. Previous research has focused on developing frameworks, including a BPMN workflow engine, capable of orchestrating OGC Web Processing Services and thus enabling construction of interoperable workflows comprised of distributed resources. These could populate, for example, the GEOSS repository. However, to date, such work has focused on executing workflows with direct access to the pre-defined data inputs and outputs, with a lack of flexibility and efficiency in semantic interoperability or data management during composition and execution. This article suggests a meta-approach based on two possible configurations that enable workflow orchestration at a meta level using a direct coupling with a metadata catalogue. The designs of a Web Processing Service profile and of a BPMN profile are presented as potential approaches to abstract the data interchange between the processing steps of a workflow. The paper concludes with a discussion of how these approaches may be extended and how complete workflows might be registered and managed using catalogues services.

1 Introduction

Use of interoperable standards enables seamless sharing of data and processing between different systems. Following a flow of tasks, chaining of services in an automated way or at least though a manual step by step execution is common practice for data producers or researchers using a GIS desktop environment. In geospatial processing, workflows may be composed of many different tasks (for the processing steps) requiring selection and configuration of parameters as data inputs. For example, a processing workflow for modelling the land and hydrographic characteristics of a region might use a sequence of analysis steps based on digital elevation, samples of rainfall, vegetation and soil data, see Figure 1. The use of a workflow representation to chain web services of these models together helps manage this process and also aids in documentation and reuse of a processing sequence.

However, despite the benefits in the adoption of a standard such as Business Process Modelling Notation (BPMN)[1], with a workflow engine consuming the BPMN2.0 XML, significant effort is required in the customisation of the engine for application with Open Geospatial Consortium (OGC) services [8]. Furthermore, on-going effort is required in order to document the datasets produced, while interim results generated part way through a sequence may be lost following a workflow's completion. Moreover, if a workflow is executed many times with the same data sources, this becomes inefficient.

This paper proposes two solutions in order to minimise this redundancy, facilitate architecture implementation, alleviate computation cost, increase the use and generation of metadata for composition support, and to aid data discovery. The two solutions are defined as profiles of the BPMN and WPS standards with both approaches operating as coupling of the workflow service (editor and engine) and a local metadata catalogue. For both solutions the workflow is instantiated using metadata links in order to resolve the syntactic and semantic binding at late as possible in the orchestration (BPMN profile) or in the execution itself (WPS profiling).

2 Background

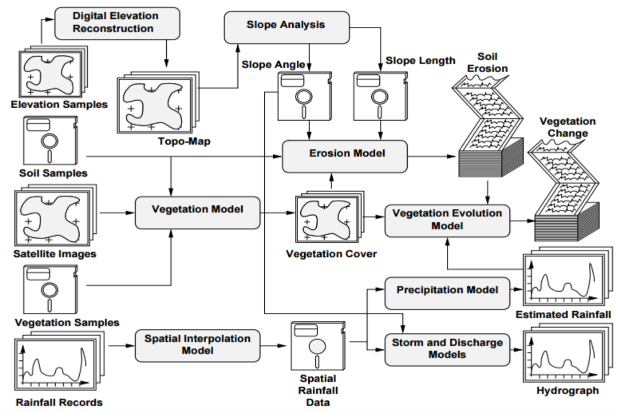
Execution of workflows and chaining of processing tasks is undertaken in many areas of geospatial science. Simple chaining of tasks is available in systems such as QGIS processing modeller and ArcGIS ModelBuilder. These enable creation of geoprocessing workflows based on a library of operators provided by each system. The OGC Web Processing Service (WPS) offers a method for exposing processing tasks according to a standardised interface. Several of these processing tasks may be chained to form a complete workflow. Chaining of Web Processing Services may be

[1] www.bpmn.org

achieved by various methods. The WPS standard itself identifies three ways to chain services [11]: 1) to use a BPEL engine to define and execute the workflow (such as [6,18]); 2) wrap a sequence of WPS calls within another WPS (such as [2,4]); 3) encode a chain of services within the execute query to form a cascading request.

