WPS Application Profiles for Generic and Specialised Processes

Edward Nash[1]

[1]Chair of Geodesy and Geoinformatics, Institute for Management of Rural Areas, Faculty of Agricultural and Environmental Sciences, Rostock University
edward.nash@uni-rostock.de

Abstract. The recently-published OGC Web Processing Service (WPS) standard provides the ability to define and use WPS Application Profiles for defining processes in a reusable way. The use of such profiles for commonly-used processes in an information community should, together with appropriate service registries, ease semantic problems in finding and orchestrating WPS. This paper presents some initial ideas as to how the profiles could be structured, considering particularly the cross-domain nature of many fundamental GIS operations compared to the highly specialised nature of WPS intended for use in one particular domain, and makes suggestions as to how the profile mechanism may be extended to enhance the utility of these profiles.

1 INTRODUCTION

The Open Geospatial Consortium (OGC) recently approved the Web Processing Service (WPS) 1.0.0 as an OpenGIS® Implementation Standard (Schut, 2007). The WPS standard defines a generic web-service interface to data processing operations. Mechanisms for both an OWS-standard HTTP Get/Post (REST-architecture) and WSDL/SOAP operation are specified. The core of the standard is the specification of processes by process identifier and the data inputs and outputs which may be used with each process. Using the standard OWS-style operation, a list of available processes (identifiers and free-text abstracts) is included in the capabilities document whilst the details of each process may be accessed through a `DescribeProcess` operation.

As a means to standardise semantically-driven service discovery and orchestration, the standard defines a mechanism whereby “WPS Application Profiles” defining standard processes may be created. These should use a URN in the OGC namespace to uniquely identify the process and provide, at a minimum, a reference DescribeProcess response. This process description defines the data input and process outputs by name, their cardinality and data types/formats. A WPS Profile would therefore effectively standardise process `urn:ogc:a` as having inputs `B` and `C` which may be supplied in data format (MIME-type) `D` or `E` and that the process outputs are `F` and `G` in data formats `H` and `I`.

According to the standard, "Geospatial infrastructures can establish a geospatial processing web by specifying a repository that contains a semantically defined hierarchy of processes, each identified by a URN. A WPS Application Profile can define each unique process within the repository, and each WPS instance can refer to that URN" (Schut, 2007, p7). For generic spatial data infrastructures (SDIs), a single repository specifying generic GIS processes such as polygon overlay, map algebra, transformations, etc., may be feasible. Specialist SDIs which include data and standard processes which are only of interest to a particular information community, and the semantics of which should be well-known within that community, will however also require services from generic SDIs (Nash & Kofahl, 2006). To avoid duplication of WPS application profile definitions, such SDIs will therefore either require a cascading repository which binds further repositories to present a unified result set to the user or to specify which external repositories should be used for which process definitions.

This paper presents some initial ideas for some generic WPS application profiles as well as some specialist ones applying to the agricultural domain. Some suggestions for clarifications and further development of the WPS profile mechanism to overcome limitations identified in this work are also made.

2 SUGGESTED WPS APPLICATION PROFILES

In this section some potential WPS application profiles and a basic hierarchical structure are suggested. Both generic and specialist processes are considered, demonstrating how information communities may structure their own profiles whilst re-using common profiles. The specialist processes presented here are described in more detail in Nash et al (2007a and 2007b); in general, the specialist services automate geoprocessing workflows by presenting an interface to a service chain to the user. Further examples of specialist services from other domains are presented e.g. by Stollberg and Zipf (2007 and 2008). The chain is orchestrated either using statically-bound services (pre-defined) or dynamically based on a service catalogue. Further examples of generic processes are based on those presented in Heier and Kiehle (2006), Kiehle et al. (2006) and Stollberg et al. (2007).

2.1 Generic Processes

Generic WPS services provide fundamental GIS functionality such as polygon overlay, map algebra, transformation or interpolation. As such functionality may be required in many information domains, it is suggested that globally-standardised WPS application profiles may be produced for such processes. The WPS standard has some basic multi-language support whereby the server may specify the natural languages supported and the `DescribeProcess` operation may specify the language to use from the available set, but the data structures (title and abstract) within a process or input/output data description support only a single language. How a multi-linugal application profile is to be specified using the current mechanism is therefore unclear.

