SUOMEN GEODEETTISEN LAITOKSEN JULKAISUJA
VERÖFFENTLICHUNGEN DES FINNISCHEN GEODÄTISCHEN INSTITUTES
PUBLICATIONS OF THE FINNISH GEODETIC INSTITUTE
N:o 138

REAL-TIME CONTENT TRANSFORMATIONS IN A WEB SERVICE-BASED DELIVERY ARCHITECTURE FOR GEOGRAPHIC INFORMATION

Lassi Lehto

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Surveying for public examination and debate in the Auditorium M1 at the Helsinki University of Technology (Espoo, Finland) on December 14th, 2007, at 12 noon.

KIRKKONUMMI 2007

ISBN-13: 978-951-711-271-0 (paperback)
ISBN-13: 978-951-711-272-7 (PDF, http://lib.tkk.fi/Diss/)
ISSN 0085-6932

Helsinki University Press
Helsinki 2007

ABSTRACT

The proliferation of location-aware consumer applications is increasing the demand for Web service-based delivery of geospatial content. Significant national and European-wide initiatives have been launched to develop Spatial Data Infrastructures (SDI) aiming at more efficient dissemination of public-sector geodata resources. These frameworks largely depend on the availability of online services and datasets. At the same time, geodata-related network solutions are becoming more sophisticated, due to the gradual introduction of feature-based representations of geographic information into the Web application domain. Along with the growing number of different application areas and content resources involved, the diversity challenges related to client platforms and source databases have also increased.

The interoperability opportunities provided by the broad adoption of Web standards, on one hand, and the societal pressure created by the diverse needs of the wide consumer use base and public sector efficiency requirements, on the other, create the framework within which the present study was conducted. One central goal is the broad application of general-purpose Web methodologies to problems encountered in the geospatial computing domain. The research aimed to resolve the heterogeneity challenges by real-time content transformations, applied in a network service-based delivery architecture. The present thesis focuses on three particular applications of real-time content transformation: multi-purpose publishing, on-the-fly generalisation and schema transformation. These processes are carried out in a distributed multi-tier service architecture framework, designed in the study.

The results of the study confirm the applicability of generic stylesheet-based mechanisms for platform-specific geodata visualisation. On-the-fly generalisation is found to be a valuable tool for creating highly personalised map displays in which dynamic context parameters, such as the current location of the user, can be taken into account. The study proved real-time schema transformations to be a feasible solution for ensuring the dynamic service-level interoperability of heterogeneous source databases. The declarative way used to configure the transformations proved to be especially beneficial, enabling two-way query and data transformations to be defined using the same base methodologies.

A service architecture framework for advanced geospatial Web applications was designed in the study. This six-tier architecture consists of the following conceptual service layers, each with a well-defined responsibility in the service provision: a data service layer, data integration layer, data processing layer, portal layer, value-adding service layer and client layer. Open standardised service interfaces are used throughout the architecture framework.

The three real-time content transformation types were successfully implemented in the extensive service platform developed in the EU-funded GiMoDig project. In this test environment topographic base datasets from four countries (Denmark, Finland, Germany, Sweden) are integrated by means of a two-way real-time schema transformation. The resulting harmonised dataset is further processed by an on-the-fly generalisation process, which adjusts it to the actual presentation scale and use case. Finally, the dataset is transformed into a visual map by an online multi-purpose publishing process, taking the particularities of the used client device into account. The final result is a seamless cross-border map service capable of real-time integration of heterogeneous source datasets and of dynamic adaptation to various different use platforms.

Keywords: Internet GIS, transformation service, schema transformation, service architecture

TIIVISTELMÄ

Paikkatietoon liittyvien käyttösovellusten yleistyminen lisää tarvetta paikkatietosisältöjen saatavuudesta Web-pohjaisina verkkopalveluina. Merkittäviä kansallisia ja Euroopan laajuisia aloitteita on käynnistetty kehittämään paikkatietoinfrastruktuureja, joiden tavoitteena on tehostaa julkishallinnon ylläpitämien paikkatietoaineistojen jakelua. Kehitettävät paikkatietoinfrastruktuurit rakentuvat oleellisilta osin verkkopalveluina saatavilla olevien paikkatietojen varaan. Samanaikaisesti verkkopalveluihin pohjautuvat paikkatietosovellukset ovat kehittymässä entistä monipuolisimmiksi. Tähän kehitykseen on vaikuttanut erityisesti kohdemuotoisten paikkatietojen vähittäinen käyttöönotto Web-sovelluksissa. Paikkatietoja hyödyntävien sovellusalueiden ja sovelluksissa hyödynnettävien tietovarastojen lisääntyminen on aiheuttanut merkittäviä haasteita käytettävien sovellusalustojen ja tietokantajärjestelmien heterogeenisyyden takia.

Tämä tutkimus sijoittuu kahden merkittävän tekijän muodostamaan viitekehykseen. Yhtäällä ovat ne mahdollisuudet, joita laajalti sovelletut Web-standardit tarjoavat paikkatietojen yhteiskäyttöisyydelle. Toisaalla vaikuttavat yhteiskunnalliset paineet liittyen laajentuneen paikkatietojen käyttäjäkunnan tarpeisiin ja julkishallinnon tehokkuustavoitteisiin. Tutkimuksen keskeinen periaate on yleiskäyttöisten Web-metodien laaja soveltaminen paikkatietoalan ongelmissa. Tavoitteena on ratkaista em. heterogeenisyysongelma tietoverkkopalveluihin pohjautuvassa sovellusarkkitehtuurissa suoritettavien tosiaikaisten sisältömuunnosten avulla. Tutkimus keskittyy kolmeen sisältömuunnoksen soveltamisalueeseen: monikanavajulkaiseminen, tosiaikainen yleistys ja skeemamuunnos. Sisältömuunnoksia varten on tutkimuksessa kehitetty monitasoinen palveluarkkitehtuuri.

Tutkimuksen tulokset osoittavat, että yleiskäyttöisiä tyylisivuteknologiaan perustuvia mekanismeja voidaan soveltaa paikkatietojen visualisointiin, niin että käytetty laiteympäristö otetaan huomioon. Tutkimuksen mukaan tosiaikaista yleistystä voidaan onnistuneesti soveltaa personoitujen karttojen tuottamiseen. Lähestymistapa mahdollistaa dynaamisten tekijöiden, kuten käyttäjän paikka, huomioimisen. Tosiaikainen skeemamuunnos osoittautui tutkimuksen mukaan hyödylliseksi työvälineeksi dynaamiseen, verkkopalveluiden tasolla tapahtuvaan heterogeenisten aineistojen integrointiin. Erityisen hyödylliseksi osoittautui sovellettu deklaratiivinen tapa muunnoksen määrittelyssä, koska se mahdollistaa sekä kyselyjen että sisältöjen muuntamisen samaa perusmekanismia käyttäen.

Osana tutkimusta määriteltiin sovellusarkkitehtuuri, joka tukee edistyksellisten verkkopohjaisten paikkatietopalvelujen rakentamista. Tämä kuusitasoinen palveluarkkitehtuuri koostuu seuraavista käsitteellisistä palvelukerroksista, joista kullakin on hyvin määritelty osuutensa kokonaispalvelun muodostumisessa: aineistopalvelukerros, tietojen integrointikerros, prosessointikerros, portaalikerros, lisäarvon tuottajan palvelukerros, asiakassovelluskerros. Avoimia standardoituja palvelurajapintoja sovelletaan laajasti arkkitehtuurin eri osissa.

Väitöskirja perustuu suurelta osin EU-rahoitteiseen tutkimusprojektiin nimeltä GiMoDig. Tässä projektissa implementoitiin kaikki kolme esiteltyä tosiaikaisten sisältömuunnosten sovellusesimerkkiä. Kehitetty palveluprototyyppi mahdollistaa neljän valtion (Saksa, Suomi, Ruotsi, Tanska) kansallisten topografisten tietokantojen integroinnin kaksisuuntaisia kyselynaikaisia skeemamuunnoksia soveltaen. Muunnosten tuloksena syntyvää tietoaineistoa prosessoidaan edelleen tosiaikaisen yleistyksen keinoin, sovittaen aineisto käytössä olevaan tarkastelumittakaavaan ja käyttötarkoitukseen. Lopulta paikkatietoaineisto muunnetaan visuaaliseen muotoon ottaen huomioon käytetty sovellusalusta monikanavajulkaisemisen periaatteiden mukaisesti. Lopputuloksena käytettävissä on kansalliset rajat ylittävä karttapalvelu, joka integroi heterogeenisiä aineistoja tosiaikaisesti ja sovittaa visualisointituloksen käytetylle sovellusalustalle dynaamisesti.

Avainsanat: Internet GIS, muunnospalvelu, skeemamuunnos, palveluarkkitehtuuri

ACKNOWLEDGEMENTS

This study has been conducted at the Finnish Geodetic Institute (FGI) over the last ten years. The FGI provides excellent conditions for carrying out academic research. I would like to express special thanks to Professor Risto Kuittinen, Director General of the Institute, for actively supporting this great tradition.

I am grateful to Professor Kirsi Virrantaus (Department of Surveying, Helsinki University of Technology), the supervisor of my thesis, for her encouragement and advice during the last phases of the study. The reviewers, Professor Lars Bernard and Dr.-Ing. Andreas Donaubauer, are warmly acknowledged for their contributions to the thesis.

I am happy to have worked with several pleasant and supportive individuals at the Department of Geoinformatics and Cartography, FGI. First of all I would like to express my gratitude to Professor Tapani Sarjakoski, the head of our Department. My research career would never have reached this level, had it not been for Tapani's stubborn insistence and relentless encouragement. You are highly appreciated for convincing the most reluctant candidate in history to carry on!

I would like to thank Dr. Tiina Sarjakoski, who at the very beginning encouraged me to start working on the general subject field of the thesis, the Internet GIS. Over these ten years Tiina has always been supportive, ready to help in all the delicate nuances of academic writing, and full of fresh ideas for the next stage of the study. I owe you a big thank you, Tiina!

In the practical daily research work I have been very happy to work with Jaakko Kähkönen. I would like to thank you, Jaakko, especially for the relaxed atmosphere you have brought into those innumerable working hours we have spent together – discussing and developing, coding and configuring, struggling and succeeding. I could not have hoped for a better closest colleague!

Special thanks go to all the other people at the Department with whom I have had the privilege to work. I would like especially to name Annu-Maaria Nivala, Dr. Juha Oksanen, Mari Laakso, Ulla Pyysalo and Kirsti Filén. You have all contributed to the wonderful working environment, which has given me such pleasure to come to every morning!

This study is largely based on the EU-funded research project GiMoDig. In the context of this project I have had an opportunity to work with a number of pleasant colleagues from Finland and abroad. My special thanks go to Professor Bengt Rystedt, Dr. Lars Harrie, Professor Monica Sester, Dr.-Ing. Andreas Illert, Mark Hampe, Sabine Afflerbach, Reino Ruotsalainen, Flemming Nissen, Jørgen Münster-Swendsen and Andreas Hvas.

My most heartfelt thanks go to my family; to my beloved wife, Helena, and to our wonderful series of T's: Teemu, Tommi and Tuuli. You are the joy of my life and provide endless motivation for all kinds of struggles and undertakings in life – even this one!

Kirkkonummi, Nov 12th, 2007
Lassi Lehto

ORIGINAL PUBLICATIONS

This thesis is based on the following previously published articles:

Paper I Kähkönen, J., Lehto, L., Kilpeläinen, T. and T. Sarjakoski, 1999. Interactive visualisation of geographical objects on the Internet. *International Journal of Geographical Information Science, Special Issue*, **13**(4):429-438.

Paper II Lehto, L., 2003. A Standards-Based Architecture for Multi-purpose Publishing of Geodata on the Web. In: Peterson, M.P. (ed.), *Maps and the Internet*, Elsevier Science, pp. 221-230.

Paper III Lehto L. and T. Sarjakoski, 2005. XML in Service Architectures for Mobile Cartographic Applications. In: Meng, L., Zipf, A. and T. Reichenbacher (eds.), *Map-based Mobile Services – Theories, Methods and Implementations*, Springer Berlin Heidelberg New York, pp. 173-192.

Paper IV Lehto, L. and T. Sarjakoski, 2004. Schema Translations by XSLT for GML-Encoded Geospatial Data in Heterogeneous Web-Service Environment. Proceedings of the XXth ISPRS Congress, July 2004, Istanbul, Turkey, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, **XXXV**(B4/IV):177-182, also CD-ROM.

Paper V Lehto, L. and L. T. Sarjakoski, 2005. Real-time generalization of XML-encoded spatial data for the Web and mobile devices. *International Journal of Geographical Information Science, Special Issue*, **19**(8-9):957-973.

Paper VI Lehto, L., 2007. Schema Translations in a Web Service Based SDI. *Proceedings of the 10th AGILE Conference on Geographic Information Science*, 'The European Information Society: Leading the way with geo-information', May 8-11, 2007, Aalborg, Denmark, CD-ROM.

In Paper **I**, Lehto contributed to the design phase of the study, developed major parts of the prototype software and wrote most of the paper.

In Papers **III** and **IV**, the co-author acted as a supervisor.

In Paper **V**, the co-author contributed to the parts related to general aspects of real-time generalisation.

LIST OF ACRONYMS AND ABBREVIATIONS

AJAX	Asynchronous JavaScript and XML
AS	Application Schema
CORBA	Common Object Request Broker Architecture
DEM	Digital Elevation Model
DOM	Document Object Model
ESDI	European Spatial Data Infrastructure
ETRS	European Terrestrial Reference System
FE	Filter Encoding
GI	Geographic Information
GIS	Geographic Information System
GML	Geography Markup Language
HTML	Hypertext Markup Language
INSPIRE	Infrastructure for Spatial Information in Europe
ISO	International Organization for Standardization
KML	Keyhole Markup Language
MDA	Model Driven Architecture
MOF	Meta Object Facility
OGC	Open Geospatial Consortium
POI	Point of Interest
RIA	Rich Internet Applications
SDI	Spatial Data Infrastructure
SFS	Simple Features Specification
SLD	Styled Layer Descriptor
SOA	Service Oriented Architecture
STX	Streaming Transformations for XML
SVG	Scalable Vector Graphics
TC	Technical Committee
UI	User Interface
UML	Unified Modeling Language
URL	Uniform Resource Locator
VAS	Value-Adding Service
W3C	World Wide Web Consortium
WFS	Web Feature Service
WFS-X	Web Feature Service, Transforming
WMS	Web Map Service
XHTML	Extensible HyperText Markup Language
XMI	XML Metadata Interchange
XML	Extensible Markup Language
XSLT	Extensible Stylesheet Language Transformations

TERMINOLOGY

The following terms and their definitions are taken from the ISO TC211 Terminology Glossary (ISO TC211 2006).

