Standardized Geoprocessing - Taking Spatial Data Infrastructures one Step Further

Christian Kiehle[1], Klaus Greve[1] & Christian Heier[2]

[1] University of Bonn, Department of Geography
Bonn, Germany
{kiehle, greve}@giub.uni-bonn.de

[2] University of Aachen, Research Institute for Water and Waste Management
Currently employed by Wupperverband, GIS Department
Wuppertal, Germany
che@wupperverband.de

SUMMARY

Spatial data infrastructures (SDI) have been widely accepted to share geospatial data among organizations. Today SDI's main focus lies on distributed data storage in the form of spatial web services, the retrieval through catalogues, and the visualization in form of web map services. The hypothesis presented in this paper takes SDI's one step further by providing a method to process geodata, in an Open Geospatial Consortium (OGC) compliant way, into information. It addresses the problem of service chaining by providing a system architecture to process data into information using Business Processes.

The proposed methods utilize spatial standards provided by OGC, International Standardisation Organisation (ISO) and 'mainstream IT' standards provided by the World Wide Web Consortium (W3C) to achieve highly complex service chains with geoprocessing capabilities. In companion with semantic description of service's metadata, a self-organizing net can be established. Two case studies present the potential of standardized geoprocessing services.

The result is a generic service for providing common geoinformation capabilities (e.g. spatial algorithms, Map Algebra etc.) for using in next level SDI.

KEYWORDS: *distributed geodata, service chaining, Business Process, Service Oriented Architecture, Business Process Execution Language, self-organizing nets*

INTRODUCTION

Recent developments in Geographical Information Science, namely the concept of spatial data infrastructures (SDI), provide access to globally distributed *data* through network technologies (Nebert 2004a). The SDI concept enhances mainstream Information Technology's (IT) widely accepted principle of Service Oriented Architectures (SOA) to include spatial features. Thus, SDI seem to be an adequate tool to handle and access large collections of spatially distributed, georeferenced data inventories (Beaumont et al., 2005, Bernard et al., 2005, Tait, 2005). The basic idea of SOA – providing functionality as a set of independent services – is based on dynamic integration and composition. Several well-established standards of the Open Geospatial Consortium (OGC) foster interoperability between data and services in order to set up a spatially enabled SOA.

The value-added component in spatial data handling is the acquisition of information through processing and concatenation of data. This process involves the acquisition of problem-specific data, the application of specific computations (e.g. normalized differentiated vegetation index (NDVI), spatial intersection, spatial buffering, etc.), and the visualisation of results (usually as maps or map-like presentations). It turns data into *information*. We distinguish between information retrieval as selection of the adequate data for our questions or problems in the data store and information processing as finding of hitherto unknown information through concatenation of data from the store.

Currently, SDI open standards predominantly support the retrieval and visualisation of data through concatenated web services, while the process of information acquisition through data processing and concatenation is normally driven by human actors with more or less monolithic applications. If we try to transform this process to web services, we face several problems:

- It is not guaranteed, that defined processes can be reused to answer new questions without modification.
- Service chains, in order to peform highly complex problem solution tasks, have to be defined manually.
- There is no publicly known standard on how to implement information generating processes in an OGC-like way.

The hypothesis presented in this paper is that the next level of SDI will have to deal with highly complex service chains in order to process *data* into *information*. The process of turning data into information is not yet specified in an open and widely accepted standard. Atomic process units do not allow for complex spatial calculations (e.g. modelling groundwater vulnerability assessment etc.), therefore the chaining of single services into more complex processing units has to be established. The process of service chaining has to be defined in a human as well as machine legible way. Utilizing the service metadata capabilities (as known from common OGC web services in form of the 'getCapabilities' interface) services can dynamically interact without human interference. Under the condition that services can automatically adapt to other problem-specific services (e.g. by utilizing semantic description as a discovery mechanism) the proposed methodology is able to establish self-organizing nets for information generation.