2.1.1 Namespace Structure

In the literature there are various attempts at classifying GIS operations e.g. Goodchild (1991) and Albrecht (1996). Although the latter claims that 20 data-model independent universal GIS operations can be identified, the data structure is admitted as dictating differences in functionality, and the former groups operations primarily based on the data structure used. A basic distinction in GIS processing operations is between those working primarily with raster data and those working with vector data. It is therefore suggested that the first part of the namespace structure for generic processes could reflect this division, with a third branch for operations converting between data structures. Further branches, e.g. for processes operating on topological data structures, isolines, etc., may be introduced as necessary. Within each branch a functional grouping could occur. Based on the examples considered here, a basic, and necessarily incomplete, sketch of the proposed structure is shown in Figure 1. Although the OGC publish a URN policy (OGC, 2007), it is not clear from the WPS standard in which branch WPS application profiles should be defined. A common URN root of `urn:ogc:def:wpsProfile` is therefore assumed.

2.1.2 Example Generic Processes

A standard set of operations for raster data is defined by Tomlin's map algebra (Tomlin, 1990). These may be split into local, focal, zonal and global operators, with the local operators further grouped into arithmetical, Boolean and comparison operators. Some selected examples of local and focal map algebra operations are shown in Table 1. A further common operation is the transformation of a raster e.g. through resampling or coordinate system transformation which is commonly required in order to overlay multiple raster datasets.

Vector-raster conversion, e.g. through interpolation or kriging is also considered here with two simple interpolation algorithms. In this case, the ability to create a hierarchy of application profiles would be useful as all interpolation operations require the same set of basic parameters (those for nearest neighbour interpolation in Table 1), which could be specified in an abstract "superprofile", together with some algorithm-specific ones which may be specified in each concrete profile.

Finally two vector operations are considered, namely the creation of a buffer based on a geometry property of a given set of features, with the necessary parameters based on Heier and Kiehle (2006), and the spatial join operation using the geometry properties of two sets of features, for which the suggested inputs are based on Stollberg et al. (2007). The first of these two operations is here not allocated to a sub-category of vector operations since it is considered to be a unique operation. The latter is allocated to a sub-category of overlay operations which could also include other processes where two input layers are required with the output being a function of their geometries, e.g. cropping.

2.2 Example Specialist Processes

It is expected that the majority of specialist services will implement defined geoprocessing workflows either locally or through orchestrating further services, both specialist and generic, to automate domain-specific calculations, models, etc. and solve interoperability problems within that domain (Stollberg and Zipf, 2007). Nash et al. (2007a, 2007b) describe processes specific to the precision agriculture domain, with two complex workflow-based processes (generation of site-specific nitrogen fertilisation maps and of in-field management zones) and two domain-specific simple processes (normalisation of raw yield date to 14% moisture and calculation of nitrogen loss due to a crop). Although the WPS standard states that all application profiles should be defined using a URN in the OGC namespace, it is questionable whether this is appropriate for domain-specific profiles. It is therefore suggested that information communities should use, and if necessary define, their own URN namespace for such WPS profiles, for which they should be responsible for maintaining the structure and registry. For the agricultural processes presented in Table 2 the fictional `preagro` URN namespace is used.

Table 1: Summary of selected generic WPS application profiles – see Table 3 for details of proposed data types

