Computers, Environment and Urban Systems 34 (2010) 79–88

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/compenvurbsys

Map generalization and schema transformation of geospatial data combined in a Web Service context

Theodor Foerster[a,*], Lassi Lehto[b], Tapani Sarjakoski[b], L. Tiina Sarjakoski[b], Jantien Stoter[a]

[a] International Institute for Geo-Information Science and Earth Observation (ITC), P.O. Box 6, 7500AA Enschede, The Netherlands

[b] Department of Geoinformatics and Cartography, Finnish Geodetic Institute, Geodeettinen Laitos, PL 15 (Geodeetinrinne 2), 02431 Masala, Finland

ARTICLE INFO

Article history:
Received 13 August 2008
Received in revised form 22 June 2009
Accepted 22 June 2009

Keywords:
Automated generalization
Schema transformation
Content transformation
Web Service classification
OGC WPS

ABSTRACT

The integration of multiple geodata sets is a major challenge for developing Spatial Data Infrastructures (SDIs). Currently, this integration is achieved using schema transformation processes. However, as SDIs mature and the need for more complex transformation processes increases, generalization provides appropriate tools for supporting complex transformations for the integration of different data at different scales. Additionally, if processes for generalization and schema transformation are both available as Web Services, it becomes feasible to combine these two types of processes in Web Service chains. To establish such chains successfully, interoperability is a crucial issue. This paper presents a common service classification addressing the issue of interoperability based on former classifications for generalization and schema transformation processes. The applicability of establishing such processing chains and the applicability of the classification are demonstrated by two process scenarios involving generalization and schema transformation. The feasibility of both process scenarios is studied by implementing them in a Web Service architecture. The presented architecture uses the OGC Web Processing Service (WPS) interface specification.

1. Introduction

The dissemination of geoinformation on the Web has become increasingly important in an information society and in particular in domains such as urban planning to support e-governance and public participation (Carver & Peckham, 1999; Kingston, Carver, Evans, & Turton, 2000; Nedovic-Budic, Feeney, Rajabifard, & Williamson, 2004). To disseminate geoinformation within national Spatial Data Infrastructures (SDIs) and to support the notion of “collect data once, use it many times,” it is important to derive information automatically (e.g., using maps at different scales and adapting them for specific applications). One solution to provide automatically derived spatial information on the Web is Geoprocessing Services, which have been intensively studied over the past few years as network bandwidth and computational capabilities have increased. Thus, it is relevant to study the feasibility of applying Geoprocessing Services to derive geoinformation at a specific scale and for a specific purpose from Web-based geodata.

For deriving geoinformation automatically, the two most important types of geoprocesses are generalization (Mackaness, Ruas, & Sarjakoski, 2007) and schema transformation (Lehto, 2007b). They share a common task, namely *content transformation*. Schema transformation processes transform the thematic characteristics of geodata, whereas generalization processes transform the spatial properties, for example, the transformation of polygonal roads in the source dataset to a road network in the target dataset (Fig. 1). Schema transformation has its origin in general database applications (e.g., renaming an attribute), whereas generalization has its origin in geospatial applications. However, the two types of processes overlap considerably, and a clear separation is not possible in some cases.

Interoperable Web Service interfaces have made it possible to combine schema transformation and generalization processes into a Web-based process. Combining these two processes is required when integrating different data models of different scales to facilitate the reuse of data within and across SDIs (Williamson, Rajabifard, & Feeney, 2003). Reusing data reduces the costs of data maintenance and capturing. More importantly, it assures consistency and enables applications in a cross-national context, such as physical planning (Nedovic-Budic et al., 2004), risk management and environmental monitoring (Williamson et al., 2003).

This paper studies the feasibility of establishing a Web-based process that combines schema transformation and generalization for data integration purposes. First, schema transformation is applied to transform the thematic properties of one dataset to meet

* Corresponding author. Tel.: +31 53 4874373.
E-mail addresses: foerster@itc.nl (T. Foerster), lassi.lehto@fgi.fi (L. Lehto), tapani.sarjakoski@fgi.fi (T. Sarjakoski), tiina.sarjakoski@fgi.fi (L.T. Sarjakoski), stoter@itc.nl (J. Stoter).

0198-9715/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compenvurbsys.2009.06.003

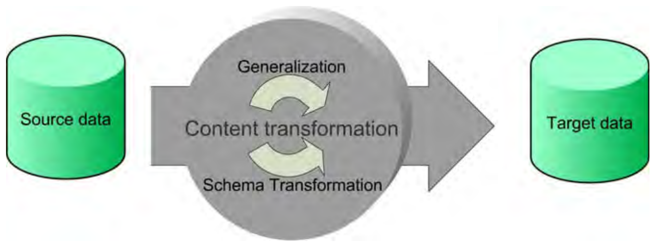

Fig. 1. Generalization and schema transformation combined as a content transformation process.

the thematic properties of the other dataset. Second, the spatial properties are transformed according to the requirements of the target dataset.

To establish such a content transformation process on the Web, we propose a Content Transformation Service. For this service, we introduce a Web Service taxonomy that classifies services according to two criteria: service granularity and service functionality. Service granularity indicates the level of complexity of the exposed process, i.e., whether the process is implemented through a single service or is composed of a set of fine-grained services (Haesen, Snoeck, Lemahieu, & Poelmans, 2008). Service functionality refers to the type of transformation (schema transformation or generalization). This classification is suggested to improve the semantic interoperability of such Web-based processes. Based on this classification, some design guidelines are given for selecting input parameters for the process interfaces.

To prove the feasibility of combining generalization and schema transformation in a Web Services environment, this research applies the proposed approach to a case study. In this case study, a building dataset of the municipality of Helsinki is transformed according to the data model of the Finnish national SDI. The transformation is realized in two alternative implementations and involves both transforming the spatial properties (i.e., generalization) and transforming the thematic properties (i.e., schema transformation).

The presented processes are deployed using the OGC Web Processing Service (WPS, (OGC, 2007)) that is OGC's interface specification for Geoprocessing Services. The specification has been used in several research projects, for instance, in the context of chaining geoprocesses (Schaeffer & Foerster, 2008) and processing large volumes of geodata on the Web (Granell, Diaz, & Gould, 2007). The WPS interface specification also plays an important role in SDI development by providing derived geoinformation on the Web in a standardized way. This has been demonstrated by projects related to, for example, water resource management and fire risk assessment (Diaz, Costa, Granell, & Gould, 2007; Friis-Christensen et al., 2006).

The implementation of the process scenario described in this research is based on the GeoServer application server (GeoServer, 2008) and the 52°North WPS framework (North Geoprocessing Community, 2007).

The following section is an introduction to Geoprocessing Services, specifically in the context of Generalization and Schema Transformation Services. Section 3 discusses the conceptual links between the two types of processes and introduces the classification. Section 4 describes a case study in which both types of services are combined to perform a common task of content transformation. Additionally, the architecture to implement the case study is presented. The paper ends with conclusions and an outlook for further research.

2. Introduction to Geoprocessing Services

This section introduces the basic idea of Geoprocessing Services and related concepts. As Geoprocessing Services are a type of Web Service, this section starts with an introduction about Web Services and their capabilities, as well as the current obstacles to their use. This is followed by a review of related work on Web Generalization Services and Schema Transformation Services. Finally, this section will describe the conceptual relationship between those two types of services.

