(Draft) Actas de las II Jornadas Ibéricas de Infraestructuras de Datos Espaciales (JIIDE'2011), Barcelona, 9-11 de noviembre de 2011. 2011

Implantation of OGC geoprocessing services for Geoscience

Francisco J. Lopez-Pellicer [1], Walter Rentería-Agualimpia[1], Rubén Béjar[1], Juan Valiño, F. Javier Zarazaga-Soria[1], Pedro R. Muro-Medrano[1]

[1]Computer Science and Systems Engineering Department
Universidad Zaragoza

{fjlopez, walterra, rbejar, juanv, javy, prmuro}@unizar.es

Abstract

This paper presents an investigation about the availability of standard geoprocessing services and its use in the geosciences domain. To do that, the web has been crawled in March 2011 to find the servers available conforming to the Web Processing Service interface specification published by the geospatial standards organization Open Geospatial Consortium (OGC), which gives support to standard Web-based geoprocessing. The research goals are (i) to provide a reality check of the availability of Web Processing Service servers, (ii) to provide quantitative data about the use of different features defined in the standard that are relevant for a scalable Geoprocessing Web (e.g. long-running processes, Web-accessible data outputs), and (iii) to test the capability for finding Web services in the geoscience domain.

Keywords: Service discovery, Web crawler, Web services, WPS.

1 Introduction

Online geoprocessing systems offer access to spatial computations (e.g. coordinate transformations, orthorectification), thematic computations (e.g. geoparsing, geocoding), temporal computations (e.g. change detection, temporal proximity) and metadata computations (e.g. statistical calculation, geographic annotation). The complexity of those processes varies from typical large-scale geospatial computations involving substantial amounts of data to very simple transformations of data. Wrapped as Web services, geoprocessing systems can offer additional features supported by other components, such as persistent storage of outputs and transference of data over networks. The geospatial community realizes that providing standard Web access to geoprocessing systems is a natural step in the evolution of the geospatial cyberinfrastructure [1,2]. Moreover, standardization should encourage implementation of new geoprocessing endpoints, which is beneficial not only in geoscience research areas but also in other areas [3].

The purpose of this paper is the evaluation of the implementation of a standard for Geoprocessing Web services: the Open Geospatial Consortium (OGC) *Web Processing Service* (WPS) specification. Since 1994, the OGC has been providing standards that ease the use and integration of geospatial Web services [4,5]. However, few studies are focused on providing data about the number of servers implementing them [3,6]. In the case of the WPS specification, the evaluation requires first the gathering of a collection of WPS servers. The traditional approach relies on the search of metadata records about them in geospatial catalogues [7]. This is the scenario analysed in the existing literature with references to the discovery of WPS servers [8,9]. However, the discovery of Web services using catalogues assumes that service providers register and keep updated metadata about their services. Although this assumption is frequently met for other geospatial resources, such as geospatial datasets [10], Web services seem to be an exception [11]. Nowadays, information sources are no longer restricted to catalogues. The literature gives evidence of the use of search engines and focused crawlers to discover general-purpose Web services [14] and geospatial Web services [6,15]. Therefore, it is worthwhile to analyse the current state of the implementation of the WPS specification using systems that index the Web.

2 The WPS specification

The version 1.0.0 of the WPS standard was released in June 2007 [16]. The specification describes a service instance that offers a simple HTTP interface for publishing processing algorithms, especially geoprocessing algorithms, and for requesting their execution. Companies (e.g. Galdos Systems — Canada, PCI Geomatics — Canada), government agencies (e.g. GeoConnections — Canada, Wupperverband — Germany) and universities (e.g. Autonomous University of Barcelona — Spain, University of Münster — Germany) have participated in its development. According to OCG, 15 products implement the WPS 1.0.0 specification at the time of the analysis. The implementing products include open source projects, such as *degree*, *PyWPS*, *52º North*, and *ZOO Project*, and products developed by software vendors, such as ERDAS and Intergraph.