Term	Definition	In this thesis
application	manipulation and processing of data in support of user requirements	
application schema	conceptual schema for data required by one or more applications	mostly refer to GML application schema
base standard	ISO geographic information standard or other information technology standard issued as a source from which a profile may be constructed	
client	software component that can invoke an operation from a server	
concept	unit of knowledge created by a unique combination of characteristics	
conformance	fulfilment of specified requirements	
coordinate	one of a sequence of n numbers designating the position of a point in n-dimensional space	
coordinate transformation	change of coordinates from one coordinate reference system to another coordinate reference system based on a different datum through a one-to-one relationship	used as a general term also covering *coordinate conversion*
coordinate conversion	change of coordinates, based on a one-to-one relationship, from one coordinate system to another based on the same datum	see previous
data	reinterpretable representation of information in a formalised manner suitable for communication, interpretation, or processing	
data interchange	delivery, receipt and interpretation of data	
data transfer	movement of data from one point to another over a medium	
data type	specification of a value domain with operations allowed on values in this domain	
dataset	identifiable collection of data	
definition	representation of a concept by a descriptive statement which serves to differentiate it from related concepts	
encoding	conversion of data into a series of codes	
encoding rule	identifiable collection of conversion rules that define the encoding for a particular data structure	

Term	Definition	In this thesis
encoding service	software component that has an encoding rule	
feature	abstraction of real world phenomena	
feature attribute	characteristic of a feature	corresponds to term *property*
geographic feature	representation of real world phenomenon associated with a location relative to the Earth	
geographic information	information concerning phenomena implicitly or explicitly associated with a location relative to the Earth	
geographic information service	service that transforms, manages, or presents geographic information to users	
geographic information system	information system dealing with information concerning phenomena associated with location relative to the Earth	
implementation	realization of a specification	
information	knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning	
inheritance <UML>	mechanism by which more specific elements incorporate structure and behaviour of more general elements related by behaviour	mostly applies to XML Schema inheritance
instance	object that realizes a class	
interface	named set of operations that characterize the behaviour of an entity	
interoperability	capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units	
layer	basic unit of geographic information that may be requested as a map from a server	also used to refer a level in n-tier computing architecture
object <UML>	entity with a well-defined boundary and identity that encapsulates state and behaviour	
operation	specification of a transformation or query that an object may be called to execute	
portrayal	presentation of information to humans	
portrayal service	generic interface used to portray features	corresponds to portal service
position	data type that describes a point or geometry potentially occupied by an object or person	mostly used in more generic sense

Term	Definition	In this thesis
profile	set of one or more base standards or subsets of base standards, and, where applicable, the identification of chosen clauses, classes, options and parameters of those base standards, that are necessary for accomplishing a particular function	
raster	usually rectangular pattern of parallel scanning lines forming or corresponding to the display on a cathode ray tube	mostly used in more generic sense
reference data	data accepted as representing the universe of discourse, to be used as reference for direct external quality evaluation methods	topographic base datasets, to which other thematic data can be referred
request	invocation of an operation by a client	
resource	asset or means that fulfils a requirement	a network-resident resource
response	result of an operation returned from a server to a client	
schema	formal description of a model	
server	a particular instance of a service	
service	distinct part of the functionality that is provided by an entity through interfaces	
service chain	sequence of services where, for each adjacent pair of services, occurrence of the first action is necessary for the occurrence of the second action	
service interface	shared boundary between an automated system or human being and another automated system or human being	mostly applies to computer-computer communication
service metadata	metadata describing the operations and geographic information available at a server	relates to the response message of the GetCapabilities -request
simple feature	feature restricted to 2D geometry with linear interpolation between vertices, having both spatial and non spatial attributes	
spatial attribute	feature attribute describing the spatial representation of the feature by coordinates, mathematical functions and/or boundary topology relationship	approximate synonyms are *geometric property, spatial property*
spatial object	object used for representing a spatial characteristic of a feature	not used, approximate synonym is *geometric object*
transfer protocol	common set of rules for defining interactions between distributed systems	
vector	quantity having direction as well as magnitude	
vector geometry	representation of geometry through the use of constructive geometric primitives	

CONTENTS

1 INTRODUCTION

1.1 Background

The Internet has become the most important means for disseminating geographic information (GI) (Peterson 1997, Peterson 1999,). The World Wide Web (henceforth: Web), especially, is widely recognized to be an indispensable delivery platform for geospatial data and applications (Kraak & Brown 2000, Plewe 2007). Extension of the Web computing framework to mobile devices is making digital maps part of the layman's everyday life, thus emphasizing the importance of network-based solutions for geospatial applications (Gartner 2001, Peng & Tsou 2003). In future, digital geographic information will be used to support a wide array of human decisions – ranging from strategic global issues, such as how to respond to the challenges of climate change, to the daily details of an individual citizen's life, such as where to find a vacant parking space. All this will be supported by geospatial services residing in computer networks.

Geographic Information Systems (GIS) have come a long way from the data centres of the 60s, travelling via desktop PC applications to Web services and via mobile solutions to ubiquitous computing platforms. In this process, the number of potential users of digital geodata has increased all the time and consequently the number of different application areas where geospatial information is being used has soared. This development trend is illustrated in Figure 1.

First-generation GI applications on the Web mostly deal with simple pre-styled map visualisations. As geospatial Web applications mature, however, more sophisticated approaches must be developed. These include facilities for the dynamic configuration of map services to support the desired usage- and user-specific visualisation modes and, increasingly, processes involving real data content in feature-based representations (Wei et al. 2001). Even the applications that predominantly aim at visual, map-based presentation will increasingly rely on live geospatial databases as a basis for their visualisation process.

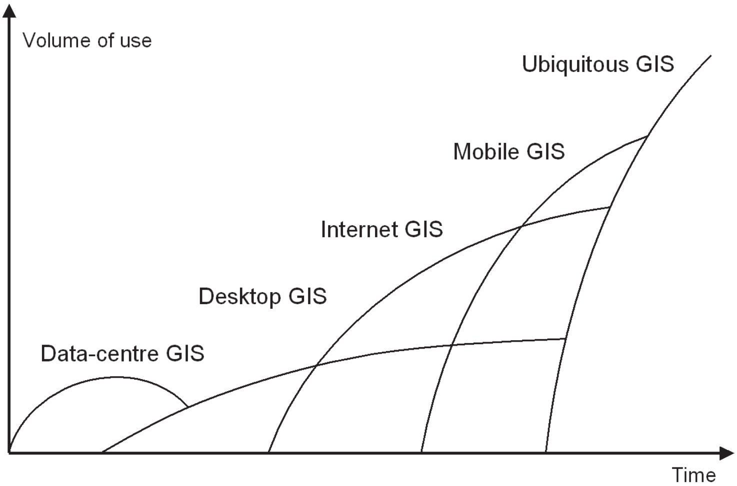

Figure 1. The five consecutive phases of GIS development

1.2 Motivation for the research

The growth of Web-based processing of GI has brought about some confusing diversity on at least two important fronts. As the number of geospatial datasets accessible through Web content services increases, the challenge of processing them together in a particular Web application is also growing (Díaz et al. 2008). Similarly, as the number of user communities interested in geospatial Web services increases, so does the challenge of supporting all the different client devices and platforms used to access the services (Lehto et al. 2002). These two aspects of diversity can be approached as a chance to reuse valuable resources, thereby maximising the investment made in their capture, and as an opportunity for integrating content, thus creating added value in the form of innovative new products and services. The two aspects of the diversity challenge are illustrated in Figure 2.

Adoption of strictly standardised Web methods and technologies in geospatial applications promotes harmonisation of the related computing practices. Consequently, introduction of feature-based GI into the Web domain has created an opportunity for standardisation of their modelling, encoding and processing methods, also motivating further research to promote full exploitation of the improved interoperability and data-sharing facilities (Peng 2005, Lehto 1998b).

As well as being motivated by the diversity challenges and opportunities provided by technology and standardisation, the current study is given further justification by important governmental initiatives aiming at the creation of national and European-wide Spatial Data Infrastructures (SDI). These processes aim to advance the exploitation of existing geospatial data sources by creating a Web-based service infrastructure in which these resources can be easily accessed and utilised in various application areas. One of the most important initiatives is the INSPIRE programme led by the European Commission, which aims to create a European infrastructure for spatial data related

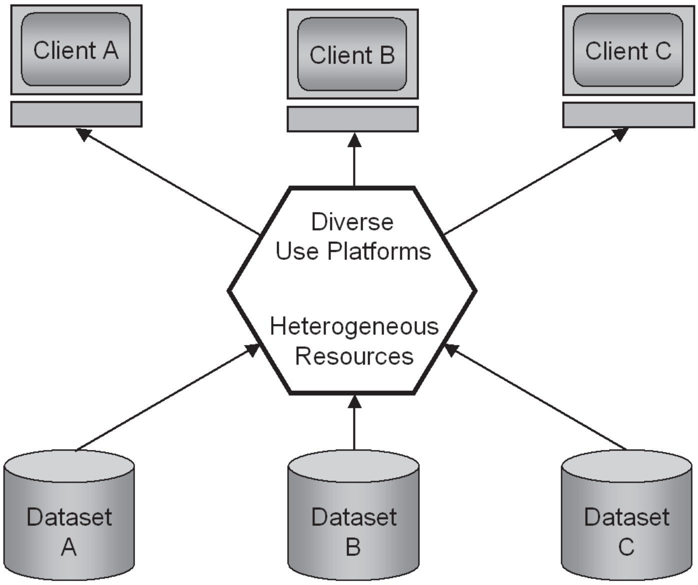

Figure 2. The two diversity challenges faced in the Internet GIS.

to protection of the environment (European Parliament and Council 2007).

Two major principles of the INSPIRE programme are particularly relevant in the context of the present study: 1. European-wide spatial data supply would be based on a network of seamless, standardised Web services, 2. These services are expected to rely on existing data resources, without needing new data capture or extensive database re-engineering. The above principles require content transformations to be applied during the data delivery process. These could take the form of an off-line batch process or an online, real-time procedure. This thesis focuses on real-time transformations carried out in a network service-based computing environment. The general framework in which the study was conducted is illustrated in Figure 3.

What follows briefly explains the main terms and concepts of the thesis. The terminology used in the related literature is not completely consistent. Section 1.3 outlines the terminological conventions used in the study. The relevant official ISO-standardised terms are listed in Section 'Terminology' on page vii-ix.

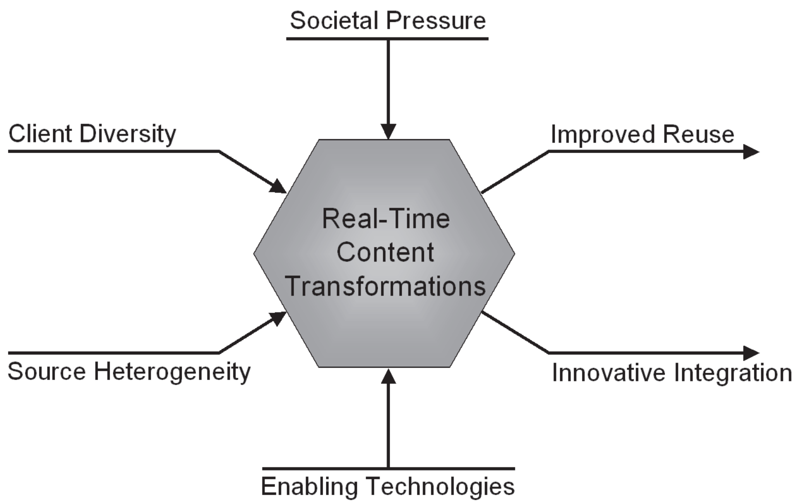

Figure 3. General framework of the study: transformation of the two challenges (client diversity and source heterogeneity) into two opportunities (improved reuse and innovative integration) by means of real-time content transformations, motivated by the related societal pressure and enabling Web technologies.

1.3 Real-time content transformations

The transformations discussed in this thesis are carried out in **real time**, i.e. as an integral part of a synchronous request-response dialogue in a network service-based computing environment. Typically this dialogue takes place between a calling client application and a responding server component in a traditional two-tier client-server computing architecture. In this thesis, the approach is extended to include a middle-layer component, responsible for the transformations required (see Figure 4). Later in the study this architecture is further expanded to include more computing layers, each with a distinct role in the dialogue.

Transformations as understood in this thesis process geospatial **content**. The term 'content' has been selected intentionally to cover both GI in its visualised form, i.e. maps, and the real geographic data, used as a source for the visualisation. The study mostly deals with maps as **vector-formatted** visual representations of geographic data. Geographic data is treated as a set of **features**, corresponding to **objects** or **phenomena** recognizable in the real world. The conceptual structure, in which the data components related to a given real world object are to be represented, is called the **model**, or **data model** of the object. An exact description of the model, expressed using some formal modelling language, is called the **schema** of the object.

The term **schema transformation** as used in the study actually refers to the process, in which geospatial data is transformed from one schema (source schema) into another (target schema). The process can be categorized as a

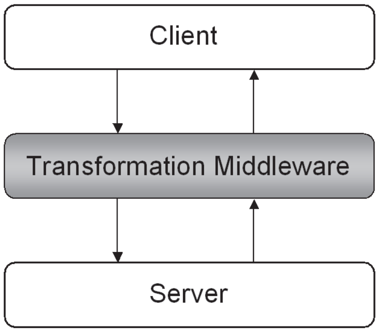

Figure 4. Three-tier service architecture involving a middleware layer responsible for real-time transformations..

schematic transformation. The other possible transformation categories include **syntactic transformation** (changing the syntax of the data representation) and **semantic transformation** (schema transformation in which no exact semantic correspondences between source and target schema can be found). When schemas are transformed from one modelling language to another, the term **model transformation** is also used. In this thesis, the focus is on schematic transformations.

A schema transformation can be divided into two main phases: defining the transformation (**configuration time**) and carrying out the transformation (**run time**). In the literature, the configuration phase of a schema transformation is also called **schema mapping.** The run time phase is usually called schema transformation or **schema translation**. In reference to a schema transformation carried out on a query sentence, the term **query rewriting** is also used.

The study also deals with other types of transformation apart from schema transformations. These include visualisation transformations, aimed at supporting varying presentations of the same source dataset on different visualisation platforms. **Multi-purpose publishing** of GI is the term used in this thesis to refer to these transformations. Another type of real-time transformation, closely related to multi-purpose publishing, is on-line, or **on-the-fly generalisation** of geographic data.

Content transformations are approached in this study in the context of two different tasks: as a mechanism enabling several schematically heterogeneous datasets to be integrated into a seamless whole (**data integration**), and as a tool to facilitate access to a single data source applying various presentation platforms and environments (**data reuse**). In the former case, the study applies an approach in which all the component schemas are mapped into a single common reference schema called a **global schema**. The process of creating this schema is referred to as **schema integration**.

The various processes discussed above are considered in the study as components in a Web service-based computing environment. The reference framework in which these components are organised and related to each other is called the **service architecture**. A commonly used approach to service architecture is to see it as a **stack of conceptual layers**, each with a distinct responsibility in the service provision.

1.4 Goal of the study

The main goal of the study described in this thesis is to develop Web standards-based methods for real-time content transformations, aiming to alleviate the two diversity challenges described above. The study deals specifically with the transformations related to **multi-source integration** and **multi-platform display of GI**. At the same time, it aims to develop a flexible **service architecture framework** that would support geospatial solutions relying on real-time transformations in a network service-based computing platform.

Based on the above introduction, the hypothesis for the present study can thus be formulated as follows:

Real-time transformations applied to the feature-based representation of geographic information in the Web environment using generic standardised Web methodologies will significantly advance the interoperability and reuse of geodata resources.

In what follows, the identified research questions are presented, the results yielded by the study discussed and the final conclusions drawn. The study deliberately excludes a few potentially interesting issues from the discussion. It does not deal with methodologies aimed at automating the schema-mapping phase of the transformation, the most important of these being the ontology-based mechanisms for semantic matching of source and target schemas. No issues related to dynamic real-time schema matching are discussed, nor are schema transformations between modelling languages. The study does not consider problems related to service chaining or dynamic service composition. In the discussion of the architecture, issues related to service metadata and service discovery are not covered.

The thesis is structured as follows. Section 2 presents a review of relevant previous research. Section 3 discusses the research questions the thesis aims to resolve. The research methods used are presented in Section 4. Section 5 introduces the six papers covered by the thesis. The results yielded by the study are discussed in Section 6. The study ends with a conclusion and a discussion of prospects for further work in Section 7.

2 RELATED RESEARCH

The areas of research and development relevant for the thesis include the work of standardisation bodies, especially the Open Geospatial Consortium (OGC) (OGC 2007) and the World Wide Web Consortium (W3C) (W3C 2007), studies dealing with various different transformation types and their applications in the GI domain – in particular the research done on transformations in the context of Web services – and work on service architecture. The following discussion also includes a section on relevant developments in the GIS industry.

2.1 Standards development

The standards developed by the W3C are particularly important in the area covered by this thesis. The most important standards include the Extensible Markup Language (XML) Specification (W3C 2004a), Extensible Stylesheet Language Transformations (XSLT) (W3C 1999b), and XML Schema (W3C 2001a;b). XML specification provides the foundation for encoding feature-based GI on the Web, XSLT is the main tool used for content transformations in the study, and XML Schema provides a mechanism for pre- and post-transformation validations. From the map visualisation point of view, the specification defining an XML-based encoding for vector-formatted images, Scalable Vector Graphics (SVG) (W3C 2003), is essential.