Process Identifier (`urn:ogc:def:wpsProfile:*`)	Inputs		Data Type	Output	Data Type
`conversion:toRaster:interpolate:inverseDistance`	1..1	Features	features	Interpolated	raster
	1..*	Attribute	XPath		
	1..1	TargetGrid	rectified grid		
	1..1	PowerFactor	number		
	1..1	MaxDistance	distance		
	1..1	MaxFeatures	+ve integer		
`conversion:toRaster:interpolate:nearestNeighbour`	1..1	Features	features	Interpolated	raster
	1..*	Attribute	XPath		
	1..1	TargetGrid	rectified grid		
`raster:mapAlgebra:focal:applyMatrix`	1..1	Data	raster	Result	raster
	1..1	Matrix	2D matrix		
`raster:mapAlgebra:local:arithmetic:addition`	2..*n*	Addend	raster	Sum	raster
`raster:mapAlgebra:local:arithmetic:subtraction`	1..1	Minuend	raster	Difference	raster
	1..1	Subtrahend			
`raster:mapAlgebra:local:boolean:and`	2..*n*	Data	raster	And	raster
`raster:mapAlgebra:local:comparison:lessThan`	1..1	LValue	raster	LessThan	raster
	1..1	RValue	raster		
`raster:transform`	1..1	Data	raster	Transformed	raster
	1..1	TargetGrid	rectified grid		
	0..1	Interpolation	*c* Interpolation		
`vector:buffer`	1..1	Features	features	Buffered	features
	1..*	Attribute	XPath		
	1..1	Distance	distance		
	0..1	CapStyle	*c* CapStyle		
	0..1	Quantisation	distance		
`vector:overlay:spatialJoin`	2..*	Features	features	Joined	features
	2..*	Attribute	XPath		
	1..1	Predicate	*c* Predicate		

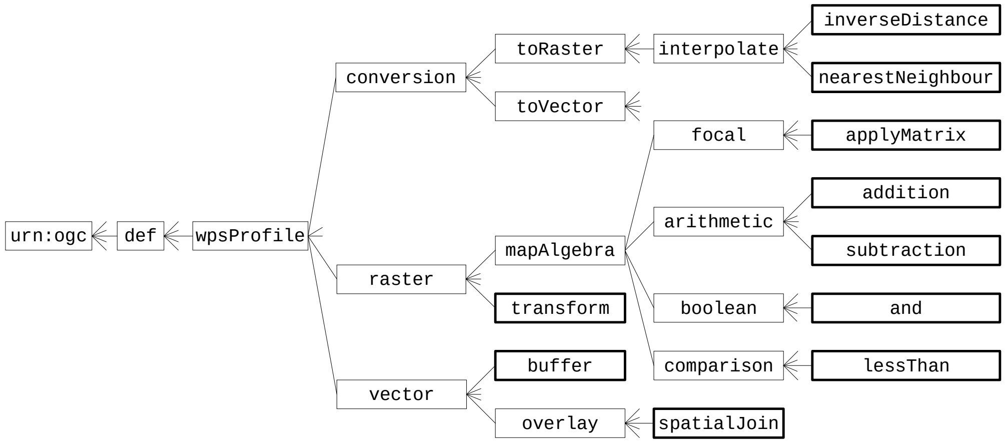

Figure 1: Outline structure of URN namespace for generic processes with selected profiles. Individual processes are bold.

Table 2: Summary of selected specialist processes for precision agriculture – see Table 3 for details of proposed data types

Process Identifier (`urn:preagro:wpsProfile:*`)	Inputs	Data Type	Output	Data Type
`managementZones:generate`	1..1 Field	field	ManagementZones	zoning
	1..1 InputDataset	raster		
	1..1 NumZones	+ve integer		
	1..1 MinZoneSize	+ve integer		
	1..1 Raster	window		
`siteSpecific:applicationMap:totalNitrogen`	1..1 Yield	yield data	Fertilisation	recommended fertilisation
	1..1 Field	field		
	1..1 SoilNitrogen	soil data		
	1..1 Crop	code		
`siteSpecific:dataNormalisation:yield:moisture`	1..1 Yield	yield data	NormalisedYield	yield data
	1..1 Field	field		
`siteSpecific:nutrientRemoval:nitrogen`	1..1 Yield	yield data	NitrogenRemoval	yield data
	1..1 Field	field		
	1..1 Crop	*c* Crop		

2.3 Data Types

As can be seen from Table 1 and Table 2, many data types for input and output values are common to many processes. The current WPS profile mechanism whereby the profile is defined using a WPS `ProcessDescription` would however require that each data type and the accepted values and formats must be defined anew for each process. It is therefore suggested that a mechanism whereby data types and acceptable standard formats (e.g. GeoTIFF and/or GML `RectifiedGridCoverage` for raster) may also be specified using a profile mechanism and allocated a URN, which may be referenced from WPS application profiles and `ProcessDescription` documents, is also required.

