



LUND UNIVERSITY

Methods to Improve and Evaluate Spatial Data Infrastructures

Toomanian, Ara

2012

[Link to publication](#)

Citation for published version (APA):

Toomanian, A. (2012). *Methods to Improve and Evaluate Spatial Data Infrastructures*. [Doctoral Thesis (compilation), Centre for Geographical Information Systems (GIS Centre)]. Department of Physical Geography and Ecosystem Science, Lund University.

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Methods to Improve and Evaluate Spatial Data Infrastructures



LUND
UNIVERSITY

Ara Toomanian

Centre for Geographical Information Systems (GIS Centre), Department of
Physical Geography and Ecosystem Science, Lund University, Sweden

2012

A doctoral thesis at a university in Sweden is produced either as a monograph or as a collection of papers. In the latter case, the introductory part constitutes the formal thesis, which summarizes the accompanying papers already published or manuscripts at various stages (in press, submitted or in preparation).



Copyright © Ara Toomanian

Cover:© Iman Khayatan

Centre for Geographical Information Systems (GIS Centre), Department
of Physical Geography and Ecosystem Science, Faculty of Science

ISBN 978-91-85793-26-6

Printed in Sweden by Media-Tryck, Lund University
Lund 2012

Abstract

This thesis mainly focuses on methods for improving and evaluating Spatial Data Infrastructures (SDIs). The aim has been threefold: to develop a framework for the management and evaluation of an SDI, to improve the accessibility of spatial data in an SDI, and to improve the cartography in view services in an SDI.

Spatial Data Infrastructure has been identified as an umbrella covering spatial data handling procedures. The long-term implementation of SDI increases the need for short/middle term feedbacks from different perspectives. Thus, a precise strategic plan and accurate objectives have to be defined for the implementation of an efficient environment for spatial data collection and exchange in a region.

In this thesis, a comprehensive study was conducted to review the current methods in the business management literature to approach to an integrated framework for the implementation and evaluation of SDIs. In this context, four techniques were described and the usability of each technique in several aspects of SDI implementation was discussed.

SDI evaluation has been considered as one of the main challenges in recent years. Lack of a general goal oriented framework to assess an SDI from different perspectives was one of the main concerns of this thesis. Among a number of the current methods in this research area, we focused on the Balanced Scorecard (BSC) as a general evaluation framework covering all perspectives in an SDI.

The assessment study opened a window to a number of important issues that ranged from the technical to the cartographic aspects of spatial data exchange in an SDI. To access the required datasets in an SDI, clearinghouse networks have been developed as a gateway to the data repositories. However, traditional clearinghouse networks do not satisfy the end user requirements. By adding a number of functionalities, we proposed a methodology to increase the percentage of accessing required data. These methods were based on predefined rules and additional procedures within web processing services and service composition subjects to develop an expert system based clearinghouses.

From the cartography viewpoint, current methods for spatial data presentation do not satisfy the user requirements in an SDI environment. The main presentation problem occurs when spatial data are integrated from different sources. For appropriate cartography, we propose a number

of methods, such as the polygon overlay method, which is an icon placement approach, to emphasize the more important layers and the color saturation method to decrease the color saturation of the unimportant layers and emphasize the foreground layer according to the visual hierarchy concept.

Another cartographic challenge is the geometrical and topological conflicts in data shown in view services. The geometrical inconsistency is due to the artificial discrepancy that occurs when displaying connected information from different sources, which is caused by inaccuracies and different levels of details in the datasets. The semantic conflict is related to the definition of the related features, i.e., to the information models of the datasets. To overcome these conflicts and to fix the topological and geometric conflicts we use a semantic based expert system by utilizing an automatic cartography core containing a semantic rule based component. We proposed a system architecture that has an OWL (Web Ontology Language) based expert system to improve the cartography by adjusting and resolving topological and geometrical conflicts in geoportals.

Svensk sammanfattning

Denna avhandling är inriktad på metoder för att förbättra och utvärdera infrastrukturer för geografiska data (eng. Spatial Data Infrastructure). Syftet har varit trefaldigt: att utveckla ett ramverk för hantering och utvärdering av en infrastruktur, att förbättra tillgängligheten av geografiska data i en infrastruktur, och för att förbättra kartografi i visningstjänster i en infrastruktur.

En infrastruktur för geografiska data har identifierats som ett paraply som täcker hela hanteringen av geografiska data. Det långsiktiga genomförandet av en infrastruktur ökar behovet för återkopplingar från skilda perspektiv både på kort och medellång sikt. I grunden behövs en strategisk plan med konkreta mål för effektiv datainsamling och bra utbyte av geografiska data.

Den första studien i denna avhandling granskar hur metoder inom företagsutveckling kan användas för utveckling och utvärdering av en geografisk infrastruktur. Sammanlagt har fyra tekniker studerats och deras relevans för arbete med en geografisk infrastruktur diskuterats. För utvärdering av en geografisk infrastruktur rekommenderas användning av Balanced Scorecard (BSC). Fördelen med denna utvärderingsmetod är att den bildar ett ramverk där man kan studera infrastrukturen från flera perspektiv och även studera hur olika utvärderingsparametrar kan påverka varandra över tiden. Detta studeras närmare i den andra studien i avhandlingen.

Ett huvudsyfte med en geografisk infrastruktur är att skapa en portal för distribution av geografiska data till användarna. Till denna portal, som ofta benämns geoportal, kopplas producenternas karttjänster, dvs. en användare ska kunna nå alla producenternas data via portalen. I den tredje studien utvecklade vi metoder för att bearbeta producenternas data för att bättre svara mot slutanvändarnas behov. Dessa metoder integrerades i en portalmiljö. På så sätt skapas en geoportal som har ett större utbud av data.

En användare ska kunna studera geografiska data i en geoportal. Detta leder ofta till problem i de fall då data kommer från olika producenter; data är helt enkelt inte anpassade till sampresentation. I de två avslutande studierna utvecklar vi ett antal kartografiska metoder för sampresentation av data. I fjärde studien inriktar vi oss på problemet hur tillämpningsdata kan överlagras en baskarta utan informationsförlust. I den femte studien

utvecklas en metod för korrekt geometrisk och topologisk integrering av data.

Acknowledgments

During my study there were many people who, without their help, I could not be in this situation and finalize my PhD. Therefore, I would like to take the opportunity to express my deepest gratitude to all of the people who supported me in my study and research work.

First and foremost, I offer my sincerest gratitude to my first supervisor, Lars Harrie. I greatly appreciate him because he improved and shaped my research skills. Without his encouragement, I could not have finished my research. Truly, I greatly admire his style of supervision. I am also thankful for my second supervisor, Ali Mansourian. He opened a new window in my scientific life, which is continuously improving even now. Ali, you are very creative and always full of positive energy which increased my productivity. I will never forget your kind and responsible manner. Additionally, I would like to express my deepest gratitude to my third supervisor, Petter Pilesjö. Aside from being my co-supervisor, he is the one that supports me from A to Z! He always has a solution for any problem and I think that he is one of the heroes of Lund University!

This study was financed by the European Union funding program, Erasmus Mundus “External cooperation Window- lot 7” and Lund University. I am thankful for their support.

I would like to express my gratitude to my parents, my lovely sister, her husband, my dear nephew and my cute niece for supporting me for my entire life. Without their unlimited help, I could not have finished my research.

I am deeply indebted to my lovely wife and her family for their cooperation and support. I will never forget their positive influence on my work during the time that I was far from home.

I want to thank my colleagues and fellow PhD students at the department of physical geography and ecosystem sciences, the GIS Center staff (Karin, Micael, Roger, Mitch, Hanna, Andreas, Ulrik, Jing and Abdulghani). GIS Center has a positive atmosphere where any scientific mission can be accomplished successfully due to the brilliant and professional synergy among the staff. I also appreciate the department kansli staff for their support, with special thanks to Petra Andersson.

Special thanks also to all co-authors (Lars, Ali, Petter, Per-Ola, Anders, and Ebrahim), and organizations and authorities (Lantmäteriet, Lund and

Helsingborg municipalities), which without their contributions, the research was impossible to be conducted.

I would like to give special thanks to my friends and fellow country mates; without your spiritual support, it would have been difficult for me to stay abroad. I had a very nice time with you during my four years of study in Sweden and learned many things from you, which is a valuable asset for me.

Contents

1	Introduction	1
1.1	Background	1
1.2	Problem statement	2
1.3	Objectives	2
1.4	Thesis structure	3
2	Spatial Data Infrastructure fundamental concepts	5
2.1	Introduction	5
2.2	Spatial data	5
2.3	User requirement of spatial data	7
2.4	Spatial data sharing	10
2.5	Spatial Data Infrastructure -SDI	11
2.6	SDI components and nature	13
2.7	Clearinghouse networks	15
2.8	INSPIRE: European SDI initiative	18
3	SDI management and evaluation	20
3.1	Introduction	20
3.2	Overview of SDI evaluation research	20
4	Spatial web service composition	24
4.1	Introduction	24
4.2	Web processing service applications	25
4.3	Applications of service composition	26
4.4	Semantic web services	27
4.5	Methods to increase accessibility	28
5	Cartographic aspects of SDI	31
5.1	Fundamentals of cartography	31

5.2	Map design and symbolization	31
5.3	OGC standards for web cartography	36
5.4	Cartographic challenges in view services	37
6	Summary of papers	41
6.1	Paper I	41
6.2	Paper II	43
6.3	Paper III	45
6.4	Paper IV	46
6.5	Paper V	48
7	Conclusions	49
8	References	51

1 Introduction

1.1 Background

With the development of technology, digital spatial data have become increasingly important. The public sector, private companies and the public attempt to collect and produce spatial data a decision making tool. Strategic planning, risk analysis, site selection, and route finding are some of the major applications.

The collection and production of spatial data are time consuming and costly. Additionally, the production of spatial data from spatial raw data requires professionals, experts and advanced skills. A number of important challenges in this area influence spatial data collection, such as lack of coordination and collaboration, duplicate or parallel activities, and spatial data collection within two or several organizations. These types of duplicated work waste both financial and human resources.

To overcome different aspects of the current challenges in spatial data handling, a Spatial Data Infrastructure (SDI) has been introduced. An SDI is a collaborative environment for managing, storing, and exchanging spatial data. An SDI includes technical, social, institutional and political issues as well as financial challenges; hence SDI is an umbrella concept covering the entire spatial data handling domain (Groot and McLaughlin, 2000).

Currently, SDIs are implemented in numerous countries due to the high demand for spatial data exchange. Moreover, SDI implementations cross national borders and a number of regional borders. In several cases, continental agreements for regional level SDIs exist.

One of the major characteristics of an SDI is the long-term implementation procedure, which needs proper short/middle term feedbacks from different perspectives. In many countries, even one or two decades are not sufficient to implement all of the SDI components. Therefore, a precise strategic plan must be defined, with proper objectives and initiatives to reach the goal

according to the needs of the country by the evaluation and the refinement of the progress of an SDI.

In this thesis, we propose new ideas to evaluate and improve spatial data infrastructures from different aspects to implement a more effective and operative spatial data sharing framework for a wide range of users.

1.2 Problem statement

The SDI evaluation has been considered one of the foremost challenges in recent years. The main problem is the lack of a general goal oriented framework to assess an SDI from different perspectives. Among a number of the current methods in this research area, we focused on business management methods from a strategic viewpoint to investigate problems with performance measurement issues in the context of SDI implementation and evaluation. During the assessment study, we found that various problems exist from several technical and cartographic aspects of spatial data exchange in an SDI.

Current methods for spatial data presentation are not sufficient for an SDI environment. In particular, problems occurs spatial data stems from different sources. In this case, semantic, geometrical and topological heterogeneities must be resolved.

1.3 Objectives

The general aim of this thesis is to improve spatial data infrastructures. There are three specific objectives:

1. Develop a framework for the management and evaluation of an SDI.
2. Improve the accessibility of spatial data in an SDI.
3. Improve the cartography in the view services in an SDI.

1.4 Thesis structure

This thesis has a summary part followed by five papers. Due to the diverse nature of SDIs and also the broad range of topics for this research, a number of subjects are described in the summary part. The second chapter provides a general overview of spatial data infrastructure concepts. The third chapter is a literature review of current evaluation and implementation methods. In this chapter, we discuss the relevant subjects for Papers I and II. The fourth chapter contains an overview of methods used for spatial data exchange workflow. This chapter reviews the methods of web service processes that publish spatial data. The fifth chapter provides a general overview of the cartographic background of spatial data. The chapter includes theory regarding visualization techniques and several of the current challenges from a cartographic viewpoint. Chapter six provides a summary of the papers. Finally, chapter seven presents the conclusions of the thesis.

The papers are sorted according to their subjects:

- Paper I Toomanian, A., Mansourian, A., (2009). An Integrated Framework for the Implementation and Continuous Improvement of Spatial Data Infrastructures, In *SDI Convergence. Research, Emerging Trends, and Critical Assessment*. B. van Loenen, J.W.J. Besemer, J.A. Zevenbergen (Editors). Nederlandse Commissie voor Geodesie, Netherlands Geodetic Commission 48, pp 161-173, 2009.
- Paper II Toomanian, A., Mansourian, A., Harrie, L., Rydén, A., (2011). Using Balanced Scorecard for Evaluation of Spatial Data Infrastructures: a Swedish Case Study in accordance with INSPIRE. *International Journal of Spatial Data Infrastructures Research*, 6, pp 311-343.
- Paper III Mansourian, A., Omid, E., Toomanian, A., Harrie, L., (2010). Expert System to Support Functionality of Clearinghouse Services. *Computers, Environment and Urban Systems*, 35(2), pp159-172.
- Paper IV Toomanian, A., Harrie, L., Olsson, P., (2012). Automatic symbolization methods for geoportals. *The Cartographic Journal* (In Press).
- Paper V Toomanian, A., Harrie, L., Mansourian, A., Pilesjö, P., (2012). Automatic integration of spatial data in viewing services using a semantic based Expert System (Submitted).

In Paper I, the author made the main part of the study and the writing. The idea behind the study is shared with the co-author.

In Paper II, the author made the main part of the data collection and analysis from the authorities, and together with the co-authors investigated the Swedish NSDI implementation progress in accordance to the INSPIRE directive.

In Paper III, the author contributed the literature review, defined the case studies and different scenarios and prepared the theoretical background of the implementation.

In Paper IV, the author studied different methods for map visualization and implemented the cartographic methods. The study and writing were performed together with the co-authors.

In Paper V, the author conducted the implementation and wrote the manuscript. The ideas were developed with co-authors.

2 Fundamental concepts of Spatial Data Infrastructure

2.1 Introduction

This chapter addresses the fundamental concepts of Spatial Data Infrastructure (SDI). An SDI includes the rules, laws, standards and the data that are used to improve the access of spatial data. SDI indicates a type of soft infrastructure. The soft infrastructure is a main complement to the concept and is as important as but not the same as hard infrastructure, which addresses physical installations such as roads and water pipelines. In soft infrastructure as such, spatial data are crucial to reach a sustainable development.

The structure of this chapter is as follows: First, we provide an overview of spatial data. Next, we discuss the current state of spatial data sharing. The next section concerns the basic components of an SDI and SDI hierarchy. Finally, this chapter discusses important parameters of SDI development and current models for SDI funding.

2.2 Spatial data

More than 80% of the data used in most organizations and institutions activities, planning and management have a spatial nature and characteristics (Budic et al., 1999; Lemmens, 2001; Rhind, 1999). Spatial data, also known as geospatial data or geographic information, are defined as data associated with a specific location on the earth, particularly information regarding natural phenomena, and cultural and human resources (Williamson et al., 2003). Spatial data include maps, aerial and satellite images.

There are three types of spatial data in the geo community. Traditionally, the main part of the spatial data have been collected and produced by the

public sector. In the last decade, due to substantial financial benefits in the spatial data market, companies have made great investments that have generated the second type of commercial spatial data. Finally, in recent years, a third type of spatial data sets joined this community, which is volunteered spatial data. These data are the results of web 2.0 techniques and attempts to overcome the licensing issues for public sector and commercial data. This thesis focuses on public sector data.

Spatial data in conjunction with geographic information systems are utilized for visualization and analysis. Spatial data have wide ranging uses because any human activity takes place within a specific location or particular area, and therefore, the impacts of activity are more meaningful and more practical from a spatial viewpoint. Finally, spatial data are crucial for the improvement and development of economic and financial situations and for the protection of natural resources (Executive Order, 1994).

A substantial portion of all decisions made by national and local agencies are dependent on a location or have a spatial impact (Albaredes, 1992). When an analyst has a proper background about the study region, a more efficient and well-organized result is reached, and the decision-making process is improved for any location-based research. Additionally, spatial knowledge allows the user to connect various related information existing in the same location and collects the data in an integrated spatial database.

In this regard, logistics study and planning, environmental management and protection, society planning, crisis management and road network design are several applications of spatial data on local, national and international scales (Bernard et al., 2005; Williamson et al., 2003; Masser, 1998).

As a result, spatial data and related techniques lead to spatial knowledge that is important in the decision making processes. Using spatial data and techniques, the results of any location-based analysis can be properly understood. Additionally, collaboration and positive synergy increases among various domains across governmental, private and academic sectors. Furthermore, spatial knowledge has a direct and effective impact on economic, social and environmental development and is regarded as one of the major elements for sustainable development (Mansourian, 2006).

Spatial data collection and maintenance are expensive and require complex techniques and team work. Moreover, spatial data or service production from spatial raw data requires advanced skills and proficiency (van Leonen, 2003). Therefore, lack of coordination and collaboration, and thus duplicate or parallel activities in spatial data collection from two or several

organizations, is stated as one of the major challenges in this area (Chan et al., 2001; Nebert, 2001; When de Montalvo, 2000). These types of duplicate work waste both financial and human resources.

Proper documentation of spatial data is an essential aspect of data sharing. Similar datasets with different spatial and temporal accuracies without any documentation make data selection more challenging for any user. Also, other major problems are less attention to metadata production from the spatial data producers' side as well as standards for metadata in producing procedures (Nebert, 2001; van Leonen, 2003).

Applying different local standards for spatial datasets is another major problem. Several challenges involve the use of various standards to produce and store spatial data and the use of inappropriate standards without considering the user requirements. In this context, there are a number of problems:

- Challenges in the spatial data integration produced by different organizations.
- The spatial data produced may be inappropriate for the end users because it does not fit the user requirements.

2.3 User requirements for spatial data

Several challenges in spatial data collection, storage and distribution influence spatial data usage. To analyze and describe the current bottlenecks in using spatial data, the status of spatial data required for end users are discussed according to the data functions theory and modern theory for decision flow. This theory is categorized into four categories: *availability*, *accessibility*, *applicability* and *usability* (Mansourian, 2006; Nedovic-Budic et al., 2004; Feeney and Williamson, 2003):

2.3.1 Availability

Availability addresses with the existence of spatial data. The main concern is whether the required spatial data can be found. Other important parameters affecting the result of end users analysis are the quality and the specifications of the available data.

In reality, a number of scenarios emerge from the availability viewpoint:

- The data may be available in the same organization where the user works or another organization that produces the data.
- The data exist in multiple sources or do not exist at all.
- There exist old versions of data or exist for some parts, and
- Inaccurate data exist.

Occasionally, the situation is even worse because of a combination of the mentioned probabilities, e.g., the data exist in multiple sources with less accuracy and do not cover the entire region (Mansourian, 2006).

2.3.2 Accessibility

This term describes the context end user authentication and limitations regarding data access. The required data may be available in an organization but are not accessible. In this context, numerous factors can hindrance data accessibility (Feeney and Williamson, 2003):

Administrative constraints: To obtain a dataset, complex, time consuming, and bureaucratic procedures occasionally must be overcome in many organizations on the part of both data producers and data holders.

Inappropriate announcement: Occasionally, data holders do not announce available data, end users are not aware of spatial datasets and thus do not ask about them. Furthermore, the insufficient amount of advertisements from data holder due to the smaller revenue and benefits that is available in these markets.

Cultural issues: Some organizations do not distribute available data to keep quality and accuracy problems within the organization. Some organizations keep data to maintain power in negotiations and inter-organizational discussions. Finally, lack of the copyright regulations can cases cause inappropriate data sharing by data producers.

Security: In a number of situations, there are limitations for releasing data because of military usage or national security threats caused by spatial data.

Pricing policies: In some cases, the cost of a spatial dataset is too high and it is not cost effective for end users to purchase the data.

2.3.3 Applicability

Applicability is the percentage of accessible spatial data that is compatible with current standards and end user needs (Feeney and Williamson, 2003). The data should be applicable for high performance and in hardware and software environments. Occasionally, available and accessible data encounter problems regarding format, geometrical structure, context, data definition and classification, quality, etc. Consequently, the data they do not fit the end user requirements, and therefore, the user must perform time-consuming procedures to edit and prepare the data for use.

Several parameters affect the spatial data applicability:

Standards and data characteristics: Differences among standards in various datasets can cause complexity in data integration. Additionally, the existing classes and categories for features in addition to the dataset scale sometimes do not fit the user requirements. Finally, in many cases, the accessible data are not topologically well structured.

Data quality: Low quality data can affect any project, and sometimes inaccurate data are not used by the end user. Additionally, the quality of an update data sometimes does not satisfy the end user in compare to the quality of previous versions.

2.3.4 Usability

Usability refers to the amount of usage and the quality of use for the end user. In many cases, applicable data are not used in an efficient way.

Two primary factors affect the degree of spatial data usage. If the available or accessible data do not satisfy the needs of end users, the users do not use the data in their analysis. In this context, lack of metadata and the low awareness of data characteristics are other bottlenecks that influence the amount of data usage.

Organizational culture also changes the degree of spatial data usage. Many organizations act as both data producers and end users. These organizations minimize data sharing because of organizational resistance or other limitations. Additionally, inter-organizational conflicts and disagreement among employees sometimes lead to chaotic conditions and as a result the data are not useful for any purpose. Lack of awareness concerning the advantages of using datasets in decision making issues,

planning and analysis is one of the main restrictions on spatial data usability (Thellufsen et al., 2009).

According to the aforementioned descriptions and concepts, in an ideal situation, 100% of the spatial data requirements are available, accessible, applicable and usable for the end users. However, due to parallel projects, limitations and other challenges, only a small percentage of the required data are usable for various applications (Figure 2.1). The primary goal of a spatial data infrastructure is enhance the status of spatial data usage in a certain situation.