The main aspects addressed in this paper are:

- Current standards lack a definition of processing capabilities. The recently developed OGC Web Processing Specification (WPS) (Heier & Kiehle, 2005, Schut, 2005) tries to overcome this limitation by providing a service specification according to OGC reference model (ORM) (Percivall, 2003). WPS is introduced and two case studies present the potential of standardized processing services.
- Single processes do not answer complex queries. The processes have to be chained in order to establish complex information-generation tasks. A meta-language for business processes – the Business Process Execution Language (BPEL) – is considered exemplary to chain atomic services into complex service chains.
- The emergence of the semantic web will lead to automatic data integration and information processing. Thus, self-describing and easily discoverable processes like WPS are easily integrated into semantically enabled SDI.

The paper utilizes methods derived from software engineering, OGC web services (OWS), and integrates workflow management facilities (i.e. BPEL interpreter) in order to archive a high level of abstraction in service chaining. Thus, the proposed methodology allows the integration of processing capabilities in spatial data infrastructures.

FROM LOCALLY GENERATED INFORMATION TO GLOBALLY GENERATED INFORMATION

Spatial data infrastructures support the utilization of distributed (geo)data. Interoperability on a technical (i.e. syntactical) level is established by adhering to common standards by OGC or, in case of "mainstream" IT (as opposed to spatial Information Technology), to standards provided by e.g. the World Wide Web Consortium (W3C). The most widespread standards adopted in web oriented spatial Information Technology are:

- Geographic Markup Language (GML) for encoding spatial features in an eXtensible Markup Language (XML) dialect (Lake et al., 2004)

- Web Map Service (WMS) for publishing cartography through internet technologies (de La Beaujardiere, 2004)
- Web Feature Service (WFS) (Vretanos, 2002) for the distribution of vector based data
- Web Coverage Service (WCS) (Evans, 2003) for the distribution of raster based data
- Catalogue Service (CAT) (Nebert, 2004b) for the enquiry of spatial data and spatial web services
- Web Coordinate Transformation Service (WCTS) for the transformation of data into several geodetic reference systems (Whiteside et al., 2005)

Except for the WCTS these standards only provide access to geodata, geo-enabled catalogues, and the capabilities to visualize data in the form of maps. Our aim is to provide processing capabilities through spatial web services technology to enable acquisition of information through concatenation of globally distributed data.

Users need information to answer a query or support a decision. In most cases, the required information cannot be found directly from the web. The user has to locate relevant data, assemble data from several sources / servers, process the data and then visualize the results. In most SDI, users are supported by Catalogue Services to find relevant data, and Web Map Services to visualize the required information. Web Feature Services transmit spatial feature described in GML. For some problems or questions the user will find service chains of cascading services that process and analyse data. These service chains are not very flexible and only suit predefined tasks. In most cases the user has to download the data to his local machine and process it with a stand-alone Geo Information System (GIS). The concept of WPS is to overcome these restrictions and enable web services to fullfil geo-processing tasks such as: spatial intersection of features, vector-raster-conversion, application of spatial modelling to analyse data and to produce topical information. WPS are easily accessible and flexible libraries of geoprocessing algorithms in a web service environment (like a stand-alone GIS contains a library of geoprocessing algorithms in a stand-alone environment). To function in a web service environment the WPS has to offer self-describing atomic geoprocessing tasks, which can be accessed and used by humans and other web services. The vision is to implement a self-describing and self-organizing net of Web services to solve problems and to answer geographical data queries.

In order to demonstrate the concept, an Interoperability Experiment (IE) for the definition of an OGC compliant Web Processing Service took place during 2005 (Heier & Kiehle, 2005, Schut, 2005). The result is a draft specification for the definition of a processing service according to the OGC Reference Model. The specification defines an interface for the retrieval of service metadata, inherited from the *getCapabilities* interface, which must be implemented by all OWS. This yet simple interface is presumably the most powerful. Besides metadata about the service provider, metadata about the available processes are also provided. The proposed processes can be of spatial, as well as non-spatial nature. The structured service metadata (well-formed XML, valid against an open XML Schema) are legible by humans as well as by machines. This allows the automated integration of services according to their self-description.

