# Do we know how to orchestrate on the GeoWeb platform?

Jan Růžička[1], František Klímek[1], Michal Šeliga[2], Martin Prager[3]

[1]Institute of geoinformatics, VSB-TU of Ostrava, 17. listopadu 15,
708 33, Ostrava – Poruba, Czech Republic
jan.ruzicka@vsb.cz, frantisek.klimek.hgf@vsb.cz

[2]T-MAPY spol. s r.o., Vědecko-technologický park Ostrava, a. s., Technologická 372/2,
708 00 Ostrava - Pustkovec, Czech Republic
mise@tmapy.cz

[3]NESS Slovensko, a.s., Zvolenská cesta 14
974 05 Banská Bystrica, Slovak Republic
martin.prager@gmail.com

**Abstract.** The research project “GeoWeb services orchestration” GA 205/07/0797 is close to the last year of its financial support from the grant agency. The main result from the project should be a methodology that leads a company when it decides to use a service oriented architecture advanced technique named an orchestration for building (or design) geographic information system. Nearly two years of a research discovered more problems that they solved, although we believe that we know how to do not realize an orchestration. The paper is going to present first version of the methodology. The designed methodology is based on practical experiences and hundreds of hours of a research in the wide area of an orchestration and related areas focused on GeoWeb platform usage. The paper should describe problems that has been discovered and that are not solvable at a moment (in a conditions of the project financial and time capacities) and should be solved in following projects.

**Keywords.** GeoWeb, geoinformatics, web services, orchestration, BPEL, methodology, SOA.

**Název.** Víme jak dirigovat na platformě GeoWeb?

**Abstrakt.** Výzkumný projekt „Orchestrace služeb pro GeoWeb" GA 205/07/0797 se blíží do posledního roku svého řešení. Hlavním výstupem projektu by měla být metodika popisující jak postupovat v případě, že se společnost rozhodne geografické informační systémy navrhovat s využitím pokročilé techniky servisně orientované architektury zvané orchestrace. Téměř dva roky výzkumu objevily více problémů než jich vyřešily, přesto si troufáme říci, že již víme jak si při orchestraci určitě nepočínat. V rámci příspěvku bude prezentován návrh metodiky, která zohledňuje praktické zkušenosti a stovky hodin výzkumu týmu badatelů v široké oblasti orchestrace a příbuzných oblastí se zaměřením na využívání platformy GeoWeb. Zmíněna budou také místa, která byla výzkumem objevena jakožto v současné době (za současných kapacit projektu) neřešitelná, a která bude jistě vhodné řešit v dalších navazujících projektech.

**Klíčová slova:** GeoWeb, geoinformatika, webové služby, orchestrace, BPEL, metodika, SOA.

## 1 Introduction

When we have started to study possibilities of an orchestration in the area of GeoWeb there were not any articles that describe any theoretical or practical solution. We have started our research in 2005. A first practical result from the research was system WSCO (Web Services Catalogue for Orchestration). This catalogue has covered several metadata items, that could be useful for orchestration. After that we have started with analyses of languages for orchestration. We spent more than half a year on research of WS-CDL language to finally find out, that this language is not suitable for orchestration. Then we have obtained support from Grant Agency of Czech republic to continue our research. During two years of research we have analysed several languages that can be used for an orchestration. We did several tests and prepared several orchestras. Now we are ready to summarize our experiences into methodology for orchestration in the area of GeoWeb.

GeoWeb is a platform, that consist of standards, open web services, people and technical equipment that can be used for building flexible geographic information systems (GIS). GeoWeb is nowadays at the beginning of standard usage. We have a lot of standards available, but only several of them are in a common use. To be honest, there is only one standard used in such scope, that can declare it as a common standard. This standard is WMS (Web Map Service). Unfortunately this standard is too simple to base complex GIS only on itself. Of course there are other standards, that are well known, but their usage is reserved to top development companies or to research institutes. Most of this standards are quite old, but not yet mature. They must be proofed by practical usage. There are new standards or technologies growing every day (like GeoRSS or MapAPI) and we believe that GeoWeb can be a platform for building GIS of the future.