The OGC standards relevant for this study include the Simple Features Specification (OGC 2006, OGC 1999), which establishes a restricted data model for representing feature-based GI. The Geography Markup Language (GML) (OGC 2004a) specification defined by the OGC is used in the study as a base standard for encoding GI in XML. The OGC service interface specifications used in the study include the Web Map Service (WMS) (OGC 2004b) specification, the OpenLS Presentation Service (OGC 2005a) specification, and the Web Feature Service (WFS) (OGC 2005b) specification. Various studies have been published on issues related to the use of XML-based standards in geospatial applications (Dorninger 2003, Zhang & Li 2005).

The OGC has investigated the issue of schema translation in its interoperability programme, the Critical Infrastructure Protection Initiative (CIPI). The concept of a Translating Web Feature Service (WFS-X) was studied in the context of the Transportation Pilot of the US national Geospatial One Stop (GOS) initiative. The service developed in the project delivers transportation-related data from two heterogeneous sources, transformed into the common application schema in two real-time schema translation processes. The transformations are based on XQuery- and XSLT-technologies. Clients express their queries in the published common schema. So far, the project results are available only as an internal OGC draft (Bishr & Buehler 2003, Woodsford 2005).

The interoperability project ORCHESTRA funded by the European Commission has identified the 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 repository interface is used to manage (create, delete, get, set) the schema mappings in the service. The main mapping interface enables schema transformations to be carried out on the data instance level. Data sources can be either outside the schema mapping service (loosely coupled service) or can be provided by the service itself (tightly coupled service). The service interface also enables a concatenation operation for integrating several sources into a single process. The specification does not define a schema mapping language, but instead provides a mechanism by which the client application can indicate the applied language.

2.2 Content transformations

2.2.1 General

One of the first geodata-related studies discussing interoperability based on real-time transformations in a networked environment is described by Pascoe & Penny (1995). The paper considers schema translations on three different abstraction levels: the conceptual, implementation and physical. Interoperability between systems is assumed to be achievable via a sequence of schema translations, performed by software modules called communicating interfaces. Data transfer between the two participating interfaces is based on the use of a common representation of the data content. The development of a widely supported standard representation is mentioned as a future goal.

An early work by Abel et al. (1999) presents an architectural design for Web-based GI content and processing provision. This approach relies on distributed, heterogeneous services that are described in a common registry. Access to the services is based on a common query language and mediator components that translate between the query language and the native commands used in the services. The architecture consists of three layers: customer, translating middleware and services.

Various studies have investigated schema translation problems in the context of relational-to-XML or XML-to-relational content transformations. Curtis and Müller (2006) discuss these issues in the geospatial application domain, identifying various types of atomic transformation components that typically occur in the relational-to-XML translation. These include simple one-to-one translation, grouping of columns into a complex element, transforming table joins into child-parent relationships, fan-out (i.e. creating element types based on columns values), fan-in (i.e. translating multiple tables into a single element type), filling in constants, concatenating various cell values into single element content, and value translations, e.g. from one reference system to another.

2.2.2 Schema mapping

Braun (2004) discusses issues involved in matching two schemas with each other (schema mapping). The proposed approach is based on a formal language enabling correspondences between the schemas to be expressed. Another study concentrating on schema mapping is described by Friis-Christensen et al. (2005). The study identifies approaches for schema mapping, such as linguistic correspondences, similarity of attribute naming at schema level and attribute value at instance level, structural matches between schemas and constraint similarities. Manoa et al. (2004) propose a schema matching approach based on instance-level evaluation of the geometric feature properties. Devogele et al. (1998) explore schema mapping in the context of a database integration process.

Balley (2005) describes a study concentrating on an interactive user interface for defining a schema transformation. Schema transformation is approached as a pre-process in a data ordering transaction aimed at improving data usability. The transformation is first defined on a view of the original schema. Actual transformation instructions are produced separately for data and metadata content after the transformed schema has been validated. A description of how the transformation could be carried out as a Web service is given by Balley et al. (2006).

An approach for defining schema mapping on the conceptual level according to the principles of Model Driven Architecture (MDA) is presented by Donaubauer et al. (2006), who later discuss how to implement this functionality in the context of OGC-compliant Web services (Donaubauer et al. 2007). The approach relies on the use of XMI-encoded UML models for the source and target conceptual schema indication and on MOF Query/View/Transformation-based schema mapping instructions. The actual schema translation is implemented as an extension to the Web Feature Service (WFS) interface definition called model-driven WFS (mdWFS). The user can select among the source schemas available in an mdWFS, and 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 Gnägi et al. (2006).

2.2.3 Data integration

The early work related to data integration in the geospatial domain was carried out in a database environment. The computing systems in this field are commonly referred to as federated databases. An example of this approach, called Virtual GIS, is described by Abel et al. (1998). The system consists of a single virtual database and several autonomous component databases. The virtual part is responsible for maintenance of the federated database and its schema. In the described system the two-way schema translations are carried out by the local processing modules of the component databases.

Practical examples of rule-base schema transformation include IMKICH-project in Netherlands (de Vries & Tijssen 2004), in which the transformation is done in database environment by employing the database view technology.

Grønmo et al. (2000) describe one of the first studies to use XML as a tool for improving interoperability in the geospatial Web service environment. This research followed a model-driven approach, based on ISO TC211 principles. The XML-encoding of GI is based on the rules defined in ISO standard 19118. The three-tier interoperability architecture relies on a common data model and on two-way mappings between this model and the proprietary models used by the client and the server. The related model transformations are hard-coded and do not follow any standards. The research pre-dates most of the XML-related interoperability specifications of the OGC.

The early work done on Web technologies, especially the XML data encoding mechanism as a solution for improving interoperability in the geodata context, include MIX (Mediation of Information using XML), which introduces an approach based on wrapper/mediator middleware (Zaslavsky et al. 2000). In this 3-tier architecture, legacy databases are covered by wrappers that translate queries expressed in the system's common language into queries native to the database, then transform the results yielded by the database into the commonly agreed language. A mediator processes the user request, generating appropriate sub-queries that go to the relevant data sources (actually to their wrappers). Once the mediator has received all the results, it integrates the individual pieces together, finally returning the answer to the user application.

A comprehensive platform for semantic data integration called BUSTER is presented by Visser and Stuckenschmidt (2002). BUSTER is a wrapper/mediator-based interoperability middleware solution providing seamless access to semantically heterogeneous geospatial data resources. The middleware service is first configured in a semi-automatic process in which the transformation rules are set up. During the query phase these rules are applied by the mediator component of the system so as to achieve semantic integration of the source data sets. The semantic mapping is based on the concept of context, i.e. the transformation is made from source context to target context, applying predefined functions as re-classification schemes.

In another related study, a mediator-based data integration solution, VirGIS, was developed for a Web environment (Essid & Boucelma 2004). The mediator module in this example employs a special XQuery- and GML-based query language called GQuery. The mediator forwards the individual sub-queries to the data services involved as WFS queries. The schema transformation approach is based on the use of an integrated global schema. The particularities related to query processing in VirGIS are presented by Essid et al. (2004).

2.2.4 Data visualisation

The use of an XSLT-based process for the visualisation of geographic data has been described in several research papers. A project aiming at good cartographic results is detailed by Mathiak et al. (2004). The process is not designed for a service environment; rather, it was developed as an XML-based batch process for the production of high-quality maps.

A generic approach to XSLT-based bi-directional GML-SVG transformation has been presented by Merdes et al. (2005). The aim in this research was to develop a generic process for transforming geographic data, expressed in an arbitrary GML application schema, into SVG for visual editing and at the same time enabling the edits to be restored to the database via a reverse SVG-GML transformation. Misund and Vålerhaugen (2004) aim to achieve similar functionality by means of cascading schema analysis and loose data integration.

2.2.5 Real-time generalisation

The idea of dynamic on-the-fly generalisation has recently gained interest among the geospatial research community. The topic has been mostly discussed in the context of multi-resolution or multiple-representation database research and usually taken as an auxiliary tool helping to bridge the gaps between the pre-processed scale levels. The field of research is referred to by the researchers as the on-the-fly, real-time, online or on-demand generalisation.

The topic of real-time generalisation has been discussed in a few research papers. Van Oosterom (1995) presents an approach, which combines supportive, pre-computed data structures with on-the-fly map generalisation. A prototype system demonstrating satisfactory processing times for multi-scale visualisation is described by van Oosterom and Schenkelaars (1995). Another research based on the combined use of multi-scale data structures and online generalisation has been reported by Jones at al. (2000). Burghardt et al. (2005) discuss issues related to the implementation of generalisation functionality in the service environment.

2.2.6 Tools

Some software vendors have developed solutions for schematic content transformations. Snowflake Software has GML-based data delivery with the related schema translation as the main functionality of its GO product line (Curtis 2005). The research carried out by Snowflake Software explores ways of defining the schema transformation at the semantic level and then deriving the corresponding syntactic mapping rules automatically, thus improving reuse of the transformation in different technology environments.

The transformation engine available from Safe Software Inc. is able to read an arbitrary GML document and interpret its structure via type discovery based on an analysis of the corresponding XML Schema (Murray & Chow 2005). The schematic translation from the source data model to the engine's internal representation is configured using a declarative XML-based mapping language developed by Safe Software. Further sophisticated modifications, such as coordinate system transformations or language translations, can be defined using an interactive graphic user interface. Other examples of schema transformation tools available from commercial GIS vendors include the GeoMedia Schema Remodeling tool by Intergraph (Seeley 2003) and the ESRI Data Interoperability software (ESRI 2007).

3 RESEARCH QUESTIONS

The main research questions of this study are formulated around the main theme: real-time content transformations on the Web. Firstly, as an important precondition, the feature-based representation of GI on the Web must be considered. XML-based representations of feature-level geodata have been studied in various research projects (Zaslavsky et al. 2000; Grønmo et al. 2000). An extensive international effort has been invested in developing standardised ways to handle geodata in the framework of the emerged general-purpose Web technologies (see Section 2.1). However, a comprehensive approach for adapting these methodologies in a consistent way for geodata access, transformation, validation and visualisation is still largely missing.

The following discussion on research questions continues with the issues related to the main application areas of real-time content transformation – multi-purpose visualisation, on-the-fly generalisation and data integration. The existing works related to the application of the generic Web mechanisms for geodata visualisation do not consider them in the context of multi-purpose publishing and multi-client support – a task for which they could most appropriately be adapted (see Section 2.2.4). The early research considering on-the-fly generalisation pre-date the introduction of the Web as a major platform for geodata delivery (van Oosterom 1995). Even those research activities that take generalisation as a Web service, mostly regard it as an interactive expert service, not as an intermediate process in a data delivery chain. Data integration approaches that can be found in the reports of the previous research also mostly disregard facilities offered by the emerging generic Web mechanisms (see Section 2.2.3).

Most of the early work on schema transformations focuses on database environment, without paying much attention to issues related to service-oriented environments. Some research papers with a service-oriented approach present results that are based on closed, project-specific solutions without wider applicability in the generic Web computing framework (de Vries & Tijssen 2004; Visser and Stuckenschmidt 2002). The main topic of the research presented in this thesis, real-time content transformations in the Web environment, has not been comprehensively dealt with in any of the papers discussed earlier. Thus a research question is raised, whether schema transformation could be considered as an independent Web service.

The architectural issues did not receive much attention in the related research. The early works in this area rely on a rather simple three-layer model (Abel et al. 1999). This is the starting point also for the current thesis (see Figure 4). Obviously, a more detailed service architecture structure is needed, if all the types of content transformations considered above are to be included. The last research question focuses on the design of the service architecture in which real-time content transformations could be most favourably carried out.

Within this general framework, established on the basis of the existing relevant research, the six detailed research questions of this study can be formulated as follows:

1. What are the main benefits gained by bringing object-based geographic information into the Web application domain?

The first examples of geospatial applications on the Web were based on the simplest possible representation of GI – a static map in the form of a raster image. This approach has actually remained popular to the present day, even though object-oriented computing technologies have become mainstream both on the database level and in desktop applications. Advanced processes aimed at somehow modifying geospatial content, an essential function in context-sensitive services, heavily depend on the availability of detailed object-specific information. The first solutions for feature-based, vector-formatted geodata on the Web have started emerging very recently.

Raster image-based representation of geodata has several serious limitations. The visual properties of the image are difficult to change locally in the client application. Individual map themes cannot be dynamically selected for display. Interaction with the image is severely limited – for instance, the user cannot be provided with visual feedback of mouse-controlled actions on the image. Also, the local scaling of the image is possible only in a very restricted scale range.

A map can potentially be used as a graphical user interface to all kinds of spatially referenced information. However, a user interface represented as a single raster image gets extremely difficult to handle, as there are no distinguishable UI components that could be individually managed. In this sense, the use of geographic features as points of access to spatially related information in various forms could yield significant benefits, specifically in Web applications that often involve arbitrary, non-conventional data sources.

2. How should geospatial visualisation services be organised to support multiple different client applications and device platforms?

The constantly spreading use of public networks for disseminating GI poses several new challenges for data providers. One is the fact that the same content resources are being utilized in many different environments. In addition to the still active use of maps on paper and the already well-established Web use, content providers need to take into account the vast and constantly changing array of mobile devices. Specifically with regard to big national datasets, maintaining separate content repositories individually designed for specific user platforms becomes impossible.

Many of the existing Web map services rely on raster-based, pre-styled datasets. This approach is becoming increasingly problematic with the constant introduction of new geodata-related user applications. The need to integrate base reference datasets with auxiliary content provided by a third party requires flexible ways of adapting background maps to the particular overlay information. On the other hand, if the background map is delivered as a vector image, the need to support a given encoding language becomes a new challenge.

The traditional flat, two-tier server-client architecture has proved inadequate for services that require support for multiple different client environments. The tight integration of visualisation properties with information content that prevails in many two-tier solutions is a problem that effectively prohibits flexible adaptation to varying visualisation requirements. The proprietary, closed nature of existing service architectures also makes it difficult to adapt the service to user preferences or specific application needs. Particularly few research projects have approached the problem of multi-client support systematically from a service architecture point-of-view.

3. Can map generalisation be treated as a real-time content transformation?

One of the most significant research topics in the development of mobile Internet applications has been the personalisation of services to better support individual user preferences and usage requirements. Personalisation usually focuses on aspects like the language, colours and general layout used in the user interface of the application. The service content may also be dynamically adapted to the actual usage context, e.g. using an appropriate unit of measure conversion.

Map-related mobile applications involve several challenging research questions concerning service personalisation. How should map content be modified to provide the best possible support for the user with his unique usage situation and individual information requirements? What kind of new map personalisation facilities can be used when the most important context parameter for map use, the actual location of the user, becomes widely available? How can map personalisation be realized in a dynamic, service-oriented delivery architecture framework?

The research question in focus here is the applicability of real-time, on-the-fly generalisation to map personalisation in a mobile geospatial application domain. What benefits could potentially be achieved when the generalisation is carried out in real-time? How could real-time generalisation be organized in a network service-based computing environment?

4. What is the most feasible mechanism for integrating heterogeneous geospatial datasets in the Web service environment?

As geospatial Web services mature, the focus of development is shifting from pure map visualisation issues to more sophisticated tasks involving real source datasets. So far, the emphasis in service-building projects has been on establishing individual content services providing access to geospatial databases maintained by a single organisation. The next logical steps in the evolution of geospatial Web services include the introduction of analysis and processing-oriented services and meeting the growing demand for data integration.

Several significant governmental development initiatives have been launched to improve the use of GI in decision-making at different levels of the public administration. A prominent example is the INSPIRE process led by the EU Commission, which aims at more efficient use of spatial information in tasks that relate to protection of the environment. One of the fundamental goals of the INSPIRE process is to facilitate European-wide integration of spatial datasets. This objective should be achievable in a network service-based interoperability architecture which relies on existing data resources.

Both Web service evolution in general and parallel governmental data-sharing initiatives call for a solution for real-time integration of heterogeneous geospatial datasets in the Web service environment. A mechanism has to be found allowing seamless, homogeneous content services to be built on top of structurally and semantically varying source datasets. In many cases, permanent, database-level data harmonisation is not possible, effectively making real-time transformations the only solution for achieving interoperability.