A detailed description about input and output of one specific process is accessible through the *describeProcess* interface. Using the aforementioned service metadata interfaces, an automated service binding can be established in order to allow services to interact dynamically. The *execute* interface provides the underlying functionality of the service, e.g. a spatial processing algorithm like intersection, union, dissolve, etc. Figure 1 summarizes the WPS interface. A service providing a spatial intersection of two GML themes is available via *http://geotech.lih.rwth-aachen.de/wpsclient/*.

The OGC WPS specification allows the definition of atomic process units which can – due to their defined interfaces – be chained easily and flexibly, as demonstrated in the next section.

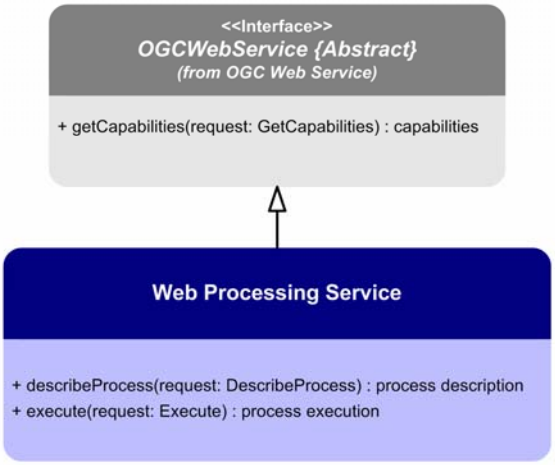

Figure 1: Web Processing Service interface definition

SERVICE CHAINING BY WEB SERVICE ORCHESTRATION

The term 'web service orchestration' describes the way web services interact to form a higher level business process (e.g. the visualization of growth of population would require both web services hosting data about population, and a business process to chain the services involved according to an algorithm in order to compute the growth of population). The participating services are selected and coordinated by a central orchestration service. In opposite to the choreography of web services, where each service knows his successor and ancestor, web service orchestration leaves the services loosely coupled (Veerawarana et al., 2005).

The orchestration of web services can be archived manually, by feeding the output of a web service as input into another web service, semi-automatically by defining the order of web service-interaction in a configuration file, or fully automatically by providing capabilities to establish a self-organizing net.

While recent papers announce the advent of the semantic geospatial web, this paper proposes already available, commonly accepted and adopted open standards to build up a next generation SDI, paving the way to the semantically enabled spatial web. However, first the obstacle of ad-hoc problem-specific service chains has to be overcome.

