WEB-BASED SHARING OF A GEO-PROCESSING CHAIN: COMBINATION AND DISSEMINATION OF DATA AND SERVICES

Swati Gehlot $^{(1,2)}$ and Edward Verbree $^{(1)}$

⁽¹⁾ OTB Research Institute for Housing, Urban and Mobility Studies, Section of GIS Technology, Delft University of Technology, Jaffalaan 9, 2628 BX Delft, The Netherlands ⁽²⁾ Delft Institute of Earth Observation and Space Systems, Delft University of Technology, Kluyverweg 1, 2629 HS Delft, The Netherlands Email: (S.Gehlot, E.Verbree)@tudelft.nl

Commission IV

KEY WORDS: Internet GIS, Geo-processing chain, Process sharing, ModelBuilder, OGC-WPS

ABSTRACT:

The recent integration of the internet with geo-information systems (GIS) has evolved the concept of Internet Distributed Geo-Information Services. Numerous research and implementation studies of standards of metadata, geo-libraries, data storage and retrieval etc. are being done giving way to exchange and share of geo-information. However, very little effort has gone into making it possible to share process objects, or representations of processes of GIS use. Currently in most cases, the innovative algorithms for a certain geo-processing technique are developed for a dedicated purpose with a limited testing on the required prototype. One of the main reason of these problems is that we do not have a standard way to package geo-processes and making them available for testing [Regnauld, 2006]. If the geo-processing model itself is shared, and made understandable to the research communities, the risk of duplications and ambiguities in geo-processing can be taken care of. The model sharing would be highly useful to track the sequence of operations, exact criteria of conversion and transformation of spatial data, input/output specifications of processes, model constraints etc. Here we present a framework of a combination of geo-processing services for a dedicated application of land subsidence data visualiza-

tion and interpretation. We combine different geo-data sources to correlate the local data (for example, cadaster, land use, water levels etc.) with PS-InSAR land subsidence data. An ESRI ModelBuilder tool is used to clearly describe the subsidence data classification with respect to the cadaster and height data. The geo-processing model to carry out this visualization and interpretation of PS-InSAR data is described step by step in form of a sequential model.

The future vision of the research is to follow and implement the proposed standards of OGC-WPS architecture to suit our algorithms and eventually share the data and the geo-processing services.

1. INTRODUCTION

The World Wide Web offers a fast and straightforward environment for the sharing of scientific knowledge. When the scientific knowledge is in the form of geographical information, GIS is an outstanding medium of storing, manipulating, visualizing and analyzing the spatial data. The recent integration of internet technology and GIS has produced an expanding area of research called Internet Distributed GIServices. The term GIServices is often used to describe a GI function offered by a server for use by any user connected to the internet [Goodchild et al., 2001]. Most of the latest GIServices are dedicated to spatial data access/dissemination (e.g., gazetteer service of Alexandria digital library at http://www.alexandria.ucsb.edu/), spatial data exploration/transformation/processing (e.g., geocoding services like www.geocode.com), and spatial data catalogue services (e.g., www.geographynetwork.com). Moreover, the increasing availability of online tools such as Google Earth and NASA World Wind are changing the way we interact with the spatial data. According to a recent study by some reporters of the Nature magazine, the virtual globe provided by GoogleEarth is becoming increasingly popular amongst millions of people and a variety of researchers. Ranging from environmental scientists to disaster management teams, this free internet tool of data meshing and visualization is proving to be a boon to share valuable data and increase awareness of GIS's potential.

However, most of the research and the implementation studies of standards of metadata, geo-libraries, data warehouses etc. are devoted to facilitate the sharing of geographic data. The sharing and exchange of "process models" or the "process representations" is ad-hoc and unorganized till date. Numerous geo-processing models are created and used but there are very few implementations to try to connect these process models. In many cases the actual code or program dedicated to a particular application is very short lived due to advances in the scripting language, completion of the project itself, etc. Éventually, in the research community the re-invention of the wheel takes place even when a similar application has been created and used long ago. This in turn leads to the inherent interoperability of the service with many other already available tools to add to the geo-processing chain. Moreover, in many cases, when these repetitions in the processing occurs and the different users use the already available codes without knowing the actual process models, strange outcomes of the geoprocessing is expected.