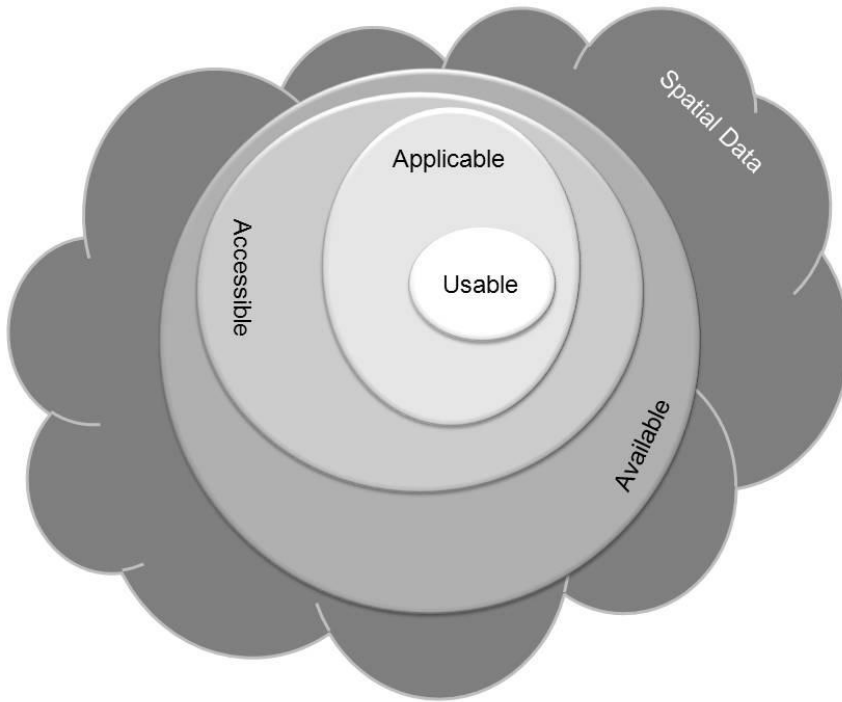


Figure 2.1. User requirements for spatial data (cf. Feeney et al., 2003).

2.4 Spatial data sharing

Many GIS projects require a large amount of spatial data for which the collection and production are time consuming and costly. Spatial data sharing is essential to avoid unnecessary costs from duplicate production procedures. Spatial data sharing is also required to increase the benefits of

spatial data usage and to encourage multiple usages of spatial data that have been collected for a specific purpose. Data sharing also increases data quality when many people work on a specific task and try to discover and edit errors within the data (Williamson et al., 2003). Hence, data sharing promotes both financial and human resources savings.

The concept of data sharing affects decisions and development issues through better harmonization and coordination on the management level. Spatial data in an organization are considered as an infrastructure for other organizations for better collaboration and cooperation in future projects.

Spatial data sharing has many advantages, but technical, social, economic, legal, political and organizational obstacles must be circumvented. Spatial Data Infrastructure is the key to facilitate spatial data sharing by providing essential collaboration and cooperation among different organizations.

2.5 Spatial Data Infrastructure -SDI

2.5.1 Definition

The history of Spatial Data Infrastructure (SDI) is long, but SDI became important in the 1980s due to great demand for cooperation and spatial data sharing. National SDI discussions began among academics in the US. The subject was officially confirmed with the President's executive command in 1994 (Executive Order, 1994). Statistics show that more than 120 countries have implemented a national SDI (Crompvoets et al., 2004; Crompvoets and Bregt, 2001). Currently, more countries are developing this infrastructure due to improvements and demand for spatial data.

The definition and interpretation of SDI differ among nations due to the specific conditions in each country. A number of definitions in the literature from major contributors in this area exist (e.g., Stojanovic et al., 2010; GSIDI, 2009; CGDI, 2004; Lemmens, 2001; Chan et al., 2001; Masser, 1998). According to the Office of Management and Budget in the US, "SDI is a framework of spatial data, metadata, users and tools that are interactively connected to use spatial data in an efficient and flexible way" (OMB, 2002). The Federal Geographic Data Committee defined National Spatial Data Infrastructure (NSDI) as: "An umbrella covering policies, standards, organizational procedures and technologies where is used to use, manage, and produce spatial data" (FGDC, 1997, p. 106).

2.5.2 Specifications

The SDI concept is generally used to avoid gaps within spatial datasets and duplication in data production in addition to other well-known spatial data problems (van Leenen and Kok, 2004). SDIs are also an innovative way that aims to design an environment for collaboration and cooperation. Specifically, an SDI provides a dynamic internal and external cooperation among organizations that responds to the collaboration needs of data providers and end users. An SDI is also a method to prepare and expand a mechanism for sharing and developing spatial datasets. In this way, data holders communicate with the current technologies to achieve various degrees of political organizational purposes in an efficient way (Chan et al., 2001). This environment is created via mechanisms that facilitate the sharing, access and use of spatial data in different communities.

SDIs are used to overcome the user requirement limitations mentioned in previous section by facilitating the availability, accessibility, applicability and usability of spatial data. Users can use the proper networks to find the available data easily within an SDI and can follow common procedures to access to the data. Moreover, with an SDI, the data may become more applicable with respect to the quality, format and other specifications. Finally, an SDI helps to increase the collaboration and cooperation among organizations in addition to the effective use of data for any application. In this regard, SDIs provide several benefits:

- SDIs remove unnecessary tasks, duplicate activities and parallel procedures.
- SDIs create an appropriate spatial data market.
- SDIs facilitate process-based management for spatial data.

In recent years, SDIs have been implemented as platforms or basic spatial data frameworks in numerous countries. SDIs are used not only to facilitate spatial data access but also to integrate spatial data in numerous situations. SDIs facilitate the integration of spatial datasets and aim to design a proper interface for data sharing among organizations, private sectors and end users. Another major goal of SDIs is to integrate and harmonize the spatial data collected from different sources and/or stored in different databases. This framework decreases both access time and cost for end users. Finally, all technical and organizational solutions help and facilitate SDI component relationships to benefit society (Rajabifard et al., 2003a; Rajabifard et al., 2003b).

A variety of aspects must be considered, such as organizational restructuring, legislation frameworks, cultural improvements, economic considerations and business factors. Consequently, SDI implementation requires skills and specific experiments (Remkes, 2000). Furthermore, SDI development needs detailed collaboration and communication among the different levels of governmental authorities and private sectors, which creates new types of responsibilities for any type of institute. Consequently, SDIs effectively provide the opportunity for all collaborators to access all current spatial datasets for use in internal, local, national and regional decisions. Finally, SDIs are mechanisms to support the results of spatial data activities and benefit from cooperation.

2.6 SDI components and nature

According to Rajabifard et al. (2003a), an SDI has five core components: access networks, policy, standards, data and people. In this context, an SDI is an infrastructure developed for organizations and various users to produce and use spatial datasets. To enable this infrastructure, proper access methods in addition to standards and policies to control and define a framework for any cooperation and collaboration among the organizations must be used. Figure 2.2, represents the relationship among the SDI components.

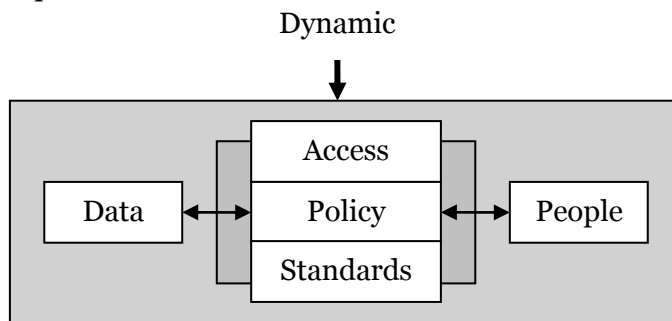


Figure 2.2. SDI Components (adopted from Rajabifard et al., 2002, p. 14).

SDI activities are implemented at various levels within the local, national, regional and global scales. Most activities in one level affect the other levels. Many levels of SDI are closely related and can influence other levels (Rajabifard et al., 2000a). As an example, the key parameters that build a regional SDI are dependent on the neighboring countries. Therefore, in these countries, there is a common connection and interaction for

exchanging knowledge and experiences from previous related work in spatial datasets.

During SDI formation, some negative relations may influence the implementation process. To overcome these problems, Spatial Hierarchy Reasoning (SHR) theory is used for SDI execution. This concept aims to divide a complex procedure and problem to several simple problems by fulfilling the connections among the problems to solve the total procedure. SDI hierarchy uses the same method to simplify problems and improve the current implementation of an SDI (Rajabifard et al., 2000b). Figure 2.3 illustrates the SDI hierarchy model, which is designed according to the internal relations among local, national, regional (multi-national) and global SDIs.

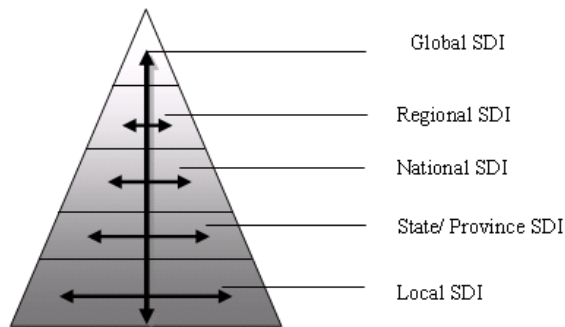


Figure 2.3 SDI Hierarchy, adopted from (Rajabifard et al., 2000b).

By using the SHR model in an SDI, all of the hierarchical specifications, such as Part-Whole, the Janus effect and community can be utilized (Rajabifard et al., 2000b). According to the part-whole characteristic, a high-level SDI (global level) contains other lower-level SDIs, such as regional SDI. In addition, a regional SDI is a whole for a region and a part for the global. The Janus effect for each element in SHR (e.g., National SDI) has two aspects. The first aspect is a view of the upper level (in this case, the high levels are the regional and global levels), and the second aspect is a view of the lower level (local and province levels), where such a relation is named a vertical relation among the SDI levels. This relation is represented with vertical two-way arrows in Figure 2.3. There is also a complex horizontal relation among the elements of the same level, such as organizational, management and political relations within an SDI, which are represented with horizontal two-way arrows in Figure 2.3.

Financial support is one of the other major challenges in SDI development. SDI funding requires broad research on documents and data analysis

according to different economic models. SDI funding also requires proper models for financial issues in addition to testing and evaluation of the model (Giff and Coleman, 2003). A number of basic issues must be discussed in SDI funding to understand the financial support for the SDI implementation (Giff and Coleman, 2002; Groot, 2001; Rhind, 2000). However, these models are beyond the aim of this thesis.

2.7 Clearinghouse networks

Considering the technical structure of an SDI, a gateway for a better data sharing interface among the data holders, users and different clients must be developed. Currently, this infrastructure plays a major role in data exchange in a spatial context due to the network access development and internet improvements.

Clearinghouse networks are established to facilitate efficient access to spatial data resources to decrease the cost caused by the duplicate collection of spatial data (Crompvoets et al., 2004; FGDC, 2000b; Rhind, 1999). The early generation of clearinghouses provided users with either information about the data, which was termed metadata, or a link to the data producer web site and hints for accessing the data (Philips et al., 1999). To search for a spatial data layer, a user sets search parameters, e.g., geographic boundary, data theme, and data layer name in the clearinghouse user interface. If any data exist according to the search criteria, data retrieval addresses were presented to the user as an output (Radwan, 2002).

The development of internet technology and the advancement of spatial web services set up a new generation of clearinghouses that are based on geoportals in addition to catalogue and spatial services (Bernard et al., 2005; FGDC, 2009; Bell, 2008). These clearinghouses provide users with standard and proper methods for searching and accessing required spatial data. In the next section, we describe the details of a clearinghouse.

2.7.1 Main components of clearinghouses

As mentioned above, the new generation of clearinghouse networks is based on geoportals as a gateway to access spatial data. A portal is a website that acts as a gateway or entrance point to the world wide web and

contains useful pages, hyperlinks, application links search engines, news and other services (Granić et al., 2011).

A portal in the spatial data infrastructure domain is called a *geoportal*. There is no definition of a geoportal in the ISO classification for spatial services, but a geoportal can be explained in the form of user interaction needed services (ISO/TC-211, 2009). One of the reasons for the recent development of SDIs is the creation of web-based metadata services, which are considered to be primary components of geoportals. These services are able to resolve user requirements to access the complementary information in spatial fields. The improvement in spatial services created web-based facilities and capabilities for better use of metadata systems and better solutions to find proper datasets (Tait, 2005).

Current clearinghouse networks have a number of basic components in addition to the geoportal (Figure 2.4). Catalogue services provide the functionality to publish metadata on spatial data resources and to search and query metadata (Bernard et al., 2005). A metadata repository is where information about spatial data is stored. Spatial services, which are connected to data servers, provide clients with the capability of viewing and/or downloading spatial data. Some of the major spatial services that are used in clearinghouse architecture consist of visualization services (Beaujardiere, 2006), download services (Vretanos, 2005), coverage services (Whiteside and Evans, 2008) and processing services (Schut, 2007). Registry services are where spatial services and catalogue services are registered to be discoverable by a geoportal.

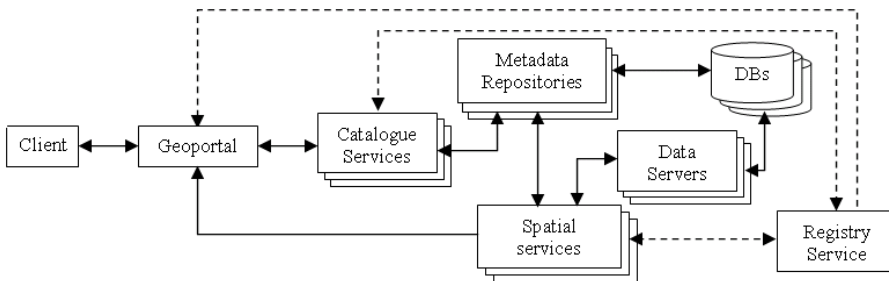


Figure 2.4. Architecture and elements of a clearinghouse (from Paper III).

As Figure 2.4 shows, in a clearinghouse, a user requests data via a geoportal and sets the search parameters. Subsequently, the geoportal searches metadata repositories through catalogue services to discover the required data if they are available. After finding the data, if the relevant data server(s) support (or provide) spatial services, such as visualization or download services, the data can be viewed or downloaded by the user.

2.7.2 Standards for gateways

The main standards utilized for current gateways in spatial data handling are currently specified by the Open Geospatial Consortium (OGC) in cooperation with the International Organization for Standardization (ISO). OGC offers several specifications for web services. In this thesis the WMS, WFS and WPS standards are used.

WMS: Web Map Service (WMS) is an OGC standard that enables a user to view a map from a remote server over the internet. Hence, this standard is used for view services. It is important to know how a map is defined in this context. OGC describes a map as “a portrayal of geographic information as a digital image file suitable for display on a computer screen”. Therefore, the output of a WMS is a visual representation of the spatial data in raster format (PNG, GIF or JPEG) or in vector based graphical formats, e.g., Scalable Vector Graphics (SVG) (Beaujardiere, 2006, pp vii).

There are three main operations in WMS which together arrange the entire rendering procedure. The *GetCapabilities* operation, delivers the metadata existing in the servers; The *GetMap* operation provides the user with the requested map. Finally, the *GetFeatureInfo* returns the specific information related to a selected points shown on a map.

Since the map is an image it cannot be used for further GIS-analyses. For a user that needs the spatial data, there are OGC standards for download services, which are Web Feature Service (WFS) for vector data and Web Coverage Service (WCS) for coverage (raster) data.

WFS: Web Feature Service (WFS) is an OGC standard that facilitates the distribution of geographic data in vector format throughout the internet. Data are distributed and encoded in the Geography Markup Language (GML). The WFS allows a client to retrieve and update spatial data encoded in GML from multiple Web Feature Services for which each server must be a WFS.

WFS contains three compulsory operations. *GetCapabilities* describes the capabilities of the existing dataset or its metadata, *DescribeFeatureType* expresses the structure of any feature type that the service can obtain. Finally, the *GetFeature* operation retrieves data upon request and fetches any relevant query results from the user (WFS, 2005). WFS also contains request for editing data, but these functions are not used in this thesis.

WPS: The OGC Web Processing Service (WPS) defines a standardized interface that facilitates the publishing of spatial processes and the discovery of and binding to those processes by users. Some examples of a

process are transformation services and analysis services. Publishing implies making machine readable information available, apart from the human readable metadata, which allows for service discovery and use (Schut, 2007; Bergenheim et al., 2009).

According to OGC specifications there are three mandatory operations in WPS procedure. Similar to the other spatial web services the first operation is *GetCapabilities* where it describes the capabilities of the server implementation as well as metadata documents; *DescribeProcess* returns the servers processes that the WPS can handle; and *Execute* which lets the client run a specific process that is implemented within WPS procedure applying relevant input parameters and getting appropriate outputs.

2.8 INSPIRE: A European SDI initiative

In this section, we present the current regional SDI for Europe. In 2002, the EU commission launched the Infrastructure for Spatial Information in Europe (INSPIRE) as the major activities related to the environmental and spatial datasets within Europe. The INSPIRE directive aims to create a European Union (EU) spatial data infrastructure (Directive, 2007). This infrastructure enables the sharing of environmental spatial information among public sector organizations and facilitates public access to spatial information across Europe (INSPIRE, 2009a).

A European spatial data infrastructure assists in policy-making across boundaries (Masser, 2010). Therefore, the spatial information considered under the directive is extensive and includes a great variety of technical themes, including not only environmental monitoring as the main purpose but also other disciplines, such as agriculture and transportation system. This activity addresses current technical standards, protocols, organizational threads and coordination, data policies such as how to access to datasets, and the production and maintenance of spatial information.

Using this infrastructure, environmental spatial datasets are available not only for the public citizens but also for governmental organizations. Also, spatial data can be shared among data producers and end users inside the European Union. This harmonized infrastructure helps to make better decisions within EU borders and attain more accurate environmental analyses.

The fundamental and common principles of INSPIRE are as follows (INSPIRE, 2009a):

- Data should be collected only once and kept where it can be maintained most effectively.
- It should be possible to combine seamless spatial data from different sources across Europe and share it among many users and applications.
- It should be possible for data collected at one level/scale to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes.
- Spatial data that are necessary for good governance at all levels should be readily and transparently available.
- It should be easy to find what spatial data are available, how the data can be used to meet a particular need, and under which conditions the data can be acquired and used.

To guarantee a common European SDI all member states are responsible for implementing their national SDI in accordance with INSPIRE. This infrastructure includes common regulations related to metadata (INSPIRE, 2009b), data attributes, characteristics and specifications, network services, sharing services and data, and finally control and reporting (INSPIRE, 2009a).

In this thesis we utilize INSPIRE as a best practice environment in Europe. The INSPIRE regulations from different perspectives assist in making a suitable framework for further development of SDIs. INSPIRE also acts as a proper basis for generating indicators for SDI assessment according to the regional and global levels.

3 SDI management and evaluation

3.1 Introduction

The development and usage of proper frameworks for the implementation, evaluation and continuous improvement of spatial data infrastructures (SDIs) are currently important research topics. A wide range of methods from different disciplines are used in this research topics (Grus et al., 2010; Georgiadou et al., 2005, Kok and van Loenen, 2005; Najar et al., 2006, Van Loenen, 2006; Luzet, 2004). In this respect, methods and techniques related to performance measurement and evaluation techniques from business management (BM) literature are not yet considered in many research areas (Paper I).

In this chapter, SDI evaluation is described. Current methods for evaluating in the SDI literature are reviewed.

3.2 Overview of SDI evaluation research

During the last decade, evaluation has been counted as one of the major challenges in the SDI field (Budhathoki and Nedovic-Budic, 2007) and researchers have suggested different models and approaches for the evaluation of SDIs (Harvey and Tulloch, 2006). There is a comprehensive literature review in Cromptvoets et al. (2008) with nine approaches. This research was proposed by a number of studies conducted such as the multi-view framework for evaluating SDIs developed by Grus et al. (2007). The researchers described the theoretical basics of the multi-view framework for SDIs assessment by expressing the need for a better understanding of the objectives, complexity, multi-faceted nature, dynamics and current uses of SDIs in the context of SDI evaluation in addition to the demands for SDI assessments and the necessity to develop a framework to evaluate

SDIs. According to these studies, SDI evaluation approaches are categorized into nine categories, which stress different dimensions of the SDI: the NSDI readiness index, clearinghouse suitability, the INSPIRE state of play, an organizational perspective, a framework based on land administration systems, SDI Performance Based Management, a legal approach, and SDI effectiveness from a user perspective.

The national SDI readiness index has been studied by Fernández et al. (2008) where they state it as an important factor to be taken into account for SDI implementation. In this regard, aside from the technological issues, organizational, informational, financial and human factors are composite integrators in the creation of the readiness index.

Another approach is the suitability of national clearinghouses which was calculated twice (2002 and 2007) based on seventeen characteristics. The result suitability index is defined as a measurement of the quality and performance of this electronic facility (Crompvoets et al., 2008). Sustainability index evaluation in different time stamps is a proper clearinghouse performance indicator that supports the managers in developing successful strategies in national clearinghouse implementation and enhances national clearinghouses and national SDIs in many countries.

Vandenbroucke et al. (2008) evaluated INSPIRE on the following six elements: organization, legal framework and funding mechanisms, spatial data, metadata, access and other services and standards, and characterization of the components of the European SDI and in particular the INSPIRE directive.

The organizational perspective approach was studied by van Loenen and van Rij (2008). The authors proposed a model that focuses on the classification of SDIs in the four stages of SDI development: stand-alone/initiation, exchange/standardization, intermediary and network. However, according to this model, SDI development has to be mature.

Steudler et al. (2008) assessed an SDI based on measuring indicators determined for five assessment areas: policy level, management level, operational level, other influencing factors and assessment of performance, which originally come from land administration systems. Such a model of comparison and evaluation provides better understanding of the various aspects, finds best practice for certain tasks of an SDI and improves the entire system.

SDI performance based management was described as a systematic approach by Giff (2008). This technique facilitates infrastructure

practitioners to operate an infrastructure to identify, analyze and manage its strengths and weaknesses. The technique uses indicators for performance improvement by developing a framework for key performance indicators within an ongoing process of establishing strategic performance objectives and measuring performance.