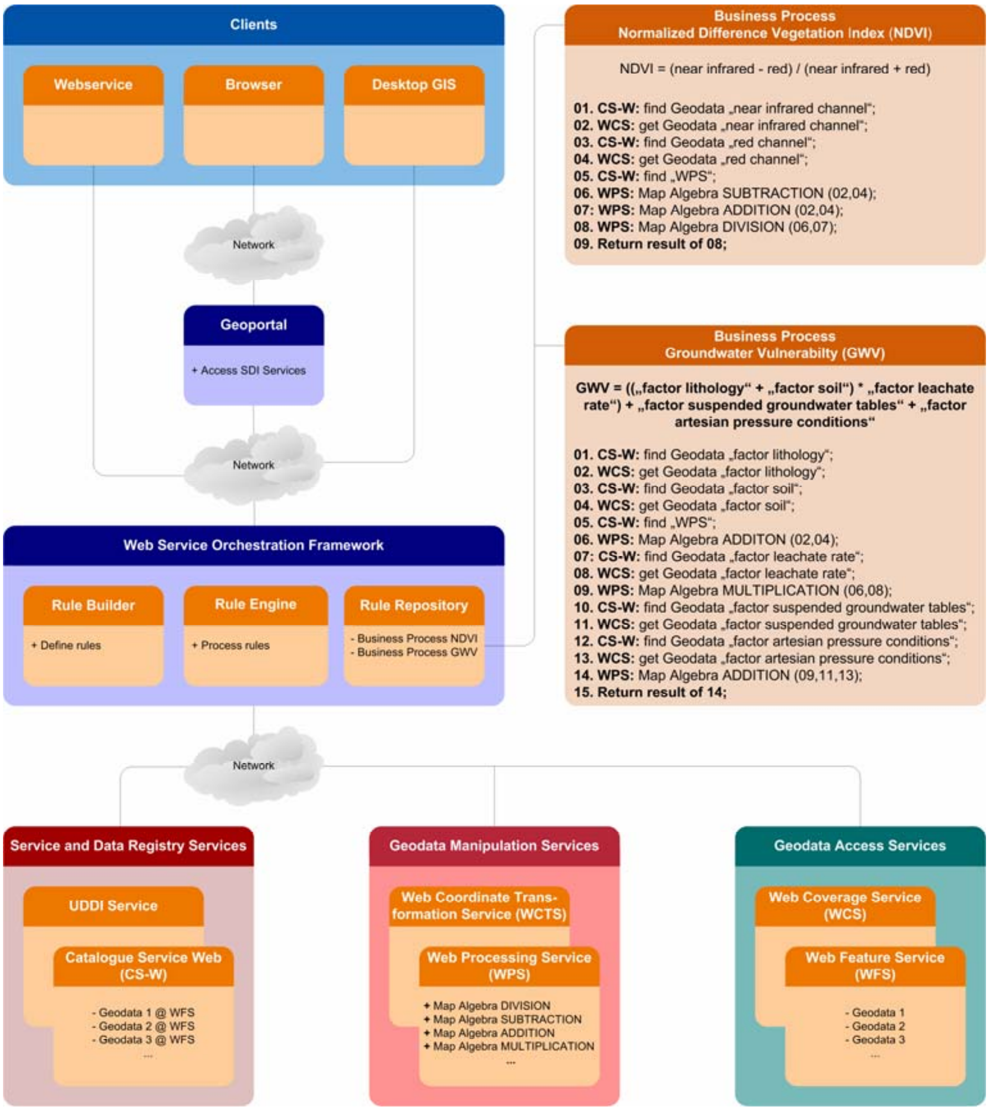

Figure 2: System architecture

Figure 2 represents the proposed system architecture to overcome the problem of service chaining. A user (human or computer) accesses a service orchestration framework through a network. He can use any available processing service in order to process data. The prototypical implementation comes along with four basic local map algebra algorithms (Tomlin 1990), namely addition, subtraction, division, and multiplication. Besides the application of map algebra on user-provided data, any distributed geodata could be used as an input. For the enquiry of the data, an OGC-compliant catalogue service is utilized.

Once the relevant data are found and provided, a single or several WPS can be bound to the user session. In the case of single map algebra tasks, the processing can be started immediately. The potential of WPS is increased by using the service chaining capabilities provided by service orchestration. To assemble complex service chains semi-automatically, the use of a Markup language like the Business Process Execution Language (BPEL) (Andrews et al., 2003) or the use of the newly added annotation facility to Java Development Environment (Gosling et al. 2005) may be utilized. As defined in figure 2, business processes describe the execution order of the services involved. Until now, two predefined business processes are stored in a process repository, one using chained map algebra invocations to compute the Normalized Difference Vegetation Index (NDVI) (Lillesand et al., 2003), and another to compute the groundwater vulnerability according to Hölting (Hölting et al., 1995). The business process documents are evaluated at runtime by the rule engine. The rule engine offers support for semantics and inference, thus allowing the integration into the semantic web.