To avoid the above stated problems in a scientific research, it is very important to consider the idea of sharing the geo-processing algorithms, so that the work proves to be useful to a number of interested researchers. In this study, we present a theoretical approach following the OGC Open Web Services [OGC, 2005] specification for sharing and interpretation of land subsidence time series data. The PS-InSAR (Persistent Scatterer Interferometric Synthetic Aperture Radar PS-InSAR) [Ferretti et al., 2001] technique is a recent application of radar remote sensing, and is considered as an efficient method of detection and monitoring of subtle deformations of the earth's surface. The PS processing is carried out by a analyzing radar time series data and extracting points with a stable and coherent phase behavior over time or, the persistent scatterers.

The interactive and dynamic 3-D visualization of a data set of public interest, such as urban land subsidence data leads to a better dissemination of scientific information to more and more users in a much easier way. Moreover it is a well known fact that the value of scientific information increases the more you link it to information that the users already are aware of. To take this opportunity, this research tries to tie up the results of urban land deformation with other common data sources such as cadaster databases, boundaries of land-use information, foundation and building types etc.

The additional geo-information sources used here are further categorized to different spatial, temporal, quantitative, qualitative, two-and three dimensional data formats, so that the scientists in different areas might relate their own available geo-datasets to test the approach. If we want to combine all possible sources of additional information to achieve our research goal of sharing land deformation data, we come across a particular geo-information process chain of certain steps of algorithm(s) and coding. In most cases, multiple services must be used together to perform a useful function. This is one of the important ongoing work of the Open Geospatial Consortium's Open Web Services (OWS) architecture [OGC, 2005]. Within such a chain, input for most of the services is an output from a previous service in the service chain, for instance, using the coordinate transformation service followed by an interpolation or query service. The future vision of the research is to follow and implement the proposed standards of OGC-WPS architecture to suit our algorithms and eventually share the data and the geo-processing services.

2. GEOPROCESSING MODELS

The complexities of the geographic data analysis for various application oriented services have lead to the concept of combination of geo-processing services for fulfillment of a particular task. Geo-processing services can be defined as services that involve analysis of geographically referenced data sets (input, output, manipulation, storage etc.). A particular geo-processing service would consist of a number of operations carried out in a particular sequence or order and involves the transfer of geo-data so as the derived output from one service serves as input to the next service and so on. The geo-processing service model can be explained as a combination of various tasks such as query service, web mapping service, web feature service, reprojection service, interpolation and extrapolation services etc. A simple example could be invoking an address matching service to perform the unique identification of a place (coordinates or bounding boxes), followed by a mapping service to provide a map corresponding to this location with combination/intersection of various data sets. Also, reprojection services might be used to transform the data from one spatial reference system to another so as to facilitate the client's requested projection.

A geo-processing service can also be described in form of "Process Model". In general, the term model is very vast and vague in its definition, but however, in the present context, we refer to a model as a graphical representation of a repetitive task or application, that is, a particular geo-processing tool. The main intention is to convey the model metadata (description of the model itself). A Geo-processing model can be described in different forms,

• **Graphical models** Graphical models are the simplest representations of geo-processing sequences. The step wise processes, the parameters for processes (input/output), manipulation and transformation steps, decision making, comparison or selection of subsets etc are shown distinctively by the graphical models. In principle, every processing step has a standardized way of representation in the overall geo-processing model. Common examples of graphical model

representations may be flowcharts, concept diagrams, Unified Modeling Language (UML) Schemas, ESRI ModelBuilder etc.

- Scripts Models in form of scripts are a compilation of geoprocessing commands written in a programming language such as C++, Python, shell scripts for unix based GIS etc.
- Other representations Command line syntaxes using Visual Basic tools in ArcGIS 9, map algebra operations etc.

3. SHARING OF GEO-PROCESSING MODELS

Till date, very little effort has gone into making it possible to share process objects, or representations of processes of GIS use. GIS scripts, models and other representation of processes are potentially valuable to many users and are well worth sharing [Goodchild et al., 2001]. Currently in most cases, the innovative algorithms for a certain geo-processing technique are developed for a dedicated purpose with a limited testing on the required prototype. Many researchers publish their algorithms, but the actual code or the executable is often short lived due to reasons like the researcher moves to another university, changes development platform, coding language becomes obsolete etc. A further big impact of this short life of algorithms leads to a limited interoperability with other data sets. The main reason of these problems is that we do not have a standard way to package geo-processes and making them available for testing [Regnauld, 2006].