A recent alternative to these approaches is to use BPMN to define and execute the workflow [8]. BPMN enables definition of workflows as XML data which helps document the precise sequence of tasks, data inputs and outputs that are generated. This capture of the workflow enables repeated execution of common processing tasks and simple redistribution of template patterns for rapid reuse as well as documenting the provenance within a metadata record.

Figure 1: Example geo-processing workflow for modelling land and hydrographic characteristics of a region [1].



The generation and management of metadata regarding data and processes are recognised as important tasks in facilitating effective spatial information management. Catalogue services provide a mechanism for achieving this allowing indexing and search of data which may be distributed across different locations [9]. For example, the satellite imagery, rainfall and soil samples, etc. referred to in Figure 1 may be registered in such a repository. Similarly, the processes and models used in Figure 1 might be registered, e.g. slope analysis would have a formal definition of the algorithm employed and the unit of measure. The OGC Catalogue Service for the Web (CSW) specification defines a standard for implementing these services [12]. Methods to aid production of metadata are being developed which automate the generation of ISO standardized elements [5,17]. Meanwhile, a number of workflow suites (workflow editor and workflow engine) are available including some that are open source, such as JBPM[2]. However, the support for composition by discovering the elements of a workflow and registering their outputs is limited. To our knowledge there is no solution integrated within a workflow environment for managing metadata.

The OGC WPS standard specifies application profiles for the purpose of defining a domain of application and to facilitate semantic interoperability. A profile may characterise a number of constraints such as the types of input and consequently the way these input types are handled within the process which are then common to all processes following this profile. BPMN - which allows full specification of the workflow - permits creation of a profile which may also be imposed at the implementation level.

3 Conceptual architecture

In this section we describe two possible approaches for handling workflow construction using metadata objects. In both cases, we propose to use the standards: BPMN, OGC CSW and OGC WPS. Even though our proposals are independent of the implementations of these standards we adopt the JBPM workflow environment, GeoNetwork CSW[3] and 52North WPS[4] for developing our concept. Currently, GeoNetwork implements the OGC-CSW 2.0.2 ISO Profile which enables cataloguing of metadata on datasets and services according to ISO19115 and ISO19119 standards [13]. The 52North WPS implements version 1 of the OGC standard [11]. The JBPM environment implements version 2 of the BPMN standard [16].
The OGC Web Processing Service definition specifies that data maybe exchanged between clients and servers in two ways: either by reference or by value. In this work, references are adopted to enable data exchange between components.

3.1 Web Processing Service profile architecture

In this configuration the workflow is executed using a customised WPS profile, instantiable using only metadata record entries points in a CSW. The WPS specification details profiles as a mechanism for defining functionality common to a set of processing tasks [11,15]. The management of the data inputs and outputs to the metadata catalogue is fully handled by each of the processing tasks following their invocation.

Figure 2 presents a UML sequence diagram of the system design using the WPS profile (the Full Meta Object WPS profile or FMO profiled WPS). We signify this profile as a WPS Wrapper which also provides a mechanism to reuse any existing developed WPS processes (an FMO profiled version of a given WPS). This allows one to show the specific handling at the 'profile level'. At the start of the workflow an execute request containing the metadata links, as instantiated by the user in the workflow editor, is made to a simple customised work item (task) for a WPS call. The WPS then executes a GetMetadata function which returns the relevant metadata record and then proceeds to extract data inputs from this object to construct a new WPS ExecuteProcess request. Once this request is executed on the processing server, the URL reference of the result is returned back to the FMO profiled WPS which is then registered as a new metadata record in the catalogue (RegisterResult) and the metadata URL is passed back to the workflow engine. The workflow engine then proceeds to the next processing task in the chain. Here, besides higher level service interoperability ensured by

[2] http://www.jbpm.org/

[3] http://geonetwork-opensource.org/

[4] http://52north.org

OGC services, the syntactic resolution is made only at the execution level within the WPS.

3.2 BPMN Customised Work Item architecture

In this configuration, the handling and registration of the input and output data is fully performed by the workflow engine. The engine is acting as client to the WPS requests and thus must get and receive the metadata, defined in the instance of the workflow, and using links contained within this metadata construct requests appropriately. The methods used to comply to the Full Meta Object profiling are now operating from the customised work item of the engine.

