WEB SERVICE FOR POSITIONAL QUALITY ASSESSMENT: THE WPS TIER

E. M. A. Xavier a, b, *, F. J. Ariza-López b, M. A. Ureña-Cámara b

a Brazilian Army Geographic Service, QGEx, 70630-901 Brasilia, Brazil – (emerson)@dsg.eb.mil.br
b Dept. of Cartographic Engineering, University of Jaen, 23071 Jaén, Spain – (fjariza, maurena)@ujaen.es

Commission II, WG II/4

KEY WORDS: data quality, quality evaluation, web service, positional control, TerraLib, WPS

ABSTRACT:

In the field of spatial data every day we have more and more information available, but we still have little or very little information about the quality of spatial data. We consider that the automation of the spatial data quality assessment is a true need for the geomatic sector, and that automation is possible by means of web processing services (WPS), and the application of specific assessment procedures. In this paper we propose and develop a WPS tier centered on the automation of the positional quality assessment. An experiment using the NSSDA positional accuracy method is presented. The experiment involves the uploading of the client of two datasets (reference and evaluation data). The processing is to determine homologous pairs of points (by distance) and calculate the value of positional accuracy under the NSSDA standard. The process generates a small report that is sent to the client. From our experiment, we reached some conclusions on the advantages and disadvantages of WPSs when applied to the automation of spatial data accuracy assessments.

1. INTRODUCTION

Actually there are lots of geospatial data sources available to generate data almost instantaneously. Imagery from aerial or satellite platforms, and the popularization of Unmanned Aerial Vehicle (UAV), or ‘drones’, has allowed generate geospatial datasets in an unmanageable way, what some authors named ‘big data’ trend (Crampton et al., 2013). Other relevant data source is the crowdsourced data, generated by volunteers almost daily (Neis and Zielstra, 2014). This overload data scenario brings new challenges for the official spatial data suppliers, or National Mapping and Cadastral Agencies (NMCA). Traditionally, these institutions create and manage authoritative datasets in a standardized way. However, today many data ‘producers’ intend to represent the same phenomena, geospatial features, following their own rules. This new scenario may lead users questioning the quality of available datasets: ‘which one does fit my purposes?’, a fitness for use issue (Servigne et al., 2006).

In these cases, few or nothing information about the quality of a spatial dataset is available, so we believe would be interesting exist a web service with the capability of assess the quality of a test dataset against a reference dataset. A data quality validation service is an appealing topic in the geospatial research agenda which has been developed in current projects (Kruse, 2014).

The geospatial data quality assessment is commonly executed by means of a coordinated set of processing instructions. Inside a Spatial Data Infrastructure (SDI) environment, where the interoperability is a requirement, those processes would be encapsulated into a standardised way, like the Open Geospatial Consortium (OGC) specification Web Processing Service (WPS) (Schut, 2007). However, some authors argue the focus of a SDI still remain in data providing instead of data processing (Hofer, 2013). Díaz et al. (2012) argued that the integration of geoprocessing functionalities in a SDI environment is an open challenge. On the other hand, Masó et al. (2012) point out the versatility of WPS in a SDI. For these authors, WPS is ready to encapsulate practically every kind of geospatial process.

Considering that quality assessment procedures for geospatial data can be published and executed by means of a standardised interface over the web, we delineated a first version of a web service for the automation of positional quality assessment (Ariza-López et al., 2015). In this paper we present the WPS part of our three-tier architecture for this web service.

According to Brauner et al. (2009), there are two types of research about geoprocessing services: research focused on generic problems (e.g. performance), and research focused on some specific application domain (e.g. spatial statistics). Our research can be classified into the second group, since we are working on an application-oriented investigation.

This paper is structured as follows. Section 2 provides a general background on geoprocessing over the web and a brief exposition of related work in the automation of quality assessment. Section 3 describes the WPS tier of our architecture for on-line quality assessment. Section 4 describes a prototype for positional quality assessment and this section also discusses the results of our experiment. Finally, section 5 brings some conclusions and further work.

2. BACKGROUND

Web Processing Service (WPS) is an OGC standard that defines an interface aiming to publish and to use geospatial processes, as well as the discovery of and the binding to those processes by clients (Schut, 2007). In the early March 2015, OGC consortium released the version 2.0 of this specification (Mueller and Pross, 2015). The WPS interface is defined by means of six operations: three mandatory and included in WPS 1.0 – GetCapabilities, DescribeProcess and Execute; two optional operations to handle with asynchronous processes – GetStatus and GetResult; and the Dismiss operation. This last operation is defined in the Dismiss extension, and it can be used in both situations:

* Corresponding author

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.