5. How can schema transformation be organised in an extensive Web service-based SDI?

With the increasing number and diversity of applications making use of geospatial data in the Web environment, the need to support various different data representations based on a given source dataset has increased. One of these diversity aspects is the proliferation of different data modelling approaches. Even when essentially equivalent datasets are involved, the details of the way the data is structured and encoded may vary significantly.

The process called schema transformation may potentially provide a solution to this problem. Schema transformation is approached here as an operation transforming an input dataset from the schema it is expressed in (source schema) into another indicated schema (target schema). In the process, the schema transformation may modify the structure, vocabulary (e.g. language), unit system and applied data typing of the source dataset. Schema transformation may, for instance, become necessary in the context of a data integration process, or because the data models used in the source dataset are different from those in the user application.

The issue of schema transformation becomes most conspicuous in the context of an extensive Web service-based SDI. The roles of individual actors in such a large information system are becoming increasingly specialized. Data content managers specialize on keeping their datasets up-to-date. Application service providers devote themselves to developing the best possible support for the user base they serve. In such a setting, the role of data transformer will also become essential.

The research question facing the SDI developers is: how can the function of schema transformation be organized in a network service-based data infrastructure? Could schema transformation be understood as a service type of its own? How can this kind of service be parameterized? In what way could the transformation service be integrated with the data content services involved?

6. What kind of service architecture is required to support an advanced geospatial Web application?

Most existing geospatial Web services are based on the traditional two-tier client-server architecture. This architecture satisfies the criteria for the basic task of delivering a simple prestyled map from the server to the client for display. However, the processes involved in Web-based SDIs are becoming increasingly complex. Many services deal with geospatial datasets, in addition to visual map representations. User applications frequently require various spatial data sources to be integrated and processed together with auxiliary application-specific overlay information. The workflow might include processes in which the service content is somehow modified. Finally, the results may potentially be rendered on platforms with vastly different presentation facilities.

In connection with complex Web applications requiring several individual services to co-operate, the question arises of how these services should be organized into a working solution. In what kind of general framework can the role of each service be considered? How the services can communicate with each other? In what way could the generic processing facilities provided by the underlying computing platform be best utilized? What kind of service architecture would be needed for a Web application involving heterogeneous data sources and data processing requirements, varying third-party application information, and several different end-user visualisation platforms?

4 RESEARCH METHODS

The work carried out in the study can be predominantly categorized as solution-driven constructive research. The approach used is adaptation of generic IT methodologies to the problems encountered in the geospatial application domain. Particular attention is given to the potential of the techniques emerging as a result of the Web-related research and development as factors enabling improved interoperability in geodata context. Another key element in the study is the constant attempt to apply standardised solutions as far as possible. The following briefly describes the most relevant generic and geodata-specific Web standards and their role in the thesis. The used XML-based standards and their relation to the discussed functions of the geospatial data delivery process are illustrated in Figure 5.

Extensible Markup Language (XML)

XML is a generic syntax and functionality specification developed by W3C for encoding arbitrary data in textual form. XML has been specifically targeted for Web use. The XML specification defines the syntactic rules and basic structural components (element and attribute) for constructing XML-documents. However, the XML base standard does not specify any concrete vocabularies designed for a specific application area. Thus, XML can be described as a meta-language. Hundreds of different XML vocabularies have been developed over the last few years for a diverse set of applications. Geography Markup Language (GML) is the most widely recognized XML vocabulary designed for geospatial applications. The XML specification also defines the responsibilities of a functional software component called an XML parser, which can be used to read and validate XML documents. Being the base standard for data, image and message encoding, XML constitutes the general technology basis for the research in this thesis.

XML Schema

XML Schema is a W3C-developed modelling language for defining XML vocabularies. XML Schema constructs can be used to define new

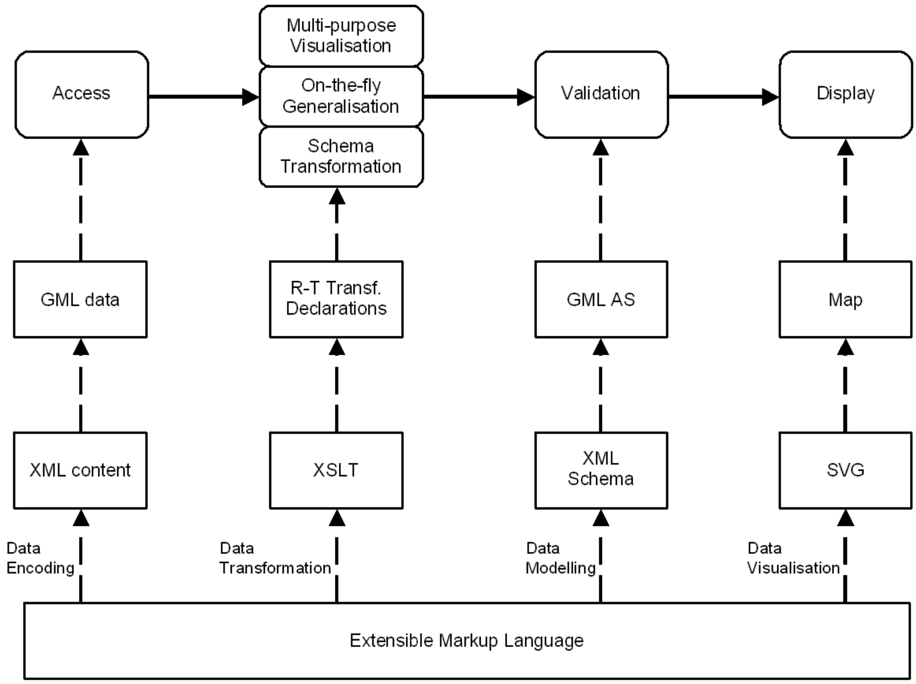

Figure 5. The used XML-based standards, their geodata-specific instances and the functions they support in the geospatial data delivery process.

XML-based languages. XML Schema contains a rich set of modelling elements for constraining both the structure and the data types available for use in a valid document. XML Schema documents are encoded in XML syntax. The corresponding XML Schema document can be referenced from within an XML instance document and subsequently read by a validating XML parser for the document's schema validity check. In this thesis, XML Schema has been used as a tool for defining the GML application schemas needed and for validating the related data streams.

Extensible Stylesheet Language, Transformations (XSLT)

XSLT specification is part of the W3C style definition mechanism, Extensible Stylesheet Language (XSL). XSLT specifies a language for defining transformations from one XML document into another XML document. The most typical example of XSLT usage is the transformation of an XML document from an application-specific data structure into an Extensible HyperText Markup Language (XHTML) conformant document for visualisation in Web browsers. XSLT is a declarative language providing a wide selection of functionalities for transforming the source document's structure and vocabulary. More complicated computations can be introduced into the language through extension functions. XSLT declarations are also expressed in XML syntax. In the research described in this thesis, XSLT plays a central role as a base methodology for declaring transformations from one representation form to another.

Geography Markup Language (GML)

GML is an OGC-specified vocabulary for encoding geodata in XML. The GML base standard introduces a wide set of generic XML elements designed to be used in geospatial applications. These include, for instance, most of the geometry types encountered in geographic datasets. However, GML does not define element types for concrete feature classes. These always have to be declared in local application schemas, following well-specified rules defined in the GML base standard. New element types introduced in local GML application schemas must be derived by type inheritance from the abstract base types defined in GML. Data source-specific GML application schemas can introduce both simple and complex properties for the declared feature types. A given feature type can have several geometric properties for storing the related geometry information. The specification also includes the concept of feature collection for packaging several features into a single set. The GML specification is approaching the ultimate status of an ISO standard under the number 19136. In the research described in this thesis, GML is widely applied as the encoding standard for geospatial data content.

Scalable Vector Graphics (SVG)

SVG is an XML-based vector image format developed by the W3C. SVG images can contain basic 2D geometric primitives, such as points, polylines and polygons, raster images and text labels. Symbol geometries can be defined once and used in multiple places in the image. SVG images can contain animated elements and application-defined scripts. Separate SVG profiles have been defined for two different categories of mobile devices (SVG Mobile, SVG Tiny). Web browsers currently provide varying support for SVG visualisation. In the current study, SVG is applied as a tool for delivering vector-formatted maps to various different client platforms. The interactive scripting features of the SVG standard are also used to attach various user controls to the map visualisations.

Web Map Service (WMS)

WMS is an OGC-developed service interface specification which is currently also available as an official ISO standard (ISO 19128). WMS defines a Web service that provides access to geographic information in visualised map form. A WMS service is expected to offer basic service metadata (GetCapabilities query), deliver maps (GetMap query) and optionally provide further information about a feature clicked on the map (GetFeatureInfo query). WMS specification declares a set of parameters for each query and defines their semantics to standardise the communication between the service and the requesting client application. WMS services typically deliver maps in the form of raster images. In the prototype service platform developed in this study, WMS was used as an access interface to the map services.

Web Feature Service (WFS)

WFS has been developed by the OGC as a service interface specification for the delivery of geospatial data. A WFS service provides access to service metadata (GetCapabilties query), to schema descriptions of individual feature types (DescribeFeatureType query) and to the feature-formatted representation of the data itself (GetFeature query). In addition, WFS interface specification also includes functionalities for updating the data store behind the interface (Transaction/Insert, Delete, Update queries). The interface contains facilities for feature locking to keep the updates consistent. By default, the response to a DescribeFeatureType query is expressed as a piece of XML Schema, and the response to a GetFeature query as a stream of Geography Markup Language –encoded data. WFS has been introduced into the official ISO process under the number 19142. In the service implementations developed as a part of this study, WFS plays an important role as an access interface standard for geospatial data services.

5 PAPER INTRODUCTIONS

Paper I: Interactive visualisation of geographical objects on the Internet

Paper I discusses the emergence of the Web environment as a platform for interactive applications, in contrast to the original approach in which the Web is seen as a distributed repository of static documents. The paper deals specifically with the object-based representation of geographic information in the Web processing environment. The importance of standards is also emphasised.

Paper I stresses the importance of interoperability in the era of network-based software solutions. The three main interoperability components covered in the paper are: Java as a tool for developing sophisticated platform-independent client applications, Common Object Request Broker Architecture (CORBA) as a solution for distributed computing in a heterogeneous environment, and the Open GIS Consortium's Simple Features Specification (SFS) as a convention for representing domain-specific information in object form. The SFS for a CORBA environment represents the early work of the OGC in the area of geospatial services. Interoperability is achieved by transferring real object instances with a well-defined behavioural interface to the client environment. A standardized mechanism for object serialization is the focal point in this approach. In the research described, this is provided by CORBA-based object distribution, conforming to the SFS specification. Wang (2000) presents another early example of a CORBA-based distributed solution for GI interoperability.

A Java applet-based object browser, NetGIS, was developed in the project. Communication with the server is based on a dialogue in which a list of available feature classes is first requested from the service and the user interface of the client application updated accordingly. This approach reflects the dynamism that has recently become popular in Web applications with the introduction of standardised OGC content service interfaces. The data model of a selected feature type can also be dynamically requested and the user interface for displaying feature properties configured accordingly. Content transformations are discussed in the context of Paper I as the ability of the client application to carry out local coordinate system transformations.

In the context of the overall study, Paper I represents early work that pre-dates most significant developments in the area of the geospatial Web. However, it reflects many of the major trends that have been spreading over the last few years. These include the idea of a generic browsing client able to connect dynamically to several previously unknown geospatial services by using standardised access interfaces, browsing of geographic information in object form with a Web browser-based client application, and the inclusion of multimedia content types as properties of geospatial objects.

In the architecture discussed, standardized object behaviour is transferred to the client platform in the form of a serialized Java object. This approach has not been given too much attention in more recent research and has been largely replaced by the idea of standardised object content representation, transferred as a stream of XML-encoded data. It remains to be seen, whether standardized object behaviour could in future alleviate problems related to the current need to translate data content between incompatible structure models. Recent developments in the area of Rich Internet Applications (RIA) suggest that this might be the case.

Paper II: A standards-based architecture for multi-purpose publishing of geodata on the Web

Paper II discusses the concept of multi-purpose publishing and the required separation of the data content from the presentation characteristics. The paper also discusses stylesheets as a mechanism for defining a transformation from data content into a given presentation form. The notion of stylesheets is here applied to the context of geospatial data and its various representations. A service architecture framework consisting of a set of conceptual layers is introduced and the relevant service interface standards for each layer described. Extensible Stylesheet Language Transformations (XSLT) is proposed as a generic information technology mechanism for geospatial data visualisation, specifically in the context of multi-purpose publishing. The need for multi-purpose publishing capabilities is emphasised specifically by the expansion of mobile Internet technologies. Introduction of the Extensible Markup Language (XML) as a major technology platform in the Web services environment is seen as an opportunity for greater integration of geospatial applications into general IT processes (Lehto 2000).

The four-tier service architecture presented in Paper II consists of the data service layer, information service layer, portal service layer and client layer. The main task of the data service layer is to provide geospatial content as a source data for various services on the information service layer. The information service layer might integrate several individual data services and add further service-specific information to offer as a result a complete service content, adapted to certain specific end-user needs.

The responsibilities of a portal service include support for service discovery from the client's viewpoint, and dynamic, real-time adaptation of the service content to the capabilities of the client device from the information service provider's viewpoint. The role of the portal service layer is seen as crucial in multichannel publishing of geospatial content. In this context, interesting opportunities are opened up for a user-centric personalisation of the service response in which even the user's current location could be taken into account. The service stack presented in Paper II represents the first version of the layered architecture approach discussed in this research. Later papers elaborate this design further and some new conceptual layers are introduced.

The importance of standardized access interfaces and content encodings is emphasised in the subsequent discussion. On the data service layer, the importance of the emerging Web Feature Service (WFS) access interface and the recently introduced Geography Markup Language (GML) is acknowledged. Other significant standardization initiatives mentioned include the Web Map Service (WMS) interface, Open Location Services (OpenLS) service interface specifications, the Scalable Vector Graphics (SVG) image-encoding standard, and the X3D language for representing 3D models in XML.

The last section of Paper II presents XSLT technology as a tool enabling services to support the principles of multi-purpose publishing in the context of geospatial data. The declarative nature of the XSLT process is emphasised as an important property, enabling separation of the transformation definition from the software implementation that carries it out. In addition to the basic functionality of changing the encoding language of the processed content from geospatial data representation (GML) into visual representation (SVG), the paper also discusses content transformation as a means of carrying out cartographic generalisation. This idea stretches the concept of multi-purpose publishing of geographic information well beyond a mere change of the encoding language, into the realm of dynamic, context-driven content personalisation (Gartner 2007).

Paper II discusses some new prospects for geospatial service provision in the Web environment. These include the potential shift from the largely provider-defined representation of geographic information towards a user-centred approach in which the end-user is controlling the way the information is presented. The evolution will be gradual and slow, largely because of the traditionally strong organizational linkage between geospatial data provision and control of its presentation details. However, increasing demands for multipurpose publishing capability will accelerate this paradigm shift.

The discussion in Paper II emphasizes one of the main principles of the research carried out in this study: adaptation of general-purpose IT solutions to the problems encountered in geospatial data processing. The specific generic technology introduced in Paper II is the stylesheet-based visualisation of domain-specific XML-encoded content in the Web environment. When the process of visualising geospatial data is seen as a transformation from one representation form to another, it becomes natural to look into generic transformation techniques for a solution to this need. XSLT emerged in this context as an appropriate tool for defining content translations that can be configured dynamically and carried out by existing, general-purpose software components. In the later papers this approach is developed further and applied in various ways, specifically in the context of geospatial Web services.

Paper III: XML in service architectures for mobile cartographic applications

Paper III concentrates on architectural issues, specially taking into consideration the new challenges posed by mobile applications. After a general introduction to XML technology, the paper presents an XML-based solution for each of the following important tasks: modelling and validation (XML Schema), encoding and linking (GML, XLink), transformations (XSLT) and visualisation (SVG). These solutions are then discussed in the context of geospatial data processing in the Web environment.