2.1. Review of Web Services

Web Services connect readily available software components on the Web in a loosely coupled way (Alonso, Casati, Kuno, & Machiraju, 2004). This enables different applications to reuse these Web-enabled software components. Moreover, as Web Services communicate with platform-independent protocols, they can be reused by any application written in any programming language and/or running on any operating system. Overall, a Web Service can be defined as a software component that provides functionality through a Web-accessible interface in a programming language- and platform-independent manner (Vaughan-Nichols, 2002). The Web Services interface is described in a computer-understandable way, which is a fundamental requirement for ensuring interoperability.

For Web Services to interact with each other, they have to be interoperable. The task of establishing interoperability between Web Services is a challenge, as they are connected in a loosely coupled way; that is, the service interaction is established during runtime and the services do not know each other in advance. The ISO standard 19119 "Geographic Information – Services" identifies two levels of interoperability for Web Services (ISO/TC 211, 2005):

- Syntactical – the Web Services use the same structure and input/output format for the information
- Semantic – the Web Services communicate based on an agreed meaning of the message parameters.

Interoperability is also crucial to enable sequencing of multiple Web Service instances, that is, combining different Web Services to achieve a designated goal. This sequencing of services is called Web Service chaining. In the context of Web Service chaining, three types of user interaction have been classified (Alameh, 2003):

- Transparent: involves full user interaction and requires prior knowledge of the user about the service and the context of the application. This is the simplest method of Web Service interaction.
- Translucent: the user is aware of interaction within a Web Service chain, but cannot alter the order.
- Opaque: the chain of services is presented to the user as one service, thus the user is not aware of the chain. The chain is static and preconfigured by the service provider.

The Open Geospatial Consortium (OGC) (OGC, 1994) specifies Web Service interfaces suitable for geospatial applications. These Web Services are mostly referred to as *Geospatial Web Services* (Di, Zhao, Yang, Yu, & Yue, 2005). The specifications of Geospatial Web Services mainly provide syntactical interoperability, as they are concerned with the encoding of the input parameters, but not with their semantics. However, full interoperability, i.e., also addressing interoperability at the semantic level, is still a subject for research and relates to the development of the Geospatial Semantic Web (Bishr, 2006). A promising approach to enabling the semantic interoperability of Geoprocessing Services is the use of ontologies and semantic service classifications, as introduced by Lemmens (2006) and Lutz (2007).

In the context of Geoprocessing Services, the OGC has standardized the WPS interface (OGC, 2007). It describes a simple way to

Web-enable processes using Web Service technology. The service interface provides a straightforward communication pattern involving three operations: *GetCapabilities* to retrieve service metadata, *DescribeProcess* to request process metadata and *Execute* to perform a specific process. In particular, the WPS service interface is more comprehensive than other Web Service standards (Gottschalk, Graham, Kreger, & Snell, 2002), as it already describes a specific means to encode process parameters and specifies basic communication patterns. Lately, Brauner, Foerster, Schaeffer, and Baranski (2009) proposed a research agenda for Geoprocessing Services, which addresses performance, semantic descriptions and service orchestration as the most important issues.

Currently, there is no OGC specification for service chaining. However, Schaeffer and Foerster (2008) present an approach for chaining OGC Web Services that applies the common IT standard for service chaining, called the Business Process Execution Language (BPEL).

There is a lot of discussion about Web Services in general. Their communication protocol is seen as storage-intensive and requires a high level of processing performance. Additionally, there is the trade-off between shipping the execution code to the client and shipping the data to the service. This trade-off results from the huge data volume that might be required to be sent to the Web Service. Nevertheless, Web Services are currently the default approach for providing information on the Web for any type of application, and they represent a promising means for building future applications on the Web (Yu, Liu, Bouguettaya, & Medjahed, 2008), due to their high potential of maintenance and software reuse. A Web Service needs to be updated and maintained only once. This becomes more relevant with the increasing number of client applications that use network-based functionality. Due to the centralized but loosely coupled nature, Web Services can provide up-to-date information that can be integrated by many applications, such as risk management (Annoni et al., 2005).

2.2. Review of Web Generalization Services

Generalization can be defined as the transformation of spatial data from a source model to a target model in relation to the level of detail or scale (Weibel & Dutton, 1999). The concept of "level of detail" is closely related to the granularity of the data content. From a cartographic perspective, the counterpart of the level of detail is the scale. However, "scale" describes a slightly different concept, as it is not only used to describe the reduction of detail (i.e., caused by zooming out), but also addresses the maximization of information regarding the application (e.g., by emphasizing specific aspects of the map).

The concept of a Web Generalization Service has been introduced by Edwardes et al. (2003). They described the desire of the research community to develop an interoperable common research platform by means of Web Services. This platform was intended to facilitate the reuse and exchange of generalization knowledge (i.e., generalization algorithms) within the generalization research community. Later on, Sarjakoski et al. (2005) and Edwardes, Burghardt, and Neun (2005) extended this idea to provide generalization functionality on the Web, either as an atomic or a complex process or even as an all-encompassing generalization process.

Edwardes et al. (2005) introduced a classification of Generalization Services according to the service granularity to improve interoperability of Web Generalization Services (Fig. 2). To each class of service a *service interface* and *a Graphical User Interface (GUI)* has been attached to indicate the possible interaction modes (i.e., computer-to-computer interaction or human-to-computer interaction). At the bottom level of the classification are the simple *generalization support services* that provide basic functionality for enriching the data with structures needed by the generalization process (such as triangulation). The next level is the *generalization operator service*. The operator service is in line with the concept of the generalization operator, derived from the first experiments with automated generalization (McMaster & Shea, 1992). Both the operator service and the support service are used by the *compound generalization service* that drives the generalization process and automatically evaluates the results. Initially, the compound generalization service only had a GUI attached. Foerster, Stoter, and Lemmens (2007) suggested extending it with an additional service interface to allow more flexible communication patterns (e.g., the operator service may call a compound service).

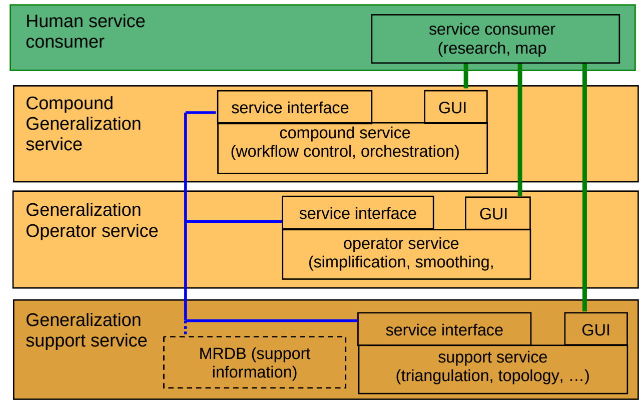

Fig. 2. Web Generalization Service classification originally adopted from Edwardes et al. (2005) and adjusted according to Foerster, Stoter, and Lemmens, (2007).