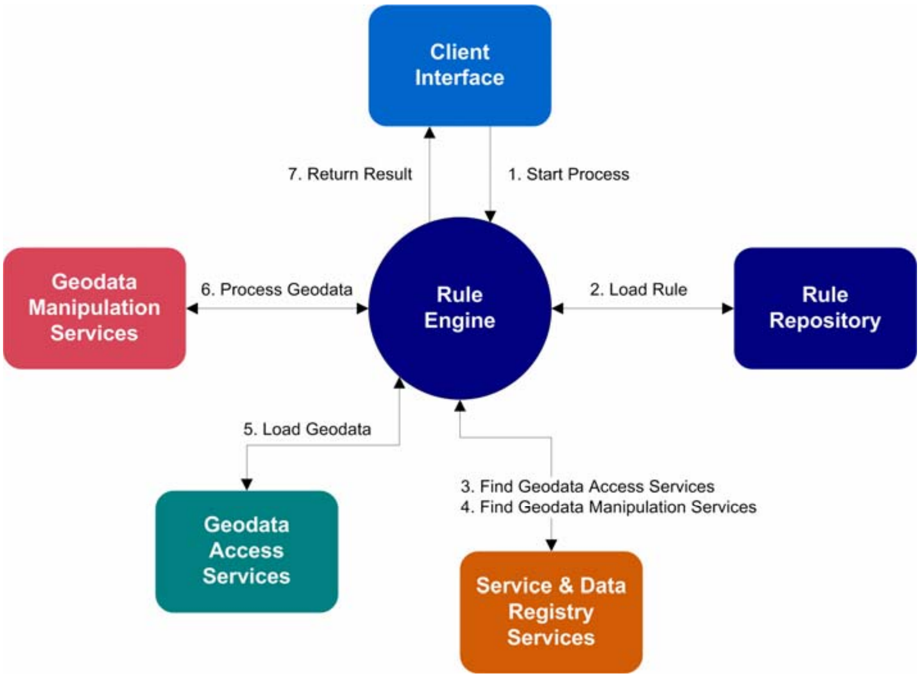

***Figure 3*:** Rule Engine as the central Business Process interpreter

A business process is a collection of activities that takes one or more types of input and creates an output that is of value to the user. By predefining business processes, the processes are available through the network, thus allowing the reuse of the implementation. The user is enabled to define their own business processes by chaining web services through a graphic user interface (GUI). A rule engine (http://drools.org) is used as a central Business Process interpreting unit, as shown in figure 3. The rule engine is the mediator between client requests, stored and newly defined rules, service and data registries, and the process of turning data into information.

The proposed system uses BPEL to chain services registered in a SDI. While the service chains could be defined manually, the discovery and integration of services will be performed by spontaneous service-binding, depending of the available services in a catalogue service.

The self-describing nature and the ability to cascade services foster the reusability of services by humans as well as other web services. OWS in general and WPS in particular, provide extensive service metadata facilities, including XML Schema (Fallside, 2001) to define custom data structures.

Thus, chained and cascaded services enable the definition of service chains to retrieve complex information and even simulation models. Self-describing services allow for more than linear chaining, but even self-organizing nets.

CASE STUDIES

The following case studies carried out by the authors illustrate the advantages of the proposed system and especially the power of WPS. Case study 1 is taken from a recently finished research and development project. Detailed information on the project is given in (Kiehle, 2006). Case study 2 describes in all brevity an implementation to compute the Normalized Difference Vegetation Index (NDVI) out of remotely sensed date (i.e. satellite imagery).

Case Study 1 – Geo web service Groundwater Vulnerability

During a research and development project, the authors developed a web-based information system to assess and map groundwater vulnerability (Azzam et al., 2003, Kiehle, 2006). Groundwater vulnerability describes "an intrinsic property of a groundwater system that depends on the sensitivity of that system to human and / or natural impacts" (Vrba & Zaporozec, 1994). The system operates with approximately 20 different data sources (vector data as well as raster data), provided by a variety of public data providers, and interlinked in an organizational SDI.

To compute the groundwater vulnerability (formula 1, S), five distinct factors have to be computed from distributed data. The factors represent the specific protectiveness of the soil (formula 1, B), the groundwater overburden including the depth to the first groundwater table (formula 1, Gm), the leachate rate (formula 1, W), and some general supplements for permanent artesian pressure (formula 1, Q) and suspended groundwater layers (formula 1, D). Computing the five factors according to formula one, the groundwater vulnerability (a dimensionless point value) can be assessed. Higher values represent a greater risk of groundwater contamination. The risk of groundwater contamination according to (Diepolder, 1995 and Hoelting et al., 1995) is determined by formula 1

$$S_g = (B + \sum_{i=1}^{n} G_i * m_i) * W + Q + D \quad (1)$$

In order to compute the overall protective effectiveness, the factors (as represented in formula 1) have to be computed. Each factor is provided by different data providers throughout North Rhine-Westphalia, Germany. An OGC-compliant catalogue service acts as a 'data broker' to distributed data services. The data is computed by a map algebra service according to formula 1. The results of this process are afterwards incorporated into a map-like display. Since all required data is raster data, the algorithm is implemented using Map Algebra.