3 Methodology

The collection of OGC Web services used in this research was collected between 1st and 5th March 2011 using a focused crawler. Each run of the crawler used as seed search results from automated queries made to Bing, Google and Yahoo!. From these results, the crawler explored the Web applying two well-documented crawling strategies: shark-search [18] and best first [19]. The queries made to search engines included mandatory terms associated with requests for OGC service metadata (e.g. “request”, “getcapabilities”, “service”) plus additional terms related to the targeted standard service (e.g. “wps”, “processing”, “profile”) or tasks related to the service (e.g. “coordinate transformation”, “interpolation”, “grid”). For example, the search results of the query “getcapabilities wps interpolation” include pages about interpolation methods with links to WPS servers, and WPS service metadata documents that provide interpolation processes. The goal of this search strategy is to discover references to OGC Web servers in HTML text, Web links, indexed research and technical papers. A complete analysis about the precision of different alternatives and strategies for searching geospatial Web services is available in [6].

4 Results

The results include an analysis of the implementation of the WPS specification worldwide, quantitative descriptions of the processes published, including information about the number of geoscience processes, long-running processes and Web-accessible data outputs, and a review of the use of features related to interoperability: application profiles, WSDL descriptions and multilingual support.

4.1 Implantation

This section analyses how widespread is the use and who is adopting the specification. Figure 1 compares the total of WPS services found with the total of services that meet any other OGC specifications about geospatial Web services. Only 58 out of 9329 OGC Web services found were WPS services. Most of them (47) support WPS 1.0.0 requests. The remaining 11 services only support deprecated versions (e.g. version 0.4.0 [20]. Table 1 presents the localization of the WPS services and the number of service providers per country. This table reveals that 84% of the services found are located in Europe.

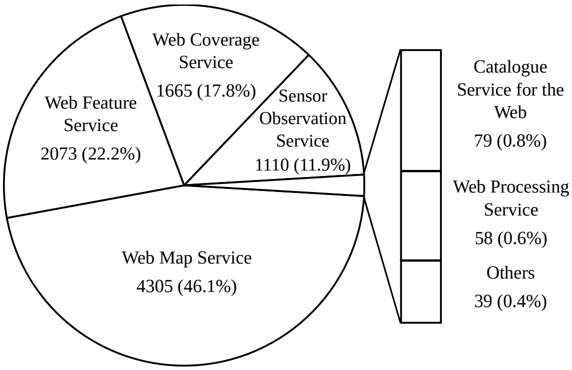


Figure 1: OGC Web service instances classified by specification.

Country	Services	Providers	Country	Services	Providers
Spain	24	7	China	3	1
Germany	13	7	Canada	3	1
France	4	3	United States	2	1
Czech Republic	4	2	Switzerland	1	1
United Kingdom	3	3	Australia	1	1

Table 1: WPS instances and service providers by country.

4.2 Processes published

The services found offer 1316 processes. Two services are considered inactive because they do not declare offered processes. The sample median of the distribution of processes offered by services is four. That is, half of the services have four or less processes. Three services offer 750 processes (57% of the processes found). One of them is a demonstration server of a software vendor. The rest are located in universities. Table 2 shows the distribution of processes per service.

Offerings	0	1	2	3	4	5-9	10-20	21-100	+100
Servers	2	8	4	3	17	8	6	7	3

Table 2: Processes per WPS instance.

The *ProcessOffering* section in the *GetCapabilities* response shall contain a brief description of the processes offered by each service. The description of each process must contain an *Identifier* and a *Title*, and may include some optional self-describing elements: an *Abstract* providing additional details, a reference to a *Metadata* document about the process, information about the *Profile* supported (discussed in Section 5.4), the location of its *WSDL* (discussed in Section 5.5) and the *processVersion*. Only 87.1% of the processes contain the mandatory fields *Identifier* and *Title*. The *Abstract* is found only in a third of the processes analysed. It is quite relevant that 21.3% of the processes contain a *Metadata* element that works either as a keyword container or as an explicit link to the respective *DescribeProcess*. None of the analysed processes contains a link to an ISO metadata document or equivalent. Table 3 summarizes the findings.