Another approach for SDI evaluation was the essential legal framework for developing an SDI developed by Janssen (2008). The assessment is not based on empirical evidence but primarily uses legislation, case law and jurisprudence. The assessment distinguishes three levels of legal assessment: compliance, coherence and quality. The final approach is the SDI effectiveness from a user perspective, which focuses on the effective use of SDIs by recognizing both the current and potential users and attempting to fulfill their needs regarding data and services by determining contextual factors and outcomes (Nedović-Budić et al., 2008).

Giff and Cromptvoets (2008) present a structured concept of SDI assessment. The authors use an in-depth analysis of performance indicators (PI) based on an eleven-step conceptual framework for designing PI for assessing SDIs in Canada. They also present and critically analyze a framework to guide SDI coordinators in the intricate task of designing PIs for their initiatives.

Another study conducted on evaluation strategy for SDIs is based on a maturity matrix (Van Loenen, 2006). The maturity of the SDI was evaluated according to several technical (e.g. data and metadata), non-technical (e.g. organizational) and policy (free data policy or cost recovery policy) measures. Based on this strategy, Van Loenen evaluated several SDIs in Europe and United States. Geudens et al. (2009) used a multi-criteria analysis to evaluate SDI policy strategies, which takes into account all the different criteria and actors involved in the complex SDI decision-making context in an integrated framework.

Despite several SDI assessment approaches, there is still lack of an integrated method which covers different aspects of an SDI assessment that can measure the progress of an SDI. Moreover, a limited number of assessment approaches are able to demonstrate whether SDIs realize the intended goals (Grus et al., 2010). Therefore, a comprehensive goal-oriented SDI evaluation should be based on the evaluation of several dimensions in a common framework, which must be defined within a long-term project to control the progress of an SDI. In this thesis, two studies are presented based on current techniques and methodologies from business management literature. The first study is an integrated framework for the implementation and continuous improvement of SDIs, which is

utilized based on Six Sigma, ABC (Activity Based Costing), BSC (Balanced Scorecard) and TQM (Total Quality Management). This study describes the business management techniques and provides an integrated framework, based on these techniques, for the implementation and continuous improvement of SDIs (Paper I). The second study is an evaluation of the SDI Balanced Scorecard (BSC) including a case study in accordance with the INSPIRE directive. A general framework for the evaluation of the Swedish NSDI according to the INSPIRE directive is depicted. The case study demonstrates that BSC is applicable for Swedish NSDI evaluation, and the results can be used by other SDIs (Paper II).

4 Spatial web service composition

4.1 Introduction

The Open Geospatial Consortium (OGC) has published a number of standards to provide the end user with spatial data, maps or spatial processes. Other types of standards exist to publish spatial processes and the discovery of and binding to those processes by users. However, with the various combined web service applications developed in recent years, there has been a demand for combining these services, and the demand for increasingly complex functions has indicated the limitations of the single operations of services. Thus, a single web service cannot handle the user requests, and a combination of services is necessary to fulfill the user requirements.

The need for combining specific functions was inevitable due to the new requirements and to the need for a chain of services for executing more complex processes. According to the definition of ISO (2005), Service chaining (composition) is referred to as a set of dependent, combined services to achieve larger tasks. Service composition is considered one of the benefits of SDIs through combining simple spatial data services generating value-added service chains (Einspanier et al., 2003). This procedure implements a workflow of automatic business processes by applying within a part of or an entire action, according to a set of rules that sequentially pass the tasks from one process to another (ISO, 2005).

Two types of general patterns for service chaining are identified by Friis-Christensen et al. (2009): *centralized* and *cascaded*. The centralized pattern contains a central component that controls the requests for the services used. Therefore, the execution workflow is controlled by a central component.

In the cascaded method, service chaining is controlled with a backward approach. In this structure, the end result service is called directly.

Subsequently, the invoked service handles the other service(s) required to retrieve the input processes/ parameter. The procedure continues in the same manner until it reaches the last service in the chain.

In this chapter, we discuss three main components of service composition, which is followed by an example of their application in this thesis (Paper III). First, a number of web processing service (WPS) applications are discussed because they play an important role in a service composition. Next, service composition usages are described in more complex applications for which a single WPS cannot handle the entire procedure. Finally, semantic web services are discussed as another component linked to the service composition chain. Any of these three components can be used for increasing the accessibility of spatial data, as discussed in the following section.

4.2 Web processing service applications

Web processing service has been broadly utilized in various applications to facilitate any type of processes that publishes spatial data. Different applications make use of a WPS for analyzing spatial data through the web. In this section, we describe several of the applications that use a single WPS. Bergenheim et al. (2009) used WPS for on line generalization. The authors implemented a real-time generalization service, so called WPS PHP Server, to dynamically generalize roads. This web service is based on an existing GIS platform (GRASS). The result shows that such a WPS-based interface is not only useful in allowing remote access to spatial data, but it is also an appropriate solution for GIS processing, specifically under limitations such as computing power for field work data collection.

From a specification viewpoint, Walenciak and Zipf (2010) proposed a WPS application profile for spatial analysis in business marketing. The authors described methods to enhance the current specification regarding the application profiles. Additionally the authors presented a specific application domain to be examined for use in spatial analysis and in the transferring of results to an application scheme.

Researchers also proposed applications for WPS in 3D processing analysis (Lanig and Zipf, 2010). They The authors classified and defined specific functions related to 3D data. In this regard, they the authors represented the domain- specific WPS application profiles for 3D city models and identified some several applications for 3D processing operations, such as disaster management.

There are more complex applications that a single WPS is not able to handle, and a combination of processing services is required for such applications. In the next section, we discuss the application of service composition as a solution for more complex usage of WPSs and other web services.

4.3 Applications of service composition

A number of studies on chaining spatial web services have been conducted in recent years. Granell et al. (2005) proposed a methodology based on the abstract description of services and workflows. The method contains three processes, which are the service abstraction process, service composition process and translation process. These processes share two aspects as an integrated component: a composition of complex services and a set of workflow patterns. The authors propose a novel model for implementing the steps of a suggested methodology using an efficient technique for developing service compositions.

Later, Lemmens et al. (2006) used the same method for ontology-based service composition approaches. The authors stated the integration of different spatial services from various resources was one of the primary challenges in geo community due to the complexity and heterogeneity of spatial data. They also identified a GIS workflow approach to use semantic and syntactic service descriptions inside a service chain. By using these types of service chains expert users can link geoservices remotely to develop complex services and the analysis of spatial data. Such methods also simplify XML based description languages to build a service reuse architecture based on ontology and service descriptions. In a follow up study (Lemmens et al., 2007) they introduced a deep service description as a semantic and syntactic service description in service chaining by combining two prototypes. One prototype addressed geoservice discovery abstract composition, and another prototype supported concrete composition and the execution of geoservices.

Kiehle et al. (2007) considered SDI an environment to exchange spatial data among organizations, in which the spatial web services play a major role. The authors address the problems of service chaining by providing a system architecture to implement complex geo processing models and workflows based on web services using web service orchestration. They also proposed methods based on geo community standards to establish a generic web service architecture to be used in all SDIs. Zhang et al. (2008)

also proposed an OGC-standard-based spatial data service chaining process.

All of the aforementioned studies indicate the complexity of applications and that service chaining can be an appropriate solution to exchange spatial data within the SDI community. In this regard, one type of web services addresses the definition and description of the web services that substantially influence for new developments in web services. In the next section, we describe semantic web services as a prominent concept for SDIs.

4.4 Semantic web services

Semantic web services (SWS) include specific description of properties and capabilities in a computer interpretable way and consequently provide interoperability between them (McIlraith and Zeng, 2001). Semantic web services are software components that can be re-used and are self-contained, independently. The services can fulfill any demand and/or combine with other web services for more complex processes. SWSs have a modular structure and can be published, located, or called through the web.

Many improvements in the semantic web services field have been occurred in the last decade. Lemmans (2006) specified four different approaches in SWSs as the OWL-Services (Martin et al., 2004), the Semantic Web Services Framework (W3C, 2005), the Internet Reasoning Service (Motta et al., 2005) and the Meteor-S (Patil et al., 2005). Ermolayev et al. (2004) presented an agent enabled framework for semantic web service composition. The authors introduced their methodology based on the semantic web as an agent capability containing proper ontological description. In this research they proposed a method to compose web services by the dynamic composition of agents, which perform any collaborative task that a service requests. They initiated a middle agent layer to conduct service request to task transformation, agent-enabled cooperative task decomposition and performance. Later, Kumar and Mishra (2008) employed the same style framework for more untouched issues and utilized cognitive parameters and quality of service (QoS) parameters in service provider selection. They used this method in education planning and admission-process for higher-education.

4.5 Methods to increase accessibility

Considering the current problems associated with clearinghouse networks, we propose a method to increase the accessibility of data in clearinghouse networks based on web processing services, service composition, semantic web services and expert systems. By utilizing this methodology, the users can have improved access to the available spatial data resources by offering similar semantic matching datasets when the data may be found under other synonyms in other disciplines. The users can also have more successful searches in a spatial data clearinghouse because candidate data are retrieved when the requested data cannot be found. Finally, users can easily access the requested spatial data, through automatic arrangements for data processing that are available to produce the data requested.

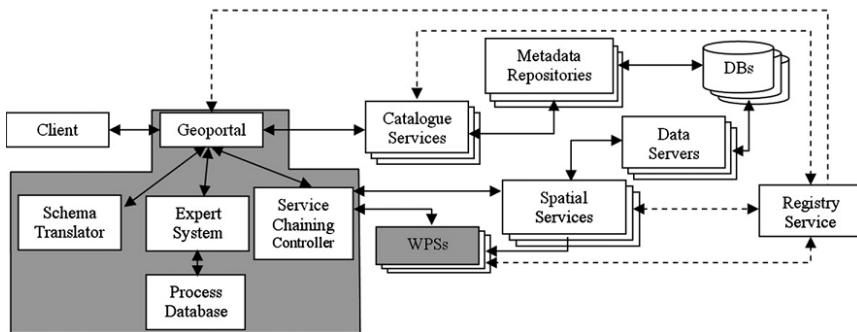


Figure4.2. General structure of an expert clearinghouse.

Figure 4.1 shows the system architecture of an expert-based clearinghouse. In this system, beside the typical clearinghouse network structure, a number of additional components are incorporated within the system :

- Schema translator: To manage the retrieval of synonyms for the search phrase and send the translated expression to the geoportal.
- Expert system: To find and define an instruction for processing the candidate data layers and the best combination to generate required data.
- Process database: To organize features of each process, including the function of each process and its inputs and outputs.
- Service chaining controller: To produce the required data using a chain of different web services. The controller manages the workflow of the chaining.

- Web processing services: To implement the instructions of expert search engines, which consist of retrieving output from the expert search engine, subsequently accessing the data servers to get data layers, and finally conducting desired processes on the data layers to generate the required data.

An expert clearinghouse follows a workflow. The user connects to a geoportal to search for the required data and sets the search parameters. The geoportal searches in the metadata repositories in catalogue services. If the required data are not discovered, the geoportal connects to schema translator to determine synonyms of the required data layer and subsequently searches through its synonym phrases (semantic matching). If nothing is found, the geoportal connects to the expert system to identify candidate data layers in the region. To do this, the geoportal passes the search results to the expert system to determine a proper combination of candidate data layers, among existing layers. The expert system also determines the required processes using the process database. The results are then sent to the geoportal. The geoportal searches for the processing services, which offer the required processes. The service chaining controller sends data processing requests to the proper processing services. Finally, the processing services retrieve the data through data services and process them to generate the required data.

For a better understanding of the methodology, we provide an example showing how to increase the accessibility of spatial data using the methods described above. In this example, two scenarios are described in which a number of services are utilized to improve the data exchange within clearinghouse networks. The first scenario is that the user searches for rain contours, but there are other synonyms for the required data in the repositories, such as rainfall data, synoptic station data and precipitation contours.

Another scenario is that the data are available but do not cover the entire extent of the region and consequently, transformation and preprocessing must be performed to generate the required data (e.g. the available data are DEM, contours and slope for certain parts of the region but not for the entire area). The more complex case is of course a combination of these two scenarios. Figure 4.2 illustrates several of the mentioned scenarios as existing challenges to the access of required data.

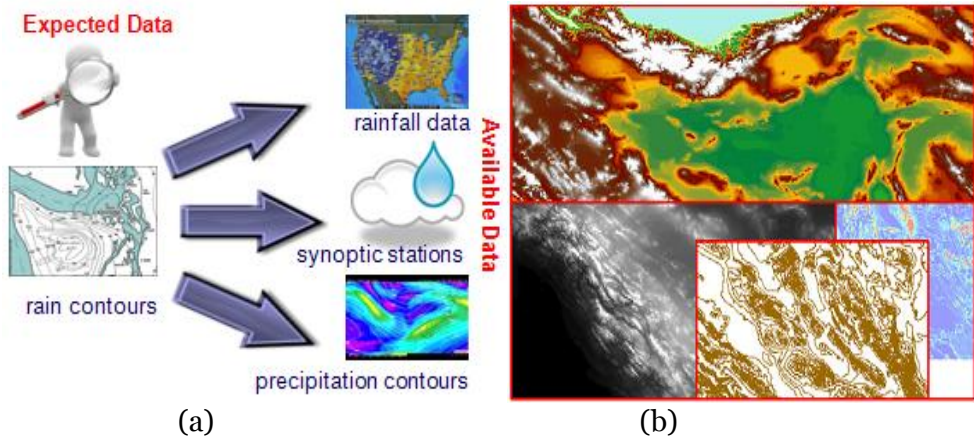


Figure 4.2. Several challenges in current clearinghouses. (a) First scenario: the user searches for rain contours, but the available data are rainfall data, synoptic station data and precipitation contours which are synonyms of the required data. (b) Second scenario: the candidate data are available for parts of the selected region, necessitating the transformation and preprocessing of data (e.g. the available data are DEM, contours and slope for some parts of the region but not for the entire area).

5 Cartographic aspects of SDI

5.1 Fundamentals of cartography

It is difficult to obtain a complete and general definition of cartography. However, according to the International Cartographic Association (ICA) cartography is “the art, science and technology of making maps together with their study as scientific documents and works of art” (ICA, 1973 p. 1). This definition implies that using certain techniques and following specific rules is not sufficient to create a “good” map. Artistry also plays an important role in cartography (Keates, 1996).

Currently, maps are specific and focus on important information visualization. Maps are application-oriented and the main purpose of a map changes the type of data and the symbol design. In a great deal of modern maps, the application-oriented, user-demanded information (e.g., navigation instructions) is emphasized, whereas less dominant data (e.g., base map) must be blurred as a base map only to convey a general perception of the region. Meanwhile, unnecessary information must not be included. In general, cartography addresses the process of selecting the essential data to be shown and symbolizing those data within a map.

From an SDI perspective cartography is important, for example for view services in geoportals. Consequently, it is vital to briefly describe the primary cartographic aspects and address the problems existing in the SDI field. In this chapter, we describe several basic components of cartography followed by a number of cartographic challenges from an SDI perspective.

5.2 Map design and symbolization

Map design refers to the layout of a map in which elements, such as the title and legend, are added (Robinson et al., 1995). Aside from the standard map elements such as the scale bar, north arrow and other basic

components that each influence the visual appearance, the symbolization of data layers affects the visual output.

Symbolization refers to the design procedure for symbols and the text used to visualize the spatial data. Symbolization is not an isolated process; creating a map is an iterative process that is repeated until the cartographer is satisfied with the map.

For a better understanding of the concept of map design and symbolization, we describe several of terms in the cartographic literature. Graphic elements and visual variables play a major role in creating maps. Depending on the scale and type of map, a feature may be represented in different ways. Consequently, a brief explanation on the basic components is useful for further discussions.

5.2.1 Visual variables

Four different types of visual variables make the graphic element of maps prominent. According to Robinson et al. (1995), these variations are called the primary variables. Several of the variables refer to object shape, but several objects have color variations:

- Orientation refers to the direction of elements, e.g., lines and elongated symbols.
- Size refers to the dimension (e.g., length, height, width) of a symbol.
- Shape refers to the form of a symbol and may be figurative or geometric.
- Color is used with different variations to invoke differences and similarities, end emphasize or deemphasize information. The three dimensions of color are (Dent, 1999):
 - color hue
 - color value
 - color saturation

Figure 5.1 illustrates the primary visual variables mentioned above.














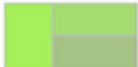
	Point	Line	Area
Orientation		NA	NA
Size			NA
Shape			NA
Color - Hue			
Color- Value			
Color - Saturation			

Figure 5.1. The primary visual variables according to Robinson et al. (1995).

Another concept for improving visualization through secondary visual variables exists, which is a type of pattern created by repeating graphic elements (Robinson et al., 1995). Patterns are varied by adjusting the arrangement, texture, and orientation (Figure 5.2).

Arrangement is used to create patterns by shaping and configuring the elements. This type of pattern may be random or systematic.

Texture is creating patterns by the resizing and the spacing of elements. Small spacing and small elements generate fine texture (e.g., thin lines).

Orientation refers to the directional arrangements e.g., of rows or points in-line.

Figure 5.2 shows the secondary visual variables with some examples. A combination of graphic elements and various visual variables is used to design any type of symbol.

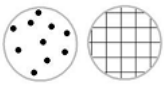

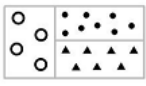
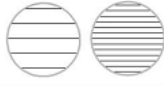





	Point	Line	Area
Arrangement			
Texture			
Orientation			

Figure 5.2. The secondary visual variables according to Robinson et al. (1995).

5.2.2 Visual hierarchy

Visual hierarchy is an important aspect of visualization in a map. Visual hierarchy is a graphical representation of the intellectual hierarchy, in which the symbols and map elements are ranked according to their relative importance (Slocum et al. 2005:220). The concept concerns emphasizing symbols that are more important and deemphasizing insignificant and base information. A proper visual hierarchy focuses the eye of the user first to the most important element and subsequently to the rest of the map elements.

Expert cartographers have the knowledge of well-designed map production. However, this knowledge is not transferred to automatic map-making programs. The situation is especially problematic when different data sources integrated in a web map services each have special visual characteristics. Recently, studies have been conducted to enhance web cartography and semantic issues, such as those of Bucher et al. (2008), Iosifescu-Enescu et al. (2009) and Chesneau et al. (2005).

5.2.3 Visual priority

The concept of visual hierarchy is an extensive term used in cartographic literature. However, due to user demands for specific applications there is the need to define new terminologies for layer priorities. In some cases, the user has to place more emphasis on certain layers and only view other layers as base map. Also, the user is requested to denote in which level in the visual hierarchy layer should be.

The following levels are defined as visual priorities (Paper IV):

- Foreground – Additional information layers that are of high relevance for the application.
- Middle ground – Data layers in the base map that are essential for the application.
- Background – Less prominent data layers in the base map.

The order of the layers in the final map follows certain rules that are derived from particular basic cartographic instructions:

- A layer in the foreground is always on top of a layer in the middle ground, and a layer in the middle ground is always on top of a layer in the background.
- Within each level (back-, middle- and foreground) point layers are on top, line layers are in between and polygon layers are at the bottom.

5.2.4 Symbolization in web cartography

Web cartography is the procedure of designing maps published on the internet. In these situations, the maps are called screen maps. In many respects, screen maps are similar to maps printed on paper, and most of the principles described above should be followed. Most traditional cartographic rules are also applicable in the web environment but a few components differ:

Point symbols: Pictorial symbols are more common for internet maps because they attract less experienced map readers (Van den Worm, 2001). However, maps may be viewed on screens with limited resolution, so complex symbols should be avoided.

Line symbols: For line symbols, the possible limited resolution of screens must be considered. Lines should be wide, and visual variables such as orientation and texture are less suitable (Van den Worm, 2001).

Area symbols: According to Van den Worm (2001) many web design programs suggest tools to design complex area symbols but due to many reasons and limitations, the file size should be kept small.

5.3 OGC standards for web cartography

In our studies of cartographic solutions for SDI, we rely on OGC standards. The two most important standards in this area are the Styled Layer Descriptor and Symbology Encoding.

5.3.1 Styled Layer Descriptor - SLD

The SLD implementation specification standardizes the process of defining feature symbolization and data coverage, which is an important standard for cartography (Müller and MacGill 2005). SLD is an XML-based description language for extending web services such as Web Map Services (WMS) and Web Feature Services (WFS). Several of the key specifications of SLD are the structuring of the style attributes and the understandability for computers and for users. Each layer is symbolized with user-defined styles.

The appearance of any map in a web map service is defined with styled layers. In this regard, every layer, depending on the design conditions has one or many styles in the case that a map contains a number of layers (SLD, 2007). More specifically, every layer can be a transparent layer, and all of the features can be styled in a selected form. Consequently, by applying an SLD for a map, each layer may have a specific graphical representation and style, which enhances the map legibility and readability.

There are three options in using the current OGC WMS standard for map styling. In the first method, cartography and styling is decided by the service provider. The second option is that the user selects a style from a number of predefined symbolization styles. Finally, in the third option, the user defines the symbolization using an SLD.

5.3.2 Symbology Encoding - SE

The SE specification is the direct follow-up to SLD (Müller, 2006). SE is the most recent OGC standard for the portrayal of spatial data and is a language that describes how a style is rendered. SE is used in conjunction with SLD in a WMS-service. SLD dictates which styles to use, and SE describes how the styles are portrayed.

5.4 Cartographic challenges in view services

5.4.1 Overview and related works

A view service is considered one of the major components in an SDI. The service facilitates the preliminary view and query access in any kind of spatial data exchange. The view service is the first step for user interaction with the available geographic data through the SDI before getting any data, which may be costly and time consuming.

A view service often uses data integration from several data sources. Figure 5.3 shows the general setup of a user request from different data sources. In this structure, the user requests a set of spatial data/products in a client (browser) through a view service. The spatial data are selected from a number of web sources (basic services). Finally, the response is produced in the form of a vector or raster graphic format.

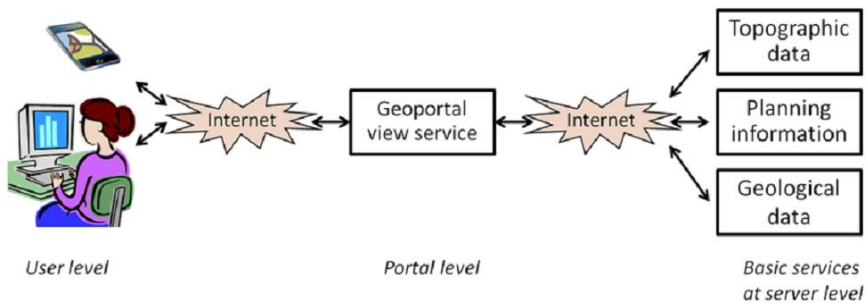


Figure 5.3. The general design for user request from distributed data using a geoportal view service (Harrie et al., 2011, p.93).