The layered service architecture is further elaborated in the latter part of the paper. A significant principle emphasized in the discussion is the adoption of open, standardized access interfaces on each of the conceptual service layers. This approach makes it possible to bypass unneeded layers and access the service level providing the best support for the calling application's needs. The standards-based solution requires XML encodings to be applied consistently in request and response messages throughout the service architecture. The above-mentioned base technologies can thus be employed in all the processing steps – an approach yielding significant savings in service implementation.

A detailed discussion of the service architecture developed in the EU-funded research and development project GiMoDig (Sarjakoski & Sarjakoski 2005) forms the core of Paper III. The role of the Value-Adding Service (VAS) layer is specially emphasized. It is the responsibility of this service layer to translate the generic context parameters coming from the client layer into an appropriate map request for portal layer and into the required additional content queries for Point of Interest (POI) data services. In the proposed service architecture, it is thus the responsibility of the VAS layer to understand the needs of the user – and the context parameters used to convey them. It must also be able to translate these into correct map and POI queries.

The role of the portal layer in the GiMoDig service architecture is to act as a mapping specialist. It is the responsibility of the portal service to translate the incoming map query into a data content query that satisfies the needs of the map visualisation process. This also includes setting the generalisation parameters appropriately for the generalisation process run by the processing service. The portal must thus understand the properties of the geospatial data content and the processes required to transform these into an appropriate visual representation.

Lots of content transformations are involved in the computing tasks the services perform during the request and response processing. In the GiMoDig prototype service, these transformations are configured consistently by a set of XSLT processing declarations. The approach is found to be especially suitable for a service provider-controlled processing environment. It is easy to test and fine-tune a transformation by editing dynamically interpreted character-encoded declarations. The standard XSLT functionalities provide support for most of the frequently needed transformation types. One drawback is the fact that developing a meaningful set of XSLT templates comprises a substantial programming task. However, in many cases the complexity and unpredictability of the process do not allow for a simpler approach.

Paper IV: Schema translations by XSLT for GML-encoded geospatial data in heterogeneous Web-service environment

In Paper IV, the focus is shifted from visualisation aspects to the topic of schema transformations, especially in the context of geodata integration. The paper presents the main findings in this research related to integration of heterogeneous data sources in a Web service environment. The discussion is largely based on experiences gained in the GiMoDig.project

The paper presents one of the main practical motivations for the research in Web service-enabled dynamic data transformations: the accelerating European integration process and the subsequent need for seamless information services with complete pan-European coverage. Many of the related datasets are too large for a centralized solution to be practical. The local data maintenance processes are often so tightly connected to the databases that full European-wide harmonization becomes impossible. The only solution to the problem seems to be data integration via dynamic Web services relying on real-time data transformations. The INSPIRE process is pointed to as a prominent European example of processes that are driving development towards Web service-enabled interoperability in spatial domain.

Based on the experiences gained in the GiMoDig project, the paper elaborates the service architecture introduced in Paper II. The main findings include the recognized need for a dedicated data integration layer with the main task of carrying out coordinate system and schema transformations from individual nation-specific forms to agreed common systems. A separate data processing layer for other kinds of data manipulations, such as generalisation or dynamic labelling, is also introduced. The portal layer is likewise presented, with the same base function of multi-purpose publishing as is discussed in Paper II.

The generic XML transformation technology XSLT is here adapted as a tool for the schema transformations needed in dynamic geospatial data integration processes. The technology has proved to be easily applicable to the tasks encountered in schema transformation, as operations such as renaming, reclassification, reordering and value conversions are readily available. XSLT extension functions are found to be a valuable tool for introducing geodata-specific operations, such as coordinate system transformation, into the process.

On the Data Integration service layer of the GiMoDig prototype, schema transformations are carried out to integrate the topographic map datasets of four countries (Germany, Denmark, Sweden and Finland) into a seamless service providing datasets in a jointly agreed common schema (GiMoDig Global Schema) (Afflerbach et al. 2004). The integration service carries out two-way transformations, handling both query and data transformations using the same base technology, XSLT.

The main conclusions yielded by the research discussed in Paper IV include the fact that dynamic data integration based on real-time schema transformations and carried out on both data queries and data content is a promising approach for seamless pan-European geodata provision. Adherence to commonly agreed international standards has proven to be a necessary precondition for a working data integration solution. For instance, in the GiMoDig integration service both input and output schemas are required to be GML Application Schemas – a restriction greatly reducing the number of schema integration operations required. The fact that in a WFS-compliant data service both the data query and the response dataset are encoded in XML resulted in another benefit of the standards-based approach: both messages can be transformed using the same base methodology.

Paper V: Real-time generalization of XML-encoded spatial data for the Web and mobile devices

Paper V discusses the concept of real-time, network-enabled content transformations in the context of map generalisation (Lehto & Kilpeläinen 2000). First, the most characteristic principles of real-time, or on-the-fly, generalisation are discussed. These include support for arbitrary result scales, active personalisation and linear process flow (no checking or iteration). The main strength, and at the same time the most severe limitation, of the approach is the real-time nature of the computation. Compared with traditional solutions for generalisation, on-the-fly generalisation enables a completely new level of adaptation to the actual needs and context of the user. Parameters such as the type of mobile device used and the position of the user can potentially be considered as input to the generalisation process. Together with the pre-configured individual user preferences regarding map content and style, these parameters enable ad hoc, user- and application-specific generalisation processes in the network service-based computing environments. An example of this kind of process is what is called vario-scale representation (Harrie et al. 2002a;b). In this approach, the map is centred on the user's location and the level of generalisation increases with the distance from the user.

As on-the-fly generalisation has to be carried out in real-time during the request-response dialogue, the generalisation operators involved have to be rather simple and predictable. However, the study reported in Paper V indicated that, even with these limitations, the real-time generalisation process could facilitate the transmission of map visualisations to mobile devices and improve the map interpretation. The generalisation operators used in the study include selection, simplification and aggregation.

Background technologies used include GML, SVG and XSLT. A Java-based extension mechanism is employed to introduce more complicated generalisation computations into the XSLT process. The approach was tested in a prototype platform based essentially on freely available open-source software components.

The main conclusion derived from the study described in Paper V is that the real-time processing of geospatial content in a network service-based delivery architecture opens up unprecedented opportunities for providing the user with personalised, highly adapted ad hoc map visualisations. Compared with traditional cartographic generalisation, the real-time approach has serious limitations in the sense that it does not allow for post-process checking and iterative improvement of the results. At the same time, it offers new potential for improving the fitness-for-use of the resulting generalised map visualisation, taking into account the user's actual usage situation and location, the device on which the map is to be rendered, and other relevant context parameters.

Paper VI: Schema translations in a Web service-based SDI

Paper VI continues the discussion of schema transformations introduced in Paper IV. New incentives for further study are provided by new technological developments such as the new category of powerful Web clients using the AJAX computing model, and by major organizational programmes such as the INSPIRE directive. The increasing number of different data models used in various client environments and ongoing harmonization activities make it difficult for a data provider to keep up with the demand. Real-time transformations are seen as a promising solution to the problem.

Paper VI approaches schema transformation as a specific type of Web service. The discussion largely focuses on issues related to the service interface and parameterization. The details concerning the service interface and its capabilities are interesting, as it is assumed that the schema transformation service is connected with standardized geospatial data content services according to generic Web service composition principles.

The study discusses schema transformation in the context of OGC-compliant Web services. As such, the input and output schema of the transformation are both treated as GML Application Schemas. This fact greatly reduces the complexity of the schema transformation process. As both schemas are based on the same principles, it becomes possible to analyse the transformation in a finite concept space. The process of schema transformation is thus categorized as a set of atomic transformation components. These are related to the conceptual layers of the GML feature model for further analysis. As a result, a table is constructed which indicates the applicability of each of the transformation components on each of the conceptual layers, together with a recommended execution sequence for the components.

The schema transformation can be seen as a data provider-controlled, product-oriented process, or as a client-controlled, demand-driven task. When taken as a provider-controlled process, the schema transformation service is actually exposed as a type of content service. In the case of the client-driven approach, the transformation service clearly belongs to the category of processing services.

An important consideration concerning schema transformation in a Web service environment is whether the service supports two-way transformations, i.e. whether it is capable of carrying out query transformations in addition to data content transformations. The prototype service developed in the study is an implementation of a two-way transformation engine. Considerable benefits were achieved by adhering to generic technologies and standardized geospatial Web solutions, as both the query and the data content transformations could be implemented using the same base methodology.

There is no standardized service interface specification for schema transformation. In the case study, the schema transformation was run as a sub-process on the conceptual service layer responsible for data integration. As such, the service layer implemented a standardized data content access interface, completely hiding the source data services from the calling application. The development of a service interface for schema transformation supporting the client-controlled processing mode is taken as a topic for further research.

The case implementations of the study include the service prototype of the GiMoDig project, involving two-way schema transformations between the national data models of the four participating national mapping agencies and the jointly agreed common schema, and vice versa. The dataset involved include the four national topographic databases in the scale range 1:10 000 – 1:20 000. The results confirmed the pre-assumption that cross-border, data-level integration of geographic information is feasible in a real-time service infrastructure. The study can be viewed as a successful early test, carried out according to the basic guiding principles of the INSPIRE process.

6 RESULTS

The main findings of the study are based on a prototypical approach. Several working software implementations have been developed during the study, the most important being the service prototype of the GiMoDig project (Lehto et al. 2004c). This prototype strictly follows the 6-tier service architecture designed in the study. Real-time transformations are applied in the prototype for various tasks, such as schema transformations (query and data), coordinate system transformations (query and data), online generalisation and visualisation. After the real-time data-level integration, on-the-fly generalisation and visualisation, the final result can be presented in a tangible form as a seamless cross-border map service.

The following presents the main results of the study. The discussion starts with a presentation of the three main applications of real-time content transformation focused on in this thesis: multi-purpose visualisation, real-time generalisation and data integration. The next main section deals with the results related to service architecture development. The third part is devoted to a discussion of schema transformation considered as a Web service.

6.1 Real-time content transformation applications

6.1.1 Multi-purpose visualisation

Related research question (2): How geospatial visualisation services should be organised to support multiple different client applications and device platforms?

An important conclusion drawn from the study on multi-purpose publishing methodologies applied in a geospatial context was that today's tightly integrated, closed map services need to make way for a more dynamic and open service architecture. The fundamental principles of this new architecture include separation of content from presentation characteristics, a shift from supplier-controlled map visualisation to predominantly user-driven mechanisms for defining presentation details of geodata and, consequently, the introduction of ad hoc, personalised, task-specific map visualisations instead of the traditional general-purpose, compromised fit-to-all maps (Sarjakoski & Nivala 2005). The user in this context does not necessarily mean the human end-user, but rather a service provider that provides value-adding information services on top of the reference map background.

According to the study, generic Web technology developments seem to provide a good platform on which geodata-related services could also be built. The results achieved in the research indicate that a generic XML-based Web processing architecture can be successfully adapted to the geospatial application domain. The most important principles include provision of the service content in an XML-based encoding and transformation of this content into the required presentation language using stylesheet-defined content transformations in real time.

An open service architecture with well-specified conceptual processing layers is a prerequisite for flexible on-demand map service provision. The results of the study indicate that this kind of service architecture consists of four distinct layers: a data service layer, an information service layer, a portal layer and a client layer. The responsibilities of these layers are, respectively: to provide access to XML-encoded data content; to integrate into this content all the necessary third-party service-specific auxiliary information and any other geospatial content; to carry out the multi-purpose transformation from content to presentation according to the requirements set by the user application and environment; and to handle the map display and user interaction.

The prototype system developed as part of the study confirms that the general-purpose, declarative XML transformation language XSLT can be successfully used as a tool for the multi-purpose publishing of geodata. In this test the generic stylesheet-based styling mechanism of the Web content was applied to the task of geodata visualisation. The results of the study suggest a future in which Web-based map visualisations are predominantly produced as client-defined, stylesheet-controlled transformations from XML-encoded geospatial data sources. In a pure XML-processing environment these visualisations are also encoded in some of the various existing and upcoming XML-based visualisation languages, such as SVG or X3D. Standardised software components will facilitate the deployment of these technologies, as suggested by Chang and Park (2006).

6.1.2 Real-time generalisation as a Web transformation

Related research question (3): Can map generalisation be treated as a real-time content transformation?

The real-time approach to generalisation yields several potential benefits compared with the traditional off-line, interactive generalisation process. When carried out in a real-time environment, the generalisation can be flexibly adapted to the actual usage situation. The most important context parameters affecting the generalisation process relate to the user's personal characteristics and preferences, to the mobile device used, and to the requirements of the specific application the map is used for.

An important user parameter significantly affecting the generalisation process is the current location of the user. Increasingly becoming available with the introduction of satellite positioning facilities on mobile devices, this parameter can fundamentally change the way the generalisation is approached in the future. An example of the location-dependent generalisation tested in the study is what is called the vario-scale map (Harrie et al. 2002a). In the real-time environment, the focal point of the vario-scale map can be centred on the current location of the user and the generalisation level of the map increased with the distance from this point. Demonstrations of a vario-scale map produced by the real-time generalisation process of the GiMoDig service prototype are shown in Figure 6.

The actual display scale and other unique characteristics of the display device, such as the screen resolution and colour-depth, can be taken into account. The end-user application can also be used as an input parameter of the generalisation process and can potentially affect the way the generalisation is carried out.

Real-time generalisation is approached in this study as a process in a standards-based Web service environment. Generalisation is seen as a content transformation carried out by an individual middleware service node in a multi-tier service platform, as also proposed by Burghardt et al. (2005). Although performance is always an issue in a real-time computing environment, the results from the prototype indicate that on-the-fly generalisation can be carried out without significant degradation of the overall service responsiveness. The generalisation operators tested in the prototype include selection, simplification and aggregation.

One significant result of the study is the successful application of general-purpose Web techniques in the prototype implementation. These include XML-based encoding of geospatial content (GML) and map visualisation (SVG). Standardised, general-purpose Web transformation methodologies were applied in the generalisation process. By applying declarative XSLT instructions, the generalisation operators can be formulated and conveyed to the service in a standardised way.

The empirical tests run on the prototype system revealed an interesting result: the transformation from the GML-encoded dataset into the SVG-encoded map image can be carried out even faster when generalisation is included in the process. This behaviour can be explained by the fact that, when it works with an already simplified dataset, the encoding transformation has to carry out far fewer coordinate manipulations than when it processes the original data.

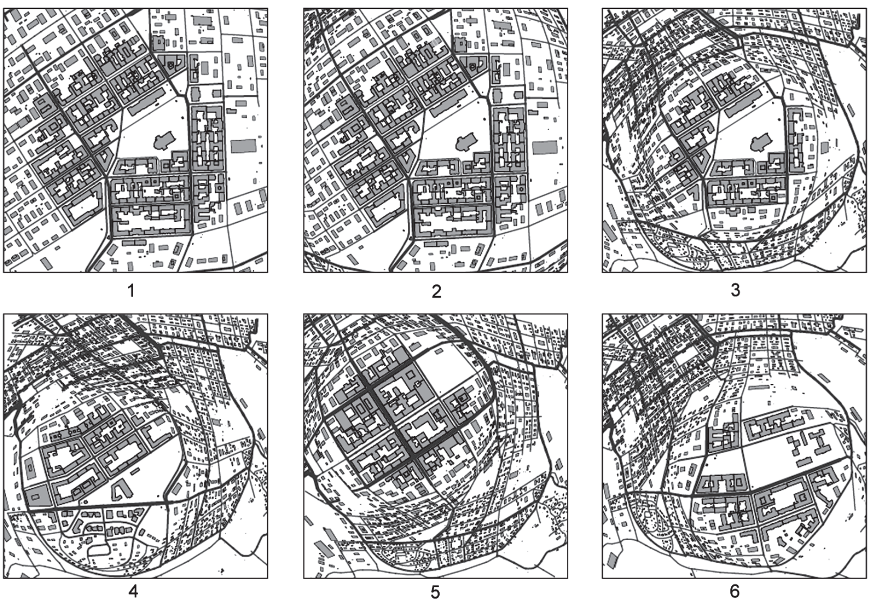

Figure 6. A series of vario-scale maps produced by the real-time generalisation process of the GiMoDig service prototype (Harrie et al. 2002a).