A common classification of generalization operators is important for interoperability between Web Generalization Services. Operators encapsulate atomic generalization functionality and provide an abstract concept representing the generalization algorithms that implement such an operator. Although there are several proposals for classifying generalization operators (Bader et al., 1999; Foerster, Stoter, & Kobben, 2007; McMaster & Shea, 1992), no classification is available that defines all available operators unambiguously. Foerster, Morales, and Stoter (2008) presented the first attempt to formalize a classification of generalization operators (proposed in Foerster, Stoter, & Kobben, 2007) using the Object Constraint Language (Warmer & Kleppe, 2003).

Various frameworks for Web Service-based generalization have been developed (Bergenheim, Sarjakoski, & Sarjakoski, 2009; Foerster & Stoter, 2006; Neun & Burghardt, 2005). Burghardt, Neun, and Weibel (2005) presented an overview of the evolution of Web Generalization Services. Since then, specification programs of the OGC and the research community have drawn more attention to Generalization Services, especially on operator services (Foerster, Stoter, Koebben, & Oosterom, 2007; Neun, 2007; Neun, Burghardt, & Weibel, 2008). Burghardt and Neun (2006) presented an example of automated service chaining. Foerster, Stoter, and Lemmens (2008) also presented a Web Service architecture, which utilizes multiple Generalization Services to provide customized base maps on the Web. Foerster, Stoter, and Lemmens (2007) provided a recent overview of previous research on Web Generalization Service.

Web Generalization Services are primarily used for a single remotely performed operation, not as chained operations to perform complex generalization involving multiple services. Automated sequencing of generalization functionality on the Web is not yet possible, because service interfaces do not support semantic interoperability. Currently the configuration of the generalization processes always requires human reasoning, because the semantic aspects of the description of the generalization algorithms are available only as readable text. Therefore, a working group, consisting of research institutes, national mapping agencies and software vendors, was recently established to specify designated data types and generalization operators using the WPS interface specification (Foerster, Burghardt, et al., 2008). In the future, the results of this working group are to be applied to different use cases of automated generalization to demonstrate the advantages of the developed framework.

2.3. Review of schema transformation services

The term schema transformation refers to the process that transforms data from one source schema into another target schema. The most typical application for a schema transformation is to provide data in an externally defined schema (e.g., at the European level) from a source dataset stored in a local schema (e.g., at the national level). A schema transformation can be divided into two main phases: defining the transformation (configuration) and performing it (runtime).

The OGC has investigated the concept of a Translating Web Feature Service in its interoperability program, the Critical Infrastructure Protection Initiative Critical Infrastructure Protection Initiative, 2002, within the context of the Transportation Pilot of the *US national Geospatial One Stop initiative* (US, 2009). The service developed for this application delivers transportation-related data from two heterogeneous sources, transformed into the common application schema by two real-time schema transformation processes. The designed processes are defined by style-sheet documents (W3C, 1999).

An approach for defining schema mapping at the conceptual level was presented by Donaubauer, Fichtinger, Schilcher, and Straub (2006), who later discussed how to implement this functionality in the context of OGC-compliant Web Services (Donaubauer, Straub, & Schilcher, 2007). The approach relies on model descriptions for the source and target conceptual schemas and schema mapping instructions. The actual schema transformation is implemented as an extension of the Web Feature Service (WFS) interface definition. The user is able to choose among the source schemas available in this specific type of WFS and can ask for a schema transformation to be registered in the service by indicating the desired target schema and the corresponding schema mapping. The result is a new WFS instance delivering data in the requested target schema. The schema mapping language used in the research is described in detail by Staub, Gnaegi, and Morf (2008).

The interoperability project ORCHESTRA funded by the European Commission has identified "schema mapping service" as a service type in its reference architecture (Lutz, 2006). The specification distinguishes between two related interfaces: the *Schema Mapping Interface* and the *Schema Mapping Repository Interface*. The Schema Mapping Repository Interface is used to manage (create, delete, get, and set) the schema mappings in the service. The Schema Mapping Interface enables schema transformations to be performed at the data instance level.

Lehto (2007a) proposed an approach for schema transformation using style-sheet documents (W3C, 1999). He also provided a general discussion on schema transformation as a Web Service. According to the study, schema transformation can be regarded as a service that exposes a data access interface, tightly integrated with the actual data source providing geospatial content in its original schema. In this approach schema transformation shields the original data source from the calling application and must thus support a two-way processing model, as both the query and the resulting dataset must be translated. In another approach the Schema Transformation Service is seen as a link in a chain of individual services. In this case the interface of a Schema Transformation Service can be seen as a type of data processing interface that can be freely connected with a desired data source. The WPS interface specification is seen as a possible standard for the proposed Schema Transformation Service (Lehto, 2007a). Lehto also introduced an operator classification for schema transformation. This classification is considered essential for enabling meaningful schema transformation processes on the Web.

2.4. Relationship between generalization and schema transformation

Since both generalization and schema transformation transform the content of spatial data, they are highly related. Generalization changes one dataset into another smaller scale dataset to meet cartographic, geometrical and topological criteria. Generalization is achieved primarily by simplifying or aggregating objects, but also by changing their type of geometry (Weibel & Dutton, 1999). Schema transformation also changes the data from a source to a target model, but does not decrease the level of detail (or scale). It changes the context of the data by, for instance, converting the data to another coordinate system or by renaming specific thematic attributes to meet the requirements of the target model.

As there is no formalization available for the concept "level of detail," schema transformation and generalization have a certain overlap. This becomes clear when comparing the generalization operators described in Foerster, Stoter, and Kobben (2007) and the schema transformation operators in Lehto (2007a). Some operators in the classifications appear to be the same (e.g., aggregation), but perform their function on different aspects of the data. Schema transformation operators address the transformation of thematic aspects only, whereas generalization operators address the transformation of spatial aspects of the data. For instance, the aggregation operator of schema transformation is the same operator as in generalization. Both aggregate certain aspects of data, but in the context of schema transformation the aggregation operator merges thematic attributes of the features, whereas, in generalization, the aggregation operator merges a set of geometries to a new single geometry.

Nevertheless, the transformation of spatial data content requires both changing the level of detail and changing the thematic context. The graph in Fig. 3 describes the concept space for content transformation processes. On the one axis the transformation regarding scale (or level of detail) is indicated, on the other axis the transformation of the context. The graph shows the relationship between schema transformation and generalization for a specific content transformation process.

In the remainder of this paper we use "content transformation" for the process that changes the data content by applying schema transformation and generalization in one process.

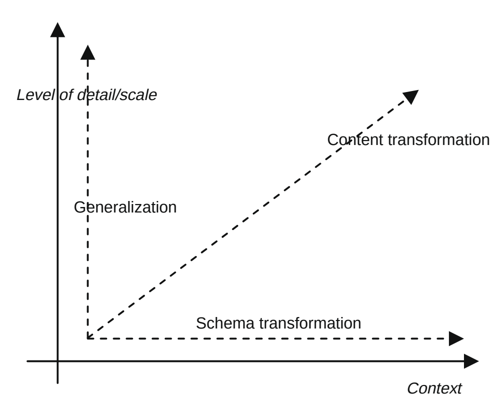

Fig. 3. The concept space for content transformation processes.