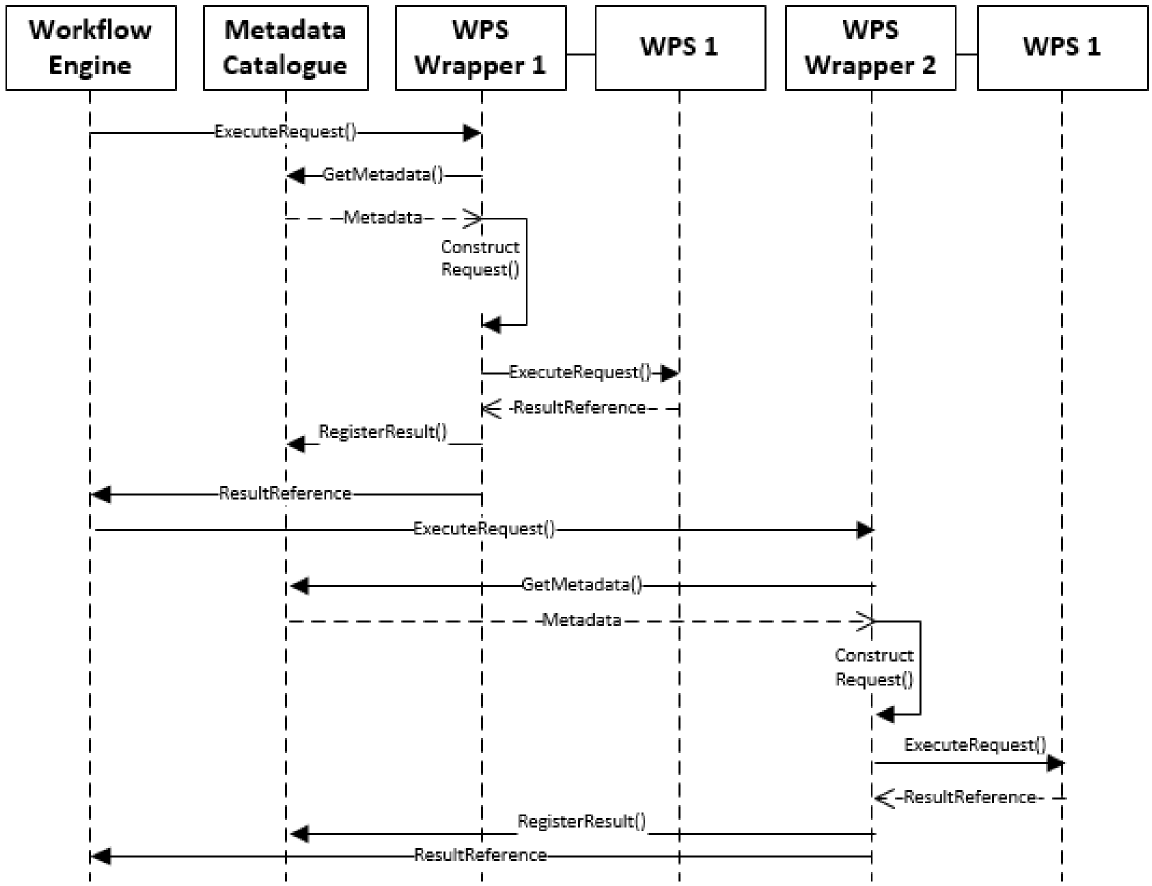
Figure 3 presents a UML sequence diagram of the BPMN profile design. At the start of the workflow execution the client makes a request to the catalogue for the metadata record for the input data set. Using the returned metadata record, the client proceeds to construct the execute request for the first task in the workflow using the reference links. Execution of this request by the processing service results in a URL reference, pointing to the output data. This result is registered within the catalogue by the workflow engine. The engine then proceeds to the next task in the sequence.