Name	Use	Found in	
Identifier	Mandatory	1311	99.6%
Title	Mandatory	1146	87.1%
Abstract	Optional	443	33.7%
Metadata	Optional	280	21.3%
processVersion	Optional	56	4.3%
Profile	Optional	2	0.2%
WSDL	Optional	0	0.0%

Table 3: Elements found in process descriptions.

Input/Output		Min	Q_1	Median	Mean	Q_3	Max
data input	total	0	2	3	4.45	6	28
	- complex	0	1	1	1.38	2	14
	- literal	0	0	2	3.01	5	22
	- bounding box	0	0	0	0.01	0	1
process output	total	0	1	1	1.17	1	10
	- complex	0	1	1	1.03	1	10
	- literal	0	0	0	0.11	0	2
	- bounding box	0	0	0	0.01	0	1

Table 4: Summary of data inputs and process outputs per process.

The *DescribeProcess* operation was used for requesting additional information about the published processes. Figure 2 summarizes the results. Some requests

(7.6%) failed due to connections errors, exception reports or even, mangled endpoint declarations. The analysis of the 1216 successful responses shows that 92% of the processes support Web access to the process results, and 75% of the processes support long running requests. Table 4 presents process signature data. Analysed processes can be characterized as having between two and six input parameters, one of them representing complex input data. The output is usually a single complex result.

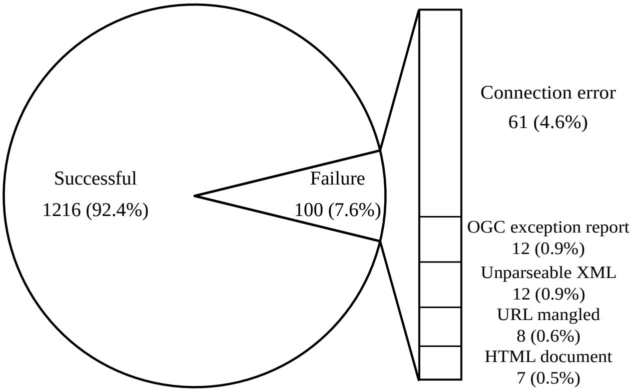
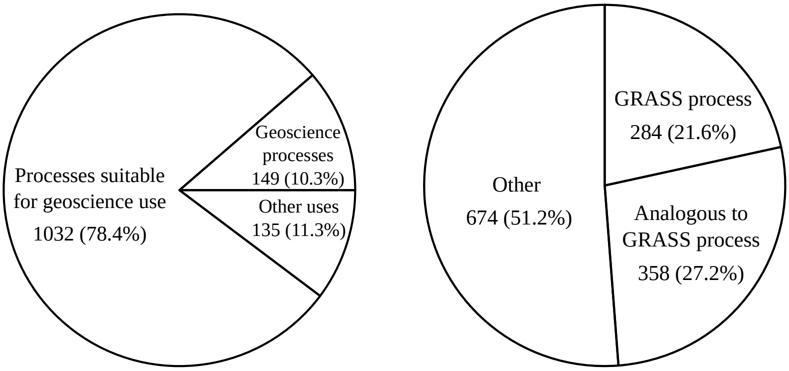


Figure 2: *DescribeProcess* responses.

4.3 Geoscience support

Two alternative manual approaches were followed to identify services related to geoscience. The first approach relies on the classification made by an expert based on the signature and the description of each WPS process. A WPS process is classified as suitable for geoscience use if the expert can guess from its signature and its description a specific geoscience process. However, if the expert has a high degree of confidence that it is a process used in some macro-area of geoscience (geophysics, geology, oceanography, etc.) using as evidence the literature on geoscience Web services (e.g. [2,21]) and other sources, the process is classified as a geoscience process. The rest of processes analysed are classified as other uses. Figure 3a contains a summary of the results. This approach classifies 88.7% of the processes as either *geoscience processes* (10.3%) or *suitable for geoscience use* (78.4%). The *other uses* category is populated mostly by data access processes, wrappers to services related to location (e.g. geocoders), demo, dummy and testing processes, and insufficiently specified processes. According to this classification, eight servers do not provide processes related to geoscience.