Spatial data integration often causes problems in the visual representation because spatial data layers are not adopted for co-visualization. Complex visualization requirements existing in different applications affect the output. This has been studied by several authors including Iosifescu-Enescu et al. (2009). The authors utilized an enriched cartographical approach for OGC standards to fulfill the complexities that stem from environmental management. In this regard, the authors used cartographic extensions to express cartographic rules with spatial operators and advanced-feature filtering for layer masking, flexible point symbolization, and patterns and gradients for all of the spatial features.

Brewer and Battenfield (2007) provide methods that can be used to create a map from a multiple-representation database. They emphasize map display changes using symbol design or symbol modification. In addition, the study comprises a demonstration of the establishment of the specific map display scales at which symbol modification should be imposed.

According to Harrie et al. (2011), there are five main issues for geoportal view services: semantic heterogeneities, geometric heterogeneities, diversity of the level of details, the inefficiency of labels, and the inefficiency of symbols. In this thesis, we primarily stress the inefficiency of symbols and geometric/semantic heterogeneities which are important in the two cartographic problems described below concerning overlay and integration.

5.4.2 Overlay problem

In map visualization of data from several sources occasionally, a portion of the data is hidden due to the spatial data overlay. The problem often occurs when data in the foreground cover the information in the background or middle ground. Figure 5.4 presents an example of the problem. In this figure, the areas selected show the existing problems for a proper visualization. The primary problem here is that the background layers are not shown due to the specific type of visual variables in the foreground.

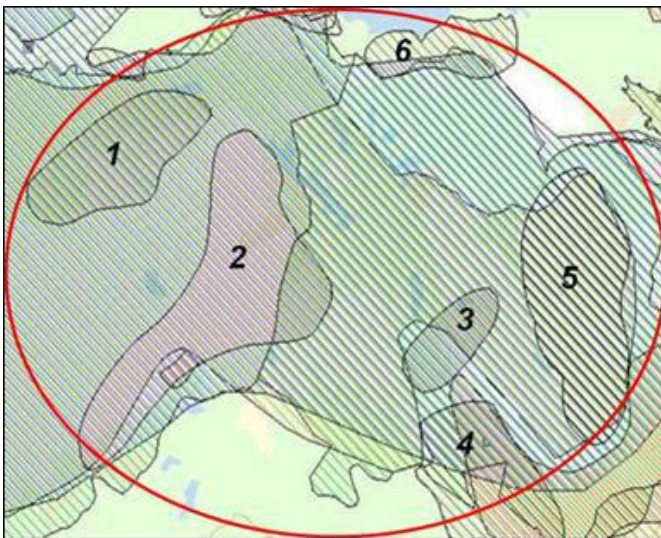


Figure 5.4. Different numbers indicate the data from different sources, for which the layer overlay has caused unnecessary coverage for the background and middle ground layers (taken from Soderman et al., 2011).

To solve the problem, alternative methods, such as using other visual variables instead of patterns, can be utilized, but the output is still poor. Transparency is another solution, which partly represents the layers but, this method makes the situation more complex. Figure 5.5 shows an alternative solution for the problem. Although such intermediate solutions solve the problem to some extent, the final result does not satisfy the needs of the user. Paper IV proposes a new method for solving overlay problems.

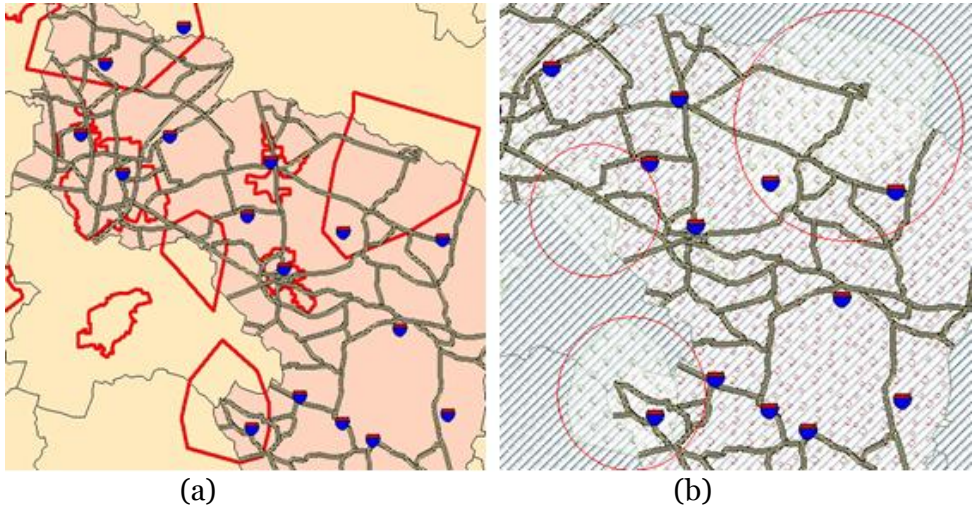


Figure 5.5. Alternative intermediate solutions for problems in map visualization (a) Using the borders and (b) transparency.

5.4.3 Problem of geometry integration

Another cartographic challenge is the conflicts in data shown in view services. The conflict concerns both geometrical and semantic inconsistencies. Geometrical inconsistency is due to the artificial discrepancy created when, displaying connected information of an extent from different sources, which is caused by inaccuracies and different levels of details in the datasets. The semantic conflict is related to the definition of the related features, i.e., to the information models of the datasets. Two examples of these conflicts are as follows: a sea layer overpassing (violating) land, for which according to the definition, there is a shoreline dividing these two features; a river overlaid on a lake, for which semantically, a river cannot be on top of a lake (the water cannot be

separated in two features). In both instances, these inconsistencies generate conflicts in the maps that affect the legibility.

A view service often requires the integration of data from several data sources. In some cases, according to the nature and semantics of the datasets, the data layers should be disjointed and not overlaid. Figure 5.6 represents existing challenges in an application of an extent; the administrative boundary and sea shore are not fully overlaid due to geometric inhomogenities.

Paper V describes these challenges and proposes methods to improve the cartography of these maps in view services. To overcome these conflicts, and to fix the topological and geometric conflicts we use a semantic-based expert system.

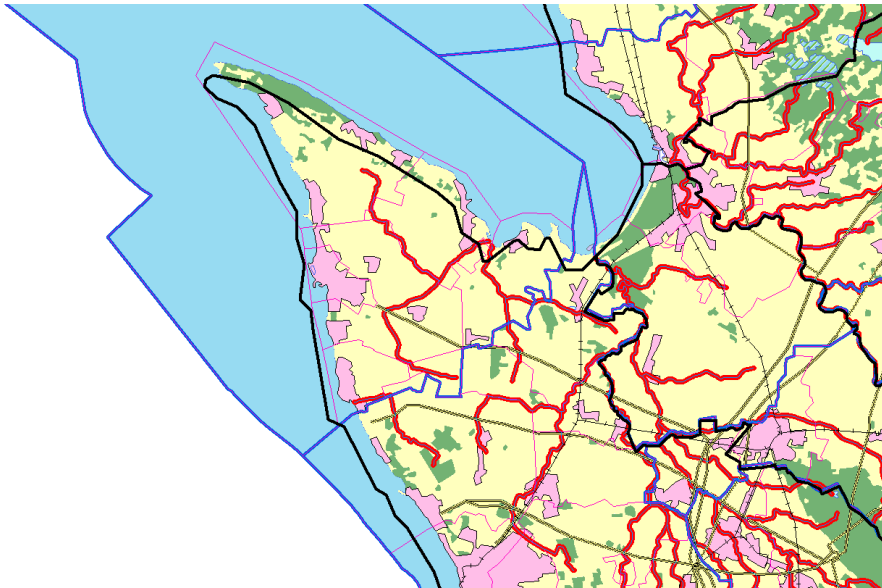


Figure 5.6. The data integration challenge from the semantic and geometric viewpoints.

6 Summary of papers

6.1 Paper I

An integration of business management concepts for the SDI implementation

The study aims to review current methods within the business management literature, which triggers to an integrated framework for the implementation and evaluation of SDIs. The applicability of each technique is described, and the usability of each technique in several aspects of SDI implementation is discussed.

In this paper, we reviewed four methods: Six Sigma (Folaron, 2003), Activity Based Costing (ABC) (Cooper and Kaplan, 1991), Balanced Scorecard (BSC) (Kaplan and Norton, 2000; Kaplan and Norton, 1996) and Total Quality Management (TQM) (Sashkin and Kiser, 1993).

The paper proposes the advantages and disadvantages of the use of BM techniques in SDI implementation. SDI is a collaborative development where various organizations and institutions are involved; therefore, teamwork and joint activities are important in the achievement of various SDI objectives. Six Sigma, as a core methodology in the integrated framework, facilitates team building and teamwork in addition to creating a collaborative environment, which is one of the main requirements of SDI development. Moreover, Six Sigma simplifies the spatial data production and updating procedures, inter-and intra-organizational data sharing, managing databases and web services, which are several examples of existing challenges within SDI implementation.

Several weak points of Six Sigma must be considered, such as the need for quality data for the measurement and prioritization of projects. However, these weaknesses are common for most evaluation and improvement methods, and because SDI implementation has a clear priority for major activities, it is not vital to use Six Sigma for SDI implementation and continuous improvement.

SDI funding is a complex task due to the diverse activities required for SDI implementation. In this respect, a proper financial framework is necessary for calculating the costs associated with each activity and relevant overheads. The framework should also monitor SDI funding for each activity, based on the mentioned estimations. ABC can satisfy these requirements. This framework is also well integrated with other continuous improvement techniques.

The weak points of ABC are related to the cost and time. In general, most monitoring and evaluation approaches have the same limitation related to the time. In addition, the ABC implementation cost is a small percentage of the financial resources comparing to the total costs required for SDI implementation.

SDI development has a complex and multi-dimensional nature, and its evaluation and monitoring must be based on a multi-perspective framework. This framework must link financial and non-financial indicators, internal and external aspects, and performance drivers and outcomes. BSC not only has the advantage of linking these factors; it can highlight inevitable trade-offs among them. Therefore, BSC can be a proper framework for the implementation and evaluation of SDIs.

In most organizations, financial measures have a higher priority than other indicators, and this issue can be counted as a weakness of BSC. However, in SDI, one of the primary goals is to benefit the society. Therefore, the non-financial benefits of spatial data use in decision making and planning are also in valuable. SDI development aims to promote society and better life for citizens.

Finally, TQM supports and encourages effective participation by involving employees in decision making process for the development of SDIs. In addition, TQM improves the quality of their work environment and provides users with a sense of value and purpose. Similar to the other methods, TQM is also a long-term procedure, and implementation of TQM requires much time and effort, which can be stated as one of the weaknesses of this method.

6.1.1 Paper contribution

The contribution of this paper is to define a structured framework for SDI implementation based on specific methods in the context of an SDI. The paper proposes Six Sigma as a core methodology. For implementing an SDI, the DMAIC (Define–Measure–Analyze–Improve–Control) approach

can be used, ABC (Activity Based Costing) can be used for the economic management of SDI, BSCs (Balanced Scorecards) can be used for monitoring the progress of an SDI and TQM (Total Quality Management) can be used for the quality management of the entire procedure of SDI implementation.

6.2 Paper II

Using BSC for Evaluation of SDI: A Case Study for INSPIRE

The aim of this study is to use the Balanced Scorecard (BSC) method, which is described in Paper I, for the development of frameworks to monitor and evaluate SDIs. According to the BSC description, the concept has a variety of elements that need to be adopted in the research. In this framework, the BSC perspectives are adapted in accordance with SDIs:

Learning & Growth: This perspective measures capacity building involved in SDI implementation at the individual (people) level.

Internal Process: This perspective evaluates internal processes for implementing SDIs. Standardization activities, data management affairs, establishing accessing networks and spatial web services, institutional arrangements and collaborative activities are some examples of the objectives to be measured.

Customer: The customer is a key factor in SDI evaluation. Investments and technological developments within an SDI ideally deliver spatial data products to the user. Therefore, customer satisfaction is an important factor to be measured in SDI evaluation (Albert, 2002; Band, 2000; Fornell, 1992; Hackl et al., 2000).

Benefit and economy: The main target of an SDI is to benefit various sections of society. Meanwhile, the economic perspective also keeps SDI active and updated according to financial challenges.

In addition to the four perspectives mentioned above, the BSC model requires other elements for the BSC framework for SDI evaluation:

Objective and description: Objectives are derived from the strategic plan and vision of an SDI. In this case study, the objectives are taken from the INSPIRE directive, the Swedish National Geodata Strategy and the general SDI goals to obtain a broad SDI evaluation framework. For each objective, several descriptions are also offered to highlight different aspects of the objective for the measurement.

Cause and effect linkage: The cause and effect linkage describes the cause and effect relationships between the objectives. The linkage initiates from the Learning and growth perspective, in which the skilled staff and managers are well-aware of SDI support and internal processes for implementing SDIs. Subsequently, proper internal processes for data management and sharing apply to a wide range of data usage and analysis by customers. Regarding user satisfaction with the data and services of an SDI, managers and decision makers use SDIs for better decision-making and planning, which results in social benefits and economic success.

Measures: Measures are quantifiable values to calculate the progress of tasks in any objectives. In this study, INSPIRE indicators and other SDI evaluation researches are used to design measures.

Targets: A target is defined as a quantifiable goal for any measure and is set during SDI strategic planning. Ideally, a combination of all the targets illustrates the general goal of an organization.

Initiatives: Initiatives are midterm programs to facilitate progression of the strategic plan. Initiatives must be defined when a strategic plan is developed. The INSPIRE directive can be considered one of the initiatives for a national SDI.

Based on the aforementioned BSC structural elements, a comprehensive model is proposed for SDI evaluation. In the data collection step, some indicators were collected directly, whereas others required extra calculation. In some cases, there were limitations and changes to the original indicators, especially if there was no obvious method for establishing a target value. Subsequently, the selected datasets are integrated into the related BSC software, and the possible results, charts and cause and effect linkages are produced as the primary output for the BSC model for SDI implementation. Using a Balanced Scorecard framework to evaluate the progress of an SDI, a clear pattern emerges from the existing situation. This pattern can be used as feedback for SDI coordinators to define strategies and set objectives, goals and visions. The adapted method provides an appropriate overview of the status of the various success factors that must be met for coordination to be successful and contribute to the development of the Swedish NSDI.

6.2.1 Paper contribution

BSC helps to evaluate SDIs from both the data producers and the users (customers) point of views. Using BSC, a general and flexible SDI

evaluation framework can be established for SDI activities. This framework considers both individual (learning and growth) and organizational levels (internal process). Finally, financial affairs and benefit achievements (benefits and economy) are an essential component of the evaluation.

The contribution of this paper is to propose a comprehensive method for SDI evaluation based on a structured business management framework for SDI implementation. As an outcome of this study a variant of this approach is operational at the Swedish national mapping agency (Lantmäteriet).

6.3 Paper III

Expert system to support functionality of clearinghouse services

The aim of this study is to use different technologies and methods to increase the functionality of clearinghouse services as a gateway to share data. Spatial data clearinghouses are considered a major component of a spatial data infrastructure (SDI). Yet, different studies indicate that national clearinghouses are not yet 100% efficient and do not function well, because the existing spatial data resources are not satisfactorily accessed or used in an optimal way. For the more efficient use of a clearinghouse, we propose an extended version of a clearinghouse, together with expert systems and semantic matching methods. The expert system aims to facilitate the identification of available data sets automatically and convert the available data to the required data based on the needs of the user. A schema translator is also used to find similar data that may be used in other disciplines or other datasets by semantic matching. We have developed a method of identifying available data and methods for data conversion according to the needs of the user. The methodology is implemented using standardized map services. In practical assess we introduce two scenarios to test, the methodology and demonstrate how an extended clearinghouse can significantly increase user satisfaction regarding accessing available data according to the requirements.

6.3.1 Paper contribution

The functionality of clearinghouses is important for a well-functioning spatial data infrastructure. This paper proposes the use of expert systems to enhance the functionality of clearinghouses. The expert system provides

the possibility to convert data in current form to a form that is sought by the user.

The practical implementation and testing of a prototype system shows that an expert spatial data clearinghouse, with the capability of identifying candidate data layers and processing them to generate users required data produces a number of benefits:

- Provides for a better use of available spatial data resources.
- Increases the number of successful searches in a spatial data clearinghouse, by suggesting candidate data to users, when the required data are not found,
- Facilitate the access of users to their required spatial data by the automatic arrangement of the processing available data to produce the required data.

The contribution of this paper is the use of an expert system for the improvement of clearinghouses which is an important step forward in building future SDIs.

6.4 Paper IV

Automatic symbolization methods for geoportals

The general aim of this study is to improve the visualization of data in an SDI environment. More specifically the study aims at improving the visualization of datasets in a view service and solving the overlay problem described in chapter 5. Visualization is often problematic when the final map contains data from different sources. In this paper, we propose the concept of layer priorities as fore-, middle-, or background and two methods to enhance the symbolization: the polygon overlay and color saturation methods.

There are different approaches to handle cartography in a view service of a geoportal. First, there are predefined symbologies available for all layers, where the end user is limited to symbologies without any changes. Second, the end-user is allowed to set the relative importance of each layer (e.g. if a layer is placed in the background, middle ground or foreground in the visual hierarchy) to provide him/ her with more capabilities. Third, there is the possibility to define several symbologies on the geoportal, and the end-user can choose these between different symbologies. The fourth approach is to allow the end-user create his own symbology for each data layer. This

paper uses the second approach, in which three priorities are defined for the user.

In this study, we implemented a prototype system of a cartographic enhanced geoportal. This implementation consists of a client, a cartographic enhanced geoportal and external web services within the java programming language in an eclipse environment. The communication between the components follows the OGC WMS standard, but because a user needs to define the visual hierarchy an additional parameter is added to the *GetMap* request. Our prototype system contains two methods: *polygon overlay* and *color saturation*. In the polygon overlay method, we utilized an icon placement approach using the displacement, distribution and removal cost functions to calculate the total cost for a random symbolization. In the second method, we decrease the color saturation of the unimportant layers to emphasize the foreground layer, according to the visual hierarchy concept. The optimization of the cost function is based on a simulated annealing approach.

By implementing the system architecture and applying the methods to different scenarios, the results show that these methods are appropriate techniques to visualize overlaying layers without data loss.

The results from two case studies show that the methods can satisfy end user requirements. We believe that these types of methods will be increasingly important to improve the cartographic quality of future view services in geoportals.

6.4.1 Paper contribution

The contribution of this paper is to improve the presentation of spatial data from a technical viewpoint in the context of SDI and data sharing in the web. In this context, a wide area of web based applications has initiated the requirement to disseminate spatial data to end users by the use of geoportals. An advantage of the proposed techniques is the possibility to overlay geospatial data layers from different sources with new symbolization methods that support visual integration. One important issue is the visual hierarchy that ranks various data according to their relative importance; that is, data layers that are more important for the application should be visually emphasized. Another important issue is that information in one layer should not obscure or hide vital information in other layers.

6.5 Paper V

Automatic web cartography enhancement using semantic based expert system

In recent years, substantial research has been conducted on improving web cartography. Special cartographic concerns must be considered, especially, for example when the data are from e.g. a view service that is taken from more than one source. In view services based on several basic services, there are various semantic, topological and geometrical heterogeneities within distributed data that hinder the final maps not only to be fully legible but also the layers to be properly overlaid. To solve the problem in current geoportals and generate high quality maps, one approach is to utilize an automatic cartography core. This system contains a semantic rule based component to fix existing conflicts automatically for the integration of spatial data. We propose a system architecture that has an OWL (Web Ontology Language) based expert system to improve the cartography by adjusting and resolving topological and geometrical conflicts in geoportals. To test the methodology, we used a case study for adding a historical border on top of a base map. The results show that the historical border is overlaid without conflicts on top of the base map and a legible map is generated as an output.

6.5.1 Paper contribution

In this study, we utilize cartographic methods implemented in the geoportal to resolve geometrical and topological conflicts. These methods are based on several principles:

1. Semantic labels of the data in the basic services
2. Semantic rule base in the portal level
3. Geometrical and topological methods in the portal level

Using these methods, the end user can obtain a proper output in view services with a good cartography when the data are from different sources. Consequently, this method leads to an increase of user satisfaction with a spatial data exchange in an SDI.

7 Conclusions

Spatial Data Infrastructure has been identified as an umbrella covering spatial data handling procedures. However, there are challenges for SDI evaluation. A major problem is the lack of a general framework to assess an SDI from different perspectives. The first objective of this thesis was to develop a framework for the management and evaluation of an SDI. We generated a goal-oriented framework for evaluating SDI progress that can assess all dimensions simultaneously. Among current methods in business management approach, Balanced Scorecard is a multi-dimensional framework that can measure the progress of implementation of an SDI according to the defined strategies, objectives and goals. As an outcome of this study, the BSC method is utilized in the Swedish NSDI, and the feedback shows a promising future for NSDI progress (Papers I, II).

The second objective was to improve the accessibility of spatial data in an SDI. From a more technical point of view, methods have been developed for data availability and accessibility in an SDI. However, traditional clearinghouse networks do not satisfy end user requirements. Consequently, we add more functionality by increasing the percentage of accessing required data. We propose methods based on predefined rules and additional procedures within web processing services and service composition subjects. The outcome gives progressive results to get required data from an expert system based clearinghouses (Paper III).

The third objective was to improve the cartography in view services in an SDI. To enhance the cartography of maps and the effectiveness of web map services we utilize a number of methods that makes the output more usable, such as the polygon overlay method, the color saturation method, and geometric/topological conflict removal methods, which are based on semantic issues (Paper IV, V).

From a data-oriented perspective, regarding to the user requirement of spatial data, this thesis conducted methods to measure and improve the availability, accessibility, applicability and usability of spatial data in an SDI environment. More specifically, in SDI evaluation, all four aspects are developed, and improved (Paper I and II). For the method development

issues, Paper III contributes by improving the accessibility, Paper IV develops the usability, and finally, Paper V proposes methods to increase the applicability and usability of spatial data.