Every dedicated geo-processing model comprises of a number of operations carried out in a particular sequence and moving the derived data forms to the next service in the chain. Based on a common geometric infrastructure (i.e. reference systems), the needed geo-information is collected, modelled and stored within geo-database management systems. Now we can deduct the desired information and an iterative process of analysis and handling, presentation and interaction is started. This process ends when some new results (new datasets, maps) are derived and that we want to exchange to other users or scientists. These others perform a similar, but for their goal intended, geo-information process chain. This will result in unavoidable duplications within the needed geo-information collection, modelling and storage. And more harmful, they could perform some analysis and data handling with the data under study without a real understanding of what is allowed to do, resulting in non-valid or odd outcomes. If these results are exchanged to the community, without giving a clue on the lineage, more strange results are to be expected, without any mechanism to identify this kind of unreliable consequences.

The need of interoperable and chained web services is becoming more and more vital on account of the factors such as growing roles of GISs in organizations and increasing availability of spatial data with its sharing capabilities [Alameh, 2003]. The maturity of the web and distributed services are required to be harnessed in order to facilitate the concept of "service chaining" and "process sharing". The term service chaining refers to an assembly of geo-processing services combining or pipelining the results to suit a dedicated application. To distribute or share the model or sequence of service chaining itself is termed as process sharing.

Furthermore, most users of traditional GIS use only a small amount of their system functionality, and the service chaining model users would use only the data and services they need without having to install, learn, or pay for any unused functionalities.

4. WAYS TO SHARE GEO-PROCESSING MODELS

An easy way to share the process models is to link these services online and describe the model as to facilitate the repetition of geo-processing or analysis in exactly the same way. However, to deal with the issues of interoperability of used geo-processing services in the service chain are quite a challenge. The design and implementation of standardized protocols to link a variety of services online is under process at the Open Geospatial Consortium [OGC, 2006].

4.1 ESRI Model Builder Tool

This utility function in ArcGIS offers the possibility to use the in-built small geo-processing programs (for example, select, intersect, buffer etc.) to construct various types of complex and dedicated process models defining a particular application. The models are represented as 'Process Flow Diagrams' The flow diagrams created with ModelBuilder are not only a convenient way to construct and modify spatial geo-analysis models but are also an excellent medium to document and present one's models to others. When complex geo-processing is carried out, in most cases, it is difficult to keep a track on the assumptions, data features, parameter values, and particular tools etc. Creation of defined models in form of data and work flow helps in carrying out a repetition of a task in exactly the same way, and hence "automates the geo-processing". The ModelBuilder (www.esri.com) lets the user save models and re-run them using different input data (for instance, calculations done for the next observation epoch). Users can copy portions of their model and smaller models can be combined to build big and complicated models. Further, these automated work flows can be exported to open source programming language such as Python, which are a collection of operations in a low level programming language.

In the present context, the most important possibility which is interesting for us is the sharing of models to others via Model-Builder. This tool allows the model sharing by means of creation of a template for processing specific type of data. The distribution of these template models allows the different users to add their own data to the model and run it using a "consistent" or prescribed modeling strategy. This feature allows the sharing of the model keeping in an account of the model constraints.

4.2 Concept of Web Processing Services (WPS)

The OGC initiative of WPS (Web Processing Services) specifications has given a new concept of process chaining and sharing across a network. A WPS can be configured to offer different types of GIS functionalities to numerous users across the internet network. These GIS functionalities can be the different small and dedicated geo-processing services with pre programmed computations and/or calculations models that can operate on spatially referenced data. The combination of these geo-processing services suited to a particular application define the Process Model for that application. The interested users can use whole or part of the process model to apply, test and customize their own applications. The OGC-WPS interface specification provides mechanisms to identify the spatially referenced data required by the calculation, initiate the calculation, manage the output from the calculation so that it can be accessed by the client. Both raster and the vector data are included in the specification [OGC, 2005].