The second approach uses a well-known geoscience package as reference for classifying processes. We use the GRASS package as reference because the literature shows that there is interest in the use of the WPS interface to publish processes of the GRASS package [22,23] The classification is performed as follows. For each WPS process, an expert looks a matching GRASS process up in the GRASS Documentation Project. If the expert has a high degree of confidence that both processes are the same (e.g. same name, same parameters, similar description, explicit references to GRASS), the WPS process is classified as GRASS process. If no match is found but the expert believes the WPS process implements an algorithm or procedure similar to some GRASS process, the process is classified as analogous to a GRASS process. This step considers additional information found in the GRASS documentation and the available literature about WPS implementations. Figure 3b presents the results. This approach classifies 48.8% of the processes as either *GRASS process* (21.6%) or *analogous to a GRASS process* (27.2%). According to this classification, 14 servers do not provide processes related to a well-known geoscience package.



(a) Signature and description. (b) Similitude to GRASS processes.
Figure 3: Classification of processes.

4.4 Application profiles

Geoscience process publishers and geoscience process consumers are not the only players in an integration scenario. Third parties may also contribute to the integration via the definition of process profiles. The use of *WPS Application Profiles* (WPS APs) is the approach of the WPS specification to standardize semantic and technical interoperability. The literature offers examples of different proposals and potential uses of WPS APs [24-26]. A WPS AP document is intended for consumption by Web service registries and by WPS clients. Users can query a Web service registry for WPS services that support a given profile using the identifier of the profile, or given a WPS service that claims to support a specific profile, to search its definition in a Web service registry. Only 13 out of 47 WPS 1.0.0 services provide a WSDL document

The survey only found two services annotated with profiles. One of the profiles is an INSPIRE Coordinate Transformation (CT) profile defined in INSPIRE Network Services [27]. This profile is based on the interface defined in the OGC draft Web Coordinate Transformation Service (WCTS). Surprisingly, the other profile is a proprietary adaptation of the INSPIRE CT profile.

5 Discussion

The use of some features considered useful for a scalable Geoprocessing Web, such as the support of long-running processes and Web-accessible data outputs, is generalized. However, results stress the fact that it is not trivial to identify or understand the nature of published processes. Human readable descriptions are only available in a third of the processes analysed. In addition, none of the analysed processes provided links to either machine processable ISO metadata documents or similar documents. It is also remarkable the slow progress in the adoption of application profiles, which are a key feature for enabling semantic interoperability. In regard to technical interoperability, it is striking that few services provide working WSDL descriptions considering the relevance of this feature for many geoprocessing workflows, such as those based in BPEL [17], [28]. From a pragmatic point of view, scarce documentation, lack of profiles, and few WSDL descriptions constitute semantic and technical barriers to the development of a Geospatial Web based on the WPS standard.

The results also show that the implementation of the specification is progressing mainly in Europe. The prevalence of Europe could be a consequence of policies related to the implementation of the European INSPIRE directive (see [29]), which promotes a European SDI with invoke and transformation services. A second relevant factor it that most of the Open Source implementations of the WPS specification are being developed in Europe (deegree, PyWPS, Northº 52, ZOO Project). In general, focused crawls are resilient to biased seeds. Our approach to mitigate this problem has consisted in (i) merging the results of many runs in a single collection, (ii) using different queries and search engines for each run, and (iii) the systematic use of queries that request pages belonging to specific top-level domains (e.g. ".edu", ".ca").

6 Conclusions

The main goals of this research are to provide a reality check of the availability of WPS servers and quantitative data about the use of different features defined in the WPS standard relevant for geospatial Web-based processing. The results show that the availability of services implementing the WPS specification is limited: only 58 out of 9329 OGC Web services found by a focused crawler were WPS instances. We should remark that these results reflect only the surface of the Geoprocessing Web.