8 References

- Albaredes, G. (1992). A new approach: user oriented GIS, Proceedings of Third European GIS Conference, 23-26 March, Munich, Germany.
- Albert, C. (2002). Service loyalty: the effects of service quality and the mediating role of customer satisfaction, *Eur. J. Market* 36, pp 810–823.
- Altuntas, B.M., Bayraktar, D., & Cebi, F. (2006). An application of expert system approach for supplier evaluation and selection. In *Technology management for the global future–PICMET 2006 conference*, .6, pp.2755–2758.
- Bagrow, L. and Skelton, R. A., 2009. History of cartography. Transaction publishers. ISBN 9781412811545.
- Band, W. (2000). Create a customer satisfaction index (CSI) to improve your performance, *Sales Market. Manag. Can.* 31 (7), 58–59.
- Beaujardiere, J. (2006). OpenGIS® Web Map Server Implementation Specification, Open Geospatial Consortium Inc., Available at: <http://www.opengeospatial.org/standards/wms>, visited in 2011.
- Bell, M. (2008). Service-oriented modeling: Service analysis, design and architecture. Wiley & Sons; 366p. ISBN:978-0470141113.
- Bergenheim, W., Sarjakoski, L.T., & Sarjakoski, T. (2009). A web processing service for GRASS GIS to provide on-line generalization. In *Proceedings of 12th Agile conference, Hannover, Germany*.
- Bernard, L., Kanellopoulos, I., Annoni, A. and Smits, P. (2005). The European geoportal--one step towards the establishment of a European Spatial Data Infrastructure, *Computers, Environment and Urban Systems*, 29(1), Geoportals, pp 15-31.
- Brewer, C. A. and Buttenfield, B.P. (2007). 'Framing guidelines for multi-scale map design using databases at multiple resolutions', *Cartographic and Geographic Information Science*, 34, pp 3–15.
- Brijesh, P., Thierry, C., and M. Peter (2008). Balancing the NHS balanced scorecard, *European Journal of Operational Research*, 185(3), 905-914.
- Budic, Z. D. and Pinto, J. K. (1999). *Interorganizational GIS: issues and prospects. The annals of regional science* (Springer-Verlag).
- Budhathoki, N.R., and Z. Nedovic-budic (2007). Expanding the Spatial Data Infrastructure knowledge base, in *research and theory in advancing Spatial Data Infrastructure concepts* by H. Onsrud (Ed.), 1st Edition, ESRI Press, Redlands, California, 294.

- CGDI (2004). Canadian Geospatial Data Infrastructure, Frequently Asked Questions, available at: <http://geogratis.cgdi.gc.ca/geogratis/en/faq.html> - visited in 2011.
- Chaichan, P. (2002). Implementation of a balanced scorecard technique for performance evaluation of a real-estate company, Master Thesis, School of Civil Engineering, Asian Institute of Technology (AIT), Bangkok, Thailand.
- Chan, T. O., Feeney, M., Rajabifard, A. and Williamson, I. P. (2001). The dynamic nature of spatial data infrastructures: a method of descriptive classification. *Geomatica*, 55(1), pp 451–462, Available at: http://www.geom.unimelb.edu.au/research/publications/IPW/4_01Chan.pdf
- Cooper, R., and R. Kaplan (1991). Profit priorities from activity based costing. *Harvard Business Review*, 130–135.
- Crompvoets, J., Rajabifard A., van Loenen B. & Delgado Fernández T. (2008). A multi-view framework to assess SDIs, Space for Geo-Information (RGI), Wageningen University and Centre for SDIs and Land Administration, Department of Geomatics, University of Melbourne, Melbourne, Australia.
- Crompvoets, J., Bregt, A., Rajabifard, A. and Williamson, I. P. (2004). Assessing the worldwide developments of national spatial data clearinghouses. *International Journal of Geographical Information Science*, 18(7), 665-689.
- Crompvoets J. and Bregt A. (2002). World Status of National Spatial Data Clearinghouses. http://www.urisa.org/Journal/Under_Review/articles_under_review.htm
- Dent, B. D. (1999). *Cartography, Thematic mapping – Fifth edition*. WCB/McGraw-Hill, 1999. ISBN: 0-679-38495-0.
- Directive 2007/2/EC (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007, Official Journal of European Union, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF>, visited in 2011.
- Epstein, M.J., and P.S. Wisner (2001). Using a balanced scorecard to implement sustainability, *Environmental Quality Management*, 11(2), 1-10.
- Ermolayev, V., Keberle, N., Plaksin, S., Kononenko, O., Terziyan, V.Y.: Towards a Framework for Agent-Enabled Semantic Web Service Composition. *Int. J. Web Service Res.*(2004) 63-87.
- Executive Order (1994). Coordinating geographic data acquisition and access, the National Spatial Data Infrastructure, Executive Order 12906, Federal Register 59, 1767117674, Executive Office of the President, USA.
- Feeney, M. E. and Williamson I. P. (2003). The role of mechanisms in Spatial Data Infrastructure development that support decision-making, *Journal of Cartography*, 3(2), pp. 21-37.
- Fernández, T.D., Fernández, M.D. and Andrade R. (2008). The Spatial Data Infrastructure Readiness model and its worldwide application In: *A Multi-view Framework to Assess Spatial Data Infrastructures*, Crompvoets,

- J.W.H.C.; Rajabifard, A.; Loenen, B. van; Delgado Fernandez, T., Melbourne, Australia : The Melbourne University Press, pp 117 – 134.
- Fernandes K.J., Raja V., and A. Whalley (2006). Lessons from implementing the balanced scorecard in a small and medium size manufacturing organization, *Technovation*, (26), 623-634.
- FGDC (2009). FGDC registry and geoportal one stop (GOS) resources, available at: <http://registry.fgdc.gov>, visited in. 2011.
- FGDC (2000a). Financing the NSDI, A report prepared by Urban Logic for FGDC, available at: <http://www.fgdc.gov/funding/urbanlogic.pdf> - visited in. 2011.
- FGDC (2000b). Questions and answers about clearinghouses. Washington, DC: Federal Geographic Data Committee.
- FGDC (1997). Framework, introduction and guide. (Washington: Federal Geographic Data Committee).
- Folaron, J. (2003). The evolution of Six Sigma. *Six Sigma Forum Magazine*, 2(4): 38-44.
- Fornell, C. (1992). A national customer satisfaction barometer: the Swedish experience, *J. Market*, (56), 6–21 (January).
- Friis-Christensen, A., Lucchi, R., Lutz, M., & Ostländer, N. (2009). Service chaining architectures for applications implementing distributed geographic information processing. *International Journal of Geographical Information Science*, 23, pp 561–580.
- GeoConnections (2007). GeoConnections Program for CGDI. available at: http://www.geoconnections.org/publications/Technical_Manual/html_e/s1_ch3.html.
- Georgiadou, Y., Puri, S.K., and Sahay, S. (2005). Towards a potential research agenda to guide the implementation of spatial data infrastructures - A case study from India, *International journal of geographic information science*, (19), 113-1130.
- Geudens, T., Macharis, C., Cromptvoets, J. and F. Plastria (2009). Assessing Spatial Data Infrastructure Policy Strategies Using the Multi-Actor Multi-Criteria Analysis, *International Journal of Spatial Data Infrastructures Research*, 4, pp 265-297.
- Giarratano J. & Riley G. (1989). *Expert Systems: Principles and Programming*, PWS-Kent Publishing Co, Boston, MA, 597 pages.
- Giff (2008). A Framework for Designing Performance Indicators for Spatial Data Infrastructure Assessment In: A Multi-view Framework to Assess Spatial Data Infrastructures, Cromptvoets, J.W.H.C.; Rajabifard, A.; Loenen, B. van; Delgado Fernandez, T., Melbourne, Australia : The Melbourne University Press, pp 211 – 230.
- Giff, G. A., and J. Cromptvoets (2008). Performance indicators a tool to support Spatial Data Infrastructure assessment, *Computers, Environment and Urban Systems*, 32, pp 365–376.
- Giff, G., and Coleman, D. (2003). Financing SDI development: examining alternative funding models, in Williamson, I.P., Rajabifard, A., and Feeney,

- M.E.F. (eds.), *Developing Spatial Data Infrastructures: from concept to reality*, London and New York: Taylor & Francis.
- Giff, G., and Coleman, D. (2002). Spatial data infrastructure funding models: a necessity for the success of SDIs in emerging countries, FIG XXII International Congress, 19-26 April, Washington, D.C., USA.
- Granell, C., Gould, M., & Ramos, F. (2005). Service composition for SDIs: Integrated components creation. In *Sixteenth international workshop on database and expert systems applications*, 26 August 2005 pp475 –479. Copenhagen, Denmark: IEEE.
- Granić, A., Mitrović, I., Marangunić, N. (2011). Exploring the usability of web portals: A Croatian case study, *International Journal of Information Management*, Volume 31, Issue 4, August 2011, Pages 339-349,
- Groot, R. (2001). Economic issues in the evolution of national geo-spatial data infrastructure, 7th United Nations Regional Cartographic Conference for the America, New York, USA.
- Groot R and J. McLaughlin (2000). *Geospatial Data Infrastructures*. Oxford, Oxford University Press.
- GSDI (2009). Global Spatial Data Infrastructure Association, available at: <http://www.gsdi.org/> - visited in 2011.
- Grus, L., Castelein, W., Cromptvoets, J., Overduin, T., van Loenen, B., van Groenestijn, A., Rajabifard, A., Bregt, A. K., (2010). An assessment view to evaluate whether Spatial Data Infrastructures meet their goals, *Computers, Environment and Urban Systems*, 35(3), pp 217-229.
- Grus, L., Cromptvoets, J., and A. Bregt (2007). Multi-view SDI assessment framework, *International Journal of Spatial Data Infrastructures Research*, 2, pp 33-53.
- Hackl, P., Kristensen, P.K., and A. H. Westlund (2000). Customer satisfaction: theory and measurement. Special Issue of *Journal of Total Quality Management*.
- Harrie, L., Mustiere, S., Stigmar, H. (2011). Cartographic Quality Issues for View Services in Geoportals, *Cartographica* 46(2), 2011, pp 92–100.
- Harvey F and D. Tulloch (2006). Local-government data sharing: Evaluating the foundations of spatial data infrastructures. *International Journal of Geographical Information Science* 20: 743–68
- Hunter, J. (2001). MetaNet – a metadata term thesaurus to enable semantic interoperability between metadata domains. *Journal of Digital Information* 1 (8). available at: <http://journals.tdl.org/jodi/article/view/33/34>. visited in 2011.
- ICA, 1973. *Multilingual Dictionary of Technical Terms in Cartography*. Franz Steiner Verlag GMBH.
- Idalina, D-S., Lucas, R., and A. Paula (2007). Developing sustainability balanced scorecards for environmental services: A study of three large Portuguese companies, *Environmental Quality Management*, 16(4), pp 13-34.

- INSPIRE (2009a). Infrastructure for Spatial Information in Europe, available at: <http://inspire.jrc.ec.europa.eu/> visited in 2011.
- INSPIRE (2009b). INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119, http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/metadata/MD_I_Rand_ISO_20090218.pdf , visited in 2011.
- Iosifescu-Enescu, I., Hugentobler, M. and Hurni, L.(2009).‘Web cartography with open standards—a solution to cartographic challenges of environmental management’, *Environmental Modelling&Software*,25,pp.988–999.
- ISO/TC-211 (2009). Geographic information/Geomatics, available at: <http://www.isotc211.org>, visited in 2011.
- ISO (2005). ISO 19119: 2005 geographic information–Services,international standard, ISOTC211.
- Janssen, K. (2008). A legal approach to assessing Spatial Data Infrastructures, In: *A Multi-view Framework to Assess Spatial Data Infrastructures*, Cromptvoets, J.W.H.C.; Rajabifard, A.; Loenen, B. van; Delgado Fernandez, T., Melbourne, Australia : The Melbourne University Press, ISBN 9780732516239, pp 255 – 272.
- Kaplan, R. S., and D. P. Norton (2000). Having trouble with your strategy? Then map it. *Harvard Business Review* (September-October): pp 167-176.
- Kaplan, R.S., and D. P. Norton (1996). Using the balanced scorecard as a strategic management system, *Harvard Business Review*, 74(1),pp 75–85.
- Kavouras M and Kokla M, 2008, *Theories of Geographic Concepts: Ontological Approaches to Semantic Integration*. CRC Press, Boca Raton, FL, USA.
- Keates, J. S., 1996. *Understanding Maps – 2nd edition*. Longman, 1996. ISBN 0-582-23927-3.
- Kiehle,C., Greve,K., & Heier,C. (2007). Requirements for next generation spatial data infrastructures – standardized web based geoprocessing and webservice orchestration. *Transactions in GIS*, 11, pp 819–834.
- Kok, B. and van Loenen, B. (2005). How to assess the success of National Spatial Data Infrastructures? , *Computers, Environment and Urban Systems*, 29(6), pp 699-717.
- Kumar, S., & Mishra, R. B. (2008). A Framework Towards Semantic Web Service Composition Based on Multi-Agent System. *International Journal of Information Technology and Web Engineering (IJITWE)*, 3(4), 59-81.
- Lanig S. & A. Zipf (2010): Proposal for a Web Processing Services (WPS) Application Profile for 3D Processing Analysis. *GEOProcessing 2010 - The Second International Conference on Advanced Geographic Information Systems, Applications, and Services*. Netherlands Antilles.
- Lantmäteriet (2010). Geodata project, www.geodata.se, visited in 2010.
- Lantmäteriet (2008). A National Geodata Strategy. http://www.lantmateriet.se/upload/filer/om_lantmateriet/Infrastruktur_geodata/Geodata_Strategy_2008.pdf, visited in 2011.

- Lemmens, R., de By, R., Gould, M., Wytzisk, A., Granell, C., & van Oosterom, P. (2007). Enhancing geo-service chaining through deep service descriptions. *Transactions in GIS*, 11, pp 849–871.
- Lemmens, R., Granell, C., Wytzisk, A., de By, R., Gould, M., & van Oosterom, P. (2006). Integrating semantic and syntactic descriptions to chain geographic services. *IEEE Internet Computing*, 10, pp 42–52.
- Lemmens, R.L.G. (2006). Semantic interoperability of distributed geo - services. PhD Thesis, Delft University of Technology, ITC Dissertation 138.
- Lemmens, M. J. P. M. (2001). A European perspective on Geo-Information Infrastructure (GII) issues, available at: <http://www.gisdevelopment.net/policy/international/interna011.htm> – visited in 2011.
- Luzet, C. (2004). Geospatial data development: building data for multiple uses, In D. Nebert (Ed.), *developing spatial data infrastructure: The SDI Cook Book*, 13-23, at: <http://www.gsdi.org/docs2004/Cookbook/cookbookV2.o.pdf>, visited 2011.
- McIlraith, S.A., Son, T.C., & Zeng, H. (2001). Semantic Web Services. *IEEE Intelligent Systems* 16(2), 46-53.
- Masser, I. (2010). *Building European Spatial Data Infrastructures*. Redlands: ESRI Press.
- Masser, I. (1998). The first generation of national geographic information strategies, *Selected Conference Papers of the 3rd Global Spatial Data Infrastructure Conference*, 17-19 November 1998, Canberra, Australia.
- Mansourian, A., Omid, E., Toomanian, A., Harrie, L., (2010), Expert System to Support Functionality of Clearinghouse Services, *Computers, Environment and Urban Systems*, 35(2), pp 159-172, Paper III.
- Mansourian, A. and Valadan Zoj, M. J. (2008.). Iran SDI initiative: study phase of NSDI, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. XXXVII. Part B4. Beijing.
- Mansourian, A., Rajabifard, A., Valadan Zoj, M.J., Williamson, I. (2006). Using SDI and web-based system to facilitate disaster management, *Computers & Geosciences*, Volume 32, Issue 3, April, pp 303-315, ISSN 0098-3004.
- Mansourian A. (2005). Development of an SDI Conceptual Model and Web-based GIS to Facilitate Disaster Management, PhD Thesis, Faculty of Geodesy & Geomatics Eng., K.N.Toosi University of Technology, Tehran, Iran
- Mapping Science Committee (1997). *The future of Spatial Data and Society*, Washington D.C.: National Academic Press.
- Martin, D., Burstein, M., Hobbs, J., Lassila, O., McDermott, D., McIlraith, S., Narayanan, S., Paolucci, M., Parsia, B., Payne, T., Sirin, E., Srinivasan, N., and Sycara, K. OWL-S: Semantic Markup for Web Services. W3C Member Submission, expository document, World Wide Web Consortium, November 2004. Available at <http://www.w3.org/Submission/2004/07/>.
- Motta, E., Domingue, J., Cabral, L., and Gaspari, M. (2003). IRS-II: A Framework and Infrastructure for Semantic Web Services. In *Proceedings The*

- SemanticWeb - ISWC 2003 Second International SemanticWeb Conference (Sanibel Island, FL, USA, 2003), J. M. Dieter Fensel, Katia Sycara, Ed., vol. Lecture Notes in Computer Science Volume 2870 / 2003, Springer- Verlag Heidelberg.
- Müller M. (2006). Symbology Encoding Descriptor Implementation Specification, Open Geospatial Consortium.
- Müller M and MacGill J.(2005). OpenGIS® Styled Layer Descriptor Profile of the Web Map Service Implementation Specification, Open Geospatial Consortium.
- Najar, C., Rajabifard, A., Williamson, I., and Giger, C. (2006). A Framework for Comparing Spatial Data Infrastructures, an Australian-Swiss Case Study In: GSDI 9 - International Conference for Global Spatial Data Infrastructure, 3-11 Nov 2006, Santiago, Chile.
- Nebert (2001). SDI Cookbook: Developing Spatial Data Infrastructure: The SDI Cookbook, Ver 1.1, Global Spatial Data Infrastructure (GSDI), Editor: Douglas D. Nebert, Technical Working Group Chair, GSDI.
- Nedović-Budić, Z., Pinto, J. K. and Budhathoki, N. R. (2008). SDI Effectiveness from the User Perspective, In: A Multi-view Framework to Assess Spatial Data Infrastructures, Crompvoets, J.W.H.C.; Rajabifard, A.; Loenen, B. van; Delgado Fernandez, T., Melbourne, Australia : The Melbourne University Press, pp 273 – 303.
- Nedović-Budić, Z., Feeney, M, F., Rajabifard, A. And Williamson, I (2004). Are SDIs serving the needs of local planning? Case study of Victoria, Australia and Illinois, USA, Computers, Environment and Urban Systems, 28, pp 329-351.
- OMB (2002). To the heads of executive departments and establishments, Circular No. A-16, August 19, 2002. available at: http://www.whitehouse.gov/omb/rewrite/circulars/a016/print/a016_rev.html, - visited in 2011.
- Patil, A., Oundhakar, S., Sheth, A., and Verma, K. (2004). METEOR-S Web service Annotation Framework. WWW 2004, May 17-22, 2004, New York, New York, USA.
- Phillips, A., Williamson, I.P. and Ezigbalike, I.C. (1999). Spatial Data Infrastructure Concepts, The Australian Surveyor, 44(1), pp 20-28.
- Radwan, M. (2002). The Development of Geographic Information Infrastructure 'GDI/SDI' to Support Access to Spatial data in Distributed Environment. Enschede, the Netherlands.
- Rajabifard A. and Williamson I. P. (2003a). Asia-Pacific Region and SDI activities, Journal of GIS Development, July 2003, Vol. 7, Issues 7.
- Rajabifard A, Feeney M F, and I. Williamson, (2003b). Spatial Data Infrastructures: Concept, nature and SDI hierarchy. In Williamson I, Rajabifard A, and Feeney M F (Eds.) Developing Spatial Data Infrastructures: From Concept to Reality. Boca Raton, FL, CRC Press: pp 17-40.

- Rajabifard, A., Feeney, M. E. and Williamson I. P. (2002). Future directions for SDI development, *International Journal of Applied Earth Observation and Geoinformation*, 4(1), pp 11-22, ISSN 0303-2434.
- Rajabifard, A., Williamson, I. P., Holland, P. and Johnstone, G. (2000a). From Local to Global SDI initiatives: a pyramid building blocks, proceedings of the 4th Global Spatial Data Infrastructure Conference, 13-15 March 2000, Cape Town, South Africa.
- Rajabifard, A., Escobar, F. and Williamson, I.P. (2000b). Hierarchical Spatial Reasoning Applied to Spatial Data Infrastructures, *Journal of spatial science*, 29 (2), pp.41-50, Australia.
- Remkes, J. W. (2000). Foreword in geospatial data infrastructure: cases, concepts and good practices, Edited by Groot, R. and McLaughlin J, New York: Oxford University Press.
- Rhind, D. (2000). Funding and NGDI, in Groot R., and McLaughlin J. (eds), *Geospatial data infrastructure concepts: cases and good practices*, New York: Oxford University Press.
- Rhind, D. (1999). Key economic characteristics of information, Ordnance Survey, U.K.
- Robinson, A. H., Morrison, J. L., Muehrcke, P. C., and Guptil, S. C. (1995). *Elements of cartography – sixth edition*. John Wiley and sons, Inc., ISBN: 0-471-55579-7.
- Sandgren, U. (2010). Lantmäteriet; project leader of the Swedish Geodata strategy. Personal communication.
- Sashkin, M. and K.J. Kiser (1993). *Putting total quality management to work*. San Francisco: Berrett-Koehler.
- Schut, P. (2007). OpenGIS Web processing service, Open Geospatial Consortium Inc., Reference number: OGC 05-007r7, available at: <http://www.opengeospatial.org/standards/wps>, visited in 2011.
- Silk, S. (1998). Automating the balanced scorecard. *Management Accounting*, 79 (11), pp 38-44.
- Shea, K. S. and McMaster, R. B., 1989. Cartographic generalization in a digital environment: When and how to Generalize. *Proceedings of 9th International Symposium on Computer-Assisted Cartography (Auto-Carto9) 1989*.
- Slocum A, McMaster R, Kessler F, Howard, H. (2005). *Thematic Cartography and Geographic Visualization*, second ed. Pearson Education Inc, 518pp.
- SLD (2007). *Styled Layer Descriptor profile of the Web Map Service Implementation Specification - Version: 1.1.0 (revision 4)*. Open Geospatial Consortium Inc., 2007-06-29. Available online: <http://www.opengeospatial.org/standards/sld> , visited in 2011.
- Soderman, A., Dahlbergm, M. (2011). Methods for visualization enhancement in multi source web mapping services using basic cartographic rules. In *Proceedings of the 25th International Cartographic Conference*, Paris, 3-8 July 2011.