3. Content transformation on the Web

This section presents a concept for Web-based content transformation process. The concept of Web Services and Web Service chains is based on a classification of Content Transformation Services that reflects the conceptual relationship between Generalization Services and Schema Transformation Services. The classification is introduced in Section 3.1. Based on the classification, it is possible to enhance Web Service interoperability for content transformation process scenarios. As Web Service chaining is still a semi-automatic process involving human expert knowledge, Section 3.2 presents guidelines for the design of content transformation processes.

3.1. Classification of content transformation services

Classifications of Web Services are meant to enhance interoperable and meaningful Web Service interaction. Web Services can be classified according to their functionality, as in the broad classification of ISO 19119 (ISO, 2005). The ISO classification might be too general for a content transformation process scenario, thus we propose a more detailed classification for Web Generalization Services and Schema Transformation Services. In our proposal services are classified with respect to their functionality (Fig. 4) and their granularity. This enables the interoperability on a concrete level (operator level) resulting in meaningful chaining of, for instance, operator services.

The UML class diagram in Fig. 4 classifies Web Generalization Services and Schema Transformation services based on their functionality. Schema Transformation Services and Generalization Services are both Geoprocessing Services. Generalization Services can be divided into Model Generalization Services and Cartographic Generalization Services, according to the model originally presented by Gruenreich (1992). However, as discussed in Section 2.4, the separation between model generalization and schema transformation is not strict, due to the missing formalization of the level of detail concept. Additionally, the operators for model generalization and schema transformation have overlap. Therefore

Table 1
Classification matrix for content transformation services.

		Granularity	
		Compound Service	Operator Service
Functionality	Model Generalization Service		
	Cartographic Generalization Service		
	Schema Transformation Service		

we propose to classify the Model Generalization Services also as Schema Transformation Services. Thus, model generalization functionality can be modeled as a specialization of schema transformation. This reflects the case that a specific thematic attribute transformation might require spatial analysis (e.g., prior clustering or spatial selection). Also selection might be considered as model generalization as well as schema transformation. Thus, model generalization functionality can be modeled as a specialization of schema transformation. The multi-inheritance of Model Generalization Services allows the service provider to choose the most appropriate implementation, by classifying the specific service either as a Model Generalization Service or as a Schema Transformation Service.

Apart from classifying both types of services by their functionality, we classify them also by their granularity (see also Haesen et al., 2008). This was previously done for Web Generalization Services by Edwardes et al. (2005). Schema Transformation Services can also be divided according to their granularity into compound and operator services. Combining the two classifications leads to a matrix, presented in Table 1. According to the matrix, a service can be classified as compound or operator service depending on its granularity, but it can also be classified according to the functionality it provides: model generalization,

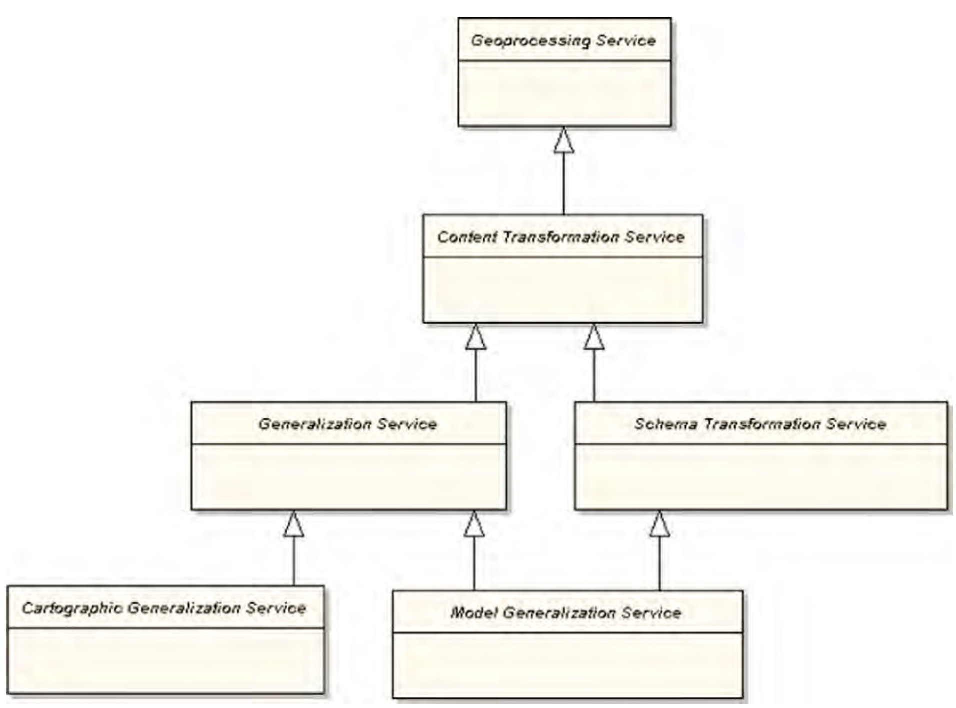

Fig. 4. UML model of the service classification according to their functionality.

cartographic generalization and/or schema transformation. Furthermore, the operator services as proposed here might be classified in more detail by using the classifications of schema transformation operators (Lehto, 2007a) and generalization operators (Foerster, Stoter, & Kobben, 2007) to assign more meaning to the operator service. For example, an operator service providing "collapse" functionality might be published in a more meaningful way through the additional classification of generalization operators. Thus, a Web Service providing collapse functionality might be described as an operator service (granularity) and as a model generalization service (functionality) and in more detail as a Web Service providing an operator according to the operator classifications of Foerster, Stoter, and Kobben (2007).

The following subsection presents guidelines to link Generalization Services and Schema Transformation Services, aiming at reusability and cost efficiency. These guidelines may offer support a complex process scenario for content transformation based on a sequence of Web Services for both schema transformation and generalization.

3.2. Guidelines for designing content transformation processes

One key aspect when designing geoprocessing chains is deciding how much knowledge is required at each node of the geoprocessing chain for optimum performance. Much unnecessary knowledge at each node increases the complexity of geoprocessing chains and must therefore be avoided.

To avoid high complexity we suggest a modular design, which prevents redundant knowledge and aims at reusability. When looking at the Generalization Services and Schema Transformation Services as well as at their granularity, the most suitable entry point to call the other service (in our case, from schema transformation to generalization) is the compound service. This type of service is able to trigger other fine-grained services. It contains the knowledge to perform the process by using other services and can be reused in multiple service chains. The design of such compound services should always follow the principle of separation of concerns, which describes the encapsulation of knowledge in modules (i.e., services) and improves their reuse in other service chains (Parnas, 1979).

Besides the proposal to follow the separation of concerns by encapsulating specific knowledge as compound services, it may increase the interoperability of Content Transformation Services to specify specific parameters for a specific type of content transformation process. However, this is an exhaustive task and a complete list of parameters may not exist, so we can only give some consideration to specific parameters. The type of parameter for generalization services relates mostly to the type of generalization operator (model or cartographic). Thus, we suggest using features (as defined in the ISO 19109 (ISO/TC 211, 2003)) for interface parameters related to model generalization and cartographic features (as defined in OGC Go-1 Application objects) for interface parameters related to cartographic generalization. Cartographic features are those that have symbolization attached for visual representation. Symbolization is an important aspect for cartographic generalization, as this describes the representation of the final object on the map. The separation of interface parameters for model generalization and for cartographic generalization has been successfully applied in Foerster, Stoter, and Kobben (2007) and Foerster, Stoter, et al. (2008) for describing the different types of generalization operators.