A business process (figure 2 "Business Process – GWV box) utilizes the provided WPS containing the four Map Algebra operations. The Business Process describes the order in which the services are chained to generate the requested information. Without affecting the services, the Business Process document could be changed at runtime to influence the data processing tasks.

Case Study 2 – Computing the Normalized Difference Vegetation Index

The Normalized Difference Vegetation Index (NDVI) is commonly used to compute a vegetation index based on satellite images (Lillesand et al., 2003). It is based on a Map Algebra formula computing the red and near infrared band from satellite images according to formula 2.

$$NDVI = \frac{nearInfrared - red}{nearInfrared + red} \quad (2)$$

In opposite to the Groundwater Vulnerability case study, the NDVI case study utilizes different data sets, but incorporates the same basic processing units, namely Map Algebra local operator for subtraction, addition and division. Figure 2 (Business Process – NDVI box) illustrates the Business Process for NDVI.

CONCLUSIONS AND OUTLOOK

As demonstrated in the case studies, the definition of standardized web processing services leads to highly reusable processing components. When included in a system architecture with support for service chaining, processing units can be chained in a fast and efficient way by defining BPEL-documents (XML-style). The core functionality (in this example: local Map Algebra operations for subtraction, addition, division, and multiplication) can be chained to perform powerful processing tasks. This leads to highly reusable services and an easy integration of new functionality.

Besides the rather simple processing units already implemented, a variety of functions,ranging from simple spatial operators like spatial intersection to highly complex model components used in global climate change science, can be encapsulated inside a WPS.

According to the self-describing nature of WPS and its processes, the inherent usability of standardized SDI is easily achieved.

Using Business Processes as service chaining units, a file format legible to both human and machine, simplifies the development and automation of service chains. The processing services can be loosely coupled with other services of the same kind (i.e. WPS), or of a different nature (e. g. WCTS). According to the stable interface definition of WPS, any kind of functionality can be incorporated inside a WPS and afterwards integrated into SDI and service chains. The completely featured Web GIS in form of a processing-enhanced SDI is at hand.

While effective service chaining utilizing Business Process has been demonstrated in the case studies, fully automatic service chaining is the next frontier in web based geoprocessing. The utilization of self-organizing nets, as opposed to classical GIS scenarios, offers significant advantages. In order to process complex analytical and information generating tasks, several sources of data and information have to be accessed and chained automatically. A spontaneous linkage of services is already possible from a technical point of view. To achieve fully automatic service chaining, besides a syntactical description of web services (in form of get capabilities and describe process XML documents), a semantic description has to be provided as well. A GIS user may know what ‘spatial intersection’ means, while a web service does not.

The future SDI scenario may feature completely automated, spontaneous service chaining: a user requests information from a geoportal to support a spatial decision. The portal then queries a semantically enhanced service registry (Klien et al., 2006) and finds potential services available for the problem solution. Without human interaction, the services are chained, the requested information is generated and delivered to the user. The technical preconditions for this scenario already exists, but a paradigm for describing spatial operations to be used in self-organising nets is still missing. This represents one of the next frontiers in Geographical Information Science.

ACKNOWLEDGEMENT

The authors would like to thank Chad Townsend for proof-reading the paper. Christian Kiehle greatly acknowledges the financial support by the Research and Development- Programme GEOTECHNOLOGIEN, funded by the German Ministry of Education and Research (BMBF) and German Research Foundation (DFG), Grant 03F0372A. This is publication no. GEOTECH – 209.

BIBLIOGRAPHY

Andrews, T., Curbera, F., Dholakia, H., Goland, Y., Klein, J., Leymann, F., Liu, K., Roller, D., Smith, D., Thatte, S., Trickovic, I., Weerawarana, S., 2003 Business Process Execution Language for web services, Version 1.1. Online: ftp://www6.software.ibm.com/software/developer/library/ws-bpel.pdf

Azzam, R., Bauer, C., Bogena, H., Kappler, W., Kiehle, C., Kunkel, R., Leppig, B., Meiners, H.G., Müller, F., Wendland, F., Wimmer, G., 2003 Geoservice Groundwater Vulnerability. Geotechnologien Science Report 2: 31-35.