6.1.3 Data integration based on a real-time two-way schema transformation

Related research question (4): What is the most feasible mechanism for integrating heterogeneous geospatial datasets in the Web service environment?

In this study, the task of data integration was approached in a dynamic new setting compared with traditional research carried out in the areas of database integration, federated databases and data warehousing. The integration process was taken as a real-time process, carried out online, during the request-response dialogue in the service-oriented architecture (SOA) platform.

The work on data integration in the study focused on real-time schema transformations. Heterogeneous data sources were transformed into a common application schema and coordinate reference system by carrying out a schema transformation on the resulting dataset during the data query processing workflow in a distributed, network service-based computing architecture. A methodology was developed for defining the transformation in a declarative manner, relying on general-purpose Web techniques.

The architectural approach taken in the study is based on the idea of conceptual service layers stacked on top of each other. Data integration is seen as one of the service layers in this architecture. A common schema was developed as a reference model in which all the processing above the data integration layer is performed. Because the data integration layer acts as a shield on top of the heterogeneous data sources, completely hiding the underlying heterogeneity from the calling service layers, data integration actually becomes a two-way transformation process. Both the incoming query and the outgoing result dataset have to be transformed.

XML-based solutions have become essential in processes involving geospatial data. Both the service requests and the content encodings follow the same base syntax. XML Schema-based data models are used to constrain the valid structure of the datasets. In the research described the GML is used as the base modelling language. Consequently, the schema transformation mechanism developed in the study uses a general-purpose XML transformation language, XSLT. Significant synergies can be achieved in the transformation definition phase, as both the queries and the data content can be dealt with using the same transformation language.

A schema transformation engine was developed in the study as part of an extensive prototyping platform in the GiMoDig project. The prototype service proved the two-way schema transformation approach to be feasible in a cross-border setting involving topographic datasets in four different countries. Satisfactory performance levels were achieved in real-time processing. The time required for the query transformation is negligible. Data transformations are much more demanding, but still represent a minor share of the total time spent in the response processing of the GiMoDig prototype. More significant phases of the process in terms of performance include data retrieval, data transmission and on-the-fly generalisation.

As a result of the transformation, the national datasets of the participating countries are available seamlessly from the integration service layer, structured according to a common schema (GiMoDig Global Schema) comprising 17 important topographic feature types and expressed in the common pan-European coordinate reference system (ETRS89).

6.2 Service architecture development

6.2.1 Feature-based processing of GI on the Web

Related research question (1): What are the main benefits gained by bringing object-based geographic information to the Web application domain?

A significant proportion of all geospatial information can logically be connected to some recognizable real-world entity, a geographic feature. Consequently, feature-based representation of reference datasets, such as topographic maps and city plans, would provide benefits for effective information retrieval (Peng 2005, Lehto 1997). The research described here represents early experimentation with feature-based treatment of GI on the Web platform.

The main benefit of the object-oriented approach was found to be the opportunity to connect rich multimedia content to spatial representation via the related geographic features (Gartner 2007). The multimedia components tested in the research prototype include images, audio files and 3D models. All these multimedia properties could be connected to the features via the commonly used URL-based Web referencing mechanism.

The prototype Web application developed as part of the research also demonstrates some useful approaches that are being used increasingly in the modern geospatial Web solutions. These include the idea of dividing the map content into raster-based background and vector-formatted, feature-based foreground. This way, the inherent strengths of both representations can be fully utilized. A simple raster map serves well as a general visual reference for the more detailed foreground information that is particularly important from the user application point of view.

Real-time content transformation processes that are in the focus of this thesis essentially require availability of the source datasets in the form of geographic features. Advanced processes aiming at dynamically changing the online representation of the geospatial content, such as multi-purpose visualisation or real-time generalisation, would not be feasible without access to detailed object-based information. As such the results achieved in this part of the research form an important prerequisite for the later work described in the thesis.

6.2.2 Conceptual service layers

Related research question (6): What kind of service architecture is required to support an advanced geospatial Web application?

The roles of the various actors involved in an advanced geospatial Web application are becoming increasingly distinct. Those of the service reference content provider and the provider of the task-specific end-user application can be clearly distinguished. The former focuses on maintaining an up-to-date reference dataset and delivering this content to the application provider either as a static background map visualisation or as a set of active features. The application provider's tasks include adding the application-specific overlay information on top of the reference content and supporting the end user with all relevant functionality.

Another important consideration is separation of the content from its presentation characteristics – a principle that becomes a necessity in many complex Web applications. If the reference content provider delivers a feature-based representation of its data holdings, the responsibility for determining the visualisation details is automatically shifted to an outside party. Specifically, integration of the reference content with other reference data sources or with application overlay data will typically require visualisation principles controlled by an external entity.

The needs of the end-user application might also set specific requirements about how the background map is to be constructed. The most obvious parameters include selection of the feature types to be present on the map, the properties of the features to be shown as map labels, the map's level of detail, and any other particularities dictated by the dynamic context parameters and user preferences. These requirements might make it necessary to carry out application-specific computations on the level of the source dataset or the derived cartographic visualisation.

Based on above considerations, certain readily distinguishable roles were identified for the provision of an advanced Web service. These include the reference data manager, reference data integrator, data content processor, map visualisation specialist, application service provider, and the end-user. To be able to separate these functions in a Web service environment, a framework was designed in which an individual task can be assigned to the party who provides the best support for it.

The layered service architecture designed in the study consists of six conceptual computing layers, each with a well-distinguished role in the overall service provision. The six layers are: a data service layer, a data integration layer, a data processing layer, a portal layer, a value-adding service layer and a client layer. The architecture is illustrated in Table 1, together with the expertise that each layer provides for satisfying the application requirements.

The service architecture designed was tested in an extensive prototype environment. Generic Web techniques were applied successfully throughout the prototype system. These

Service layer	**Role and main responsibilities**
Client layer	User interface manager Application-specific visualisations User interaction management Transforming user actions into service requests
Value-adding service layer	Application specialist Support for application needs Integration of application data Understanding of reference map needs
Portal layer	Map visualisation expert Support for multi-purpose use Transforming map requirements into content requests Understanding the content processing facilities
Data processing layer	Content processor Expertise on spatial transformations, e.g. generalisation Support for geometric and topologic operations
Data integration layer	Reference data integrator Transformation support; schema/coordinates Knowledge about available data resources
Data service layer	Reference data manager Expertise on data maintenance and management

Table 1. The six service layers and their roles and main responsibilities.

include XML-based encodings in the case of both geospatial data content and map visualisations, data modelling and validation based on XML Schema-defined application schemas, XSLT-enabled schema transformations and multi-channel presentations, and HTML-based user interfaces.

To facilitate flexible service configurations, open access interfaces are used for each of the service layers. Existing internationally developed interface standards provide adequate support on the data service layer, data integration layer and portal layer. New service interface specifications were developed for the data processing layer and valued-adding service layer.

The final tangible result of this part of the study was the provision of a seamless, homogeneous map service covering two cross-border test areas on the Danish/German and Swedish/Finnish border. The underlying datasets are integrated on the data level, generalised as required by a real-time process and visualised according to detailed context parameters determined by the needs of the user application. A map illustrating the outcome of the portal service on the Swedish/Finnish border is presented in Figure 7.

Figure 7. A map visualisation of the Swedish/Finnish border (Tornio-Haparanda) as provided by the GiMoDig portal service. The source datasets have been integrated using real-time schema and coordinate reference system transformations (Lehto & Sarjakoski 2004b).

6.3 Schema transformation service

Related research question (5): How schema transformation can be organised in an extensive Web service-based SDI?

The schema transformation-related part of the study focuses on the transformation between two GML application schemas. As such, the source and target schemas share a common conceptual framework, making it feasible to distinguish a finite set of atomic transformation components by means of which all practical transformations can be declared. These components can be further related to the principal levels of the GML feature model.

Schema transformation is approached in this study as a specific category of geospatial Web services. The study recognized two possible ways of relating the schema transformation service to the source data content service: a) an opaque, two-way transforming content access service in a static, layered service architecture framework, and b) a transparent, one-way processing service in a dynamic, chained service environment.

The main pros and cons of the alternative approaches can be summarised as follows.

Layered service architecture

Pros

- Easy for the client application as it does not need to be aware of details related to the transformation process.
- The solution is robust and works well in cases where the target data model is pre-known and stable.

Cons

- The client application is unable to affect the outcome of the translation.
- The data provider must configure the translation according to varying client needs; consequently, it must know all the data models used on the client side.

Chained service architecture

Pros

- Because the original content service is visible to the calling application, the translation service to be used can be freely selected.
- As the client application controls the transformation, it can flexibly adjust the process to the actual need.
- Because the calling application is responsible for configuring the transformation, the content provider does not need to cater for the varying needs of different clients.

Cons

- The approach is more demanding for the client-side application.
- Developing a dynamic, user-controllable schema transformation process is a challenging task.

Concerning the problem of service parameterization, the study confirmed that general-purpose Web standards could be applied for the purpose. XML Schema-based indication of the source and target data models makes it straightforward to validate input and output datasets. XSLT declarations, together with the required spatial extensions, can be used to indicate the desired schema mapping. The adherence to general-purpose Web techniques greatly facilitates the development of services, as plenty of freely available open source tools exist.

The prototype service environment developed as part of the GiMoDig project applies schema transformation as a solution for data integration in the case of cross-border mobile use. The topographic base datasets of the four participating countries (Germany, Denmark, Sweden, Finland) were integrated into a seamless data service. After the real-time schema transformation the resulting dataset is presented in the harmonized common schema and expressed in the common coordinate reference system. The schema transformation process is carried out in a layered service architecture, as part of the functions of the Data Integration service layer. The opaque, layered architecture means that the schema transformation must act as a two-way process in which both the data access queries and the corresponding result datasets have to be transformed. All of the eight schema mappings involved were defined by XSLT-based transformation declarations (Lehto et al. 2004c).

7 CONCLUSION

Following the initial map image-centric, visualisation-oriented phase of the Internet GIS, feature-based representations of GI are now gradually gaining ground. This development is bound to undermine the long-standing conception that spatial data is special data. As the same base concepts and techniques are employed in content encoding and service access – regardless of the specific application area – seamless integration of geospatial information with conventional data sources becomes much easier to accomplish.

For the first time in history all the actors participating in GI-related processes find themselves at the same table. All datasets are potentially reachable inside a single network, the global Internet. Consequently, there is unprecedented pressure to achieve interoperability of GI and the related processes. Interoperability is thus emerging in a new setting – as a strategic component of a network service-based SDI, achievable by real-time content transformations.

7.1 Main findings

The results of this study emphasize real-time content transformations as a tool for three fundamental interoperability tasks: targeting visual representations of GI for a given display platform, modifying the geometric properties of the features through generalisation in a manner appropriate for the presentation scale, and carrying out schema transformations to match the original data content into the target data model used by the receiving application.

The declarative transformation language XSLT applied in the study proved to be flexible and expressive enough to support all the requirements set by the various different schema transformations needed in the case study projects. Specifically, the approach was found useful in the context of standardised data access services, as both the queries and the data content can be transformed applying the same transformation language.

The study contributes to new understanding of the potential provided by the dynamic processing of feature-based GI on the Web.

Starting from the basic benefits achievable by feature representation, such as the ability to link the features with auxiliary multimedia and other content or to improve the interactivity of the user interface, the research concentrates on the benefits yielded by real-time transformations performed on feature-encoded geospatial data in the Web service environment. These benefits include the wider reuse of geospatial content facilitated by online multi-purpose publishing methods which rely on real-time transformations, and enhanced interoperability due to online schema transformations, potentially carried out by a dedicated processing service.

Content transformations have been recognised within the INSPIRE process as an important geospatial service type for the developing European SDI (ESDI). The methodologies for real-time content transformations developed in the present study contribute to the establishment of this infrastructure.

The service prototype built in the GiMoDig project serves as proof that seamless cross-border services can be built on top of heterogeneous sources with the help of real-time transformations – a core principle in the INSPIRE process.

The exemplary service architecture developed makes a contribution to the design of the ESDI framework by identifying the various essential conceptual roles required for a flexible service infrastructure. According to the study findings, these include the roles of the geospatial data provider, data integrator, data processing provider, map visualisation provider and user application provider. The study acts as an example of extensively distributed geospatial content service provision, in contrast to the current practice of monolithic, self-contained services, mostly maintained by a single organisation.

The findings of the thesis confirm the main presumptions expressed in the hypothesis. Firstly, real-time transformations were proven to be a useful tool for various practical tasks encountered in service-oriented solutions involving geospatial data. Secondly, introduction of standardised representation of feature-based geodata to the Web domain was found to be an important enabler for applications such as multi-purpose publishing and content integration. Thirdly, in the study generic Web methodologies were successfully used in various essential processes dealing with geospatial data. Finally, the results of the study generally suggest that the proposed approach of Web-based delivery of GI improves interoperability and reuse of valuable geodata resources.

7.2 Limitations and drawbacks

The service prototypes developed as part of the study mostly represent the provider-driven approach for controlling the transformation. This can be regarded as a limitation of the study. As the proliferation of geospatial mass-market applications continues, the need for user-controlled transformations is becoming clearer. The same declarative, general-purpose transformation language could also be applied in the case of a client-controlled schema transformation, but this would require user-friendly tools to be available to assist with the schema mapping phase.

Query transformation, or query rewriting, has been extensively discussed in database integration research. In this study, query transformations have been carried out using the same declarative method as for data content transformations. However, query transformations pose some additional, unpredictable challenges compared with content transformations. The query predicate language applied, Filter Encoding (FE), provides for the construction of arbitrarily complex queries. In addition, the conditions set for the concepts of the common global schema cannot necessarily be directly mapped to conditions for local concepts. If a query cannot be transformed, a possible workaround can be achieved by first requesting a full spatially limited subset of the source data, then transforming the dataset into the global schema, and finally applying to this dataset the original conditions expressed in the common schema. Nevertheless, this approach would involve significant performance degradation.

One of the main limitations of the transformation methodology employed in the study is the need to read the whole dataset into the run-time memory in the form of an XML DOM tree structure before the transformation can commence. This effectively means that the availability of the computing platform's memory resources restricts the amount of data that can be transformed. An alternative approach, called Streaming Transformation for XML (STX), is available, but was not evaluated in any detail in the study.

An obvious drawback in all approaches adapting XML-technologies to processes involving predominantly numeric data is the frequent need for time-consuming conversions from text to number representation and vice versa. This problem was clearly recognised in the study. Consequently, the methods developed work best on Web applications involving relatively small amounts of data.

7.3 Future work

The service architecture designed in the study and used in the pilot services that were developed is based on a layered approach in which individual service layers make use of the services provided by the layer below and provide the layer above with value-added services. While this approach works well in a stable environment, more dynamism could be achieved in a service architecture based on freely pluggable service nodes that can be chained together to form a solution for ad hoc needs. Dynamic, automation-supported service composition largely remains a topic for further research and development. The methodologies developed in this study could be tested in a service environment based on dynamic service chaining. A specific research issue in this respect would be to explore the ways by which the source data service can be connected to the transformation service.

The content transformations developed in the study are predominantly provider-controlled. The mechanism available in the standardized standard map service interfaces (WMS and OpenLS Presentation Service) for controlling the styling properties of the resulting map are rather limited. More testing could be carried out using newer styling technologies, like the Styled Layer Descriptor (SLD) mechanism, in the general framework of multi-purpose publishing. While the study exclusively concentrated on SVG as the visualisation tool for vector-formatted maps, the growing relevance of new visualisation languages such as the Keyhole Markup Language (KML) would merit further testing.

A few issues remain for further research in relation to schema transformations. These include more general questions, such as how schema transformation services could be categorised according to the transforming capabilities they support, or whether it could be feasible to develop schema transformations in the direction of a generic process able to transform previously unknown schemas in an ad hoc manner. More extensive research would also be justified regarding the relative merits of an approach to execute schema transformation on the data access level compared with its implementation as an independent service node.