Editors: A.-M. Olteanu-Raimond, C. de-Runz, and R. Devillers
doi:10.5194/isprsannals-II-3-W5-257-2015
synchronous and asynchronous jobs. These operations are presented in the sequence diagram of Figure 1.

![Sequence diagram for WPS operations.](image)

According to Mueller and Pross (2015), GetCapabilities operation returns the service metadata and a list of process available at the server. DescribeProcess provides detailed information about a list of selected processes. Execute is the key-operation of a WPS – it permits a client to execute some process given a list of parameters. GetStatus is an operation that allows a client to query status information of some asynchronous processing job. GetResult operation allows a client to recover the final result of an asynchronous job. Lastly, Dismiss operation permits a client to cancel an asynchronous job, or for finished jobs this operation will release all allocated resources, like temporary files or result files.

The research community around the Geographic Information Systems (GIS) environment has investigated some aspects of the on-line processing of geospatial data for a variety of purposes. Kiehle et al. (2007) developed an open-source WPS server, the degree WPS, in order to analyse the applicability of this specification. The authors concluded that including complex processing tasks, like a model for global climate change, can be encapsulated inside a WPS. Brauner et al. (2009) proposed three main topics for a research agenda of geoprocessing services: (1) service orchestration strategies, (2) semantic description of processes, and (3) improve the performance of these services. These authors also argued the WPS interface provides an efficient communication mechanism using its asynchronous messages capabilities.

Fris-Christensen et al. (2009) introduced the term Distributed Geographic Information Processing (DGIP) while developing an on-line application for forest fire risk analysis. Their architecture was based on OGC specifications, among them the WPS, and service orchestration provided by the Web Services Business Process Execution Language (BPEL) (OASIS, 2007). Some authors have argued the BPEL has becoming the de-facto standard for service chaining (Akram et al., 2006). Biodiversity applications also have demanded on-line geoprocessing tools. In this sense Fook et al. (2009) developed a conceptual framework that enables the collaboration in biodiversity by allowing sharing species distribution modelling experiments. Granell et al. (2010) took advantage of the standard WPS interface to develop an open architecture that allows the calibration and the running of hydrological models.

Other interesting study is due to Zhao et al. (2012). The authors introduced the term geoprocessing web as a broader concept that covers all aspects toward distributed and collaborative geoprocessing over the web. Interoperability is one of the characteristics of the geoprocessing web, and the WPS specification plays this role. Hofer (2013) evaluated the commonalities and differences between geoprocessing web (Zhao et al., 2012) and geospatial cyberinfrastructures (Yang et al., 2010). The author concluded that both concepts have the function of data analysis and knowledge generation, and also encompass the resource of distributed geoprocessing and web services.

The automation of quality control for spatial data has also shown recent works using the WPS interface. The study of Donaubauer et al. (2008) proposed a web service with the ability to generate quality information of assessed data via web services. The work used WPS to process the quality control, and ISO 19115 (ISO, 2003) for the quality report by means of metadata elements. Despite the simplicity of the quality procedure, just an overlay of previously tagged data with some quality elements, this study seemed to be the first attempt of an automatic evaluation service in the literature. A more recent study also indicated the quality evaluation can be executed through a WPS (Mobasheri, 2013).

In our research Group emerged a successful research focused on the automation of the positional accuracy evaluation, due to Ruiz-Lendínez et al. (2013). The authors proposed a solution for automatic positional accuracy assessment of polygonal features using a matching approach. The proposed methodology was able to increase significantly the number of features used in the quality evaluation procedure.

### 3. WPS TIER FOR QUALITY CONTROL

In our previous work (Ariza-López et al., 2015) we presented the first design of a web service for the automation of positional quality control of spatial datasets. We proposed a three-tier architecture as shown in Figure 2. In this section we detail the WPS tier of the proposed architecture by presenting the design concepts.

Quality evaluation procedures often evolve complex tasks and people from different organizations or departments. Facing this situation we have two design principles: interoperability and simplicity. The interoperability principle indicates the WPS tier should follow the WPS specification and schemas in order to permit a standardised way of communication. The simplicity principle leads us to avoid unnecessary issues in the processing itself, so the processing ‘part’ should be as straight as possible. The WPS tier should manage all communication issues, validation procedures, and client-server tasks. Therefore the
designer of the Evaluation tier can only focus on assessment procedures.

In addition to the classes described in the WPS specification we propose the creation of three new interfaces: AbstractProcess, AbstractComplexData, and AbstractExecuteResponse.