A last design guideline for Content Transformation Services is to distinguish between level of detail as a parameter for model generalization and a set of cartographic constraints as a parameter for cartographic generalization. For these parameters a formalized concept would be required, which does not exist (Sarjakoski, 2007).

3.3. Implementation of Web-based content transformation processes

To show the benefits of combining generalization and schema transformation in a Web-based process and to illustrate the guidelines, we used a case study. The case study transforms a building dataset of the municipality of Helsinki according to a data model deployed in the national Finnish SDI to use it in an application compliant with the Finnish SDI. To learn more about the implementations of content transformation processes, we implemented two process scenarios for the case study.

4. The case study

The transformation includes attribute renaming, coordinate transformation, selection of a subset of buildings based on minimum area and collapsing their geometry type from a polygon to

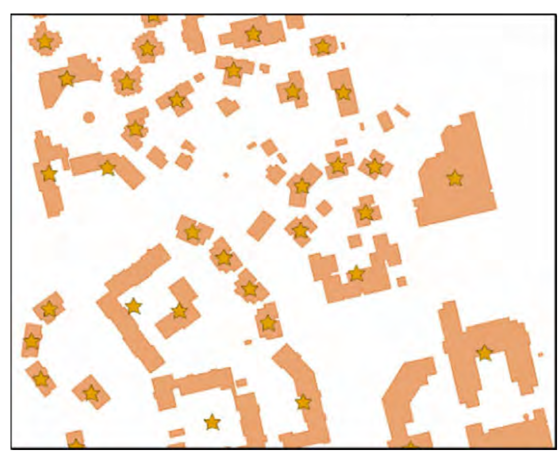

Fig. 6. An example map depicting Helsinki building polygons as source data and the symbolized process result (selected and collapsed building geometries).

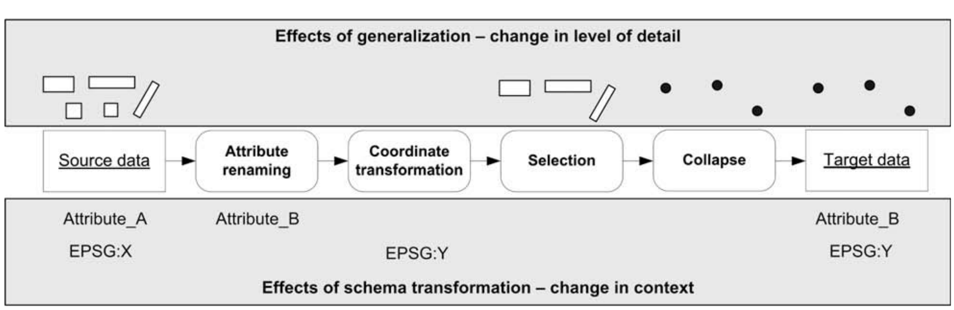

Fig. 5. Processing steps of the content transformation process applied in the presented case study.

a point. This transformation process is depicted in Fig. 5. The original data and the result of the content transformation process are shown in Fig. 6. The schema transformation addresses attribute renaming and coordinate transformation. For the generalization, first the buildings, which exceed a specific area size, are selected. Second, the collapse operator is applied on the selected buildings and returns the geometric center of each polygon. In a later stage, the generalization process might be further developed to offer a more sophisticated collapse algorithm, such as described in Li (2006).

4.1. The two implemented process scenarios

We implemented the case study in two alternative process scenarios using different Web Service instances to study how the generalization process can be combined with a schema transformation process (Fig. 7).

Process scenario I performs the generalization process within a dedicated compound Generalization Service. In this process scenario, the Schema Transformation Service is designed for a pure thematic processing task. The detailed knowledge related to the spatial processing of the input features is supposed to reside within the remote Model Generalization Service. This Model Generalization Service is designed as a compound Generalization Service that calls the Generalization Operator Services one after another. The Model Generalization Service is called with a level of detail parameter, as suggested in Section 3.2. As there is no concept available for the formalization of level of detail, we represent this parameter by using the source and target data model plus some additional transformation information (i.e., the value of the area threshold for the select by area operator). Based on the input parameters the compound generalization service decides which services have to be called and which parameter values have to be chosen.

Process scenario II incorporates the necessary generalization knowledge within the Schema Transformation Service to configure generalization functionality. The Schema Transformation Service detects a need for generalization from the respective schema mapping definition. The service then determines the generalization operators required for the schema transformation and subsequently calls the generalization operator services (select, collapse). It is important to note that the Schema Transformation Service would request the same generalization operator service instances as the compound service in process scenario I.

Both process scenarios use the concepts introduced in Section 3 and demonstrate how interoperability can be enhanced by the application of the classification for Content Transformation Services. The strategies of the two process scenarios achieve the transformation differently, regarding the principle of separation of concerns. Process scenario I applies a compound service as an intermediate step and thereby allows separating the required knowledge of schema transformation from the knowledge for generalization. Process scenario II incorporates the expertise of the generalization process inside the schema transformation itself. Regarding the configuration of services, process scenario I results in services that provide distinct functions and thereby increase the possible reuse of each service while lowering the complexity of the overall service chain. Thus, process scenario I is regarded as the more favorable one.

4.2. The implemented architecture

The architecture for the case study is illustrated in Fig. 8. To compare Web Service-based geoprocessing in different organizations, the services for our implementation ran at two locations: the International Institute for Geo-Information Science and Earth Observation (ITC) and the Finnish Geodetic Institute (FGI). Both process scenarios were implemented with WPS instances using the 52°North WPS framework (North Geoprocessing Community, 2007) running at both locations. Additionally, we wrapped the schema transformation process (Step 2, Fig. 8) (Schema Transformation Service), as a Web Map Service (WMS) (based on GeoServer application server, GeoServer, 2008) by using a special GeoServer concept for wrapping resources called *data stores*. Any WPS instance can be published as a WMS by applying the concept of data stores. This provides WPS functionality through a standardized and already well-established interface (i.e., WMS), enabling a seamless integration of this new type of service into already existing client applications.