As described in the latest specifications draft document for OGC-WPS, there are three mandatory operations (interface) that can be requested by an online client and performed by a WPS server,

- **GetCapabilities** This operation allows a client to request and receive back the information about the service metadata. The return of such a service consists of a (XML) document that describes the abilities of the specific server implementation.
- **DescribeProcess** A client can request and receive back the details of the processes that can be executed by the server using the DescribeProcess operation. Essentially, the process descriptions includes the input/output data parameters and their respective formats.
- **Execute** This operation allows a client to run a specified process implemented by the WPS, using provided input parameter values and returning the produced outputs.

5. TEST CASE STUDY

A well defined user case has been tested and partly implemented in the present work to show a dedicated example of using standard geo-processing services and their combination(s). We present a scenario of geo-processing of urban land subsidence data with other supplementary geodata sources to visualize and interpret the deformation data. The land subsidence data is computed using radar interferometric time series analysis (PS-InSAR) [Ferretti et al., 2001]. This data consists of individual locations (x,y coordinates) of radar pixels that show a coherent phase behavior over a time of 10 years. These pixels are referred as Persistent Scatterers (PS) and individual time series behavior of deformation is computed for each PS. At the end of the PSI processing chain, the available database of estimated parameters consists of locations (X, Y), relative topographic heights, linear displacement rate, ensemble coherence, displacement time series, and the atmospheric signal time series for an individual PS.

In principal, this land subsidence data consists of well defined locations that undergo slow and small deformations over long period of time. All the subsidence information about the PS is relative to an arbitrary reference point selected in the master image of interferometric analysis. However, these highly precise relative deformation measurements of PS are difficult to interpret in terms of what they represent physically on ground. One of the reasons is that the radar system resolution is 20 meters by 4 meters and the radar reflection signal is a resultant of all the objects in the resolution cells. To handle this problem of interpretation of the land subsidence data we propose to add supplementary geoinformation about the area. The first idea is to classify the land subsidence data according to the land-use or the cadaster classes of geo-data. In the present case we use following types of additional geo-data sources.

5.01 PS-InSAR Data - Time series of Persistent Scatterers The PS database is comprised of PS locations that are geocoded and projected into Dutch RD (Rijksdriehoeksmeting) coordinate system. For each individual PS, a time series of linear displacement velocity (millimeters per year) lies in the database. In principle, each PS is a point vector consisting of the attached topographic height, velocity and deformation time series as its attributes.

5.02 Large scale base map of the Netherlands (GBKN) The large scale base map abbreviated in Dutch from 'Grootschalige BasisKaart Nederland' (GBKN), gives the most detailed topographic mapping [GBKN, 2005]. These maps are a joint effort of the municipalities (Gemeente), Kadaster, and utility companies to produce a large scale base map. The scale of GBKN maps are

1:500 or 1:1000 in urban areas and, 1:2000 in rural areas. The GBKN dataset is in form of line vector defining the boundaries of buildings and built up areas. Utility companies use this map as the backdrop layer, so the accuracy is high but the content is limited to what is needed as a reference (for example, setting of the cables around buildings and roads).

5.03 Top10 Products The Top10 vector maps are maintained by the topographical department of the Dutch Land Registry Office [Top10 Vector, 2005]. This land use data is used in form of line, and polygon vectors. The boundaries of railroads and infrastructures are represented using line vector. The areas of roads, water, land use, buildings, grasslands, woods etc. are represented by polygon vectors. Although the scale of the Top10 maps are 1:10,000, and given this scale the buildings are gathered into a cluster of buildings, the Top10 data is useful as it has coverage on the area and texture information in addition to the point, line details for (rail) roads, infrastructures, water bodies etc.

5.04 Actual height model of the Netherlands (AHN) The AHN or the actual height model [AHN, 2005] of the Netherlands is very useful in terms of combination of PS. The AHN map is a high density laser scanning point set describing terrain heights, with on average at least one height per 16 square meters. As the laser points are often reflected on the roof tops or tree tops, the points in the raw data set are not referring to the terrain but mostly refer to man made features. In our research the raw dataset is of more importance as it contains the details of buildings (which are the most potential PS points as well). The data point heights are however interpolated to the resolution of 5 meters. We use this data in form of a raster grid of $5m \times 5m$.