AbstractProcess is an interface that all concrete process should implement in order to permit its use under the architecture. The interface is represented in Figure 3. The abstract class has one attribute: the description of the process using the WPS semantic. The interface has two concrete methods: `getLanguages` and `getDescription`; and two pure virtual methods: `execute` and `createDescription`. The `getLanguages` method is used for GetCapabilities operation, and the `getDescription` is used in all operations to return a summarized description for the process (in GetCapabilities), or a more complete description for the DescribeProcess operation and the Execute operation.

A concrete process should implement `createDescription` and `execute` methods. The `createDescription` returns a full description of the process, which can be hard-coded in the implementation, or can be read from a configuration file, like the deegree WPS (Kiehle et al., 2007). The `execute` method effectively runs the processing what the implementation was designed to do. It is possible to note that the `execute` method does not return an ExecuteResponse object but an array of Data objects. The goal is to avoid that the processing handles the final response, but just run its job and return the processed data. In this architecture we are using the design pattern Abstract Factory (Gamma et al., 1995) in order to manage the processes in a server. So, the processes should be registered into a 'factory' prior to be used.

AbstractComplexData is the interface for data drivers, like ESRI Shapefiles, Geography Markup Language (GML), or imagery in GeoTIFF. AbstractExecuteResponse is the interface for response to an Execute operation request. This interface has two concrete implementations: ExecuteResponse and RawDataResponse. This is necessary because the final response of a processing task may be or a standard ExecuteResponse either a raw data response, in some predefined format, if the client requests in this way. This is other reason because the execute method in AbstractProcess interface returns an array of Data instead of an ExecuteResponse.

When the WPS server receives an Execute request it acts as shown in the sequence diagram in Figure 4.

A concrete process should implement `createDescription` and `execute` methods. The `createDescription` returns a full description of the process, which can be hard-coded in the implementation, or can be read from a configuration file, like the deegree WPS (Kiehle et al., 2007). The `execute` method effectively runs the processing what the implementation was designed to do. It is possible to note that the `execute` method does not return an ExecuteResponse object but an array of Data objects. The goal is to avoid that the processing handles the final response, but just run its job and return the processed data. In this architecture we are using the design pattern Abstract Factory (Gamma et al., 1995) in order to manage the processes in a server. So, the processes should be registered into a 'factory' prior to be used.

AbstractComplexData is the interface for data drivers, like ESRI Shapefiles, Geography Markup Language (GML), or imagery in GeoTIFF. AbstractExecuteResponse is the interface for response to an Execute operation request. This interface has two concrete implementations: ExecuteResponse and RawDataResponse. This is necessary because the final response of a processing task may be or a standard ExecuteResponse either a raw data response, in some predefined format, if the client requests in this way. This is other reason because the execute method in AbstractProcess interface returns an array of Data instead of an ExecuteResponse.

When the WPS server receives an Execute request it acts as shown in the sequence diagram in Figure 4.

A concrete process should implement `createDescription` and `execute` methods. The `createDescription` returns a full description of the process, which can be hard-coded in the implementation, or can be read from a configuration file, like the deegree WPS (Kiehle et al., 2007). The `execute` method effectively runs the processing what the implementation was designed to do. It is possible to note that the `execute` method does not return an ExecuteResponse object but an array of Data objects. The goal is to avoid that the processing handles the final response, but just run its job and return the processed data. In this architecture we are using the design pattern Abstract Factory (Gamma et al., 1995) in order to manage the processes in a server. So, the processes should be registered into a 'factory' prior to be used.

AbstractComplexData is the interface for data drivers, like ESRI Shapefiles, Geography Markup Language (GML), or imagery in GeoTIFF. AbstractExecuteResponse is the interface for response to an Execute operation request. This interface has two concrete implementations: ExecuteResponse and RawDataResponse. This is necessary because the final response of a processing task may be or a standard ExecuteResponse either a raw data response, in some predefined format, if the client requests in this way. This is other reason because the execute method in AbstractProcess interface returns an array of Data instead of an ExecuteResponse.

When the WPS server receives an Execute request it acts as shown in the sequence diagram in Figure 4.
4. PROOF OF CONCEPT: THE NSSDA SERVICE

In order to validate our proposal for the WPS tier presented in Section 3, we developed a web service for quality assessment of positional accuracy using the methodology described in the National Standard Spatial Data Accuracy (NSSDA) (FGDC, 1998).