Bernard, L., Kanellopoulos, I., Annoni, A., Smits, P., 2005 The European geoportal – one step towards the establishment of a European Spatial Data Infrastructure. Computers, Environment and Urban Systems 29 (1): 15-31.

Beaumont, P., Longley, P.A., Maguire, D.J., 2005 Geographic information portals – a UK perspective. Computers, Environment and Urban Systems 29: 49-69.

de La Beaujardiere, J., 2004 OpenGIS Web Map Service (WMS) Implementation Specification (WMS). Online: http://portal.opengeospatial.org/files/?artifact_id=5316.

Diepolder, G. W., 1995 Schutzfunktion der Grundwasserüberdeckung [in german]. GLA Fachberichte (13): 5-79.

Evans, J.D. (ed.), 2003 Web Coverage Service (WCS), Version 1.0.0. OGC, 03-065r6. Online: https://portal.opengeospatial.org/files/?artivact_id=3837.

Fallside, D. C. (ed.), 2001 Schema Part-0: Primer. URL: http://www.w3.org/TR/xmlschema-0.

Gosling, J., Joy, B., Steele, G., Bracha, G., 2005 The Java Language Specification, 3rd Edition. Addison Wesley Professional.

Heier, C.; Kiehle, C., 2005 Standardisierte Geodatenverarbeitung im Internet – der OGC Web Processing Service [in german]. GIS 15 (6): 39-43.

Hölting, B.; Haertlé, T.; Hohberger, K.-H.; Nachtigall, K.-H.; Villinger, E.; Weinzierl, W.; Wrobel, J.P., 1995 Konzept zur Ermittlung der Schutzfunktion der Grundwasserüberdeckung [in german]. Geologisches Jahrbuch 63. Schweizerbart: 7-20.

Kiehle, C., 2006 Entwicklung einer Geodateninfrastruktur zur regelbasierten Ableitung von Geoinformationen aus distributiven Datenbeständen [in german]. Mitteilungen zur Ingenieurgeologie und Hydrogeologie 92.

Klien, E., Lutz, M., Kuhn, W., 2006 Ontology-Based Discovery of Geographic Information Services – An Application in Disaster Management. Computers, Environment and Urban Systems 30 (1): 102-123.

Lake, R., Burggraf, D.S., Trninic, M., Rae, L. (eds.), 2004 Geography Mark-Up Language – Foundation for the Geo-Web. Wiley.

Lillesand, T.M., Kiefer, R.W., Chipman, J.W., 2003 Remote Sensing and Image Interpretation. Wiley & Sons, 2nd edition.

Nebert, D. (Ed), 2004a. Developing Spatial data infrastructures. The SDI Cookbook. Version 2.0. Online: http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf.

Nebert, D. (Ed), 2004b OpenGIS Catalogue Service Implementation Specification (CAT). Online: http://portal.opengeospatial.org/files/?artifact_id=5929&version=2.

Percivall, G., 2003 OGC Reference Model (ORM). Online: http://portal.opengeospatial.org/files/?artifact_id=3836 .

Tait, M.G., 2005 Implementing geoportals: applications of distributed GIS. Computers, Environment and Urban Systems 29: 33-47.

Tomlin, C.D., 1990 Geographic information systems and cartographic modelling. Prentice Hall

Veerawarana, S., Curbera, F., Leymann, F., Storey, T., Fergusaon, D.F., 2005 web services Platform Architecture: SOAP, WSDL, WS-Policy, WS-Addressing, WS-BPEL, WS-Reliable Messaging, and More. Prentice Hall.

Vrba, J., Zaporozec, A. (Eds.), 1994 Guidebook on Mapping Groundwater Vulnerability. International Association of Hydrogeologist (16). Heise.

Vretanos, P. (ed.), 2002 Web Feature Service 1.0. OGC, 02-058. Online: https://portal.opengeospatial.org/files/?artifact_id=7176.

Whiteside, A., Müller, M.U., Fellah, S., Warmerdam, F., 2005 Web Coordinate Transformation Service (WCTS), Online: https://portal.opengeospatial.org/files/?artifact_id=8847