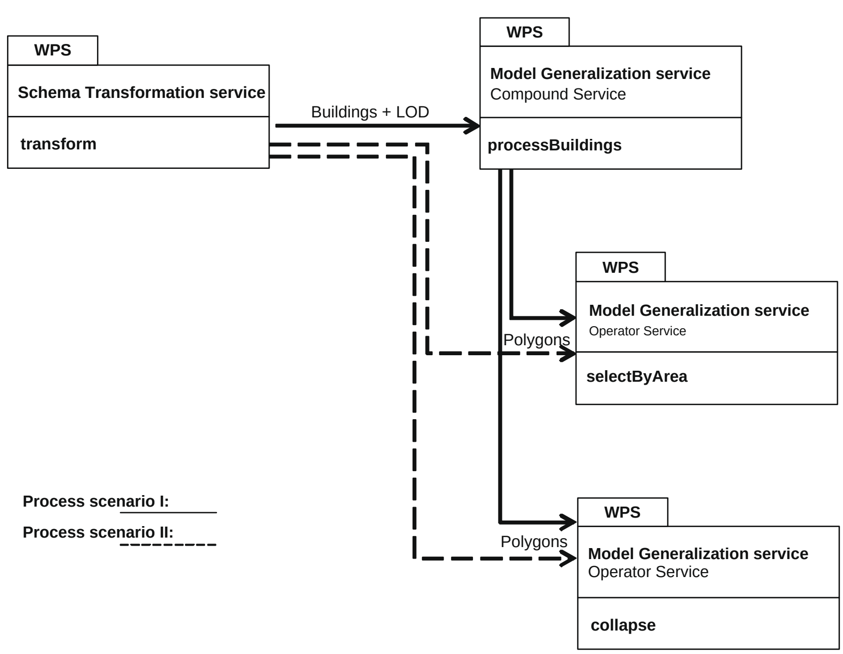

Fig. 7. Two process scenarios for the inter-service communication between a Schema Transformation and a Model Generalization service. The process scenarios reflect the classification of granularity into compound services and operator services.

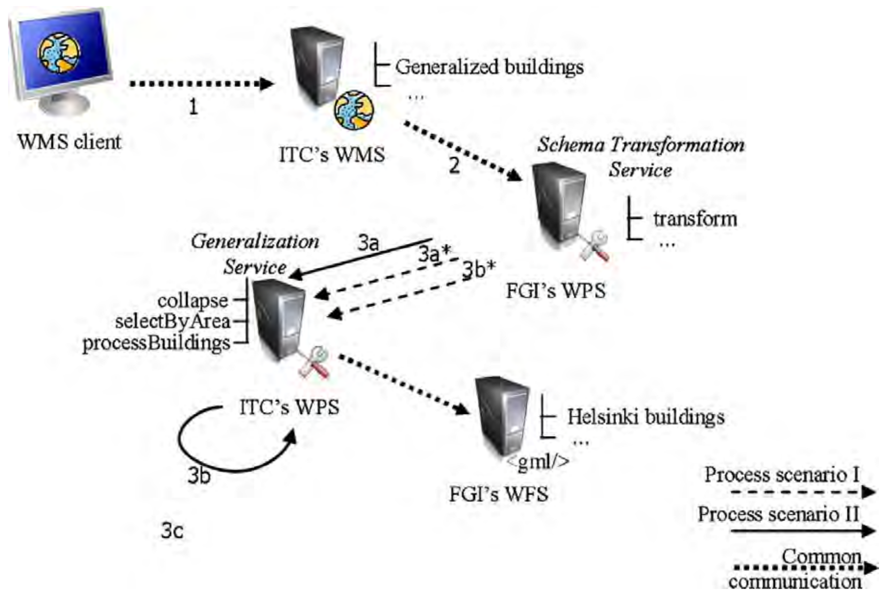

Fig. 8. The architecture of the service chain realizing the case study for content transformation.

The process data (buildings of the city of Helsinki) are served through a WFS instance. The difference of the process scenarios is realized in Step 3 (Fig. 8). In the case of process scenario I (Section 4.2), the Schema Transformation Service calls the model generalization process `processBuildings` (Step 3a) calls the generalization processes hosted at the ITC (Steps 3b and 3c). In the case of process scenario II, the FGI's Schema Transformation Service calls both generalization operator services successively (select and collapse; Steps 3a∗ and 3b∗). The WFS data are retrieved whenever required by the process workflow.

4.3. Discussion of the implementation

Deploying the architecture was a straightforward process. The 52°North WPS framework uses standard open source tools that provide a solid basis for developing more sophisticated processes. In general, all deployed services are based on open source solutions, such as the 52°North WPS framework and the GeoServer application server. The result of the processes (served through the WMS interface) can be visualized in the uDig desktop client (uDig, 2009) or in any WMS-compliant browser client. The uDig desktop client was additionally extended as a WPS client (Schaeffer & Foerster, 2008) to access the different WPS instances and to configure the different process scenarios. The configured processes, which execute instructions for the WPS, were then deployed to the GeoServer application server and thereby allowed us to visualize the process results through the WMS interface. The results of the generalization process are directly visible on the retrieved map, and the results of the schema transformation process can be queried via the WMS interface operation *GetFeatureInfo*.

The volume of the messages sent between the services was reduced to a minimum by exchanging only the reference of the data between the different services (i.e., reference to the FGI's WFS storing Helsinki building data). This capability of passing data between client and server is described as one of the features of the WPS interface specification. This feature has two main advantages. First, most of the servers are located on faster network nodes and can thereby retrieve the data efficiently. Second, by using the reference, the server can store the data once and use it multiple times, thereby reducing the communication overhead to a minimum. However, processing data on several distributed services might require rebuilding special data structures multiple times. This is a disadvantage of most of the distributed architectures and can only be solved by encoding the data in the applicable data structure. Generally such an architecture allows processing of the most recent data and thereby provides the most up-to-date information to the application. This is a major advantage for many applications such as navigation and disaster management.

It is important to note that the compound service (process scenario I) could have been implemented using workflow scripts based, for instance, on BPEL. However, for the implementation of the process scenario this would have not made any difference, as the workflow scripts would have also been created in advance, as were the WPS instances representing the compound service.

5. Discussion and conclusions

The paper demonstrates how generalization and schema transformation can be combined to implement complex content transformation processes covering both spatial and thematic transformations. By combining those two types of processes, it is possible to create transformation processes that address a change in the level of detail/scale (generalization) and a change in the context (schema transformation), making use of expert knowledge in two domains (Section 2.4). Such complex content transformation processes are required to integrate data in the context of SDIs. However, the combination of different types of processes within a Web Service environment is still a difficult task, as the processes are not fully interoperable. Therefore, we propose a classification of Content Transformation Services (Section 3.1).

This classification is the first attempt toward providing a comprehensive as well as an extensible approach for Content Transformation Services. The proposed classification supports identification and description of different types of functionality that can be modeled as separated Web Services. The classification may be further developed into a set of process profiles for Geoprocessing Services. Process profiles are currently seen as a possible solution to enhance semantic interoperability of Geoprocessing Services, as already discussed by Nash (2008). Finally, the classification can be applied in the context of the Semantic Web using the approaches of Lemmens (2006). The case study demonstrates that the principle of separation of concerns (Section 3.2) should always guide the design of geoprocessing chains, especially in highly modularized systems, such as in a Web Service environment. Based on this experience, we suggest that process scenario I is more suitable (Section 4.2). In addition, the prototypical implementation shows that it is possible to combine generalization and schema transformation to a content transformation process using Web Services. Additionally, it shows that the combination of different processes hosted at different organizations (FGI & ITC) can be achieved using Web Services in combination with the WPS interface specification. Practical approaches such as wrapping the content transformation process as a mapping service and passing the data via reference helped to build up a sustainable architecture. We gained more insight into the technological aspects by testing the two implemented process scenarios.

In the future, the compound services that are implemented as WPS instances might be replaced by workflow scripts, as has been proposed by Schaeffer and Foerster (2008). Such workflow scripts are easier to maintain by the service providers, but have also to be specified in advance and do not support automatic discovery or integration of Web Services.