4.1 Experiment

We developed our WPS tier using a set of classes and functions for Web-GIS development built over the TerraLib library (Câmara et al., 2008), an open-source software. This tier was developed taking into account the WPS version 1.0 (Schut, 2007).

The core of NSSDA service is the PointEvaluation class, an implementation of the AbstractProcess interface. Considering that the NSSDA procedure is applied over pairs of homologous points, from a reference and a test site, the first task is to perform a matching between reference and test datasets. For this purpose we adopted a simple solution using the nearest neighbour strategy taking into account only 1:1 matches. Since our objective here is assess the WPS tier, we chose this effortless matching approach for simplification purposes. After the matching, the calculus procedure runs straightforward, and the execution returns a double value (in meters) that represents the result of the NSSDA evaluation.

For this experiment we prepared two datasets of point data in the Shapefile format, with approximately 40 points in each one. Then we created a simple HTML5/Javascript client able to convert the data into a WPS execute request. The simple client was used to encode the Shapefile data into base64 encoding (Josefsson, 2006), mount the execute request, send it to server, and receive the response. Figure 5 brings an extract of the returned response in XML.

```xml
<wsps:ExecuteResponse service="WPS"/>
<wsps:Process wsps:processVersion="1.0.0">
  <ows:Identifier>PointEvaluation</ows:Identifier>
  <ows:Title>Point Evaluation - Data Quality</ows:Title>
  <ows:Process>
    <ows:Status creationTime="2015-03-27T13:59:56Z">
      <ows:ProcessSucceeded>Done</ows:ProcessSucceeded>
    </ows:Status>
    <ows:ProcessOutputs>
      <ows:Output>
        <ows:Identifier>NNSDA</ows:Identifier>
        <ows:Title>NNSDA result as real</ows:Title>
        <ows:Data>Positional accuracy value</ows:Data>
      </ows:Output>
    </ows:ProcessOutputs>
  </ows:Process>
</wsps:ExecuteResponse>
```

Figure 5. Response of an execute request to the NSSDA service.

The positional accuracy value returned by the server was calculated following the NSSDA methodology. This value represents the horizontal positional accuracy of tested data against the reference data at 95% confidence level.

4.2 Discussion

The development of the WPS tier of our architecture aroused some aspects of the specifications (1.0 and 2.0) and the applicability of WPS while a service interface facing quality evaluation. These aspects can be divided into strengths and weaknesses.

The identified WPS strengths were:

- Multiple inputs and outputs: the WPS interface does not limit anyway the quantity, type, or format for inputs or outputs. This flexibility permits that a quality evaluation service created on top of a WPS framework can be able to generate various interrelated quality outputs, for example:
  - Quality report in PDF using some template of the evaluator;
  - DQ_DataQuality or DQ_Element from ISO 19157 (ISO, 2013) encoded in XML, or according the legacy ISO 19115 (ISO, 2003);
  - Some literal value (like the NSSDA service) as a part of a quality evaluation chain.

- Ready for service chaining: the specification indicates some options and previous research pointed out its feasibility (Kiehle et al., 2007, Friis-Christensen et al., 2009). The inputs and outputs for data processing can be accessed as on-line resources, for example:
  - The reference data in a positional quality evaluation can be a file available on-line for download;
  - The test data in the same situation can be distributed by means of a Web Feature Service (WFS);
  - Some parameter in a quality evaluation procedure, like the NSSDA result, can be obtained in other WPS server.

- Process extension is relatively easy: any extension to some process can take advantage of the entire framework. We can take as example the current experiment:
  - The initial class can be split in two: Match for matching tasks; and PointEvaluation for the evaluation calculus themselves. Each one can run its own different strategies;
  - Match can use one of the many matching techniques available in the literature, like geographic context (Samal et al., 2004), belief theory (Olteanu Raimond and Mustière, 2008), or regression model (McKenzie et al., 2014).

Despite the WPS presents many advantages, in this work we have identified some open issues:

- There is no direct way to indicate local data in the processing:
  - Context: sometimes the differential of a service may be its local data, which for some reason (e.g. license,
security) cannot be available in the web. This ‘reserved’
dataset cannot be distributed, but it can be used in
processing jobs, like a reference dataset in quality
evaluation procedures;
• Status of the issue: WPS permits send data, or reference
remote data, without the provision of local data. The
latest version of specification (WPS 2.0) also does not
forecast the use of local data, but it uses the concepts of
data for value or for reference.
• There is a validation problem in the schema
wpsDescribeProcess response of WPS 1.0:
• Description: this schema does not have the attribute
’elementFormDefault’ from XML Schema specification
set to ‘qualified’, as we can expect, since many other
OGC schemas have this attribute. This leads us to an
invalid XML when sending a DescribeProcess response
using XML namespace prefixes;
• Status of the issue: there is no reference to this
behaviour in the WPS 1.0 specification. This issue was
solved with the release of WPS 2.0 schemas.