3.3 Evaluation

Note that for either solution, a customised work item must be implemented in the workflow engine (specification of a work item). However, for the WPS profiling solution this corresponds to constructing a simple http request against the WPS, whereas for the BPMN profiling the specification also must handle the interaction with the metadata catalogue. Both solutions described here offer several major benefits to the design, execution and documentation of geoprocessing workflows:

1) Workflows are generated which have greater interoperability with other workflow editors as no data types more complex than strings are used. Semantic and syntactical interoperability is better managed through direct specification of the required information via the metadata links. This can be also useful when undertaking uncertainty and sensitivity analyses as access to the metadata about data quality is required.
2) Improved efficiency through minimisation of data transfer and execution times as data travels only once from its repository to the WPS.
3) Workflow composition support (testing the adequacy of data and processes during instantiation) and workflow metadata assessment will be facilitated from dealing directly with metadata entries within a single entity (the BPMN file). Therefore, the software

Figure 2: UML sequence diagram of WPS profile workflow



support, which can be implemented either within the BPMN editor or as a plug-in or a posteriori, will have direct access to all the required information from the metadata catalogue(s).

4) The licences for some data will be handled better through the additional recording of the provenance.

Advantages and disadvantages of each of the proposed solutions:

5) Both profiling methods start with instantiating the workflow using metadata entry points for data and processes. The WPS profiling solution relies less on the BPMN editor and the work item customisation. Only a simple WPS ExecuteProcess is built after instantiation within the customised work item (tasks in the workflow engine) without specific retrieval from the metadata catalogue or the WPS beside an initial DescribeProcess request during instantiation of the workflow. On the other hand, the BPMN profiling solution needs a more involved customisation of the work items: deeper management of the metadata and registration of the results in the CSW have to be operated from each customised work item.

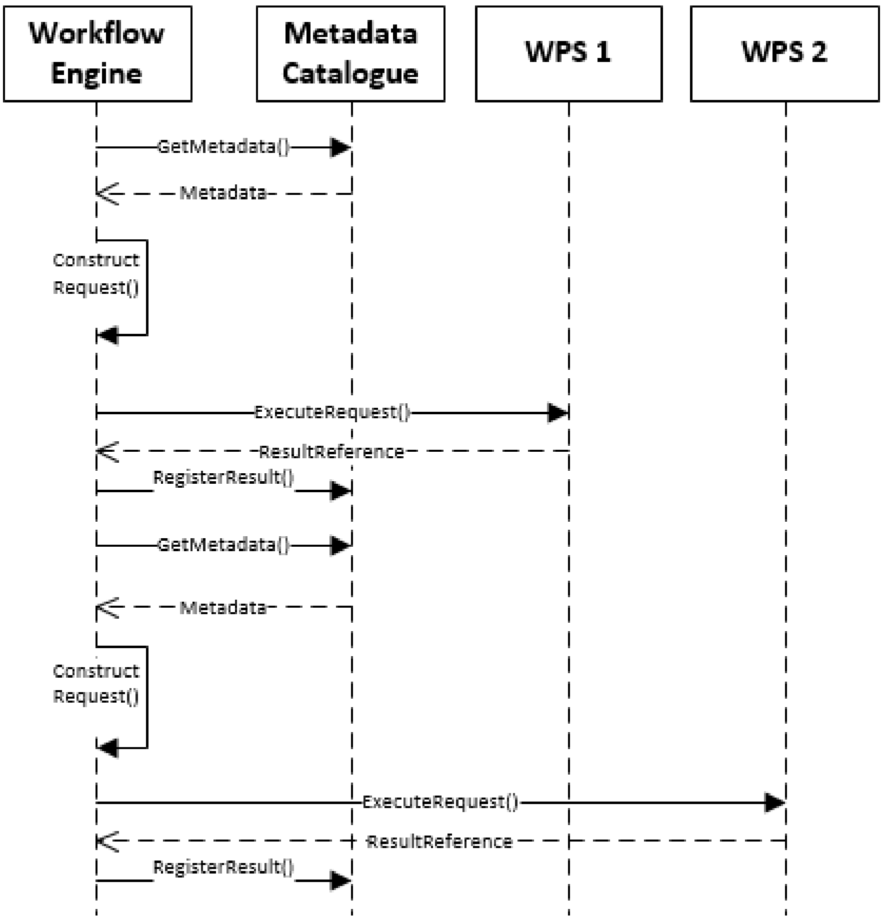
6) The BPMN profiling solution offers a tight coupling with the metadata catalogue leaving the WPS unchanged or even free of format if the metadata catalogue is also coupled with a data broker.