Data show the slow adoption of technical and semantic interoperability features of the WPS specification among practitioners. Service providers should provide WSDL descriptions of their services and develop shared application profiles. The identification of processes relevant for geoscience has been difficult due to the lack of documentation and metadata about them. Geoscience service providers should improve the description of their processes to make feasible real interoperability. What we do not know which semantic information should be added to process descriptions, apart from application profiles. The development of profiles that standardize the publication of GRASS processes could be a good starting point. Focused crawlers might help solve that issue by building collections of geoprocessing services and related resources, such as online documentation and research papers. In-depth studies of such collections could identify those missing pieces of information that are necessary for enabling semantic interoperability, especially between geoscience systems.

Acknowledgements. This work has been partially supported by Spanish Government by the projects TIN2009-10971, PET2008_0026 and "España Virtual" ref. CENIT 2008-1030 (through contracts with the National Centre of Geographic Information (CNIG) and GeoSpatiumLab S.L.) and GeoSpatiumLab S.L. The work of Walter Rentería-Agualimpia has been partially supported by a grant (ref. B181/11) from the Aragon Government.

References

[1] Kiehle, C., Greve, K., Heier, C., 2007. Requirements for next generation spatial data infrastructures-standardized web based geoprocessing and web service orchestration. Transactions in GIS 11 (6), 819—834.

[2] Yue, P., Gong, J., Di, L., Yuan, J., Sun, L., Sun, Z., Wang, Q., 2010. GeoPW: Laying blocks for the geospatial processing web. Transactions in GIS 14 (6), 755—772.

[3] Li, W., Yang, C., Yang, C., 2010a. An active crawler for discovering geospatial Web services and their distribution pattern—a case study of OGC WMS. Int. J. of Geographical Information Science 24 (8), 1127—1147.

[4] Lieberman, J. (Ed.), January 2003. OpenGIS Web Services Architecture v0.3. No. OGC 03-025. Open Geospatial Consortium, Inc., http://portal.opengeospatial.org/files/?artifact_id=1320 [accessed 23 August 2011].

[5] Whiteside, A., Greenwood, J. (Eds.), April 2010. OGC Web Services Common Standard v2.0.0. No. OGC 06-121r9. Open Geospatial Consortium, Inc., http://portal.opengeospatial.org/files/?artifact_id=38867, [accessed 23 August 2011].

[6] Lopez-Pellicer, F. J., Florczyk, A. J., Béjar, R., Muro-Medrano, P. R., Zarazaga-Soria, F. J., 2011. Discovering geographic web services in search engines. Online Information Review 35 (6), in press.

[7] Nebert, D. (Ed.), 2004. Developing Spatial Data Infrastructures: The SDI Cookbook, v2.0. Global Spatial Data Infrastructure, http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf, [accessed 23 August 2011].

[8] Friis-Christensen, A., Ostländer, N., Lutz, M., Bernard, L., 2007. Designing service architectures for distributed geoprocessing: Challenges and future directions. Transactions in GIS 11 (6), 799—818.

[9] Han, W., Di, L., Zhao, P., Wei, Y., Li, X., 2008. Design and implementation of GeoBrain Online Analysis System (GeOnAS). In: Bertolotto, M., Ray, C., Li, X. (Eds.), Web and Wireless Geographical Information Systems. LNCS 5373. Springer Berlin / Heidelberg, pp. 27—36.

[10] Nogueras-Iso, J., Zarazaga-Soria, F. J., Muro-Medrano, P. R., 2005. Geographic Information Metadata for Spatial Data Infrastructures: Resources, Interoperability and Information Retrieval. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

[11] Al-Masri, E., Mahmoud, Q., 2007. Interoperability among Service Registry Standards. Internet Computing, IEEE 11 (3), 74—77.

[12] Goodchild, M., 2007. Citizens as sensors: the world of volunteered geography. GeoJournal 69 (4), 211—22.