5.1 Use of ModelBuilder: Rotterdam, Netherlands

The city of Rotterdam in the Netherlands is studied for testing the model sharing approach. This study includes the formation of a certain geo-processing model dedicated to classification of PS data with respect to cadaster and height data. However, it is worthwhile to mention that the presented model is an example case and does not represent the "best model" to show this application. The decision making steps at various stages in the geo-processing algorithms could be subjective based on who is carrying out the analysis.

5.11 Description of geo-processes The dedicated process model for PS data classification is shown in Figure 1. As seen from the figures, there are three geo-process that are employed in the model (Process 1, Process 2, and Process 3). The following paragraphs list the description of these geo-processes in a way that is standardized and recognized by OGC specifications of WPS. Presently we explain the processes in a way recognizable for humans (i.e., natural language). These paragraphs do not show the processes in a machine understandable language (i.e., XML codes) which can be recognized by web services. The XML encoding of the below described processes is a requisite for the implementation of the WPS model. To differentiate our theoretical description comparison with OGC specified terminology, we use the term GetAbility to denote GetCapability, DescribeMethod to show DescribeProcess and Run to show Execute operation.

1. Process 1: Calculation of Near Distances

• **GetAbility** (GetCapabilities) — This operation carries out the calculation of distances of the point feature data to the nearest polyline feature.

- **DescribeMethod** (DescribeProcess) The PS point data and the GBKN building boundary data are input data for this process. The point feature data is a vector data consisting of locations and attributes of the PS deformation locations. The GBKN building boundaries is a vector shapefile of polylines. The distances of the point data to their nearest polyline feature is computed using this process and the computed distances are stored as attributes to the PS database.
- **Run** (Execute) This process is run using the proximity analysis toolset in ArcGIS 9.1. The input parameters are point and line data and the output comprises of an updated database for the point (or, the PS) data . This new PS database is stored as a "Derived Dataset 1" from the execution of Process 1.

2. Process 2: Selection Based on Attributes

- **GetAbility** (GetCapabilities)— This operation is used for selection of the input point feature data by comparison of a selected attribute.
- **DescribeMethod** (DescribeProcess)— The derived set of updated PS point data serves as an input data for this process. This point feature data is a vector data consisting of locations and attributes of the PS deformation locations. The attribute of near distances of the points to their nearest polyline feature is compared using this process. The points fulfilling the comparison criteria are selected. In the current study we select the PS points that lie within a distance of ± 2 meters from the GBKN building boundaries.
- **Run** (Execute)— This process is run using the selection based proximity analysis toolset in ArcGIS 9.1. The input parameter is the "Derived Dataset 1" and the output comprises of a selected set of PS points as "Derived Dataset 2".

It is important to mention here that in the present case, we repeat the Process 2 for further steps in the model. This data selection tool is used at various steps and from the graphical representation, the input/output data characteristics and the execution criteria are clearly stated in the model diagram.

3. Process 3: Extract Values to Point Intersection

- **GetAbility** (GetCapabilities)— This operation is used for Intersection of the selection of the input point feature data with a grid raster dataset of height model.
- **DescribeMethod** (DescribeProcess)— The "Derived Dataset 2" of selected PS point data and the height model of the city are taken as input data for this process. The selected or derived point feature data is a vector data consisting of locations and attributes of the PS deformation locations. The height model of the city comprises of a grid raster of size 5m×5m, and each grid has the height value as its attribute. This process computes the intersection of the point data and the raster grid and updates the point database by adding a nearest height values a new attribute. We consider the closest AHN grid pixel as a representative height of the PS.
- **Run** (Execute)— The raster analysis toolset is used in the execution of this geo-process. The input parameters are point locations and the raster grid values of 5×5 meters. This process basically uses the "Derived

Dataset 2" and the new input data of height model and its execution creates the output of PS point heights as "Derived Dataset 3".

The **Process 2** is used further in the model to select the "Derived Dataset 3" for PS classification. Sequentially, this dataset is compared for the height values of the PS points in "Derived Dataset 3". The PS with a height value of ± 2 meters from the NAP [RD-NAP, 2005] ground level height of the Rotterdam city are considered to be outside the buildings. Similarly, the PS points that lie outside the value of $\pm 2m$ from the ground level height of the Rotterdam city are considered as PS over the buildings. Further, these classified PS values from the "Derived Dataset 3" are selected using the Process 2 described above and with the height attribute as selection criteria. This height classification leads to the creation of two more classes of PS which is firstly, the PS on the building foot (output data class 1) or streets and secondly, the PS on the building roofs (output data class 2).