- Steudler, D., Rajabifard, A. and Williamson, I. (2008). Evaluation and Performance Indicators to Assess Spatial Data Infrastructure Initiatives, In: A Multi-view Framework to Assess Spatial Data Infrastructures, Crompvoets, J.W.H.C.; Rajabifard, A.; Loenen, B. van; Delgado Fernandez, T., Melbourne, Australia : The Melbourne University Press, ISBN 9780732516239, pp 193 – 210.
- Stojanovic, T., Green, D. R., Lymbery, G. (2010). Approaches to knowledge sharing and capacity building: The role of local information systems in marine and coastal management, *Ocean & Coastal Management*, 53(12), pp 805-815.
- Sydenham, P.H. & Thorn, R. (2005). *Handbook of Measuring System Design, Rule-based Expert Systems*, Wiley & Sons, USA, 1648 Pages.
- Tait, M.G. (2005). Implementing geoportals: applications of distributed GIS, *Computer, environment and urban systems*, 29, pp 33-47.
- Thellufsen, C., Rajabifard, A., Enemark, S. and Williamson, I (2009). Awareness as a foundation for developing effective spatial data infrastructures, *Land Use Policy*, 26(2), pp 254-261
- Toomanian, A., Harrie, L., Olsson, P., (2011), Automatic symbolization methods for geoportals, *The Cartographic Journal*, Accepted, DOI: 10.1179/1743277411Y.0000000028, Paper IV.
- Toomanian, A. Mansourian, A. , Harrie, L. and A. Rydén (2011). Using Balanced Scorecard for Evaluation of Spatial Data Infrastructures: A Case Study in Accordance with INSPIRE, *International Journal of Spatial Data Infrastructures Research*, Vol. 6, Paper II.
- Toomanian, A. and Mansourian, A. (2009). An Integrated Framework for the Implementation and Continuous Improvement of Spatial Data Infrastructures. *SDI Convergence. Research, Emerging Trends, and Critical Assessment*. B. van Loenen, J.W.J. Besemer, J.A. Zevenbergen (Editors). Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission 48, Paper I.
- Van den Worm, J. (2001). Chapter 7, Web map design in practise. In Kraak, M.-J. and Brown, A. (Eds.). *Web Cartography – developments and prospects*. Taylor and Francis Inc. ISBN 074840869X.
- Vandenbroucke, D., K. Janssen and J. Van Orshoven (2008). INSPIRE State of Play: Generic approach to assess NSDI. In Crompvoets, J., A. Rajabifard, B. van Loenen and T. Delgada (Eds.). *A Multi-View Framework to Assess National Spatial Data Infrastructures*, Melbourne University Press, Australia, pp 145 - 172.
- van Loenen, B. and van Rij, E. (2008). Assessment of Spatial Data Infrastructures From an Organizational Perspective, In: *A Multi-view Framework to Assess Spatial Data Infrastructures*, Crompvoets, J.W.H.C.; Rajabifard, A.; Loenen, B. van; Delgado Fernandez, T. (Eds.), Melbourne, Australia: The Melbourne University Press, pp 173 – 192.
- Van Loenen, B. (2006). *Developing geographic information infrastructures; the role of information policies*. Dissertation. Delft University of Technology. Delft: DUP Science.

- Van Leenen, B. and Kok, B. C. (2004). Spatial Data Infrastructure and policy development in Europe and United States, Delft University of Technology, The Netherlands: DUP Science.
- Van Leenen, B. (2003). Free charge or revenue generation? (invited reply: cost recovery policy irrational but understandable), in: GM International 2, 17, pp. 45-46.
- Vretanos, P.A. (2005). Web Feature Service Implementation Specification; Open Geospatial Consortium (OGC), OGC 04-094, available at: <http://www.opengeospatial.org/standards/wfs>, visited in 2011.
- W3C (2005). Semantic Web Services Framework (SWSF) Overview. W3c member submission, World Wide Web Consortium.
- Walenciak, G. and Zipf, A. (2010). Designing a Web Processing Service application Profile for Spatial Analysis in Business Marketing 13th AGILE International Conference on Geographic Information Science 2010 pp Guimarães, Portugal.
- WFS (2005). Web Feature Service Implementation Specification - Version: 1.1.0. Open Geospatial Consortium Inc., 3 May 2005. Available online: <http://www.opengeospatial.org/standards/wfs>, visited in 2011.
- When de Montalvo, U. (2000). Access to spatial data – what determine the willingness of organizations to share it?, 4th GSDI Conference, Cape Town, South Africa.
- Whiteside, A. and Evans, D. J. (2008). Web Coverage Service (WCS) Implementation Standard, Open Geospatial Consortium (OGC), OGC 07-067r5, 2008, available at: <http://www.opengeospatial.org/standards/wcs>, visited in 2011.
- Williamson I. P., Rajabifard, A. and Feeney, M. E. (2003). Developing spatial data infrastructures: from concept to reality, CRC Press, ISBN 041530265X
- Zhang, D.-H., Xie, B., Yu, L., & DI, L.-P. (2008). Open geospatial information services chaining based on OGC specifications and processing model. In International workshop on education technology and training & international workshop on geoscience and remote sensing, 21–22 December 2008, Shanghai, China.

Toomanian, A., Mansourian, A., (2009). An Integrated Framework for the Implementation and Continuous Improvement of Spatial Data Infrastructures, In *SDI Convergence. Research, Emerging Trends, and Critical Assessment*. B. van Loenen, J.W.J. Besemer, J.A. Zevenbergen (Editors). Nederlandse Commissie voor Geodesie, Netherlands Geodetic Commission 48, pp 161-173, 2009.

An Integrated Framework for the Implementation and Continuous Improvement of Spatial Data Infrastructures

Ara Toomanian¹ and Ali Mansourian²

¹Lund University, Sweden, Ara.Toomanian@gis.lu.se

²K.N.Toosi University of Technology, Iran, Mansourian@kntu.ac.ir

Abstract

Development and usage of proper frameworks for implementation, evaluation and continuous improvement of spatial data infrastructures (SDIs) is currently an important research topic. A wide range of methods are being researched. In this respect, methods and techniques on performance measurement and evaluation techniques from business management literature are not yet considered. Some techniques and methodologies from business management literature could be developed based on Six Sigma, ABC (Activity Based Costing), BSC (Balanced Scorecard) and TQM (Total Quality Management). This article describes these techniques and then provides an integrated framework, based on these business management techniques, for implementation and continuous improvement of SDIs.

Key Words: spatial data infrastructure (SDI), Implementation, Continuous Improvement, Six Sigma, ABC, BSC, TQM.

1. INTRODUCTION

In recent years, many countries implement and develop NSDI (Masser, 2005a). Also, scientists suggest operational platforms for the SDI implementation such as SDI business model (Wagner, 2005), SDI partnership (Warnest et al., 2002) and spatially enabling governments (Masser et al., 2007). Considering the fact that SDI implementation is a matter of technical, technological, social, institutional, political issues and also financial challenges (Nedovic-Budic et al., 2004; Masser, 2005b; Mansourian et al., 2006; Onsrud, 2007), different aspects and perspectives must be brought into attention for the progress of SDI implementation. Moreover, considering various dimensions of an SDI as spatial data production issue, data accessibility, data sharing, updating, standardisation and institutional matters, the need for a structured and integrated implementation framework is inevitable.

The next significant and essential requirement for an SDI implementation is the performance measurement and the continuous improvement due to the complexity and long term procedure of SDI implementation. In an SDI, it is important to have feedback from different dimensions and perspectives and to improve the weak points in order to have an effective and operational SDI. Such improvements may help to decrease additional costs and will lead to high quality spatial data products. Furthermore, all SDI users as well as the whole society will be satisfied with standard, accessible spatial data products and delivery within a high performance SDI.

A variety of research is conducted in accordance with SDI evaluation and performance measurement (Georgiadou et al., 2005; Georgiadou et al., 2006; Kok and van Loenen, 2005; McDougall, 2006; Van Loenen, 2006; Najjar et al., 2006; Giff and Lunn, 2008; Fernández and Cromptvoets, 2008; Lance et al., 2006; Grus et al., 2007). However, few attention is paid to business management literatures which provide proper techniques for performance measurement and evaluation.

In the business management literature, there are a variety of techniques which are used for continuous improvement of industries and/or organisational activities. Six Sigma, Activity Based Costing (ABC), Balanced Scorecard (BSC) and Total Quality Management (TQM) are some of these techniques that are also the targets of the article. Each of the mentioned techniques covers a dimension of SDI implementation. This article aims to address utilisation of these techniques as an integrated framework for implementation and continuous improvement of SDIs. Such integration will cover different aspects of SDI implementation and evaluation requirements. With this in mind, first, the techniques are reviewed briefly and then their feasible applicability for SDI implementation and evaluation is described.

2. CONCEPTUAL FRAMEWORK FOR SDI MEASUREMENT AND IMPLEMENTATION

In this section, we introduce a number of measurement methods used in the business management literature and describe their original purpose, then denote an integrated framework as an SDI implementation and evaluation procedure.

Six Sigma is a problem solving and continuous improvement method based on statistical methods where all the employees within an organisation have different roles within the entire technique. Six Sigma framework and guidelines can be used as a basic framework for SDI implementation.

Activity Based Costing (ABC) is a useful method to find the real price of the products according to the organisational costs and overheads. It also tries to assign costs to each activity and removes unnecessary and unprofitable tasks in an organisational process. ABC can be useful for estimating SDI costs as well as cost reduction and spatial data valuation.

Balanced Scoreboard (BSC) is a performance evaluation method used for evaluating and monitoring the strategic plans and the objectives. It can be used as an evaluation and monitoring method for SDIs and also for measuring the progress of SDI implementation according to different perspectives.

Total Quality Management (TQM) is a method to monitor the process quality. It deals with the entire product procedure and tries to keep the work process in a high standard. This method can be utilised as a proper technique for both quality control of SDI work process and spatial data.

Table 1 represents the usage domains and a description for various techniques discussed above. In the following sections, we will describe each method separately.

Table 1: A general overview of different business management methods.

Method	Key Premises
Six Sigma	Problem Solving and continuous improvement
ABC	Financial management and evaluation
BSC	Performance evaluation
TQM	Quality enhancement

2.1 Six Sigma

Six Sigma is one of the most effective problem solving methodologies for improving business and organisational performance. It was first originated and introduced by Motorola Company in 1987 and targeted an aggressive goal of 3.4 parts per million defects (Barney, 2002; Folaron, 2003). The background of Six Sigma method is a statistical approach where two main items are discussed:

- the roll up of characteristic behaviours, and
- the natural increase of variation for each characteristic in the long term.

Here, the sigma scale is a universal measure of how well a critical characteristic performs compared to its requirements. It works in such a way that if sigma score increase, the characteristic will be more capable (Gygi et al., 2005). Six Sigma is using a scientific, structured method for business improvement that could be used for any aspect of organisation, process or person.

Six Sigma is defined as “high-performance, data-driven approach to analyse the root causes of business problems and solving them” (Blakeslee, 1999). Other persons described Six Sigma as a disciplined and statistically based approach for improving product and process quality (Hahn et al., 2000). Also, Six Sigma refers to a business process that allows organisations to improve drastically their bottom line by designing and monitoring everyday business activities in ways that minimise waste and resources while increasing customer satisfaction (Harry and Schroeder, 2000). To achieve these aims, Six Sigma involves all employees in the organisational activities, according to their skills, and also obtains their feedback for problem solving and continuous improvement of the processes. Solving complex and strategic problems is conducted through experts and professionals and moderate tasks are carried out via medium level of skill and average trained employees. The regular transactions are conducted by other staffs.

The Six Sigma methodology has two project strategies, DMAIC (Define, Measure, Analyse, Improve, and Control) and DMADV (Define, Measure, Analyse, Design, and Verify), which are describe bellow. These strategies are a set of standardised and systematic methods that each project has to use in order to have a continuous improvement.

2.1.1 DMAIC

DMAIC is a problem solving and continuous improvement strategy for any kind of organisational strategy. It includes the following steps (Gygi et al., 2005):

- Define: writes the problem statement context and project objective setting;
- Measure: understands the process and improves the baseline performance and capability of the process or system;
- Analyse: uses data and tools to understand the cause and effect relationships in the process or system;
- Improve: determines and develops the modifications that lead to a validated improvement in the process or system and tries to implement solutions to achieve the objective statement, and
- Control: establishes plans and procedures and implements processes control methods to ensure the improvements are sustained.

To use this strategy, effective contribution of skilled and trained staffs, at different management levels, is essential. In other words, all employees have *fundamental* role for the DMAIC implementation. In addition, completing one step is a prerequisite for moving to the next step. After passing all steps successfully, a Six Sigma project is completed.

This strategy can be utilised for in an early stage of SDI implementation. In such situations, there are a number of initial tasks to start the SDI implementation procedure. Strategic plans, action plans, general objectives are some of the primary documents which have to be completed in the *define* step. Afterwards, within the implementation procedure of SDI, data production and delivery processes, collaboration among organisation for data exchange and also maintenance and standardisation of spatial data are measured and evaluated in the *measure* step. To perform this, integration of the Six Sigma measurement methods and the SDI evaluation indicators is suggested. *Analyse* deals with analysing the results of the measurements and identifying those barriers that impede SDI implementation and those positive points that facilitate the implementation. *Improve* enhances the procedures of SDI implementation regarding to the information derived from previous stage. Finally, *control* aims to check whether improvements of the previous steps caused the SDI implement in a proper way or not. As SDI implementation is a long term process, this methodology might be used many times as loop within the period of implementation.

2.1.2 DMADV

There are many similarities between DMADV and DMAIC. The major difference is in the last two letters which refer to *Design* and *Verify*. *Design* refers to either a new process or a corrective step to the existing one, eliminating the error origination that meets the target specification. *Verify* means verification by simulation of the performance of developed design and its ability to meet the target needs (Gygi, 2005). In DMADV, the processes change and redesign according to the customer's needs. Such change is needed in order to fit to the on demand requests instead of the improvement and control steps which more focus on readjusting and controlling by one way or other.

Although there are many overlaps in this strategy with the previous one, nations and societies which have already started an SDI implementation procedure and would like to extend or adjust it can use the DMADV strategy. In this strategy, the re-design of the SDI may extend or restructure the previous framework and then in the validate stage, it will be evaluated and monitored according to the new process and situation.

2.2 Activity Based Costing (ABC)

Financial aspects and cost are main features for SDI development. Even though the SDI budgets mainly stem from the government resources, these subjects are essential for the SDI managers to succeed in the spatial data market.

In the traditional way of management and accounting methods in the 1930s, corporate rules had a basic role to force companies for providing financial accounts. Although the application of strict rules was a proper way for financial accounts, management accounts were proposed as a decision-making tool in business atmosphere and therefore required more flexibility (Letza and Gadd, 1994). In such a method, production overhead was absorbed to the product cost to value the stock. Moreover, labour costs were used as a convenient overhead recovery base, although the ratio of the total labour cost was not proportional.

However, the traditional methods often fail to incorporate the final cost today. The reason is that the technological costs and other overheads have increased rapidly, due to the expansion of global competition, and the increase of interactions via communication media, development of IT and access to inexpensive information systems. Therefore, new accounting methods such as Activity Based Costing (ABC) have been introduced.

ABC was first introduced in the late 1980s by Johnson and Kaplan (1987). Scientists expanded the first initial idea and developed a method for cost drivers to calculate activity costs for each product and service. They argued that such method supplies accurate cost data needed to make proper strategic decisions for product mix, sourcing, pricing, process improvement, and evaluation of business process performance (Cooper and Kaplan, 1992; Swenson, 1995).

ABC is a costing model which determines the activities in an organisation and assigns the cost of each activity resource to products and services separately regarding to the actual usage by each. It also generates the real cost of products and services by removing unprofitable activities and eliminate lowering prices of overpriced ones. Here, an activity is defined as a discrete task that a company makes in a product or service, and uses cost drivers to assign activity costs to products, services or customers related to these activities (Cooper, 1988; Ittner et al., 2002). In this method, products use activities and the activities use resources.

ABC has two main stages to assign overhead costs to products and services (Hilton, 2005). First, based on the definition, the main activities are determined and overhead costs are assigned to the activity cost pools according to the amount of resources used by activities. The activities are often derived from information gathered from interviews, questionnaires, and time cards (Cooper and Kaplan, 1991). The second stage contains cost allocation from each activity cost pool to each product line concerning to the amount of the cost driver utilised by the product line (Bjornenak and Mitchell, 2002). In other words, at the first step, organisational resources are grouped in the different pools such as salaries, license fees, operational costs and depreciation. Then, different institutional missions and tasks are grouped into homogeneous activities such as data preparation, research and development (R&D), data delivery (Ooi and Soh, 2003). In this way, each activity will use a percentage of a single or multiple cost pools. For example, the data preparation activity will use 10% of the rental cost, 20% of the salary and 40% of the operational costs.

As ABC reveals the links between performing particular activities and the demands those activities make on the organisation's resources, it provides managers with a clear picture of how products and services both generate revenues and consume resources. The profitability picture that emerges from the ABC analysis helps managers focus their attention and energy on improving activities that will have the biggest impact on the result.

An important part of SDI implementation are the financial and economical issues. A proper financial funding model may lead the SDI coordinators to a successful and operational SDI. Furthermore, having a clear idea about the SDI cost and the way of cost reduction will also increase the efficiency of SDIs. With this in mind, using the ABC method, main activities of SDIs are determined and according to the transparent implementation tasks, unprofitable and parallel activities will be eliminated. Also, in each step, the financial resources can be predicted with respect to different contributors

whether the financial support is from the spatial data market or authorities. Moreover, for any task and process within a clear financial and economic perspective, evaluation and monitoring can be easily performed by the SDI coordinators.

2.3 Balanced Scorecard (BSC)

The success of the next generation of Spatial Data Infrastructures (SDIs) will, in part, depend on the ability of SDI coordinators to comprehend, analyse and report on the performance of their initiatives (Giff and Lunn, 2008). Therefore, it is necessary for SDI coordinators to use proper models and measuring techniques to assess and monitor the progress of SDIs.

BSC, as a technique from business management literature for strategic performance management, was introduced by Kaplan and Norton (1992) as a set of different measures that allow for a holistic, integrated view of business performance. It was a complementary solution for the traditional financial parameters to measure the performance in organisations. In other words, BSC is a performance measurement framework that provides an integrated look at the business performance of an organisation by a set of measures including both financial and non-financial metrics (Kaplan and Norton, 1992; Kaplan and Norton, 1996). Also, BSC refers to a multi-dimensional framework that uses measurement to improve an organisation's strategy.

There are some basic elements in the BSC structure which leads the strategy measurement in a proper way. A *perspective* is an element into which the strategy is decomposed to drive implementation. In most BSC structures, there are four perspectives: *financial*, *internal process*, *customer*, and *learning and growth*. As Norton and Kaplan (2000) mentioned, "Balanced Scorecards tell you the knowledge, skills and systems that your employees will need (learning and growth) to innovate and build the right strategic capabilities and efficiencies (internal processes) that deliver specific value to the market (customer) which will eventually lead to higher shareholder value (financial)". It is possible to add other perspectives or sometimes replace the mentioned perspectives according to the specific strategies. The perspective can be defined as an interpretation of the strategy in different dimensions.

The second main element of the BSC design is called *objective*. An objective is a statement of strategic intent, describing how a strategy will be made operational in an organisation. In other words, objectives are the main elements of the strategic plan and the entire strategy can be broke down into many objectives. In the BSC design, normally a limited number of objectives exist relating to one of the perspectives, which is normally described in one or two sentences.

The next basic element in a BSC design is the *cause and effect linkage*. In the BSC structures, objectives are related and depend on each other through cause and effect relationships. The cause and effect linkages are like *if – then* statements where the objectives in each perspective are linked with the graphical connectors according to the rules derived from different dimensions.

Another element of the BSC is the *measure* term, which is a performance metric one can calculate the progress of an objective. A measure must be quantifiable. In a BSC design there are reasonable numbers of measures explicitly linked to an objective. In addition, the measure concept is typically represented via mathematical formulas.

The fifth element for BSC design is called *target*. A target is a quantifiable goal for the each measure. A combination of targets on the BSC design is the general goal of an organisation. They help the organisation monitor the progress toward strategic goals, and give proper feedbacks if necessary.

Strategic initiative is the last element of a BSC design. They are action programs that drive strategic performance and the activities which will lead the organisation to achieve the strategic results. All ongoing initiatives in an organisation should be associated with the strategy in the BSC.

BSC design can be used as an evaluation and monitoring framework for SDIs. By defining performance indicators as well as desired targets, for each objective, SDI coordinator and managers can measure a current situation, compare it with the target and then evaluate the progress of an SDI. Considering the four main perspectives in the BSC structure, BSC provides a general framework for evaluating SDIs from users' and data producers view point. It also helps to evaluate internal processes, financial affairs and even capacity building at the individual level. So BSC can be regarded as a general framework for an SDI evaluation.

2.4 Total Quality Management (TQM)

SDI implementation requires intra-organisational activities which imply that there are various hierarchical management decision making steps in different levels. Having a proper tool for increasing the quality of the entire procedure leads the SDI to succeed in not only high quality data production and management, but also in facilitating data sharing and access. Therefore, applying a quality management approach for the development of SDI is essential.

TQM consists of three main concepts. *Total* refers to the organisation (e.g., SDI organisation) and includes the whole supply chain and product life cycle. *Quality* means a high degree of excellence in products and also the comparison indicators with the existing standards. *Management* is the process of planning, organising, leading, coordinating, controlling and staffing (Fayol, 1966). TQM is a collection of principles, techniques, processes, methodologies, tools and best practices that over the time have been proven effective in order to increase the internal and external customer satisfaction with a minimum amount of resources.

Sashkin and Kiser (1993) defined TQM as an intense and long-term commitment to quality implementing such a commitment requires the use of tools and techniques. The commitment is more important than the way of utilising the method. TQM is a method to change the organisational values and beliefs in order to let everyone know the most basic aim which is the quality for the customer. Also the ways of working together are determined by what will support and sustain this basic aim (Sashkin and Kiser, 1992). On the other hand, they argued such a system as a shift in the way of thinking and the culture of an organisation rather than using a specific software, technique or specific tool (Sashkin and Kiser, 1993). TQM tools include quality training, process improvement, benchmark management, Statistical Process Control (SPC), Quality Control circle (QCC) and quality information computerisation (Huang and Chen, 2002).