More detailed issues regarding schema transformation include, for instance, the question of how extensively the schema mapping language used in the study (XSLT) would serve the transformation requirements of the more extensive ESDI. It also remains unclear how easily the schema mapping language could be adopted as a standardized mechanism for transferring mapping instructions from the client to the service. A very useful research project would be the development of a standardised set of XSLT extension functions that could be used to transform geospatial data from one schema into another. The new version of the transformation language released in Jan 2007, XSLT 2.0, could also be explored for useful new functionalities applicable in schema transformations.

Some studies have recently concentrated on the idea of defining schema mapping at a conceptual level and then deriving the logical-level transformations from the conceptual mapping. One relevant research line would be to explore the feasibility of using XSLT-based transformation declarations as the logical level mapping in this scenario.

In the present study the results of real-time transformations were mostly demonstrated in the context of visualisation-centric applications. Data-level integration of heterogeneous sources, enabled by the schema transformation mechanisms developed in the study, would also support various online processing- and analysis-oriented use cases. The recent active research on geospatial Web processing services provides additional motivation for further work in this area.

The very active research on geospatial ontologies is seen as a promising source of new solutions for a schema mapping, able to overcome semantic discrepancies between the source and the target schema. The rooting of the participating schemas into a common semantic reference, expressed in the form of a domain ontology, would make it feasible to always determine the best possible match between the target and source concepts. The application of ontology-based mechanisms to schema mapping would significantly advance the prospect of automatic schema transformation. Further work towards this goal would certainly be fully merited.

8 REFERENCES

AASGAARD, R., 1992. *Automated cartographic generalization, with emphasis on real-time applications.* Phd dissertation, department of geodesy and photogrammetry, the norwegian institute of technology, 1992.

ABEL, D. J., GAEDE, V. J., TAYLOR, K. L. AND X ZHOU, 1999. SMART: Towards Spatial Internet Marketplaces. *GeoInformatica* 3: 2, pp. 141-164.

ABEL, D. J., OOI, B. C., TAN, K-L. ANS S. H. TAN, 1998. Towards integrated geographic information processing. *International Journal of Geographic Information Science*, 12: 4, pp. 353-371.

AFFLERBACH, S., ILLERT, A. AND T. SARJAKOSKI, 2004. The Harmonisation Challenge of Core National Topographic Databases in the EU-Project GiMoDig. Proceedings of the XXth ISPRS Congress, July 2004, Istanbul, Turkey. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXV(B4/IV):129-134, also CD-ROM.

ARCTUR, D., HAIR, D., TIMSON, G., MARTIN, E. P., AND R. FEGEAS, 1998. Issues and prospects for the next generation of the spatial data transfer standard (SDTS). *International Journal of Geographical Information Science*, 12: 4, pp. 403-425.

ARNOLD, K. AND J. GOSLING, 1996. *The Java programming language* (Addison-Wesley).

BALLEY, S., 2005. Improving geographical datasets usability by interactive schema transformations [online]. *Proceedings of the 8th AGILE Conference on GIScience*, Estoril, Portugal, 26-28 May, 2005. Available from: <http://www.agile-secretariat.org/Conference/estoril/papers/70_Sadrine Balley.pdf> [Accessed 18 July 2007].

BALLEY, S., BUCHER, B. AND S. LIBOUREL, 2006. A service to customize the structure of a geographic dataset. *Proceedings of the Second International Workshop on Semantic-based Geographical Information Systems (SeBGIS'06)*, Montpellier, pp1703-1711.

BADARD, T. AND D. RICHARD, 2001. Using XML for the exchange of updating information between geographical information systems. *Computers, Environment and Urban Systems (Special Issue: GISRUK 2000)*, 25, pp. 19–31.

BERNARD, L. AND A. WYTZISK, 2002. A Web-based Service Architecture for Distributed Spatio-temporal Modeling, *Proceedings of the 5th AGILE Conference on Geographic Information Science*, Palma, Spain, April 25–27, 2002, pp. 299–306.

BERNAND, L., EINSPANIER, U., LUTZ, M. AND C. PORTELE, 2003. Interoperability in GI Service Chains, *Proceedings of the 6th AGILE Conference on Geographic Information Science*, Lyon, France, April 24-26, 2003

Bishr, Y. and K. Buehler, 2003. GOS Transportation Portal, Implementation Architecture and Lessons Learned. Internal OGC Interoperability Program Report.

Braun, A., 2004. From the Schema Matching to the Integration of Updating Information into User Geographic Databases. *Proceedings of the Geoinformatics 2004 Conference*, 7-9 June 2004, Gävle, Sweden, pp. 211-218.

Burghardt, D., Neun, M. and R. Weibel, 2005. Generalization Services on the Web – Classification and an Initial Prototype Implementation. *Cartography and Geographic Information Science*, 32: 4, pp. 257-268.

Chang, Y-S. and H-D. Park, 2006. XML Web Service-based development model for Internet GIS applications. *International Journal of Geographical Information Science*, 20: 4, pp. 371-399.

Curtis, E., 2005. Schema Translation and Semantics in Data Interoperability [online]. *Snowflake Software White Paper*. Available from: <http://www.snowflakesoftware.co.uk/news/papers.htm> [Accessed 17 July 2007].

Curtis, E. and H. Müller, 2006. Schema Translation in Practice [online]. *Snowflake Software White Paper.* Available from: < http://windows.hostireland.com/eurosdr/workshops/models_2006/ discussion/Schema Translation%20in%20Practice.pdf> [Accessed 17 July 2007].

Devogele T., Parent C. and S. Spaccapietra, 1998. On spatial database integration. *International Journal of Geographic Information Science*, 12: 4, pp. 335-352.

de Vries M. and T. Tijssen, 2004. From Hybrid Mapping to Integrated Query and Processing: Two SDI Cases from the Netherlands [online]. *Proceedings of the 24th Urban Data Management Symposium*, Oct 27-29, 2004, Chioggia, pp. 15.1 – 15.9. Available from: <http://www.gdmc.nl/ publications/2004/SDI_cases.pdf> [Accesses 18 July 2007].

Díaz, L., Granell, C. and M. Gould, 2008. Spatial Data Integration over the Web. In V. E. Ferraggine, J.H. Doorn, L.C. Rivero (Eds.) *Encyclopaedia of Database Technologies and Applications*, 2nd Edition, (in press).

Donaubauer A., Fichtinger, A., Schilcher, M. and F. Straub, 2006. Model Driven Approach for Accessing Distributed Spatial Data Using Web Services - Demonstrated for Cross-Border GIS Applications [online]. *Proceeding of the XXIII International FIG Congress*, 8-13 Oct, 2006, Munich, Germany. Available from: <http://www.fig.net/pub/fig2006/papers/ts82/ts82_04_donaubauer_etal_0603.pdf> [Accessed 18 July 2007].

Donaubauer, A. Straub, F. and M. Schilcher, 2007. mdWFS: A Concept of Web-enabling Semantic Transformation. *Proceedings of the 10th AGILE Conference on Geographic Information Science,* 'The European Information Society: Leading the way with geo-information', May 8-11, 2007, Aalborg, Denmark, CD-ROM.

Dorninger, P., 2003. XML Technologies and Geodata [online]. *Proceedings of the 8th international symposion on information and communication technologies in urban and spatial planning and imapcts of ICT on physical space, CORP2003 GeoMultimedia03*, February 28, 2003, TU Vienna, Austria. Available from: < http://www.ipf.tuwien.ac.at/MarsExpress/docs/corp03/article0302.pdf> [Accessed 25 July 2007].

Edvardes, A., Burghardt, D. and M. Neun, 2007. Experiments in Building an Open Generalisation System. Chapter 8 in: Mackaness, W. A., Ruas, A. and L. T. Sarjakoski, (eds.), *Generalisation of Geographic Information: Cartographic Modelling and Applications*, Series of International Cartographic Association, Elsevier, pp. 161-175.

Edwardes, A., Burghardt, D., Bobzien, M., Harrie, L., Lehto, L., Reichenbacher, T., Sester, M. and R. Weibel, 2003. Map Generalisation Technology: Addressing the Need for a Common Research Platform. *Proceedings of the 21st International Cartographic Conference,* August 10-16, 2003, Durban, South Africa, pp. 170-179.

Einspanier, U., Lutz, M., Senkler, K., Simonis, I. and A. Sliwinski, 2003. Toward a Process Model for GI Service Composition, *Proceeding of the GI-Tage (GI Days).*

ESRI, 2007. ArcGIS Data Interoperability, Data Integration Tools for ArcGIS [online]. Available from: < http://www.esri.com/software/arcgis/extensions/datainteroperability/> [Accessed 25 July 2007].

Essid, M. and O. Boucelma, 2004. Mediated Geographic Web Feature Services. *Proceedings of the VI Brazilian Symposium on GeoInformat-*

ics – GEOINFO 2004, Campos do Jordão, Brazil. Available from: <http://www.geoinfo.info/geoinfo2004/papers/6366.pdf> [Accessed 16 July 2007].

Essid, M., Boucelma, O., Colonna, F-M. and Y. Lassoued, 2004. Query Processing in a Geographic Mediation System. *Proceedins of the 12th ACM International Workshop on Geographic Information Systems, ACM-GIS 2004*, November 12-13, 2004, Washington, DC, USA, pp. 101-108.

European Parliament and the Council, 2007. *Directive of the European Parliament and of the Council establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)*. PE-CONS 3685/06, Brussels.

Friis-Christensen, A., Schade, S. and S. Peedell, 2005. Approaches to Solve Schema Heterogeneity at the European Level [online]. *Proceedings of the 11th EC-GI & GIS Workshop*, 29 June – 1 July, 2005, Alghero, Sardinia. Available from: < http://www.ec-gis.org/Workshops/11ec-gis/papers/301peedell.pdf> [Accessed 18 July 2007].

Gartner, G., 2001. Mobile Internet: Applying Maps to Mobile Clients, *Journal of South China Normal University,* Guangzhou, China, July 2001, pp. 112-122.

Gartner, G., 2007. Development of Multimedia - Mobile and Ubiquitous, In: Cartwright, W., Peterson, M.P. and G. Gartner (eds.), *Multimedia Cartography,* Second Edition, Springer Berlin Heidelberg New York, pp. 51-62.

Gnägi, H.R., Morf, A. and P. Staub, 2006. Semantic Interoperability through the Definition of Conceptual Model Transformations [online], *Proceedings of the 9th AGILE Conference on Geographic Information Science*, 20-22 April 2006, Visegrád, Hungary. Available from: < http://www.agile2006.hu/papers/a128.pdf> [Accessed 23 July 2007].

Godby C.J., Smith D., Childress, E., 2003. Two Paths to Interoperable Metadata, *Proceedings of Dublin Core Conference, DC-2003: Supporting Communities of Discourse and Practice Metadata Re-search & Applications*, Seattle,Washington (USA).

Gould M., 1999. XML-based mechanism for open exchange of Geographic Information and Metadata, *GI/GIS Technology Watch*, JRC-SAI.

Gould, M. and A. Ribalaygua, 1999. A New Breed of Web-Enabled Graphics, *GeoWorld*, Mar 1999.

Grønmo, R., Berre, A-J., Solheim, I., Hoff, H. and K. Lantz, 2000. DISGIS: An Interoperability Framework for GIS – Using the ISO/TC 211 Model-based Approach [online]. *Proceedings of the 4th Global Spatial Data Infrastructure Conference*, Cape Town, South Africa, Mar 13-15, 2000. Available from: < http://folk.uio.no/roygr/GSDI-2000.pdf > [Accessed 18 July 2007].

Harrie L. and M. Johansson, 2003. Real-time data generalisation and integration using Java. *Geoforum Perspectiv*, February, 2003, pp. 29-34.

Harrie, L., Sarjakoski, L. T., and Lehto, L., 2002a. A Mapping Function for Variable-Scale Maps in Small-Display Cartography, *Journal of Geospatial Engineering,* 4(2), pp. 111–123.

Harrie, L., Sarjakoski, L. T. and L. Lehto, 2002b. A variable-scale map for small-display cartography. *Proceedings of the Joint International Symposium on "GeoSpatial Theory, Processing and Applications" (ISPRS/Commission IV, SDH2002)*, Ottawa, Canada, July 8–12, 2002, 6 p, CD-ROM.

Harrie, L., Stigmar, H., Koivula, T. and L. Lehto, 2004. An Algorithm for Icon Labelling on a Real-Time Map. In: P. F. Fisher (ed.), Development in Spatial Data Handling, *11th International Symposium on Spatial Data Handling, SDH-2004*, Leicester, UK, Springer-Verlag, Heidelberg, pp. 493-507.

ISO TC211, 2006. ISO TC211 Terminology Glossary, dated Apr 17, 2006 [online]. Available from: <http://www.isotc211.org/TC211_Terminology_Glossary-20060417-Published.xls> [Accessed 19 July 2007].

Jones, C.B., Abdelmoty, A.I., Lonegran, M.E., Van Der Poorten, P. and S. Zhou, 2000. Multi-scale spatial database design for online generalisation. *Proceedings of the Spatial Data Handling Symposium*, Beijing, pp. 7b.34–7-b.44.

Kilpeläinen, T., 1997. Multiple representation and generalization of geo-databases for topographic maps. *Publications of the Finnish Geodetic Institute*, No. 124, PhD dissertation, Kirkkonummi.

KILPELÄINEN, T., 2001. Maintenance of multiple representation databases for topographic data. *The Cartographic Journal*, 37, pp. 101–107.

KRAAK, M-J. AND A. BROWN, 2001. *Web Cartography, developments and prospects*. Taylor&Francis Inc, London, 209 p.

LEE D., MANI M., CHU, W., 2003, Schema Conversion Methods between XML and Relational Models. *Knowledge Transformation for the Semantic Web*, IOS Press, Amsterdam.

LEHTO, L., 1997. Spatial Objects in the WWW - a Java Approach. *Proceedings of the 6th Scandinavian Research Conference on GIS*, Stockholm, Sweden, 1-3 June, 1997.

LEHTO, L., 1998a. Java / CORBA Integration - a new Opportunity for Distributed GIS on the Web. *Proceedings of the ACSM Conference,* 28 February -5 March, 1998, Baltimore, USA, pp. 474-481.

LEHTO, L., 1998b. Interoperable Processing of GI on the Web. *Surveying Science in Finland*, 16(1-2): 3-18.

LEHTO, L., 2000. XML in Web-based Geospatial Applications. *3rd AGILE Conference on Geographic Information Science*, Hanasaari Conference Centre, Helsinki/ Espoo, May 25-27, 2000, pp. 162-167.

LEHTO, L., 2001. Multi-purpose Publishing of Geodata in the Web, *Journal of South China Normal University,* Guangzhou, China, July 2001, pp. 105-111.

LEHTO, L., 2003a. Architecture Specification. Geospatial Info-Mobility Service by Real-Time Data-Integration and Generalisation- project (GiMoDig), IST-2000-30090, Deliverable D4.4.1: Final system architecture, May 22 2003, Public EC report, 41 p. An electronic version available at http://gimodig.fgi.fi/deliverables.php.

LEHTO, L., 2003b. A Standards-Based Architecture for Multi-purpose Publishing of Geodata on the Web. In: Peterson, M.P. (ed.), *Maps and the Internet*, Elsevier Science, pp. 221-230.

LEHTO, L. AND J. KÄHKÖNEN. 1997. The Java Platform as a Tool for Web-based GI Delivery. *Den store Nordiske GIS-konference*, Kolding, Denmark, 29-31 October, 1997.

LEHTO, L. AND J. KÄHKÖNEN, 1999. Java 3D API as a Tool for 3D-GIS on the Internet. *Surveying Science in Finland*, 17(1-2):5-19.

LEHTO, L. AND T. KILPELÄINEN, 2000. Real-time Generalization of Geodata in the WEB. International *Archives of Photogrammetry and Remote Sensing*, Amsterdam, the Netherlands, 16-23 July 2000, XXXIII(B4): pp. 559-566.

LEHTO, L. AND T. KILPELÄINEN, 2001. Generalizing XML-Encoded Spatial Data on the Web, *Proceedings of the 20th International Cartographic Conference (ICC)* 2001, Beijing, China, August 6-10, 2001, 4:2390-2396.