It is important to note that, although the introduced classification improves the interoperability of Web Services, it does not solve the problem of establishing appropriate generalization processes for content transformation (e.g., the implementation of the collapse operator is very simple and topological errors are currently not addressed). This remains a subject for further research, especially in the domain of cartographic generalization.

Acknowledgements

The work described in this paper was carried out during a research visit by Theodor Foerster to the Finnish Geodetic Institute. The work has been supported by the RGI international exchange program.

References

Alameh, N. (2003). Chaining geographic information web services. *IEEE Internet Computing, 07*, 22–29. doi:10.1109/MIC.2003.1232514.

Alonso, G., Casati, F., Kuno, H., & Machiraju, V. (2004). *Web services* (1st ed.). Springer Verlag. p. 354.

Annoni, A., Bernard, L., Douglas, J., Greenwood, J., Laiz, I., Lloyd, M., et al. (2005). Orchestra: Developing a unified open architecture for risk management applications. In P. V. Oosterom, S. Zlatanova, & E. M. Fendel (Eds.), *Geo-information for disaster management* (pp. 1–17). Springer. doi:10.1007/b139115.

Bader, M., Barrault, M., Regnauld, N., Mustiere, S., Duchene, C., Ruas, A., et al. (1999). *AGENT workpackage D2 – Selection of basic algorithms*. Department of Geography, University of Zurich.

Bergenheim, W., Sarjakoski, L. T., & Sarjakoski, T. (2009). A web processing service for GRASS GIS to provide on-line generalisation. In J. Haunert, B. Kieler, & J. Milde (Eds.), *12th AGILE international conference on geographic information science*. Hannover, Germany: IKG, Leibniz University of Hanover. <http://www.ikg.uni-hannover.de/agile/fileadmin/agile/paper/126.pdf>.

Bishr, Y. (2006). Geospatial semantic web. In S. Rana & J. Sharma (Eds.), *Frontiers in geographic information technology* (pp. 139–154). Springer.

Brauner, J., Foerster, T., Schaeffer, B., & Baranski, B. (2009). Towards a research agenda for geoprocessing services. In J. Haunert, B. Kieler, & J. Milde (Eds.), *12th AGILE international conference on geographic information science*. Hanover, Germany: IKG, Leibniz University of Hanover. <http://www.ikg.uni-hannover.de/agile/fileadmin/agile/paper/124.pdf>.

Burghardt, D., & Neun, M. (2006). Automated sequencing of generalization services based on collaborative filtering. In M. Raubal, H. Miller, A. Frank, & M. Goodchild (Eds.), *Proceedings of geographic information science, IfgiPrints, IfgiPrints* (Vol. 28, pp. 41–43). Muenster, Germany.

Burghardt, D., Neun, M., & Weibel, R. (2005). Generalization services on the web – Classification and an initial prototype implementation. *Cartography and Geographic Information Science, 32*, 257–268. doi:10.1559/152304005775194665. 12.

Carver, S., & Peckham, R. (1999). Using GIS on the internet for planning. In J. Stillwell, S. Geertman, & S. Openshaw (Eds.), *Geographical information and planning, advances in spatial science* (pp. 371–390). Springer-Verlag.

Di, L., Zhao, P., Yang, W., Yu, G., & Yue, P. (2005). Intelligent geospatial web services. In *Proceedings of geoscience and remote sensing symposium, 2005* (pp. 1229–1232). IEEE International.

Diaz, L., Costa, S., Granell, C., & Gould, M. (2007). Migrating geoprocessing routines to web services for water resource management applications. In M. Wachowicz & L. Bodum (Eds.), *10th AGILE international conference on geographic information science*. Denmark: Aalborg University.

Donaubauer, A., Fichtinger, A., Schilcher, M., & Straub, F. (2006). Model Driven approach for accessing distributed spatial data using web services – Demonstrated for cross-border GIS applications. In *Proceedings of the XXIII international FIG congress*. Munich, Germany. <http://www.fig.net/pub/fig2006/papers/ts82/ts82_04_donaubauer_etal_0603.pdf>.

Donaubauer, A., Straub, F., & Schilcher, M. (2007). mdWFS: A concept of web-enabling semantic transformation. In *Proceedings of the 10th agile conference on geographic information science*. Aalborg, Denmark. <http://people.plan.aau.dk/~enc/AGILE2007/PDF/157_PDF.pdf>.

Edwardes, A., Burghardt, D., Bobzien, M., Harrie, L., Reichenbacher, T., Sester, M., et al. (2003). Map generalization technology: Addressing the need for a common research platform. In *Proceedings of the 21st international cartographic conference (ICC)*. Durban, South Africa.

Edwardes, A., Burghardt, D., & Neun, M. (2005). Interoperability in map generalization research. In *International symposium on generalization of information 2005*. Berlin, Germany.

Foerster, T., Burghardt, D., Neun, M., Regnauld, N., Swan, J., & Weibel, R. (2008). Towards an interoperable web generalisation services framework – Current work in progress. In *Proceedings of the 11th ICA workshop on generalization and multiple representation*. Montpellier, France.

Foerster, T., Morales, J., & Stoter, J. E. (2008). A classification of generalization operators formalised in OCL. In M. B. Edzer Pebezma & T. Bartoschek (Eds.), *Proceedings of GI-days 2008, Ifgi Prints* (pp. 141–156). Muenster, Germany: IFGI.

Foerster, T., & Stoter, J. E. (2006). Establishing an OGC web processing service for generalization processes. In *ICA workshop on generalization and multiple representation*. Portland, Oregon, USA. <http://aci.ign.fr/Portland/paper/ICA2006-foerster_stoter.pdf> Retrieved 19.09.08.

Foerster, T., Stoter, J. E., & Kobben, B. (2007). Towards a formal classification of generalization operators. In *Proceedings of the 23rd international cartographic conference (ICC)*. Moscow, Russia.

Foerster, T., Stoter, J. E., Koebben, B., & Oosterom, P. V. (2007). A generic approach to simplification of geodata for mobile applications. In M. Wachowicz, & L. Bodum (Eds.), *AGILE 2007*. Denmark: Aalborg University. <http://www.plan.aau.dk/~enc/AGILE2007/PDF/41_PDF.pdf> Retrieved 9.10.08.

Foerster, T., Stoter, J. E., & Lemmens, R. (2007). Towards automatic web-based generalization processing: A case study. In *10th ICA workshop on generalisation and multi representation*. Moscow, Russia.

Foerster, T., Stoter, J. E., & Lemmens, R. (2008). An interoperable web service architecture to provide base maps empowered by automated generalization. In A. Ruas & C. Gold (Eds.), *Headway in spatial data handling, lecture notes in geoinformation and cartography* (pp. 255–276). Montpellier, France: Springer.

Friis-Christensen, A., Bernard, L., Kanellopoulos, I., Nogueras-Iso, J., Peedell, S., Schade, S., et al. (2006). Building service oriented application on top of a spatial data infrastructure – A forest fire assessment example. In *Proceedings of the 9th agile conference on geographic information science* (pp. 119–127). Visegrad, Hungary.

Gottschalk, K., Graham, S., Kreger, H., & Snell, J. (2002). Introduction to web service architecture. *IBM Journal, 41*(2), 170–177. doi:10.1147/sj.412.0170.