All the above stated process combine together to complete the framework for a dedicated geo-processing application, namely the PS data classification based on building heights. Moreover, this classification tool can be made more complex by incorporating the other geo-information in hand. Another sub-part of the model could be the classification of PS data based on infrastructures data boundaries. These infrastructures are mainly the railroads, bridges, underground metro lines, tunnels, etc./ We can classify the PS points first on the basis of their location, or within a buffer of 2 meters from the boundary of infrastructures. Further these PS are checked for the height values of AHN data sets and classified as PS on structure top or foot similar to the first model description in section 5.1.1.

6. CONCLUSIONS AND OUTLOOK

The classified Persistent Scatterers using the described process model results in definition of two primary classes of the scatterers. Firstly, the scatterers that are reflections from the foot of the buildings, or in other words, from the street kerbs or base of poles. These scatterers may be a result of double bouncing signal of the radar energy reflected back to the satellite. Also, the second class of PS which is classified as PS on building roofs can be a result of single bounce reflection from the slant roof tops that are perpendicular to the radar line-of-sight.

As demonstrated in the case study, the definition of the standarized the geoprocessing steps in the application model leads to highly reusable processing components. When included in a dedicated system architecture with support for chaining of geoprocesses, the small process units can be used in fast and efficient way by defining the model itself in a standardized way. The core functionality of geo-processing toolsets in ArcGIS can be combined to perform complicated processing tasks. The provision of model documentation and process description in WPS standards lead to replicate the geo-process in exactly similar way (including process constraints, model assumptions, data accuracy, etc.). If the processing model itself is shared, and made understandable to the research communities, the risk of duplications and ambiguities in geo-processing can be taken care of. The model sharing would be highly useful to track the sequence of operations, exact criteria of conversion and transformation of spatial data, input/output specifications of processes etc. By coding the capabilities and response of the process in XML, the semantics or the description of services can be formalized.

While the current case study gives an overview of the required service unit framework, the further step in this research is an automated web-based processing. As opposed to classical GIS, this approach provides an efficient and flexible combination of various (online and offline) services chained in order to achieve a dedicated application. In a web-based framework, several sources of data and services need to be accessed and chained together automatically. To achieve this so called "fully automatic" service chaining, besides a syntactical description of (web) services in form of GetCapabilities and DescribeProcess XML Documents, a semantic description is required as well. A GIS user may know what an 'intersection' or 'selection' operation means, but a webservice does not. The technical preconditions for this scenario exists but the paradigm for describing spatial operations to be used in self-organizing nets is still missing [Kiehle et al., 2006].

REFERENCES

Ferretti A., Prati C., and Rocca F. (2001). Permanent Scatteres in SAR Interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 39(1): 8-20 pp.

Longley, P. A., Goodchild, M. F., Maguire, D. J., and Rhind, D. W. (2001) Geographic Information Systems and Science. John Wiley and Sons, 2001.

The Open Geospatial Consortium Inc.(OGC) (2005). Official Website www.opengeospatial.org

Regnauld N. (2006). Improving efficiency for developing automatic generalisation solutions. In *Joint ISPRS Workshop on Multiple Representation and Interoperability of Spatial Data*, Hannover, Germany, 22-24 February, 2006.

Alameh N. (2003) Chaining Geographic Information Web Services. *IEEE Internet Computing*, September-October 2003, 22-29pp.

GBKN The Large Scale Standard Map of The Netherlands (2005). Official Website http://www.gbkn.nl/

Top10 Vector Map, Topographical Department of the Dutch Land Registary Office (2005). Official Website http://www.kadaster.nl

AHN Actual Height Model of The Netherlands (2005). Official Website http://ahn.nl/

NAP The National Ordnance Datum of The Netherlands (2005). Official Website http://www.rdnap.nl/

Kiehle C., Greve K., and Heier C. (2006). Standardized Geoprocessing - Taking Spatial Data Infrastructures one Step Further. In *9th AGILE Conference on Geographic Information Science*, Visegrad, Hungary, 2006, 273-282 pp.



Figure 1. Schematic workflow of a combination of geo-processes using ESRI ModelBuilder in ArcGIS 9.1