LEHTO, L. AND T. KILPELÄINEN, 2001. Real-Time Generalization of XML-Encoded Spatial Data on the Web, *Proceedings of the GIS Research in the UK, 9th Annual Conference, GISRUK* 2001, University of Glamorgan, Wales, April 18–20, 2001, pp.182–184.

LEHTO, L. AND T. SARJAKOSKI, 2003. Real-Time Integration of XML-Encoded Spatial Data for Mobile Use. *Proceedings of the 6th AGILE Conference on Geographic Information Science,* Lyon, France, April 24-26, 2003, pp. 97-101.

LEHTO, L. AND T. SARJAKOSKI, 2004a. An Open Service Architecture For Mobile Cartographic Applications. In: Gartner, G, (ed.), Location Based Services & TeleCartography, Proceedings of the Symposium 2004, January 28–29, *Geowissenschaftliche Mitteilungen,* Technische Universität Wien, (66), pp. 141-145.

LEHTO, L. AND T. SARJAKOSKI, 2004b. Schema Translations by XSLT for GML-Encoded Geospatial Data in Heterogeneous Web-Service Environment. Proceedings of the XXth ISPRS Congress, July 2004, Istanbul, Turkey, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXV(B4/IV):177-182, also CD-ROM.

LEHTO, L., KÄHKÖNEN, J. AND T. KILPELÄINEN, 1997. WWW-Technology as Means of Transfer and Visualization of Geographic Objects. *Proceedings of the 18th ICA/ACI International Cartographic Conference*, Stockholm, Sweden, 23-27 June, 1997.

LEHTO, L., KÄHKÖNEN, J. AND T. SARJAKOSKI, 2001. Multi-purpose Publishing of Geodata in the Web, *Proceedings of the 4th AGILE Conference on Geographic Information Science*, Brno, Czech Republic, April 19-21, 2001, pp. 209-214.

LEHTO, L., KÄHKÖNEN, J., SARJAKOSKI, L.T. AND T. SARJAKOSKI, 2002. Multi-purpose Publishing of Geodata in the Web. *GeoInformatics,* 5(July/August 2002): pp. 18–21.

Lehto L., Sarjakoski, T., Hvas, A., Hollander, P., Ruotsalainen, R. and A. Illert, 2004a. A Prototype Cross-border GML Data Service. *Proceedings of the 7th AGILE Conference on Geographic Information Science*, Heraklion, Greece, April 29 - May 1, 2004, pp. 361-367.

Lehto, L., Sarjakoski, T., Nissen, F., Hvas, A., Illert, A., Afflerbach, S., Hampe, M., Ruotsalainen, R., Kemppainen, I., Hollander, P. and M. Forsmark, 2004c. Report of the GiMoDig Data Service Prototype. Geospatial Info-Mobility Service by Real-Time Data-Integration and Generalisation- project (GiMoDig), IST-2000-30090, Deliverable D6.5.2, July 9, 2004, Public EC report, 93 p.

Llavador M., Canos J.H., 2006. XSMapper: a Service-Oriented Utility for XML Schema Trans-formation. *ERCIM News*, No. 64, 2006.

Lutz, M., 2006. Specification of the Schema Mapping Service [online]. RM-OA: Reference Model for the Orchestra Architecture, OA Services Specifcations. Deliverable D3.4.3 of the ORCHESTRA project. Available from: <http://www.eu-orchestra.org/docs/OA-Specs/Schema_Mapping_Service_Specification_v0.7-JRC-IES.pdf> [Accessed 23 July 2007].

Mackaness W. and M. Jackson, 2000. Maps in a mobile world, *Proceeding of the GIS 2000 Conference*, London.

Manoa, S., Boucelma, O. and Y. Lassoued, 2004. Schema Matching in GIS. In C. Bussler and D. Fensel (eds.): *AIMSA 2004*, (Springer-Verlag Berlin Heidelberg), pp. 500-509.

Marotta A., Ruggia R., 1999. Designing Data Warehouses through schema transformation primitives. *Proceedings of the 18th International Conference on Conceptual Modeling (ER'99)*, Paris, France, November, 15-18, 1999.

Mathiak, B., Kupfer, A. and K. Neumann, 2004. Using XML Languages for Modeling and Web-Visualization of Geographic Legacy Data. *Proceedings of the VI Brazilian Symposium on GeoInformatics – GEOINFO 2004*, Campos do Jordão, Brazil. Available from: <http://www.geoinfo.info/geoinfo2004/papers/5738.pdf> [Accessed 18 July 2007].

Merdes, M., Häussler, J. and A. Zipf, 2005. GML2GML: Generic and Interoperable Round-Trip Geodata Editing using Standards-Based Knowledge Sources [online]. *Proceedings of the 8th AGILE Conference on GIScience*, Estoril, Portugal, 26-28 May, 2005. Available from: <http://www.agile-secretariat.org/Conference/estoril/papers/49_Matthias Merdes.pdf> [Accessed 17 July 2007].

Misund G. and H. Vålerhaugen, 2004. Integration of Heterogeneous GML Sources. *Proceedings of the GML and Geo-Spatial Web Services Conference*, Vancouver, Canada, 2004.

Murray, D. and J. C. Chow, 2005. GML – experiences from the Field [online]. *Safe Software White Paper*. Available from: <http://www.safe.com/solutions/white-papers/pdfs/gml-experiences-froom-the-field.pdf> [Accessed 17 July 2007].

OGC, 1999. OpenGIS Simple Features Specification for CORBA, Revision 1.1 [online], Open Geospatial Consortium. Available from: <http://portal.opengeospatial.org/files/?artifact_id=834> [Accessed 16 July 2007].

OGC, 2004a. OpenGIS Geography Markup Language (GML) Implementation Specification, Version 3.1.0 [online], Open Geospatial Consortium. Available from: <http://portal opengeospatial.org/files/?artifact_id=4700> [Accessed 16 July 2007].

OGC, 2004b. OGC Web Map Service Interface, Version 1.3.0 [online], Open Geospatial Consortium. Available from: <http://portal opengeospatial.org/files/?artifact_id=4756> [Accessed 16 July 2007].

OGC, 2005a. OpenGIS Location Services (OpenLS): Core Services, Version 1.1 [online], Open Geospatial Consortium. Available from: <http://portalopengeospatial.org/files/?artifact_id=8836> [Accessed 16 July 2007].

OGC, 2005b. Web Feature Service Implementation Specification, Version 1.1.0 [online], Open Geospatial Consortium. Available from: <http://portalopengeospatial.org/files/?artifact_id=8339> [Accessed 16 July 2007].

OGC, 2006. OpenGIS Implementation Specification for Geographic information - Simple feature access – Part 1: Common architecture [online], Open Geospatial Consortium. Available from: <http://portal.opengeospatial.org/files/?artifact_id=18241> [Accessed 16 July 2007].

OGC, 2007. The Home Page of the Open Geospatial Consortium (OGC) [online]. Available from: <http://www.opengeospatial.org> [Accessed 16 July 2007].

Pascoe, R. T. and J. P. Penny, 1995. Constructing Interfaces between (and within) Geographical Information Systems. *International Journal of Geographical Information Systems*, 9: 3, pp. 275-291.

Peng, Z-R., 2005. A proposed framework for feature-level geospatial data sharing: a case study for transportation network data. *International Journal of Geographic Information Science*, 19: 4, pp. 459-481.

Peng, Z-R and M-H. Tsou, 2003. *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Network*, John Wiley & Sons. 720 p.

Peterson, M. P., 1997. Trends in Internet Map Use. *Proceedings of the 18th ICA/ACI International Cartographic Conference*, Stockholm, Sweden, 23-27 June, pp. 1635-1642.

Peterson, M. P., 1999. Trends in Internet Map Use: a Second Look. *Proceedings of the 19th ICA/ACI International Cartographic Conference*, Ottawa, Canada, 14-21 Aug, pp. 571-580.

Plewe, B., 2007. Web Cartography in the United States. *Cartography and Geographic Information Science*, 34: 2, pp. 133-136.

Sarjakoski, T. and L. T. Sarjakoski, 2005. The GiMoDig public final report [online]. Geospatial Info-Mobility Service by Real-Time Data-Integration and Generalisation -project (GiMoDig), IST-2000-30090, Deliverable D1.2.31, March 18, 2005, Public EC report, 70 p. Available from: <http://gimodig.fgi.fi/deliverables.php> [Accessed 20 July 2007].

Sarjakoski, T. and L. T. Sarjakoski, 2007. A Real-Time Generalisation and Map Adaptation Approach for Location-Based Services. Chapter 7 in: Mackaness, W. A., Ruas, A. and L. T. Sarjakoski, (eds.), *Generalisation of Geographic Information: Cartographic Modelling and Applications*, Series of International Cartographic Association, Elsevier, pp. 137-159.

Sarjakoski, T. and A-M. Nivala, 2005. Adaption to Context – A Way to Improve the Usability of Topographic Mobile Maps. In: Meng, L., Zipf, A. and T. Reichenbacher (eds.), *Map-based Mobile Services – Theories, Methods and Implementations,* Springer Berlin Heidelberg New York, pp. 107-123.

Seeley, R. S., 2003. Data Schemas: When the Method Becomes the Madness: Issues in Sharing Digital Geospatial Data in a Global Environment and Schema Remodeling Tool as One Viable Solution. *Proceedings of the 21st International Cartographic Conference*, August 10-16, 2003, Durban, South Africa, pp. 1728-1732.

Sester, M., Sarjakoski, L. T., Harrie, L., Hampe, M., Koivula, T., Sarjakoski, T., Lehto, L., Elias, B., Nivala, A.-M. and H. Stigmar, 2004. Real-time generalisation and multiple representation in the GiMoDig mobile service [online]. Geospatial Info-Mobility Service by Real-Time Data-Integration and Generalisation- project (GiMoDig), IST-2000-30090, Deliverables D7.1.1, D7.2.1 and D7.3.1, Oct 4, 2004, Public EC report, 151 p. Available from: <http://gimodig.fgi.fi/deliverables.php> [Accessed 20 July 2007].

Sarjakoski, T., Sarjakoski, L.T., Lehto, L., Sester, M., Illert, A., Nissen, F., Rystedt, R. and R. Ruotsalainen , 2002. Geospatial Info-mobility Services — A Challenge for Interoperability and National Mapping Agencies. *Proceedings of the 5th AGILE Conference on Geographic Information Science*, Palma (Mallorca), Spain, April 25–27, 2002, pp. 585–589.

Sarjakoski, T., Sarjakoski, L.T., Lehto, L., Sester, M., Illert, A., Nissen, F., Rystedt, R. and R. Ruotsalainen , 2002. Geospatial Infomobility Services — A Challenge for National Mapping Agencies. *Proceedings of the Joint International Symposium on "GeoSpatial Theory, Processing and Applications" (ISPRS/Commission IV, SDH2002)*, Ottawa, Canada, July 8–12, 2002, 5 p, CD-ROM.

Sarjakoski, T., Sester, M., Sarjakoski, L.T., Harrie, L., Hampe, M., Lehto, L. and T. Koivula, 2005. Web generalisation service in GiMoDig - towards a standardised service for real-time generalisation. In: Toppen, F., and M. Painho, (eds.), *Conference proceedings, AGILE 2005, 8th Conference on Geographic Information Science*, Estoril, Portugal, May 26-28, 2005, pp. 509-518.

van Oosterom, P., 1995. The GAP-tree, an approach to 'on-the-fly' map generalization of an area partitioning. In *GIS and Generalization, Gisdata 1*, ESF, J.-C. Müller, J.-P.Lagrange and R. Weibel (Eds), Masser I. and F. Salgé (Series Eds.), pp. 120–132, Taylor & Francis.

van Oosterom, P. and V. Schenkelaars, 1995. The development of an interactive multiscale GIS. *International Journal of Geographic Information Science*, 9, pp. 489–507.

Vckovski, A. 1998, *Interoperable and distributed processing in GIS* (London: Taylor & Francis).

Visser U. and H. Stuckenschmidt, 2002. Interoperability in GIS – Enabling Technologies. *Proceedings of the 5th AGILE Conference on Geographic Information Science*, Palma de Mallorca, Spain, Apr 25-27, 2002, pp. 291-297.

Visser, U., Stuckenschmidt, H., Wache, H., & Vögele, T., 2001. Using Environmental Information Efficiently: Sharing Data and Knowledge from Heterogeneous Sources. In C. Rautenstrauch & S. Patig (Eds.), *Environmental Information Systems in Industry and Public Administration*, Hershey, USA & London, UK: IDEA Group, pp. 41-73.

W3C, 1999b. XSL Transformations (XSLT) Version 1.0 [online], World Wide Web Consortium. Available from: <http://www.w3.org/TR/xslt> [Accessed 16 July 2007].

W3C, 2001a. XML Schema Part 1: Structures [online], World Wide Web Consortium. Available from: <http://www.w3.org/TR/xmlschema-1/> [Accessed 16 July 2007].

W3C, 2001b. XML Schema Part 2: Datatypes [online], World Wide Web Consortium. Available from: <http://www.w3.org/TR/xmlschema-2/> [Accessed 16 July 2007].

W3C, 2003. Scalable Vector Graphics (SVG) 1.1 Specification [online], World Wide Web Consortium. Available from: <http://www.w3.org/TR/SVG11/> [Accessed 16 July 2007].

W3C, 2004a. Extensible Markup Language (XML) 1.0 (Third Edition) [online], World Wide Web Consortium. Available from: <http://www.w3.org/TR/2004/REC-xml-20040204/> [Accessed 16 July 2007].

W3C, 2007. The Home Page of the World Wide Web Consortium (W3C) [online]. Available from: <http://www.w3.org> [Accessed 16 July 2007].

Wang F., 2000. A Distributed Geographic Information System on the Common Object Request Broker Architecture (CORBA). *GeoInformatica* 4:1, pp. 89-115.

Waters, N., 1999. Is XML the answer for Internet-based GIS? *GeoWorld*, June 1999.

Wei, S., Joos, G. and W. Reinhardt, 2001. Management of Spatial Features with GML. *Proceedings of the 4th AGILE Conference on Geographic Information Science*, Brno, Czech Republic, April 19–21, 2001, pp. 370-375.

Wiegand, N., Zhou, N., Ventura, S. and I. Cruz, 2003. Extending XML Web Querying to Heterogeneous Geospatial Information, *Proceedings of the National Conference on Digital Government Research*, Boston MA, May 18-21, 2003.

Wikström L., 2006. Road Network Information Model [online], Deliverable 6.3, EuroroadS project, Available from: <http://www.euroroads.org/php/Reports/D6.3 Road Network Information Model.pdf> [Accessed 23 July 2007].

WiscLINC, 2006. Wisconsin Land Information Clearinghouse: Schema Transformation [online]. Available from: <http://www.sco.wisc.edu/wisclinc/transform.php> [Accessed 23 July 2007].

Woodsford, P. A., 2005. NMCA's and the Internet II – eDelivery and Feature Serving: Report on Joint EuroSDR/EuroGeographics Workshop [online], *Proceedings of the 11th EC-GI & GIS Workshop, ESDI: Settingthe Framework*, Alghero, Sardinia, 29th June - 1st July 2005. Available from: < http://www.ec-gis.org/Workshops/11ec-gis/papers/3019woodsford.pdf > [Accessed 23 July 2007].

Zaslavsky, I., Gupta, A., Marciano, R. and C. Baru, 2000. XML-based Spatial Data Mediation Infrastructure for Global Interoperability [online]. *Proceedings of the 4th Global Spatial Data Infrastructure Conference*, Cape Town, South Africa, Mar 13-15, 2000. Available from: <http://www.npaci.edu/DICE/Pubs/gsdi4-mar00/gsdi_iz.html> [Accessed 17 July 2007].

Zhang C. and W. Li, 2005. The Roles of Web Feature and Web Map Services in Real-time Geospatial Data Sharing for Time-critical Applications. *Cartography and Geographic Information Science,* 32: 4, pp. 269-283.

Zhang, L. and Lin, H., 1996. A client/server approach to 3D modeling support system for coast change study. *Proceedings of the GIS/LIS Conference* (Denver CO: ASPRS/ACSM/AAG/URISA/ AM-FM) pp. 1265 - 1274.