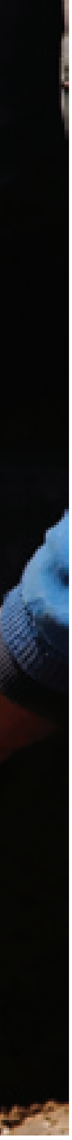
WEB SERVICES TEST BED

DISCOVERY AND INVOCATION OF SCHEMATIZATION SERVICES: A USE CASE FOR OGC-EUROSDR-AGILE PERSISTENT TEST BED FOR EUROPE

Historically, GIS have been standalone desktop applications. The emergence of open web mapping APIs such as OpenLayers, Virtual Earth and Google Maps has enabled the embedding of geospatial capability into websites and portals. At the same time that web mapping has become popular, the geospatial community has developed specifications that make it possible to manage, process, and disseminate geospatial data on the web.

As an emerging technology, web-based GIS does not currently address all the requirements fulfilled by desktop GIS - particularly in terms of secure data access and management of very large data sets. To facilitate research into geospatial interoperability and web-based GIS, a project to establish a persistent testbed for geospatial research and teaching has been commissioned by EuroSDR, the Open Geospatial Consortium (OGC), and the Association of Geographic Information Laboratories for Europe (AGILE). The key objectives of this project are:

- To act as a research test-bed for collaborative European research in geospatial interoperability.
- To aid the assessment of the current standards for geospatial interoperability in terms of research compatibility, completeness, consistency, ease of use and extensibility.
- To provide an environment for teaching standards and techniques for geospatial interoperability.
- To provide a resource to AGILE/EuroSDR/OGC for the coordination of research requirements as well as definition, testing, validation and development of open standards.



The persistent testbed has *web services* as its primary buildings blocks. A web-service may be considered to be a modular unit of remotely-hosted "black box" functionality with observable behaviour as specified by its *service contract*.

Approximately 30 academic institutions from across Europe responded to the call for expressions of interest in the initiative and a number have contributed web services based on specifications of the Open Geospatial Consortium (OGC). The contributed services include:

- Web Feature Services (WFS): for publishing vector data
- Web Map Services (WMS): for publishing rendered maps
- Web Processing Services (WPS): for publishing geocomputational functionality
- Web Coverage Services (WCS): for publishing raster data
- Catalogue Service for the Web (CSW): for publishing metadata about geospatial data, services and other resources.

Four use-cases have been designed by members of the testbed project. The use-cases cover a variety of thematic applications including unified portrayal, schematization and semantics for service discovery. The Centre for Geospatial Science at the University of Nottingham and the International Institute for Geoinformation Science and Earth Observation, ITC, the Netherlands are leading the schematization use case. This article shows how discovery and invocation of automated schematization services can be applied to spatial conflict reduction in water networks.

Basic processes in water network schematic map generation

Water pipeline networks are a good example of the practical application of schematic maps. Most schematic maps of water networks are currently produced (at best) semi-automatically using computer graphics packages – a somewhat labour-intensive and time-consuming process. The ability to generate schematic maps entirely automatically is therefore of clear benefit in utility mapping. Though connectivity details are explicitly recorded in the attribute table associated with the data, it can be difficult to visualize the true connectivity in geographic terms as there are instances where the pipelines are close or overlaid but not physically connected. The schematization service therefore acts as a visualization tool for water engineers and network modellers to help understand and analyze the hydraulic conditions of a network.

The quality of a derived schematic map is heuristically determined by the weighted sum of a number of constraints, including (i) topology – ensures that original map and derived schematic map are topologically consistent; (ii) angularity – edges should lie in horizontal, vertical or diagonal direction; (iii) minimum edge length – edges should have length greater than some minimum length and (iv) clearance - the distance between disjoint features should be greater than some minimum distance so that connectivity is apparent (Swan et al, 2007).

The schematization process then operates by iteratively perturbing vertices of the network and evaluating the quality of the resulting schematic map.

Summary of the schematization discovery and invocation process

The use case scenario is one in which a burst pipe in a water utility network has been reported. Maintenance personnel need to view a schematized map of the network in order to determine any likely effect on other parts of the network. To identify relevant web services, they search a catalogue of geospatial resources published through a CSW in order to identify a suitable service from the hundreds of geospatial web services available on the web. Unlike conventional search engines such as Google, a CSW can identify resources according to the geographic extents of the resources present. The CSW adopted for this use case is able to query a registered service to retrieve the list of available layers or processes, something conventional search engines do not do

Figure 1 depicts the deployed services, which are utilized to carry out the schematization process. It demonstrates the cross-organizational effort, as services from both