## 2 Orchestration

An orchestration is a process where are modelled processes (real or abstract) in a way of formalized description. A process modelling is a technique that uses several description tools, mainly schemas or diagrams, to describe usually real processes inside enterprise. The processes can lead across several organizations.

Each process has the start point and the end point. There can be inputs from several sources incoming into process and these inputs are manipulated in a time of processing. Some manipulation can be done in an iterative way. A process can be divided into several ways (operations), that can be processed simultaneously. Ways can be selected according to conditions.

For example we can describe process of creation of flooded areas.

1. A customer (a source) defines conditions how should the map looks, e.g. what area, what kind of flood (100 year, 50 year, real from 1997 year), what accuracy.
2. The analytic (a source) selects appropriate data sources according to defined conditions.
3. The analytic gets the data sources.
4. The analytic does analyse and extracts flood areas.
5. A cartographer (a source) uses flooded areas from the analytic, selects another data sources for assembly the map.
6. The cartographer assembles the map.
7. The map is delivered to the customer for an evaluation.
8. Comments from the customer are evaluated. According to results of the evaluation is the process flow returned to a new analyse (points 2-4) or the process is returned only to changes of the map (points 5-6).

The process can be more complicated e.g. it can contains parts such as agreement, payment, delivery.

A process modelling offers tools how to describe such process in a formal way. The formal description allows to control, test and visualise the process in a form that is understandable for common participant of the process.

Two languages are usually used for process modelling, UML (Unified Modelling Language) and BPMN (Business Process Modelling Notation). Each offers different possibilities for description of a process, but both of them (or others) can be used for a process description.

A common example of the process description inside an enterprise is ISO 9001 usage. E.g. at VSB-TUO, Faculty of Mining and Geology is published a set of descriptions such as Ph.D. degree study process.

When the description is enough detailed, we can start to automate such processes or their parts. If an individual task of a process can be solved using a computer than it is described in a way of algorithm and implemented as a web service (when working in the area of Web SOA).

A model of a process is transformed from abstract languages (BPMN, UML) to a form that can be directly run on a computer. In this area of runnable models of processes is the most known BPEL (Business Process Execution Language). A process run means reading inputs, invoking web services, deciding according to results, repeating some parts of the process and other necessary operations.

A process modelling offers possibilities how to formally describe processes inside an enterprise, to find duplicate processes, to find processes that are not optimised, etc. A process modelling helps with processes optimisation and with sources management optimisation. When it is possible, than the description is available in a form of BPEL- like language and processes can be directly invoked.

GeoWeb services orchestration can be done in many ways. The GA 205/07/0797 team has researched the two ways of possible orchestration.

### 2.1 Simple orchestras

The first way is based on orchestras where set of used services is same in a meaning of data sources and algorithms. During the building orchestra instance are searched only services that use the same data source and the same algorithms for data source and input manipulation. Data source content can change only on spatio-temporal extent of the working area. We can speak about services replication (or distribution in a horizontal plane). Current instances of the services that are connected to the orchestra are selected according to a current state of the services, such as performance, speed or provider.
These services differ on physical binding. These kind of orchestras is focused on optimisation of orchestras run. For these kind of orchestras is not needed any specific manipulation. There is only necessary to identify same services using some key. For our testing purposes we use common identification, based on standardisation organisation identification, standard identification, service identification. Such identification is described on the following example. http://www.cuzk.cz/ogc/wms/1.1.1/ZABAGED/0.1. Items are defined by url. First item is domain of the service type guarantee. Second item is abbreviation of standardisation organisation name. Third item is abbreviation of standard name. Fourth item is a version of the standard. Fifth item is abbreviation of the service type. Last item is a version of the service type.

### 2.2 Advanced orchestras

The second way is based on orchestras where current instances of the services can be just similar to each other in a meaning of data sources and algorithms. For example we can use service that uses railways data source where tracks are just simple lines between stations or we can use service that uses railways data source where tracks are modelled by real headway. We can switch between these sources in many cases, such as routing (finding the best routes) where the main parameter for routing is time. This type of orchestras is more difficult to manage than the first one.
Our research shows that usually the first type of orchestras will be used, but there are still situations when a system for orchestration should be able prepare second type of orchestras. There are two ways how to handle this problem.
The first solution is simple, but difficult to manage in a meaning of long time term, because this solution is rather static than dynamic. There must be simple database (no matter how it is organised – relational, XML) where are defined relations between data sources (services). Related services can be named group of similar services.
The second solution is based on data source evaluation based on metadata analyses. In [Růžička 2008a] is described this solution.