5. CONCLUSIONS

Nowadays the geographic information community faces a huge
availability of geospatial data. The automation of quality
assessment for these data is a challenge because manual
procedures are becoming infeasible. With this goal in mind we
are working in the design of a web service for automatic
positional quality assessment.

In this paper we present the WPS tier of our architecture for the
quality evaluation web service. The main contribution of this
work is to confirm that a WPS server can be used to automate
some positional quality evaluation procedures, in our
experiment the NSSDA methodology. Other contribution is
presenting and discussing some WPS advantages and
drawbacks while its use in quality assessment, resulting from
our experiment.

The WPS tier is part of a larger research focused on the
automation of quality evaluation. Many questions remain open
to the next stages of our research. An immediate problem
concerns to deal with more complex matching approaches,
notably for linear (Mustière and Devogele, 2008) and areal
(Ruiz-Lendínez et al., 2013) features.

With the new release of WPS 2.0 specification in March 2015,
we plan to adapt our architecture to this new model. Despite
there are many similarities between versions, probably some
adjust will need to be done. In the new version is possible to
note that our AbstractProcess view of a WPS process is
approaching of the current view where the process model was
widely decoupled from the WPS protocol.

Despite of we are initially focused on the positional quality
assessment, we believe this architecture can also be applied to
assess other quality component of geospatial data. Further work
includes: (1) investigate data matching techniques; (2) include
other point-based methodologies (e.g. ASCE, 1983, ASPRS,
2015); (3) implement line-based methods of positional
assessment (e.g. Goodchild and Hunter, 1997); (4) embody
other results to the current service (e.g. number of assessed
points); (5) develop other quality outputs (e.g. ISO 19157).

ACKNOWLEDGEMENTS

This work has been funded by the Ministry of Science and
Technology (Spain) and the European Regional Development
Fund under grant no. BIA2011-23217. The authors also wish to
acknowledge the support of the Brazilian Army's Department
of Science and Technology, and the support of the Regional
Government of Andalucia (Spain) for the Research Group
"Ingenieria Cartográfica" (TEP-164).

REFERENCES

Akram, A., Meredith, D., and Allan, R., 2006. Evaluation of
BPEL to scientific workflows. In: Sixth IEEE International
269–272.

2015. Proposal of a web service for positional quality control
of spatial data sets. In: International Workshop on Spatial Data
and Map Quality, Valletta, Malta.

ASCE, 1983. Map uses, scales and accuracies for engineering
and associate purposes. American Society of Civil Engineers
(ASCE).

ASPRS, 2015. ASPRS positional accuracy standards for digital
geospatial data. Photogrammetric Engineering & Remote

Towards a research agenda for geoprocessing services. In: 12th
AGILE International Conference on Geographic Information
Science, Leibniz, Germany, pp. 1–12.

Cámara, G., Vinhas, L., Ferreira, K.R., Queiroz, G.R., Souza,
R.C.M., Monteiro, A.M.V., Carvalho, M.T., Casanova, M.A.,
for large-scale environmental and socio-economic applications.
In: G.B. Hall and M.G. Leahy, eds. Open Source Approaches in
Spatial Data Handling, Springer Berlin Heidelberg, pp. 247–
270.

Crampton, J.W., Graham, M., Poorthuis, A., Shelton, T.,
Stephens, M., Wilson, M.W., and Zook, M., 2013. Beyond the
geotag: situating ‘big data’ and leveraging the potential of the
geweb. Cartography and Geographic Information Science, 40
(2), pp. 130–139.

Diaz, L., Remke, A., Kauppinen, T., Degbelo, A., Foerster, T.,
Stasch, C., Rieke, M., Schaeffer, B., Baranski, B., Bröring, A.,
and Wytzsk, A., 2012. Future SDI - impulses from
goinformatics research and IT trends. International Journal of

quality aware web processing service. In: Proceedings of the
6th Geographic Information Days, Münster, Germany, pp.
16–18.

FGDC, 1998. Geospatial Positioning Accuracy Standards - Part
3: National Standard for Spatial Data Accuracy. Federal
Geographic Data Committee (FGDC).

Fook, K.D., Monteiro, A.M.V., Cámara, G., Casanova, M.A.,
and Amaral, S., 2009. Geoweb services for sharing modelling
379–399.