There are many scientists working to improve the TQM method. Edwards Deming (1986, 1993) introduced fourteen management principles as requirements to remain competitive in providing products and services. These include management commitment and leadership, statistical process control, removing barriers to employee partici-

pation and control of their own quality, and continuous improvement of processes. Juran (1989) emphasised planning and product design, quality audits, and orienting quality management toward both suppliers and customers. Crosby (1984) focused on such organisational factors as cultural change, training, and leadership, and the ongoing calculation of quality costs. Important extensions to the TQM framework have included the development of customer-based specifications in the design of a product or process (Taguchi and Clausing, 1990), and benchmarking or the measuring of products/services and processes against those of organisations recognised as leaders (Camp, 1989).

TQM can be used as a general instrument for quality control of the SDI implementation procedures. To utilise such a technique in SDI, the fourteen step approach of Deming can be used in the SDI implementation procedure.

3. DISCUSSION

To investigate the applicability of the mentioned techniques for improving the development and maintenance of SDIs, this section investigates the pros and cons of each technique and their affects on SDI. Table 2 summarises the strength and weak points of each technique.

SDI is a collaborative effort: various organisations and institutions are involved in the development and implementation of SDI. Thus, team work and joint activities have a major role in arriving at the objectives of an SDI. One of the strengths of Six Sigma relates to team building and facilitating team working (see Table 2). This technique can be used for creating the collaborative environment, which is required for the development of SDI. In addition, to develop an SDI, different procedures (spatial data production and updating during daily businesses, inter-and intra organisational data sharing, managing databases and web services) have to be diffused within the organisations. Integrating human elements (culture change, user focus, spatial data-related responsibilities) with process elements (process management, measurement system analysis) can facilitate such diffusion. As highlighted in Table 2, 'integration of human and process elements' is another strength point of Six Sigma, which makes it a suitable technique in the work with implementing SDI.

With respect to the weak points of Six Sigma, 'the need for high quality data for the evaluation' can be considered as the weakness of the most evaluation and improvement methods. Also, since the priority of major activities for implementing an SDI is generally clear, so 'the prioritisation of projects' (Table 2) is not too critical for using Six Sigma for SDI implementation and continuous improvement.

Financial management of SDI is a complex task. Due to diversity of activities required for implementing an SDI, calculating the costs associated for each activity as well as relevant overheads calls for adopting proper financial frameworks. The framework should also provide the possibility of monitoring SDI funding for each activity, based on the mentioned estimations. ABC with the advantage of 'clarification and calculation of the real cost for the products, services, processes and distribution channels' and 'supporting performance measurement' (Table 2) can satisfy such an SDI's requirement. ABC is also easy to understand and well integrated with Six Sigma.

Table 2: Strengths and weaknesses of discussed methods according to the SDI.

Method	Strengths	Weaknesses
Six Sigma	<ul style="list-style-type: none"> – Team Building and Facilitation – Integration of the human and process elements 	<ul style="list-style-type: none"> – Requiring quality data available for the measurement – Prioritisation of projects is critical
ABC	<ul style="list-style-type: none"> – Easy to understand – Accurate measurement of costs – Well integration with Six Sigma and other continuous improvement tools – Supports performance measurement and scorecard – Enables costing processes 	<ul style="list-style-type: none"> – Time consuming for data collection – ABC implementation cost
BSC	<ul style="list-style-type: none"> – The ability to link: <ul style="list-style-type: none"> o Financial and non-financial indicators o Internal and external aspects o Performance drivers and outcomes – Organising disparate data, and providing benchmarks for management discussion and operations. – Highlighting inevitable trade-offs 	<ul style="list-style-type: none"> – Higher weight of financial measure
TQM	<ul style="list-style-type: none"> – Encourages effective participation 	<ul style="list-style-type: none"> – Requires much time and effort

Two weak points of ABC, mentioned in Table 2, might not be critical from an SDI perspective as:

- ‘time consuming for data collection’ is the limitation of the most monitoring and evaluation approaches, not specifically for ABC, and
- ABC implementation cost will be a small percentage of the financial resources required for the SDI implementation.

Due to complex and multi-dimensional nature of the SDI development, its evaluation and monitoring should be based on a multi-view framework linking financial and non-financial indicators, internal and external aspects, and performance drivers and outcomes. BSC not only has the advantage of linking the mentioned factors, but also can highlight inevitable trade-off among them. Therefore, BSC can be a proper framework for the implementation and evaluation of SDIs.

Regarding the weakness of BSC, from an organisational perspective, a financial measure has much greater organisational weight than its new non-financial sibling. However, in SDI, besides financial benefits of spatial data sharing, non-financial benefits of spatial data usage in decision making and planning is also of high value. Furthermore, social benefit gained from SDI has more weight than any financial indicator. Governments spend much money for SDI development to promote the society and better life for citizens, so the financial perspective is important, but not the most significant dimension of SDI implementation.

Finally, TQM encourages effective participation by involving people in the decision making process for development of SDI and improving the quality of their work environment provides them with a sense of value and purpose. Similar to the other methods, TQM is also a long-term procedure and implementation of TQM takes too much time and effort.

With respect to this description, the mentioned techniques can be used for different aspects of SDI implementation, monitoring and improvement.

4. CONCLUSIONS

This article proposes instruments and frameworks from the business management field for the implementation and evaluation of SDIs. We first reviewed different strategic and continuous improvement methods including Six Sigma, ABC, BSC and TQM. Then the applicability of each technique for the implementation of SDI was investigated. The primary investigation shows that each of these techniques can be used in some aspects of SDI implementation. In a nut shell, an integrated general framework for the SDI implementation consists of the Six Sigma as a core methodology. For implementing an SDI, the DMAIC (Define–Measure–Analyse–Improve–Control) approach can be used; ABC (Activity Based Costing) for economic management of SDI; BSC (Balanced Scorecards) for monitoring the progress of SDI and TQM (Total Quality Management) for the quality management of the entire procedure of SDI implementation.

Table 3 illustrates the summary of the usages and value of each method for SDI implementation and continuous improvement.

Table 3: Summary of the usage and value for SDI from discussed methods.

Methods	Value for SDI (Where it can be used)	Usage Domain
Six Sigma	A general framework for the SDI Implementation	Core methodology
ABC	Economic management and evaluation of SDI	Define, Measure, Improve
BSC	Monitoring the progress of SDI	Measure, Analysis, Control
TQM	Quality management of the whole procedure of SDI implementation	Measure, Analysis, Improve, Control

It is worth to be noted that the discussed techniques are originally used for business management and continuous improvement within an organisation. However, also for the implementation of SDI with its collaborative and intra-organisation nature, applying these techniques may be beneficial and worthwhile to be considered by the SDI community.

ACKNOWLEDGMENTS

The authors would like to acknowledge the European Union funding program, Erasmus Mundus “External cooperation Window” (EM ECW lot7) for the financial support. They would also like to thank the members of the Lund University’s GIS Centre, especially Dr. Lars Harrie, for support and valuable comments in preparation of this article.

REFERENCES

- Barney, M. (2002). Motorola’s second generation. *Six Sigma Forum Magazine*, 1(3): 13-16.
- Bjornenak, T. and F. Mitchell (2002). The development of activity-based costing journal literature, 1987-2000. *European Accounting Review*, 11(3): 481-508.
- Blakeslee, J. (1999). Implementing the Six Sigma solution, *Quality Progress*, 32(7): 77-85.

- Camp, R. C. (1989). *Benchmarking: The search for industry best practices that lead to superior performance*. Milwaukee, WI: Quality Press.
- Cooper, R. and R. Kaplan (1992). From ABC to ABM, *Management Accounting*, 74: 54-57.
- Cooper, R. and R. Kaplan (1991). Profit priorities from activity based costing, *Harvard Business Review*, 69(3): 130-135.
- Cooper, R. (1988). The rise of activity-based costing – part two: when do I need an activity-based costing system? *Journal of Cost Management* (fall) 2(3): 41-48.
- Crosby, P. B. (1984). *Quality without tears: The art of hassle-free management*. Milwaukee, WI: Quality Press.
- Deming, W. E. (1993). *The new economics for industry, government, education*. Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study.
- Deming, W. E. (1986). *Out of the Crisis*. Cambridge, Massachusetts Institute of Technology: Center for Advanced Engineering Study.
- Fayol, H. (1966). *Administration industrielle et générale - prévoyance organisation-commandement, coordination – contrôle*, Paris: Dunod.
- Fernández, T.D. and J. Cromptoets (2008). "Evaluating Spatial Data Infrastructures in the Caribbean for sustainable development", *Proceedings of the 10th GSDI conference*, February 25-29, 2008, St. Augustine, Trinidad.
- Folaron, J. (2003). The evolution of Six Sigma. *Six Sigma Forum Magazine*, 2(4): 38-44.
- Georgiadou, Y., Rodriguez-Pabón, O. and Lance, K.T. (2006). Spatial data infrastructure SDI and e-governance: a quest for appropriate evaluation approaches, *URISA journal*, 18(2): 43-55.
- Georgiadou, Y., Puri, S.K. and S. Sahay (2005). Towards a potential research agenda to guide the implementation of spatial data infrastructures - A case study from India, *International Journal of Geographical Information Science*, 19(10): 1113-1130.
- Giff, G. and R. Lunn (2008). "Designing Performance Indicators for Local SDI Assessment: A City of Fredericton Case Study", *Proceedings of the 10th GSDI conference*, February 25-29, 2008, St. Augustine, Trinidad.
- Grus, L., Cromptoets, J.W.H.C. and Bregt, A.K. (2007). Multi-view SDI assessment framework, *International Journal of Spatial Data Infrastructures Research*, 2: 33-53.
- Gygi, C., DeCarlo, N., Williams, B., Stephen and R. Covey (2005). *Six Sigma for Dummies: For Dummies*.
- Hahn, G.J., Doganaksoy, N. and R. Hoerl (2000). The evolution of Six Sigma, *Quality Engineering*, 12(3): 317-26.
- Harry, M.J. and R. Schroeder (2000). *Six Sigma: The Breakthrough Management Strategy Revolutionizing the World's Top Corporations*, New York: Doubleday.
- Hilton, R.W. (2005). *Managerial Accounting: Creating Value in Dynamic Business Environment*, New York: McGraw-Hill Irwin.

- Huang, F. and Y.T. Chen (2002). Relationships of TQM philosophy, methods and performance: A survey in Taiwan, *Industrial Management+Data systems*, 102(3/4): 226-234.
- Ittner, C. D., Lanen, W. N. and D. F. Larcker (2002). The association between activity-based costing and manufacturing performance, *Journal of Accounting Research*, 40: 711-726.
- Johnson, H. and R.S. Kaplan (1987). *Relevance Lost: the Rise and fall of Management Accounting*, Boston, MA: Harvard Business School Press.
- Juran, J. M. (1989). *Juran on leadership for quality: An executive handbook*. Wilson, CT: Juran Institute.
- Kaplan, R. S. and D. P. Norton (2000). Having trouble with your strategy? Then map it, *Harvard Business Review*, 78(5): 167-176.
- Kaplan, R.S. and D.P. Norton (1996). Using the balanced scorecard as a strategic management system, *Harvard Business Review*, 74(1): 75-85.
- Kaplan, R.S. and D.P. Norton (1993). Putting the balanced scorecard to work, *Harvard Business Review*, 71(5): 43-2.
- Kaplan, R.S. and D.P. Norton (1992). The balanced scorecard: measures that drive performance. *Harvard Business Review*, January–February: 171-9.
- Kok, B. and B. van Loenen (2005). How to assess the success of National Spatial Data Infrastructures? *Computers, Environment and Urban Systems*, 29: 699-717.
- Lance, K.T., Georgiadou, Y. and Bregt, A. (2006). Understanding how and why practitioners evaluate SDI performance, *International Journal of Spatial Data Infrastructures Research*, 1: 65-104.
- Letza, S. and K. Gadd (1994). Should activity-based costing be considered as the costing method of choice for total quality organizations? *The TQM Magazine*, 6(5): 57-63.
- Mansourian, A., Rajabifard, A., Valadan Zoje, M.J. and I. Williamson (2006). Using SDI and web-based systems to facilitate disaster management, *Journal of Computers and Geosciences*, 32: 303-315.
- Masser, I., Rajabifard, A. and I. Williamson (2007). Spatially enabling governments through SDI implementation, *International Journal of Geographical Information Science*, 22(1): 5-20.
- Masser, I. (2005a). "Some priorities for SDI related Research", *FIG Working Week 2005 and GSDI-8*, Cairo, Egypt April 16-21, 2005.
- Masser, I. (2005b). "The Future of Spatial Data Infrastructures", *ISPRS Workshop on Service and Application of Spatial Data Infrastructure*, XXXVI (4/W6), 14-16 Oct 2005, Hangzhou, China: pp. 7-15.
- McDougall, K. (2006). A Local-State Government Spatial Data Sharing Partnership Model to Facilitate SDI Development, The University of Melbourne.
- Najar, C., Rajabifard, A., Williamson, I. and C. Giger (2006). "A Framework for Comparing Spatial Data Infrastructures: An Australian-Swiss Case Study", *Proceedings of GSDI 9 - International Conference for Global Spatial Data Infrastructure*, 3-11 Nov 2006, Santiago, Chile.

- Nedovic-Budic, Z., Feeney, M.E.F., Rajabifard, A. and I. Williamson (2004). Are SDIs serving the needs of local planning? Case study of Victoria, Australia and Illinois, USA, *Computers, Environment and Urban Systems*, 28: 329-351.
- Onsrud, H. (ed.) (2007). *Research and Theory in Advancing Spatial Data Infrastructure Concepts*, ESRI Press: Redlands, California.
- Ooi, G and C.Soh (2003). Developing an activity-based costing approach for system development and implementation, *SIGMIS Database*, 34(3): 54-71.
- Sashkin, M. and K.J. Kiser (1993). *Putting total quality management to work*. San Francisco: Berrett-Koehler.
- Sashkin, M. and K.J. Kiser (1992). *Total quality management assessment inventory*, Seabrook, MD: Ducochon Press.
- Swenson, D. (1995). The benefits of activity-based cost management to the manufacturing industry, *Journal of Management Accounting Research*, 7: 167-180.
- Taguchi, G. and D. Clausing (1990). Robust quality, *Harvard Business Review*, 68(1): 65-75.
- Van Loenen, B. (2006). Developing geographic information infrastructures, the role of information policies. PhD thesis, Delft University of Technology, Delft, the Netherlands.
- Wagner, R.M. (2005). "Need for a Framework SDI Business Model", *Joint EuroSDR/ EuroGeographics Workshop*, BKG Frankfurt, Germany, 23-25 February.
- Warnest, M. and Rajabifard, A. and I. P. Williamson (2002). "Snapshot of SDI Development in Australia: Models, Partnerships and Lead Agencies Advancing Implementation", *Proceedings of Joint AURISA and Institution of Surveyors Conference*, Adelaide, South Australia.

Paper II

Toomanian, A., Mansourian, A., Harrie, L., Rydén, A., (2011). Using Balanced Scorecard for Evaluation of Spatial Data Infrastructures: a Swedish Case Study in accordance with INSPIRE. *International Journal of Spatial Data Infrastructures Research*, 6, pp 311-343.

Using Balanced Scorecard for Evaluation of Spatial Data Infrastructures: a Swedish Case Study in accordance with INSPIRE*

Ara Toomanian¹, Ali Mansourian^{1,2}, Lars Harrie¹, Anders Rydén³

¹ Dept. of Earth and Ecosystem Sciences, Lund University, Lund, Sweden,
ara.toomanian@nateko.lu.se, lars.harrie@nateko.lu.se

² Faculty of Geodesy and Geomatics Engineering, K.N. Toosi University of Technology, Tehran, Iran, mansourian@kntu.ac.ir

³ Swedish Mapping, Cadastre and Land Registration Authority, Gävle, Sweden,
anders.o.ryden@lm.se

Abstract

Spatial Data Infrastructures (SDIs) have been identified essential for environmental management and development activities around the world. Meanwhile, development of frameworks to monitor and evaluate SDIs is currently an important research area. This study proposes Balanced Scorecard (BSC) as a framework for evaluation and monitoring the implementation of SDIs. The concept and advantages of BSC for strategy implementation is described in the paper. Furthermore, a general framework for the evaluation of Swedish NSDI in line with the INSPIRE (Infrastructure for Spatial Information in Europe) directive is depicted. The case study shows that BSC is applicable for evaluating Swedish NSDI where the SDI implementation is defined as a long-term project.

Keywords: Spatial Data Infrastructure, Evaluation, Balanced Scorecard, INSPIRE

*This work is licensed under the Creative Commons Attribution-Non commercial Works 3.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.

1. INTRODUCTION

A Spatial Data Infrastructure (SDI) supports a wide variety of users, including environmental managers, to access, retrieve, and disseminate spatial data in a simple and secure fashion. It aims to establish the relationship between people and data through appropriate policy-making, standardization activities and the creation of accessing networks (Williamson et al, 2003; Masser, 2010). SDI is also an integrated, multi-level hierarchy of interconnected infrastructures based on collaboration and partnerships among different stakeholders (Rajabifard et al, 2003; Harvey and Tulloch, 2006; Vandenbroucke et al, 2009).

The development of Spatial Data Infrastructures (SDIs) has evolved as a central driving force in the management of spatial information over the last decade (Williamson et al, 2005). Crompvoets et al (2004) indicate that more than half the world's countries are involved in some form of SDI development. Budhathoki and Nedovic-Budic (2007) claim that such a wide interest in developing SDIs is due to functional SDI being an important asset in societal decision making and policy making, effective governance, citizen participation processes and private sector opportunities.

"Monitoring and evaluation of SDIs" is identified as one of the new research topics in current SDI literature that needs to be further expanded and developed from both theoretical and operational perspectives. Georgiadou et al (2005) also clarified that the downside of all SDI initiatives has been that there have been few instruments to monitor its progress and objectives and few frameworks to evaluate the degree of its success or failure. In other words, there is not a clear approach for SDI evaluation and monitoring. With this in mind, in recent years various researchers have embarked on an initiative to define SDI assessment and recommend tools to assist the measurement of SDI performance (Giff and Crompvoets, 2008; Crompvoets et al, 2008).

In this study, we describe a novel approach for evaluating and monitoring SDIs based on *Balanced Scorecard* (BSC). BSC aims to present management with a concise summary of the measured values of a business, and to facilitate alignment of business operations with the overall strategy. In recent years, BSC has become a successful performance measurement method not only in developed countries (Brijesh et al, 2008) but also in developing countries (Luu et al, 2008).

This paper explains how BSC can assist SDI managers and coordinators to evaluate the degree of success of an SDI both from a producers' perspective by assessing the organizations involved, and from the users' perspective by analysing their willingness of to use spatial products. It also describes how to facilitate the identification of factors that hinder progress of an SDI, or the driving

forces that motivate this. In this regard, we apply the BSC concept with the *Infrastructure for Spatial Information in Europe* indicators (INSPIRE, 2008a). To study the applicability of the approach, a case study has been carried out on the Swedish national SDI, which is defined as a long-term project in the Swedish National mapping agency.

The paper is organized as follows. It opens with the requirement of evaluation in INSPIRE and a review of SDI assessment methods and implementations. In section 3 the Balanced Scorecard (BSC) is described. Our proposal for an evaluation framework based on BSC is explained in section 4. This section also includes implementation details and a case study of the Swedish NSDI. The final sections consist of a discussion, future research and conclusions.

2. SDI EVALUATION

2.1. INSPIRE Directive

INSPIRE is a Directive of the European Parliament and the Council (Directive 2007/2/EC, 2007), aiming to assist policy-making in relation to policies and activities that may have a direct or indirect impact on the environment. INSPIRE is based on SDIs that are created by Member States and that are made interoperable with common implementation rules. This directive establishes the legal framework for setting up an operational European SDI.

Different parts of the Directive, directly and indirectly, emphasize the need of the evaluation and monitoring of individual SDIs by the relevant member state via national measures and supplementary measures at the community level. To support this, INSPIRE (2008b) has published some indicators for monitoring the implementation of the INSPIRE Directive and the use of the infrastructure. The indicators in accordance with monitoring are: existence, accessibility and conformance of metadata; extent, accessibility and conformance of spatial data sets; use and conformance of spatial data services. There are also 3-yearly reports required on various aspects from the member states regarding their progress (INSPIRE, 2008b).

2.2. Previous Studies on SDI Evaluation

In recent years, SDI evaluation has become a major challenge and researchers have suggested different models and approaches for evaluating SDIs. Crompvoets et al (2008) review several approaches proposed by a number of researchers such as the multi-view framework for assessing SDIs developed by Grus et al (2007). They described the theoretical basics of the multi-view framework for SDIs assessment by expressing the need for a better understanding of the objectives, complexity, multi-faceted nature, dynamics and the current use of SDI in the context of SDI assessment as well as the demands for SDI assessments and the necessity to develop a framework to assess SDIs.

They group the current SDI assessment approaches into nine approaches that stress different dimensions of the SDI: the NSDI readiness index, clearinghouse suitability, the INSPIRE state of play, organizational perspective, a framework based on land administration systems, SDI Performance Based Management, legal approach and SDI effectiveness from a user perspective. There have been several studies based on most of these approaches.

The NSDI readiness index has been studied by Delgado-Fernández et al (2008) where they state it as an important factor to be taken into account for SDI implementation. In this regard, beside the technological issues also organisational, informational, financial and human factors are a composite integrator to create the readiness index.

The suitability of national clearinghouses was calculated twice (2002 and 2007) based on seventeen characteristics where the result suitability index is defined as a measurement of the quality and performance of this electronic facility (Crompvoets et al, 2008). Sustainability index evaluation in different time stamps is a good clearinghouse performance indicator that supports the managers to develop successful strategies in national clearinghouse implementation and enhances national clearinghouses and national SDIs in many countries.

Assessing the INSPIRE state of play of SDIs was based on six relevant elements, namely *organization, legal framework and funding mechanisms, spatial data, metadata, access and other services* and *standards*, characterizing the components of the European SDI and in particular the INSPIRE directive (Vandenbroucke et al, 2008).

The organizational perspective approach was studied by van Loenen and van Rij (2008). They proposed a model which focuses on the classification of SDIs on the four stages of SDI development: stand-alone/initiation, exchange/standardization, intermediary and network. However, according to this model, SDI development has to be as 'mature' as possible that is aimed at. Steudler et al (2008) assess Spatial Data Infrastructure based on measuring indicators determined for five assessment areas: policy level, management level, operational level, other influencing factors and assessment of performance which are originally come from land administration systems. Such a model of comparison and evaluation helps for better understanding the different aspects as well as finding best practice for certain tasks of SDI and improving the whole system.