Table 3: Summary of data types required for selected processes. Codes/enumerations are prefixed with *c*. Default values and formats are underlined where applicable.

Data Type	**Format**	**MIME Type / {allowed values} /** ***note***
+ve integer	*LiteralValue*	-
2D matrix	MathML matrix	text/xml; subtype=mathml/matrix
distance	*LiteralValue*	*uom attribute should be specified*
features	GML FeatureCollection	text/xml; subtype=gml/featurecollection
field	agroXML Field	text/xml; subtype=agroxml/field
number	*LiteralValue*	-
raster	GML RectifiedGridCoverage GeoTIFF	text/xml; subtype=gml/rectifiedgridcoverage image/tiff; subtype=geotiff
rectified grid	GML RectifiedGrid	text/xml; subtype=gml/rectifiedgrid
recommended fertilisation	agroXML FertilisationRecommendationDoc	text/xml; subtype=agroxml/ fertilisationrecommendationdoc
soil data	agroXML SoilDataDoc	text/xml; subtype=agroxml/soildatadoc
window	WMC Window	text/xml; subtype=context/window
yield data	agroXML YieldDataDoc	text/xml; subype=agroxml/yielddatadoc
XPath	*LiteralValue*	*namespace bindings should also be specified*
zoning	agroXML Zoning	text/xml; subtype=agroxml/zoning
c CapStyle	*LiteralValue*	{butt, round, square}
c Crop	*LiteralValue*	*values from agroXML dictionary; codeSpace should be specified to indicate version*
c Interpolation	*LiteralValue*	{nearest, bilinear, bicubic}
c Predicate	*LiteralValue*	{disjoint, meet, overlap, inside, contains, covers, coveredBy, equal}

The WPS standard also states that complex data types should be identified by their MIME type. For many data types identified for the selected processes there is however no widely-recognised MIME type. Particularly where a single XML element from a schema is required then even when a MIME type for the schema is known (e.g. `text/xml; subtype=gml/3.2.1` for GML) then this does not uniquely identify the expected data type. Similarly for GeoTIFF, `image/tiff` does not uniquely identify a GeoTIFF as opposed to a standard TIFF. As a potential solution to this problem, Table 3 specifies MIME types using the "; subtype=" mechanism whereby the content type of this subtype is the schema or subformat and the subtype is the element.

3 SUGGESTED EXTENSIONS TO WPS APPLICATION PROFILES

The previous sections of this paper have identified some weaknesses in the WPS Application Profile mechanism as currently specified in the WPS 1.0.0 standard. Some extensions to this mechanism are therefore suggested here which would help address these weaknesses. In general it can be stated that these problems stem from the use of the standard `ProcessDescription` document as a means to define the profile. Whilst this provides an obvious basis for the definition, it is suggested that some extensions to this basic format are required.

3.1 Multilingual Profiles

Although the WPS specification allows for servers supporting multiple natural languages for the free-text aspects of the process and input/output descriptions (i.e. title and abstract), the current profile mechanism does not as only a single title and abstract can be defined for each element. This could be easily solved by allowing multiple titles and abstracts in the profile definition, with a requirement in the schema (specified using e.g. Schematron) that each must have a unique `xml:lang` attribute.

3.2 Process Hierarchies and Inheritance

The current profile specification foresees a structuring of profiles through the URN namespace mechanism. Whilst providing a classification, this ignores the strong relationships between many processes where e.g. multiple algorithms exist to solve a single problem, and these algorithms share a set of basic parameters to which are added some algorithm-specific parameters. This is illustrated by the two interpolations shown in Table 1. It is suggested that WPS profiles should therefore support an `extends` mechanism similar to that used in XSD. A child process would through this inherit the inputs and outputs of the parent, meaning that they must only be defined once for a set of like processes, potentially in an abstract process definition. A similar mechanism could also be used for individual instances to indicate that they extend a given profile by supporting e.g. additional optional input parameters or additional input/output data formats.

3.3 Data Type Profiles

Many processes require inputs or deliver outputs in the same data type, e.g. a standard feature collection or raster. A means to standardise names and the supported values and formats for each of these data types using a profile would reduce the redundancy in WPS profiles. Additionally, the use of MIME types for identifying complex data types has the problem that not all common geographic data types or XML elements have a well-known MIME type which uniquely identifies them. A way of using the MIME subtype mechanism to work-around this problem was discussed in section 2.3; a standard solution is however required.