Granell, C., Diaz, L., & Gould, M. (2007). Managing earth observation data with distributed geoprocessing services. In *Geoscience and remote sensing symposium* (pp. 4777–4780). IEEE. doi:10.1109/IGARSS.2007.4423928.

Gruenreich, D. (1992). ATKIS – A topographic information system as a basis for GIS and digital cartography in Germany. From digital map series to geo-information systems. *Geologisches Jahrbuch Reihe A*(Heft 122), 207–216.

Haesen, R., Snoeck, M., Lemahieu, W., & Poelmans, S. (2008). On the definition of service granularity and its architectural impact. In *Proceedings of the 20th international conference on advanced information systems engineering (CAiSE'08), lecture notes in computer science* (pp. 375–389). Montpellier, France: Springer Verlag. <http://www.kuleuven.be/sbd/upload/CAiSE08-published.pdf>.

ISO/TC 211. (2003). *Geographic information – Rules for application schema.* ISO Standard 19109, International Organization for Standardization.

ISO/TC 211. (2005). *Geographic information – Services* (p. 67). ISO Standard 19119, International Organization for Standardization.

Kingston, R., Carver, S., Evans, A., & Turton, I. (2000). Web-based public participation geographical information systems: An aid to local environmental decision-making. *Computers Environment and Urban Systems, 24*(2), 109–125. doi:10.1016/S0198-9715(99)00049-6.

Lehto, L. (2007a). Schema translations in a web service based SDI. In M. Wachowicz & L. Bodum (Eds.), *10th Agile international conference on geographic information science 2007*. Denmark: Aalborg University.

Lehto, L. (2007b, December 14). *Real-time content transformations in a web service-based delivery architecture for geographic information*. PhD thesis, Helsinki University of Technology – Department of Surveying. <http://lib.tkk.fi/Diss/2007/isbn9789517112727/isbn9789517112727.pdf>.

Lemmens, R. (2006). *Semantic interoperability of distributed geo-services*. PhD thesis, Delft University of Technology. <http://www.ncg.knaw.nl/Publicaties/Geodesy/pdf/63Lemmens.pdf>.

Li, Z. (2006). *Algorithmic foundation of multi-scale spatial representation* (1st ed.). CRC Press (Taylor & Francis). p. 282.

Lutz, M. (2006). *Specification of the schema mapping service*. OA services specifications. Deliverable D3.4.3, Joint Research Centre of the EU Commission.

Lutz, M. (2007). Ontology-based descriptions for semantic discovery and composition of geoprocessing services. *GeoInformatica, 11*(1), 1–36. doi:10.1007/s10707-006-7635-9.

Mackaness, W. A., Ruas, A., & Sarjakoski, L. T. (Eds.). (2007). *Generalisation of geographic information: Cartographic modelling and applications*. Elsevier.

McMaster, R. B., & Shea, K. S. (1992). *Generalization in digital cartography*. American Association of Geographers.

Nash, E. (2008). WPS application profiles for generic and specialised processes. In E. Pebezma, M. Bishr, & T. Bartoschek (Eds.), *Proceedings of GI-days 2008* (pp. 69–79). Institute for Geoinformatics.

Nedovic-Budic, Z., Feeney, M. F., Rajabifard, A., & Williamson, I. (2004). Are SDIs serving the needs of local planning? Case study of Victoria, Australia and Illinois, USA. *Computers, Environment and Urban Systems, 28*(4), 329–351. doi:10.1016/S0198-9715(03)00042-5.

Neun, M. (2007). *Data enrichment for adaptive map generalization using web services*. PhD thesis, University of Zurich.

Neun, M., & Burghardt, D. (2005). Web services for an open generalisation research platform. In *8th ICA workshop on generalisation and multiple representation*. A Coruna.

Neun, M., Burghardt, D., & Weibel, R. (2008). Web service approaches for providing enriched data structures to generalisation operators. *International Journal of Geographical Information Science, 22*(2), 133–165.

OGC, (2007). *Open GIS web processing service*. OGC implementation specification, open geospatial consortium. <http://www.opengeospatial.org/standards/wps>.

Parnas, D. L. (1979). On the criteria to be used in decomposing systems into modules. In E. N. Yourdon (Ed.), *Classics in software engineering, ACM classic books series* (pp. 139–150). Upper Saddle River, NJ, USA: Yourdon Press.

Sarjakoski, L. T. (2007). Generalization of geographic information cartographic modelling and applications. In W. A. Mackaness, A. Ruas, & L. T. Sarjakoski (Eds.) (pp. 11–35). Elsevier.

Sarjakoski, T., Sester, M., Sarjakoski, L. T., Harrie, L., Hampe, M., Lehto, L., & Koivula, T. (2005). Web generalisation services in GiMoDig – Towards a standardised service for real-time generalisation. In F. Toppen, & M. Painho (Eds.), *AGILE 2005* (pp. 509–518). Estoril, Portugal.

Schaeffer, B., & Foerster, T. (2008). A client for distributed geo-processing and workflow design. *Journal for Location Based Services, 2*(3), 194–210. doi:10.1080/17489720802558491.

Staub, P., Gnaegi, H. R., & Morf, A. (2008). Semantic interoperability through the definition of conceptual model transformations. *Transactions in GIS, 12*(2), 193–207. doi:10.1111/j.1467-9671.2008.01095.x.

Vaughan-Nichols, S. (2002). Web services: Beyond the hype. *IEEE Computer, 35*(2), 18–21. doi:10.1109/2.982908.

W3C, (1999). *XSL transformations (XSLT)*. W3C recommendation, W3C. <http://www.w3.org/TR/xslt> Retrieved 22.01.09.

Warmer, J., & Kleppe, A. (2003). *The object constraint language* (2nd ed.). Addison Wesley. p. 206.

Weibel, R., & Dutton, G. (1999). Generalising spatial data and dealing with multiple representations. In P. Longley, M. Goodchild, D. Maguire, & D. Rhind (Eds.), *Geographic information systems – Principles and technical issues* (2nd ed., Vol. 1, pp. 125–155), John Wiley & Sons.

Williamson, I., Rajabifard, A., & Feeney, M. F. (2003). Future directions in SDI development. In *Developing spatial data infrastructures: From concept to reality* (pp. 301–312). Taylor & Francis.

Yu, Q., Liu, X., Bouguettaya, A., & Medjahed, B. (2008). Deploying and managing web services: Issues, solutions, and directions. *The VLDB Journal – The International Journal on Very Large Data Bases, 17*(3), 537–572. doi:10.1007/s00778-006-0020-3.

Web References

GeoServer (2008). <http://www.geoserver.org> 23.01.09.

52° North, 52° North Geoprocessing Community (2007). <http://www.52north.org/wps> 23.01.09.

OGC (1994). Open geospatial consortium. <http://www.opengeospatial.org> 23.01.09.

Critical Infrastructure Protection Initiative (2002). <http://www.opengeospatial.org/projects/initiatives/cipi2> 21.01.09.

US national Geospatial One Stop initiative (2009). <http://gos2.geodata.gov/wps/portal/gos> 21.01.09.

uDig (2009). User-friendly desktop internet GIS. <http://udig.refractions.net> 23.01.09.