## 3 Methodology

The methodology will contain several examples and should have all necessary parts, but for this paper are selected only main ideas and a first draft of the methodology is described in a compact form. The methodology should lead an user from the beginning of an orchestration process to a implementation of an orchestration. The methodology is going to be constructed such as steps that advice how to decide during the orchestration process.

### 3.1 Initial decision

At the beginning of the orchestration is a decision if the usage of a new methodology will be efficient for the company. This is a same problem when deciding to implement GIS for a company. An orchestration has useful impact on management of business processes. But as usual an innovation

cost a lot of sources, including money. We are sorry, but we do not have any advice for this step based on experiences. We are going to include basic advices that comes from outside area of the GeoWeb (GIS) and we are going to include some comments from those that already did a transformation. One advice from Jari Oredevi (Karhu meeting, Pori, 2005) can be simple: “Do not believe that orchestration can help you if your process management is in a mess. First you have to change people of your management, then you can try to use a new techniques of management”.

### 3.2 Selecting an orchestration language

Spend a lot of time on languages evaluation. When you decide to use one, there is very difficult way to change it. Languages are quite complicated and it is quite difficult to use more than one for process description. Several languages will be described in the annex of the methodology. When you choose BPEL language follow the following steps. When you choose another language then there is no methodology (or we do not know about it) how to use it in the area of GeoWeb.

### 3.3 Designer and server evaluation

Spend a lot of time on evaluation of designer and server for BPEL. Each of the designers extends BPEL set of tags. Models exchange between engines or designers is not possible in a simple way. We have an experience with two environments and those will be described in annexes of the methodology. We are going to prepare standard comparison table for purposes of evaluation of another environments.

### 3.4 Proxy

BPEL does not enable to orchestrate common HTTP/GET/POST web service. That is a problem that we can not solve in any simple way. Some kind of proxy service must be included to proxy such services. We are working on one engine, that should be extensible enough to handle any type of service.

### 3.5 Cache

Basically there is not a need of a caching for an orchestration, but it is efficient to use cache servers. Unfortunately there is not any available that can handle GeoWeb services requests. You can not use simple HTTP cache servers. One server that can handle GeoWeb services requests is going to be developed for purposes of this methodology. Technical reference will be available in an annex of the methodology. The server is going to be tested only on one type of service.

### 3.6 Types of orchestras

Use only simple orchestras, where services instances are replicated. This kind of orchestras are simply to manage and there are not need any additional components described for advanced orchestras.

### 3.7 Advanced orchestras

In some cased you will need to build advanced orchestras (if not you are lucky). In this case follow the instructions:

- Do not use metadata of geodata to build groups of similar services.
- Use experts' evaluation of the orchestras results to create groups of similar services.
- Update groups of similar services according to new results evaluation.

### 3.8 Monitoring of the services

Monitoring of the services is a necessary part of the orchestration process. We have prepared a set of tests that should be provided on the services. Results of these tests should be stored in a catalogue for dynamic orchestras creation. Some of the tests can be done without cooperation with a service provider. There should be provided as well some internal tests that must be done with cooperation with a service provider. Some tests are described in [Kaszper 2008] and whole bunch of tests will be described inside the methodology.