[13] Turner, A., 2006. Introduction to neogeography. O'Reilly Media, Inc.
[14] Al-Masri, E., Mahmoud, Q. H., 2008. Discovering Web Services in Search Engines. IEEE Internet Computing 12 (3), 74—77.
[15] Refractions Research, October 2006. OGC Services Survey. white paper, Refractions Research, Victoria, Canada. http://www.refractions.net/expertise/whitepapers/ogcsurvey/, [accessed 23 August 2011].
[16] Schut, P. (Ed.), June 2007. OpenGIS Web Processing Service v1.0.0. No. OGC 05-007r7. Open Geospatial Consortium, Inc., http://portal.opengeospatial.org/files/?artifact_id=24151, [accessed 23 August 2011].
[17] Chen, N., Di, L., Yu, G., Gong, J., 2010. Geo-processing workflow driven wildfire hot pixel detection under sensor web environment. Computers & Geosciences 36 (3), 362 — 372, doi:10.1016/j.cageo.2009.06.013.
[18] Hersovici, M., Jacovi, M., Maarek, Y. S., Pelleg, D., Shtalhaim, M., Ur, S., 1998. The shark-search algorithm. an application: tailored web site mapping. Computer Networks and ISDN Systems 30 (1-7), 317—326.
[19] Cho, J., Garcia-Molina, H., Page, L., 1998. Efficient crawling through URL ordering. Computer Networks and ISDN Systems 30 (1-7), 161—172.
[20] Schut, P., Whiteside, A. (Eds.), September 2005. OpenGIS Web Processing Service v0.4.0. No. OGC 05-007r4. Open Geospatial Consortium, Inc., http://portal.opengeospatial.org/files/?artifact_id=12184, [accessed 23 August 2011].
[21] Williams, M., Cornford, D., Bastin, L., Jones, R., Parker, S., March 2011. Automatic processing, quality assurance and serving of real-time weather data. Computers & Geosciences 37 (3), 353—362.
[22] Brauner, J., 2008. Providing GRASS with a Web Processing Service Interface. In: GI-Days 2008. No. 32 in IfGIprints. IfGI, Münster, http://www.gi-tage.de/archive/2008/downloads/acceptedPapers/Papers/Brauner.pdf [accessed 23 August 2011].
[23] Li, X., Di, L., Han, W., Zhao, P., Dadi, U., aug. 2009. Sharing and reuse of service-based geospatial processing through a Web Processing Service. In: 17th International Conference on Geoinformatics, 2009. pp. 1-5.
[24] Göbel, R., Zipf, A., 2008. How to define 3D geoprocessing operations for the OGC Web Processing service (WPS)? Towards a classification of 3D operations. In: Gervasi, O., Murgante, B., Laganà, A., Taniar, D., Mun, Y., Gavrilova, M. (Eds.), Computational Science and Its Applications — ICCSA 2008. Vol. 5072 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg, pp. 708—723.
[25] Janowicz, K., Schade, S., Bröring, A., Keßler, C., Stasch, C., 2009. A transparent semantic enablement layer for the Geospatial Web. In: Terra

Cognita 2009 Workshop In conjunction with the 8th International Semantic Web Conference (ISWC 2009), October 26.

[26] Nash, E., Jun. 2008. WPS application profiles for generic and specialised processes. In: Pebesma, E., Bishr, M., Bartoschek, T. (Eds.), GI-Days 2008 : Proc. of the 6th Geographic Information Days. IFGI, University of Münster.

[27] INSPIRE Network Services, 2010. Technical guidance for INSPIRE coordinate transformation services v2.1. Draft, INSPIRE Drafting Team "Network Services", http://inspire.jrc.ec.europa.eu/documents/Network_Services/INSPIRE_Draft_Technical_Guidance_Coordinate_Transformation_Services_(version_2%201).pdf [accessed 23 August 2011].

[28] Kang, L., Wu, Q., Yuan, Y., 2010. A distributed LIDAR processing model based on OWS and BPEL. In: IEEE Int. Geoscience and Remote Sensing Symposium (IGARSS 2010). IEEE Computer Society, pp. 3628—3631.

[29] Masser, I., 2007. Building European Spatial Data Infrastructures. Esri Press.