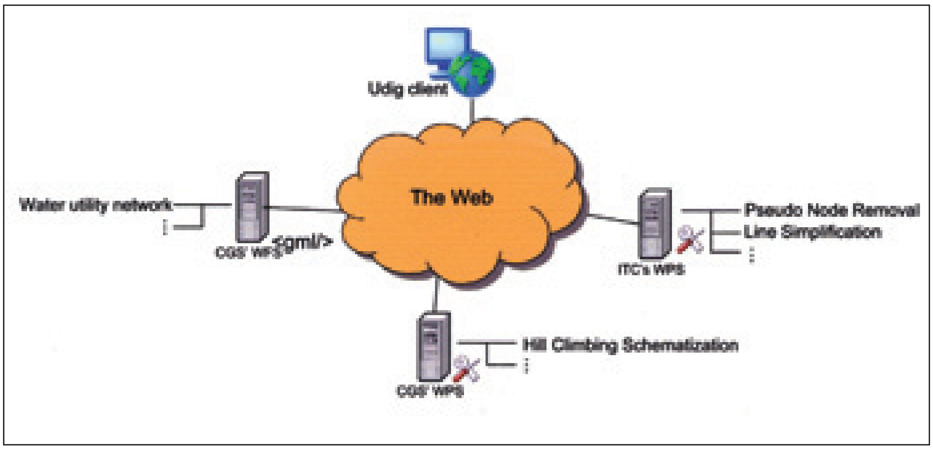


FIG. 1. Deployed service architecture components

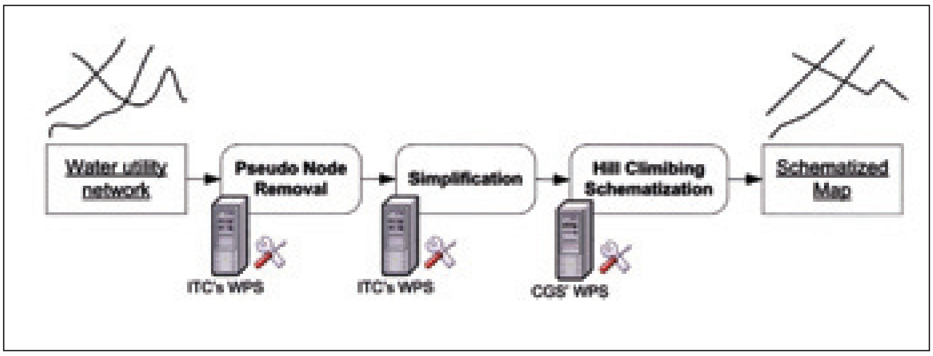


FIG. 2. Deployed process workflow

partners (ITC & CGS) are involved and demonstrates the benefits of interoperability, as information can be exchanged seamlessly using common standards.

Regarding the data applied in the schematization process, the nodes and edges of the network are stored as vector data in an Oracle database and published as features through a WFS. Checking topological consistency is the most computationally-expensive aspect of the schematization process - ensuring that all edge crossings are consistent takes time proportional to the square of the number of edges. It is therefore desirable to simplify the network geometry. In this use case, redundant vertices are removed by two Processing Services providing a special Node Removal function and a line simplification function (Douglas-Peucker algorithm). The node removal function removes all those nodes on a line network that connect only two edges (also known as pseudo nodes). The Douglas-Peucker simplification algorithm removes points on lines by preserving the original shape of the line. The combination of the two functions is necessary due to the line segmentation of the water network data, which consists of a lot of pseudo nodes and prevents the Douglas-Peucker algorithm from performing meaningfully. Thus, to simplify the data, it is necessary to first employ the node removal function. The complete workflow is illustrated in Figure 2.

There is a significant benefit of achieving simplification via (one or more) WPS over and above performing it "in situ" and storing the result in the database: since we can't know in advance how many schematization requests will be made, we don't want to incur the cost of database storage (and particularly not the attendant database-management issues such as maintaining a policy on expiry of the schematized map data).

After simplification, schematization is achieved via a 'Hill Climbing Schematization' WPS. For maps with up to 200 edges, the current implementation produces a locally-optimal schematic map in under two minutes on a Sun V20z with dual 2.4 GHz AMD Opteron processors. Further schematizing algorithms offering alternative tradeoffs between execution-time and schematization quality will be added in the near future.

Figure 3 shows the original water pipe network data and the results calculated by the process service chain.

This modularization of components is in accordance with the principles of service-oriented architecture: the processing services for the pseudo node removal function and Douglas-Peucker simplification are hosted at ITC in the Netherlands; whereas the WFS and the Hill Climbing Schematization process are WPS, hosted by the Centre for Geospatial Science at the University of Nottingham in the UK. The results of the complete workflow may then be visualised in any desktop GIS that supports OGC-conformant web services. In implementation terms, the Geonetwork CSW was used to host service metadata, with the 52°North WPS (the de-facto Java implementation of OCG-compliant geoprocessing) layered on Geotools (similarly for OpenGIS standards) used to implement the individual processing service. WFS-compatible data-access was via Geoserver and the web service client used for visualisation was uDig. All applied software packages are available through open source licenses.