### 3.9 Catalogue

A catalogue has two purposes. Instances of services must be described inside a catalogue or set of catalogues. Prepared instances of BPEL processes should be stored inside a catalogue to be available for further invocation.
Catalogue interface should allow search based on location, keywords, type of services, organisation and status of services.
Common interface UDDI does not include such set of parameters. It can be extended, but not in a simple way. We do not recommend to use UDDI for GeoWeb services orchestration.
CSW interface includes all necessary tools to allow search based on the specified parameters. What needs to be extended is the ISO 19119 (ISO 19139) profile to handle information about service status. An another way is to combine CSW and basic ISO with external service that covers additional information about a service.
Wide possibilities of CSW are not necessary for GeoWeb services orchestration. To be honest there is not available any implementation of CSW that can be directly used. We believe that the reason is its extent. That's why we are going to specify minimal needs for CSW interface to cover GeoWeb services orchestration needs. We name it CSW-Lite, because it will be compatible with CSW specification. The CSW-Lite implementation should be simpler than implementation of the whole CSW. CSW-Lite specification should be described in an annex of the methodology and we would like to implement it in a minimalistic way to show that anybody can include this interface inside any catalogue.

### 3.10 Conditions for BPEL creation

There is not space for static orchestras. Even simple orchestras must be created according to actual state of services instances. The most sufficient way that should work generally for all engines is BPEL parametrisation. We can name it template driven orchestration (TDO). Every BPEL process can read at the beginning of the invoke a set of parameters. This set of parameters can include a list of current services instances, that are at the time of invoking a process the most suitable. This kind of BPEL process description is named a BPEL template. Another way is to create an own BPEL engine, that can search each service instance right before its invocation. To build own BPEL engine is not simple enough to recommend this way as an efficient way.

### 3.11 Results evaluation

Evaluate all results. Does not matter which kind of orchestras do you use. Even simple orchestras can run into wrong results. Evaluation should be based on expert knowledge and should live with dynamically created orchestras. A methodology for results evaluation is now under development and is going to be tested during the summer 2009.

## 4 Problematic areas

We have found a lot of areas that are somehow problematic and there will not be probably enough time to solve them in the last year of the project solving.

We did not use any type of security and reliability layers. Theoretically all security and reliability

operations can be handled by proxy service, but there is too many available ways how to possibly solve this task.

We have prepared scenario for orchestras testing, but we are going to spend only a few months to test our scenario for testing. The testing of orchestras is crucial for well done orchestration. This part of the methodology should be updated according to several years of testing.

We have decided to do not recommend using metadata for geodata in a process of dynamically created orchestras. The purpose of this decision is deeply described in [Růžička 2008a]. The situation will not change generally, but individual enterprises can have their metadata in a better state than is a general state. For this kind of enterprises should be the methodology extended to cover working with metadata.

We have focused on BPEL, but we know that there are other available languages. For example BPMN language described in [Prager 2009] looks suitable for many enterprises. We did not use it because of complicated results of transformation into BPEL, that can not be simply used for template driven orchestration.

## 5 Conclusion

The methodology is not even close to be ready, but we believe that it will be available at the end of the 2009 year. The methodology should help to developers in a case when they decide to try use orchestration as a platform for their flexible GIS. We have evaluated many blind ways and the methodology should help to avoid such ways. At the other hand the methodology is focused only on BPEL language and a solution based on this language must not be the best one for a particular enterprise. If you have any suggestions or comments to our first proposal than please do not hesitate to contact us. We are opened to any ideas.

## Reference

[Kaszper 2008] KASZPER, R. Sledování stavu služeb pro GeoWeb. Diplomová práce. VŠB-TUO, Ostrava. 2008.

[Prager 2009] PRAGER, M.; KLÍMEK, F.; RŮŽIČKA, J. GeoWeb Services Orchestration Based on BPEL or BPMN? Proceedings from GIS Ostrava 2009. VSB-TUO, Ostrava.

[Růžička 2008a] RŮŽIČKA, J. ISO 19115 for GeoWeb services orchestration. Proceedings from Geoinformatics on CTU 2008. CTU, Prague. Available at: http://geoinformatics.fsv.cvut.cz/wiki/index.php/ISO_19115_for_GeoWeb_services_orchestration

[Růžička 2008b] RŮŽIČKA, J. Modelování procesů. Internetový časopis G Portál. Květen 2008. Available at: http://gisak.vsb.cz/gportal/index.php?option=com_content&view=article&id=910:%20&catid=232:interoperabilita

## Support

The article is supported by Grant agency of the Czech republic (GACR) as a part of the project **GA 205/07/0797 GeoWeb services orchestration**.