7) The WPS profiling solution has the advantage of a less complex BPMN implementation with blind metadata orchestration (for data and processes) with a simple WPS execution request. There is a looser coupling of the BPMN editor to the metadata catalogue.

4 Dicussion

The aim of orchestrating geospatial workflows using metadata objects is to facilitate easier and more effective management of a large scientific model represented by its workflow. The two profiling methods proposed in this paper approach this task in two ways: a WPS profile and a BPMN profile. Both profiling requirements are the same but operated in different components of a workflow architecture using WPSs. When using an architecture focusing on WPS profiling, automatic wrapping of existing WPS can be performed using a process broker [3]. This can be thought of as a way to add capabilities

Figure 3: UML sequence diagram of BPMN profile workflow



to a platform implementing workflow authoring using these principles.

Besides facilitating workflow composition, as conveying within one entity all metadata information, the full meta objects (FMO) profiling principle (at BPMN or WPS level) enables simple sharing of scientific models. Integration of additional services allowing critical analysis of the scientific model being developed is also facilitated conceptually and in their implementations, e.g. easier visualisation of the data, its ontological properties and its quality together. Enhanced descriptions attached to the metadata for the processes also become readily accessible. For example uncertainty analysis and sensitivity analysis can be plugged-in to the workflow environment without much added effort. Classical error propagation might be achieved using a data sampling method added and called from the FMO profile before execution. Or this can be achieved easily from bypassing each task with a sampling WPS. The meta-propagation method for error analysis [7] can be also executed from the workflow using quality information directly extracted from the metadata, and can be triggered in either FMO profiling approaches.

This work proposes the coupling of a workflow architecture with an OGC CSW-ISO metadata catalogue in order to increase the interoperability and the flexibility needed when dealing with workflows. Such a catalogue is designed to carry two sorts of artefacts: datasets and services. However, although a completed workflow can be registered as a WPS (i.e. as a service) with its metadata (i.e. the BPMN XML document), such an approach is not suitable for registering the complex organization of the workflow itself. An alternative standard to OGC CSW-ISO which does offer more flexibility is the OASIS ebXML (electronic Business XML) and particularly ebRIM (electronic business Registry Information Model) [10]. This model has been integrated as a profile of the OGC CSW standard and is identified as CSW-ebRIM [14]. This provides an alternative to the CSW-ISO profile which supports a rich set of artefacts and the ability to define custom object types, properties and associations which is not possible with the CSW-ISO profile.

The significance of this in relation to cataloguing workflows is that it offers the flexibility to define the workflow first as a class object and to associate it with source data either directly or by source data type or class, and a target dataset (result). A CSW-ebRIM compliant catalogue will allow the BPMN definition itself to be captured and version managed. It also allows its function, inputs and outputs to be classified (as these are available in the workflow document) and then searched for. If a similar classification is applied to data, it is possible to infer which processes could be applied to the datasets stored in the catalogue or whether a given dataset could be produced based on the available sources and the available processes. Therefore, this is as relevant during composition of the workflow as well as for discovering the output datasets of the workflow resulting from being triggered on the fly when retrieving the associated BPMN.

As an example, if a CSW-ebRIM compliant catalogue was configured to hold metadata about datasets as well as processes that could act upon them, then it would be possible for that catalogue to offer not only source datasets but also all datasets that could be derived from the source (by automatically inferring the possible datasets that could be generated). A request to retrieve those datasets could trigger processes which would use the source data and the process defined in BPMN to produce the data on the fly.