Discussion

Advantages of using such an architecture for schematization include the loosely-coupled nature of the framework. This allows a given service to be hosted at a number of different locations. In the event of a service being unavailable (perhaps due to a power supply failure) an alternative service might be used. Another advantage is that organisations with high-performance computing facilities could offer to host computationally expensive processes. For developers and

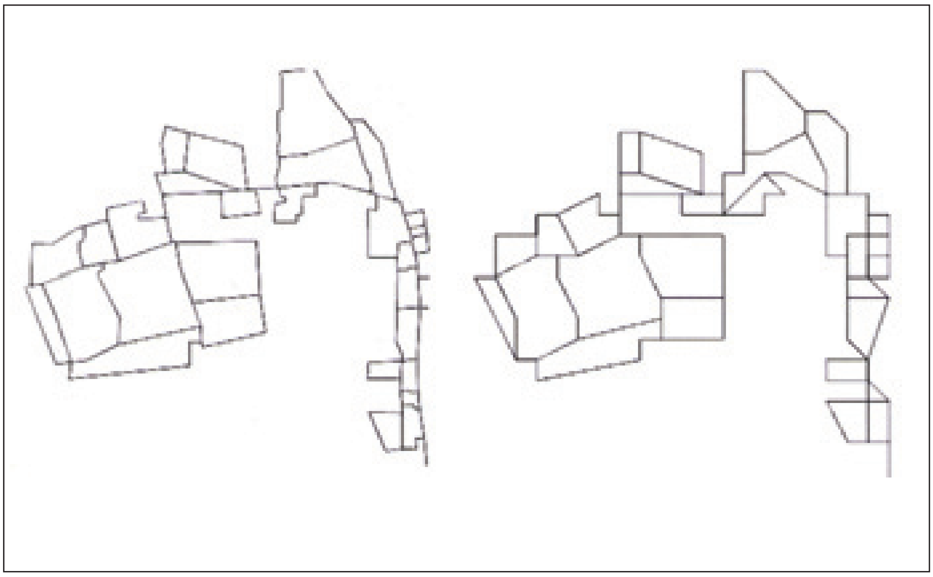
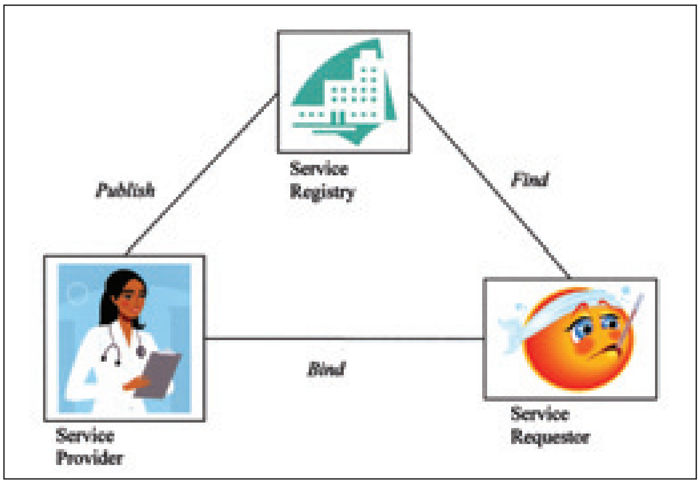


FIG.3. Water pipe network before (leftside) and after (right side) the applied simplification and schematization process]



The Publish-Find-Bind paradigm
for web services

software vendors, the service-oriented architecture offered by the persistent testbed provides a platform for testing new products and evaluating the degree of interoperability they offer. In addition, data providers can regulate and monitor access to their data by allowing only machines with specific IP addresses or by requiring a password for authentication.

One disadvantage is that, for very large datasets, performance can be affected by network bandwidth. In a workflow with several nodes (WFS, WPS etc) this can reduce efficiency as data is being transmitted from one service to the next. However, as bandwidth becomes more affordable, transmission delays in such cases will become less significant. Additionally, as most of the data is referenced and retrieved via URLs, there is the possibility of applying caching strategies for minimizing data transfer between different service nodes.

It should be noted that, at present, our processing is semi-automatic in the sense that no process chaining capabilities (such as BPEL) are applied. These additions are scheduled for the next few months.

Conclusions

This article has described a means of discovery and invocation of water network schematic maps. Dynamically-generated schematic maps will be very helpful for modellers to visualize water network connectivity. Future work will concentrate on enhancing visualization and usability via the application of additional heuristic constraints.

The AGILE GI testbed portal is http://www.gitestbed.eu. Organizations with an interest in participating in testbed development are invited to contact Dr Jerry Swan at the Centre for Geospatial Science.

References

Swan. J, Anand.S ,Ware J.M ,Jackson M, (2007) "Automated schematization for web service applications" , Springer-Verlag Lecture Notes in Computer Science 4857, pp216-226
52°North Geoprocessing Community: www.52north.org/wps.
uDig website: udig.refractions.net.
GeoServer website: www.geoserver.org.
GeoNetwork website: www.geonetwork.org.

Acknowledgements - *The authors express thanks for DfiD KaR project for the data used in this work.*

Article by Jerry Swan, Theodor Foerster#, Rob Lemmens#, Gobe Hobona*, Suchith Anand*, Mike Jackson* who are at: * Centre for Geospatial Science, University of Nottingham and # International Institute for Geoinformation Science and Earth Observation, Enschede, the Netherlands.*