4 CONCLUSIONS

This paper has presented some initial ideas as to how WPS Application Profiles as introduced in the WPS 1.0.0 standard could be structured and made some suggestions of generic processes for which global profiles could and should be created. The distinction between generic processes which have application in many fields and specialised processes which are only relevant for one community was introduced, and it was suggested that profiles for these specialised processes should not necessarily be defined in the OGC namespace or with the same hierarchy as for generic processes.

For generic processes it was noted that the structure of the OGC namespace which is to be used for WPS profiles is currently unclear. A suggestion was made for a hierarchical structure based on data structure and then functional groupings. Furthermore, some potential for improvements in the currently specified profile system were identified and solutions suggested.

Despite the weaknesses identified in the published profile mechanism, the author of this paper believes that such profiles presently present the best chance of producing truly interoperable web processing services by removing many of the semantic barriers implicit in finding and binding “foreign” services. They should therefore play an important role in enabling dynamic service chaining and workflow management for geospatial data processing.

5 ACKNOWLEDGEMENTS

The work leading to this paper was carried out as part of the *preagro* project funded by the German Federal Ministry of Research and Education (BMBF) under grant number 0330663.

6 REFERENCES

Albrecht, J. (1996). Universal Analytical GIS Operations. PhD Thesis. ISPA Mitteilungen Heft 23, University of Vechta, Germany.

Goodchild, M. F. (1991). Spatial Analysis with GIS: Problems and Prospects. Proceedings of GIS/LIS'91, Atlanta, GA, USA. 1: 40-48.

Heier, C. and C. Kiehle (2006). "Automatisierte Liegenschaftsauskunft mittels OGC Web Processing Service." GIS Zeitschrift für Geoinformatik 7/2006. 12-16.

Kiehle, C., K. Greve and C. Heier (2006). Standardized Geoprocessing – Taking Spatial Data Infrastructures One Step Further. Proceedings of AGILE 2006, Visegrád, Hungary. J. Suarez and B. Markus (Eds.). University of West Hungary. 273-282.

Nash, E. and M. Kofahl (2006) Specialised SDIs to Support Business Processes. Proceedings of AGILE 2006, Visegrád, Hungary. J. Suarez and B. Markus (Eds.). University of West Hungary. 72-80.

Nash, E., J. Bobert, K.-O. Wenkel, W. Mirschel, R. Wieland (2007a). Geocomputing Made Simple: Service-Chain Based Automated Geoprocessing for Precision Agriculture. Proceedings of GeoComputation 2007, Maynooth, Ireland. U. Demšar (Ed.). National University of Ireland, Maynooth.

Nash, E., P. Korduan and R. Bill (2007b). Optimising data flows in precision agriculture using open geospatial web services. Proceedings of the 6th ECPA, Skiathos, Greece. J. V. Stafford (Ed.). Wageningen Academic Publishers. 753-759.

Schut, P. (2007). OpenGIS® Web Processing Service. Open Geospatial Consortium, Wayland, MA, USA.

Stollberg, B. and A. Zipf (2007). OGC Web Processing Service Interface for Web Service Orchestration. Proceedings W2GIS 2007, Cardiff, UK. J. M. Ware and G. E. Tayler (Eds.). Springer-Verlag, Heidelberg. 239-251.

Stollberg, B. and A. Zipf (2008). Geoprocessing Services for Spatial Decision Support in the Domain of Housing Market Analyses. Proceedings AGILE 2008, Girona, Spain. L Bernard, A. Friis-Christensen, H. Pundt and I. Compte (Eds.). University of Girona.

Stollberg, B., M. Lutz, N. Ostländer, L. Bernard (2007) "Geoprozessierung in Geodateninfrastrukturen – Aufgaben für die nächste Generation." GIS Zeitschrift für Geoinformatik 4/2007. 22-27.

Tomlin, C. D. (1990). Geographic Information Systems and Cartographic Modeling. Englewood Cliffs, NJ, USA, Prentice-Hall.