5 Conclusions

This paper presented two alternative approaches to constructing and orchestrating workflows using metadata objects: Full Meta Objects profiling. Both solutions, BPMN profiling and WPS profiling are based on workflow instantiation from metadata entries in a catalogue for data and processes. This full meta objects principle makes the workflow BPMN encoded in XML a complete reference to the provenance and the knowledge that generated the scientific model. The main difference between the two solutions is the desired level of coupling in the architecture.

Besides the basic principles of the full meta object architecture, we discussed usage and the added flexibility that can make a workflow environment central to seamless data and model prototyping and sharing. This includes visual exchange of model diagrams, critical analysis via quality assessment, enhanced architecture via a brokering system and the potential of complex enrichment of the metadata relations via an ebRIM catalogue model which would contribute to a composition controller added capability.

It is believed that such solutions optimising the interoperability settings can be central to establishing research platform environments based on workflows, as much on the enhancement possibilities for the workflow management as on the simpler software architecture development itself. We are currently preparing the two prototyped solutions for further testing.

References

[1] G. Alonso, C Hagen. Geo-Opera: Workflow concepts for spatial processes. In M. Scholl and A. Voisard, editors, In *Proceedings of 5th International Symposium on Spatial Databases (SSD '97)*. Berlin, 1997.

[2] C. Bielski, S. Gentilini, M. Pappalardo, Post-Disaster Image Processing for Damage Analysis Using GENESI-DR, WPS and Grid Computing, *Remote Sensing* 3:1234–1250, 2011.

[3] E. Boldrini, Papeschi, F, Santoro, M, Nativi, S. Enabling interoperability in Geoscience with GI-suite. In EGU General Assembly Conference Abstracts, vol. 17, p. 12199. 2015.

[4] J. Eberle, C. Strobl, Web-based geoprocessing and workflow creation for generating and providing remote sensing products, *Geomatica.* 66:13–26, 2012.

[5] G. Giuliani, Y. Guigoz, P. Lacroix, N. Ray, A. Lehmann, Facilitating the production of ISO-compliant metadata of geospatial datasets, *International Journal of Applied Earth Observation and Geoinformation*. 44:239–243. 2016.

[6] G. Hobona, D. Fairbairn, H. Hiden, P. James, Orchestration of grid-enabled geospatial Web services in geoscientific workflows, *IEEE Transacations on Automation Science and Engineering,* 7:407–411. 2010.

[7] D.G. Leibovici, a. Pourabdollah, M.J. Jackson, Which spatial data quality can be meta-propagated?, *Journal of Spatial Science*. 58:3–14. 2013.

[8] S. Meek, M. Jackson, D. Leibovici, A BPMN solution for chaining OGC services to quality assure location-based crowdsourced data, *Computers & Geoscience*s. 87:76-83. 2016.

[9] J. Nogueras-Iso, F.J. Zarazaga-Soria, R. Béjar, P.J. Álvarez, P.R. Muro-Medrano, OGC Catalog Services: A key element for the development of Spatial Data Infrastructures, *Computers & Geoscience*s. 31:199–209. 2005.

[10] OASIS, ebXML Registry Information Model Version 3.0, 2005.

[11] OGC, Open GIS Web Processing Service, 2007.

[12] OGC, OpenGIS Catalogue Services Specification, 2007.

[13] OGC, Open Geospatial Consortium OpenGIS Catalogue Services Specification 2.0.2 - ISO Metadata Application Profile. 2007.

[14] OGC, CSW-ebRIM Registry Service - Part 1: ebRIM profile of CSW, 07-110r4. 2009.

[15] OGC, OGC WPS 2.0 Interface Standard. 2015.

[16] OMG, Business Process Model and Notation (BPMN) Version 2.0, 2011.

[17] S. Trilles, L. Díaz, J. Gil, J. Huerta, Assisted Generation and Publication of Geospatial Data and Metadata, *International Journal of Spatial Data Infrastructures* Res. 9, 24–27. 2014.

[18] G. (Eugene) Yu, P. Zhao, L. Di, A. Chen, M. Deng, Y. Bai, BPELPower—A BPEL execution engine for geospatial web services, *Computers & Geosciences*, 47. 87–101. 2012