The *SDI Performance Based Management* was described as a systematic approach, (Giff, 2008). This technique facilitates infrastructure practitioners to operate an infrastructure to identify, analyze and manage its strengths and weaknesses. It uses indicators for performance improvement by developing a

framework for key performance indicators within an ongoing process of establishing strategic performance objectives and measuring performance.

The next approach was the essential legal framework for developing an SDI focused by Janssen (2008). The assessment uses is not based on empirical evidence but makes use mainly of legislation, case law and jurisprudence. The assessment distinguishes three levels of legal assessment: compliance, coherence and quality. The final approach is the SDI effectiveness from a user perspective where it focuses on the effective use of SDIs by recognizing both the current and potential users and trying to fulfil their needs regarding data and services by determining contextual factors and outcomes (Nedović-Budić et al, 2008).

Giff and Cromptvoets (2008) present a structured concept of SDI assessment. They use an in-depth analysis of Performance Indicators based on an eleven-step conceptual framework for designing performance indicators (PI) for assessing SDIs in Canada. They also present and critically analyze a Framework to guide SDI coordinators in the intricate task of designing PIs for their initiatives. There are other studies conducted in this field such as evaluation strategy for SDIs based on a maturity matrix (Van Loenen, 2006). The maturity of the SDI was evaluated according to several measures such as technical (e.g. data and metadata), non-technical (e.g. organizational) and policy (free data policy or cost recovery policy). Based on this strategy, Van Loenen evaluated some SDIs in Europe and United States. Geudens et al (2009) used a multi-criteria analysis to evaluate SDI Policy strategies that takes into account all the different criteria and actors involved in the complex SDI decision-making context in an integrated framework.

Despite of several SDI assessment approaches, described above, there is still lack of an integrated method which covers different aspects of an SDI assessment that can measure the progress of an SDI. Moreover, there are a limited number of assessment approaches that are able to demonstrate whether SDIs indeed realize the intended goals (Grus et al, 2011). Therefore, a comprehensive goal-oriented SDI evaluation should be based on the assessment of several dimensions in a common framework that has to be defined within a long-term project to control the progress of an SDI. In next section we describe one such framework that can be appropriate for evaluating and monitoring SDIs.

3. BALANCED SCORECARD: A FRAMEWORK FOR STRATEGIC PERFORMANCE MEASUREMENT

In strategic management, performance measurement aims at achieving a goal. Performance measurement provides managers with concrete data to compare the progress of tasks with organizational objectives. Strategic performance

measurement has three main roles in the strategic chain (Chaichan, 2002): to lead managers in the right direction, to motivate managers and to help top managers identify *critical processes* and *critical success factors* (CSFs). A critical process is a series of activities that directly affects the achievement of goals. CSFs are a limited number of factors that must be measured in order to assess the degree of goal achievement (Mard et al, 2004). A CSF is measured by a set of *key performance indicators* (KPIs). KPIs are quantifiable measures that reflect the critical success factors of an organization.

Balanced Scorecard (BSC) is a well-known framework which has been widely used during the last decade for strategic performance measurement in different disciplines (Lee et al, 2008; Luu et al, 2008; Idalina et al, 2007; Lawson, 2006; Epstein and Wisner, 2001). It has been observed that most of the successful organizations are either adopting BSC or are familiar with it (Silk, 1998; Malmi, 2001; Rigby, 2001; Fernandes et al, 2006).

BSC is a performance measurement framework that provides an integrated view of the business performance of an organization by a set of measures, both financial and non-financial metrics (Kaplan and Norton, 1992, 1996). It also refers to a multi-dimensional framework that uses measure to describe an organization's strategy.

In a BSC approach, the strategy is broken down into different perspectives (Figure 1). The main four perspectives in BSCs are: *Benefit & economy (financial)*, *customer*, *internal process* and *learning & growth*. Kaplan and Norton (2000) state: "Balanced Scorecards tell you the knowledge, skills and systems that your employees will need (learning and growth) to innovate and build the right strategic capabilities and efficiencies (internal processes) that deliver specific value to the market (customer) which will eventually lead to higher shareholder value (financial)".

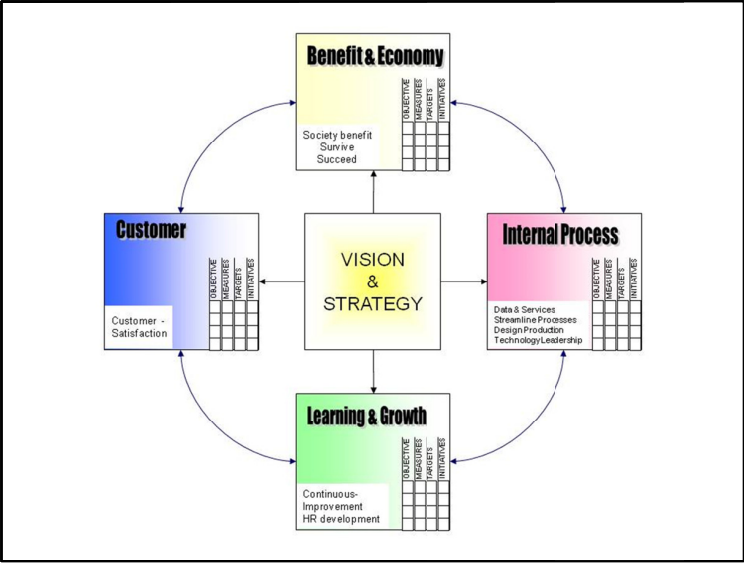
Each perspective is described by five elements. The first element is the *objective* which is a statement of strategic intent. It describes how a strategy will be made operational in an organization. Objectives are the main elements of the strategic plan and the entire strategy can be broken down into many objectives. For each objective, one or more sub-objectives are provided which express different aspects of the objective in more detail.

The next element in a BSC design is the *cause & effect linkage*, which describes the cause and effect relationships between the objectives. Kaplan and Norton (1996b, p. 149) defined the strategy as a set of hypotheses about cause and effect'. So a proper BSC design should contain outcome measures and the performance drivers should be linked together in cause-and-effect relationships (Kaplan and Norton, 1996b p. 31). It is a graphical representation of the influence

of different perspectives that start from the lowest layer of the design table to the top layer. In some cases, internal linkages exist between objectives of a perspective.

Another element of the BSC is the measures using KPI. Measures are performance metrics for the objectives. Generally, in a BSC design there are reasonable numbers of measures explicitly linked to an objective. The measure concept is typically represented by mathematical formulas. Sometimes it is not possible to define a quantifiable measure for an objective. In such cases, it is necessary to utilize some techniques or models to quantify the measure. *Customer satisfaction index* (CSI) can be used to quantify customer satisfaction level. CSI was built by Fornell (1992) and later developed in both theory and applications (Albert, 2002; Band, 2000; Hackl et al, 2000). CSI helps to measure various qualitative parameters and a number of customer-oriented objectives through questionnaire surveys. Normally, the result of such surveys is a level of satisfaction or the percentage of success generated by the products. A rigorous CSI model can be a good measure of quantifying customers' ideas for improving services. CSI can also be used to measure various objectives, from the customer's perspective.

Figure 1: Basic BSC Elements and Their Interaction with Vision and Strategy



The fourth element of BSC design is called *target*. A target is a quantifiable goal for each measure. An ideal situation is that, a combination of targets on the BSC

design is the general goal of an organization. Targets help to monitor progress toward strategic goals by comparing the results of measures with the relevant targets, and can provide good feedback if necessary.

Strategic initiative is the last element of BSC design. Initiative is the action program that drives strategic performance and activities which will lead the organization towards achieving strategic results. All ongoing initiatives in an organization should be associated with the BSC strategy.

4. BSC FOR SDI EVALUATION: A CASE STUDY OF THE SWEDISH NATIONAL SDI

4.1. The Swedish National SDI

To investigate the applicability of BSC for SDI evaluation we carried out a case study on the Swedish National SDI. The implementation of NSDI in Sweden is coordinated by Lantmäteriet (the Swedish Mapping, Cadastre and Land Registration Authority) in which the national contact point (MSCP) is the NSDI unit (Geodata strategy). NSDI is defined as a long-term project by Lantmäteriet. The objectives of the geodata strategy are: to create a national infrastructure for the geodata sector, to contribute to the development of Swedish public administration (e-governance) and to promote close cooperation between the public and private sectors. The strategy should also foster a favourable environment for the creation of value-added geodata by the private sector.

The Swedish government requires evaluation of the national geodata strategy. During 2008, this evaluation was partly made through a cost/benefit analysis. A framework is required for future annual evaluations; this framework should include methodology and recommended indicators (Sandgren, 2008). A major aim of this study is to contribute to the development of such a framework.

4.2. Framework Design

To investigate the applicability of BSC for SDI evaluation, the Swedish NSDI was selected as the case study. Since the Swedish NSDI adhere to the INSPIRE directives the case study is generally applicable for other European SDI. As the first step of the research, different literatures in the field of SDI evaluation were studied. In addition INSPIRE directives and the proposed indicators for SDI assessment were reviewed. Swedish NSDI and the attempts for evaluation of that were studied as well. The concept of BSC and its use for strategic performance measurement were also reviewed. The results of these studies have been described earlier in sections 2 and 3. At the second step, an SDI evaluation framework was prepared as a draft. Then some meetings were held with the Swedish NSDI coordinators to get their opinions about the framework and to share the ideas about the proposed objectives, goals and measurements. Within

these meetings, the framework was finalized. In the third step, the Swedish NSDI coordinators and the authors collected information in accordance to the proposed KPIs for implementing BSC and for evaluating the NSDI.

Table 1 shows the BSC framework, which is designed for SDI evaluation. The framework has the following perspectives:

Learning & Growth: This perspective deals with capacity building at the individual (people) level. Considering the role of skilled employees and highly aware managers for a successful SDI, learning & growth is an important evaluation perspective for SDIs

Internal Process: As the name indicates, internal processes for implementing SDIs are measured in this perspective; this includes standardization activities, data management factors (production, updating, storage, etc.), establishing accessing networks and spatial web services, institutional arrangements and collaborative activities.

Customer: The customer perspective is a key standpoint in SDI evaluation. All the investments and technological developments are to deliver the spatial data products to the user. Therefore, customer satisfaction is a crucial factor that must be measured when evaluating an SDI.

Benefit and economy: The main idea of an SDI is to benefit various sections of society. Moreover, the economic perspective also tries to keep SDIs active and updated with respect to the financial situation. The benefit of SDI can be considered for not only the data producers but also for the public as well as the end users.

The following elements are considered for each perspective:

Objective and sub-objectives: In general, objective terms are taken from the strategic plan and vision of an SDI. In this case study, the objectives are derived from the INSPIRE directive and Swedish National Geodata Strategy as well as general SDI goals in order to maintain a broad SDI evaluation framework. For each objective, a number of sub-objectives are also stated.

Cause and effect linkage: The cause and effect linkage initiates from the Learning & Growth perspective where professionals support internal processes for implementing SDIs. Appropriate internal processes cause a wide use of data and analysis by customers. Afterwards, satisfied customers use data and services in decision-making and planning which result in benefits to society and economic success.

Measures: Measures are quantifiable values describing the progress of SDI activities in relevant objectives. We used the indicators proposed by INSPIRE and other SDI evaluation studies for measures.

Targets: The quantitative goal for each measure is set for the SDI strategic planning. We have not mentioned targets in Table 1, because most of them were not clearly defined. However, this does not affect the overall aim of the research.

Initiatives: These are midterm programs to facilitate fulfilment of the strategic plan. In this study, four strategic initiatives are stated for each perspective. Proposing initiatives was not within the scope of this study, the initiatives here are just sample initiatives which may be considered for an SDI such as INSPIRE.

Table 1: BSC Model for NSDI Evaluation in Accordance with INSPIRE Indicators

	Objective	Sub-objective	Cause & Effect linkage	Measure (KPI)	Initiative
Benefit and Economy	Benefits to society	Effect of usage of spatial data and services	Benefits from the implementation of spatial data and economic success	KPI _{BE} 1: Monetary benefits of spatial data and service usage for the society (NMAs, governmental organizations, authorities, private sectors, public, end users).	Economic adjustment initiative program
	Cost of SDI	Funds required for implementing SDI		KPI _{BE} 2: Activity Based Costing to estimate SDI cost	
	Market Fund	Funds received from SDI market		KPI _{BE} 3: Funds received from SDI market by calculating the income of spatial data products and services sale in a year	
	SDI Fund	Funds from government and individual stakeholders intended for SDI		KPI _{BE} 4: Funds for SDI in a year, divided by the funds planned for that year	
Customer	Customer Satisfaction	Accessibility of data	Wide range of visualization and more analysis made by the customer	KPI _C 1: Satisfaction level of the accessibility of data KPI _C 2: Satisfaction level from different data services (discovery, view, download, processing) *	Enterprise CRM program
		Data characteristics		KPI _C 3: User Satisfaction level of data quality	
		Spatial data standards		KPI _C 4: Inconsistencies between data standards and user's requirements	
	Usage of Spatial Data Services	Use of spatial data services by customers		KPI _C 5: NSrd/NS KPI _C 6: NSrv/NS; KPI _C 7: NSrw/NS	
Internal Process	Data & Services	Availability and usage of spatial data services	Proper Spatial Data and Service Production in accordance with existing standards with high level metadata	KPI _{IP} 1: NSm/NS; KPI _{IP} 2: NScm/NS; KPI _{IP} 3: NSd/NS; KPI _{IP} 4: SDd/SD; KPI _{IP} 5: SDv/SD; KPI _{IP} 6: SDw/SD; KPI _{IP} 7: NSp; KPI _{IP} 8: NSc/NS;	Enterprise clearinghouse system
		Spatial data sets		KPI _{IP} 9: SDm/SD; KPI _{IP} 10: SDcm/SD; KPI _{IP} 11: SDc/SD; KPI _{IP} 12: SDe/E	Continues improvement

	Design Production	Quality control and quality assurance of work procedures		KPI _{IP} 25: Following ISO 9000 Guidelines	
		Geo-data Standardization		KPI _{IP} 26: Usage of standard for metadata (ISO 19115) KPI _{IP} 27: Semantic and synthetic interoperability standards KPI _{IP} 28: Usage of coordinate reference system	
Learning & Growth	HRD	Skills formation for spatial data management and usage	Capacity Building	KPI _{LG} 1: Number of skilled employees in SDI department per year	HR training program
	Skill Formation & Culturing	Employee empowerment		KPI _{LG} 2: Number of annual training courses Attended for employees KPI _{LG} 3: Number of annual workshops and seminars organized	
		Data quality		KPI _{IP} 13: Quality of data using measures from ISO 19138	
		Maintenance of data		KPI _{IP} 14: Lead time for updating spatial database KPI _{IP} 15: Planned lead time for updating spatial databases	
	Streamline Processes	Collaboration		KPI _{IP} 16: AG; KPI _{IP} 17: Relationship with third parties; KPI _{IP} 18: DS; KPI _{IP} 19: LBN; KPI _{IP} 20:LBR	
		Institutional process		KPI _{IP} 21: OS; KPI _{IP} 22: PR; KPI _{IP} 23: OR; KPI _{IP} 24: DB	
NS: Number of all spatial data services NSm: Count of all spatial data services that have metadata NScm: Count of all spatial data services that have conformant metadata NSd: Count of all spatial data services for which a discovery service exists NSrd: Sum of the annual number of service requests for discovery services NSrv: Sum of the annual number of service requests for view services NSrw: Sum of the annual number of service requests for download services NSp: Number of spatial processing services NSc: Count of all spatial data services that are conformant			SD: Number of spatial data sets SDm: Count of spatial datasets that have metadata SDcm: Count of spatial datasets that have conformant metadata SDc: Count of spatial datasets that have conformant metadata and are conformant SDd: Count of spatial datasets for which a discovery service exists SDv: Count of spatial datasets for which a view service exists SDw: Count of spatial datasets for which a download service exists SDe: Sum of the actual area covered by all the spatial data sets E: Sum of the relevant area of all the spatial data sets		
AG: Number and type of agreements between national and local authorities DS: Data sharing arrangements that have been, or are being, created between national and local authorities LBN: List of barriers that inhibit the sharing of spatial data and services between national authorities LBR: List of barriers that inhibit the sharing of spatial data and services between national and local authorities			OS: Existence of an organizational structure within public authorities (data producers) for intra-organizational spatial data management and coordination PR: Current procedures/mechanisms within public authorities (data producers) for offering spatial data to users OR: Organizational regulations for data sharing DB: Current processes for producing and updating spatial data during daily operations in organizations		

4.2.1. *Evaluating SDI from Learning and Growth Perspective*

Groot and van der Molen (2000) define capacity building as “The development of knowledge, skills and attitudes in individuals and groups of people relevant in design, development, management and maintenance of institutional and operational infrastructures and processes that are locally meaningful”. In general, capacity building is a concept related to education, training and human resource development (HRD). Moreover, capacity building at the individual level has been identified as an essential and basic requirement for a successful SDI.

Two objectives have been considered with respect to the learning and growth perspective (Table 1): *skill formation & culturing* and *HRD*. The skill formation and culturing will bring about HRD for SDIs. The more skilled employees in SDI, the better services delivered within an organization. The second objective relates to training courses, seminars and workshops for skills formation and increasing employee awareness in the SDI. Being aware of SDIs and their advantages and having thorough knowledge of production, maintenance and usage of spatial data leads better support from stakeholders and data custodians. The indicators are given in Table 1 (KPI_{LGS} 1, 2, 3).

4.2.2. *Evaluating SDI from the Internal Process Perspective*

Internal process is a significant part of an implementation. Therefore, defining objectives for this perspective depends on the status of the SDI implementation level (scale). The objectives used for internal processes are (Table 1): *design production*, *streamline processes* and *data & services*.

The objective *design production* relates to *standardization, quality control and quality assurance of work processes (SDI activities)*. In this group, *usage of standards for metadata, semantic and syntactic interoperability and usage of an appropriate coordinate (geodetic) reference system* (KPI_{IPs} 26, 27, 28) are the key indicators of standardization category (INSPIRE (2009)). In addition, measures from *ISO 9000 guidelines* (KPI_{IP} 25) should be utilized for quality control of work procedures, including maintenance of the infrastructure for spatial information.

Another objective is *streamline processes* which relate to the *institutional process* and *collaboration* environment for data sharing. Here, the primary institutional arrangements that a data producer needs to have for active participation in NSDI as well as the barriers that inhibit the sharing of spatial data and services are monitored. Institutional processes must be measured through investigation (KPI_{IPs} 21, 22, 23, and 24):

- existence of a proper organizational structure within public authorities (data producers) for inter-organization spatial data management/coordination;

- current organizational processes/mechanisms for offering data to users;
- intra-organizational regulations for data sharing; and
- current intra-organizational processes for producing and updating spatial data during daily activities of an organization (Luzet, 2004).

Creating a collaborative environment for spatial data management and sharing is one of the main aims of SDIs. This aim can be measured through a number of collaborative activities between organizations. In the context of a joint project, partnership efforts or any other form as follows (KPI_{IP}s 16, 17, 18, 19, and 20):

- number and type of agreements between public authorities and municipalities;
- relationships with third parties;
- data sharing arrangements that have been, or are being, created between public authorities and local institutions; and
- list of the barriers that inhibit the sharing of spatial data and services between public authorities.

Data & services is the last objective in the internal process perspective. This objective has been emphasized by the INSPIRE Monitoring and Reporting Implementing Rule (INSPIRE, 2008b), in which different indicators have been proposed for the measure. This paper proposes the following sub-objectives: *availability and usage of spatial data services, spatial datasets, data quality and maintenance of data*.

Technically, spatial data services facilitate data discovery, visualization, access and modification for users. Therefore, *the ratio of the count of all spatial data sets for which a service exists and the number of spatial data sets* is an appropriate indicator for measuring the availability of spatial data services. This indicator can be used for measuring discovery, view and download services individually (KPI_{IP}s 4, 5, 6). Meanwhile, the *number of processing services* is another indicator (KPI_{IP} 7). Spatial data services should also be evaluated with *availability of metadata for the services* (KPI_{IP} 1), *availability of conformant metadata for the services* (KPI_{IP} 2), *existence of a discovery service for spatial data services* (KPI_{IP} 3) and *conformance of the spatial data services* (KPI_{IP} 8). Finally, spatial data sets can be evaluated with the *existence of metadata for the data sets* (KPI_{IP} 9), *the conformance of metadata* (KPI_{IP} 10), *the conformance of datasets* (KPI_{IP} 11) and *the extent of spatial datasets* (KPI_{IP} 12).

To evaluate data quality, measures from ISO 19138 (KPI_{IP} 13) are proposed. To monitor maintenance of data, *the current and planned lead times for updating spatial database* (KPI_{IP} 14, 15) is suggested as an indicator. *Lead time* is the time duration between when a feature is changed in reality (e.g. a building is built) and when the changes are reflected in the spatial database. For each data set an

appropriate lead time should be defined and then the measure conducted, based on the existing and planned lead times. Comparing these two values is also useful to determine the extent that data providers accept their data updating responsibilities.

4.2.3. *Evaluating SDI from the Customer Perspective*

In an SDI, a customer refers to the users of spatial data and services, which may be governmental, private or academic organizations as well as the public. Through investigating customers' needs and feedback, managers can review processes and procedures to achieve that aim.

For SDI managers and coordinators, *customer satisfaction* is the most significant objective and should be measured to evaluate an SDI. Measuring customer satisfaction levels comparison with defined targets is an appropriate method for identifying the degree of success of an SDI and monitoring its progress. Customer satisfaction is provided in indicators KPI_C 1-4 in Table 1.

Usage of spatial data services, as proposed by INSPIRE (2008b), is the final objective from the customer perspective. It can be measured through determining the sum of the annual number of service requests for individual discovery, view and download services divided by the number of all services (KPI_C s 5, 6, 7).

4.2.4. *Evaluating SDI from the Benefit and Economy Perspective*

Benefit and economy is the last perspective to be considered for SDI evaluation in a BSC framework. In this stage, the intention is to test the general framework and find proper measures for the benefit and economy perspective which can be computable according to the current available datasets. The main objectives proposed for this perspective are: SDI fund, market fund, cost of SDI and society benefits.

SDI fund refers to the amount of money that government and public authorities spend on establishing an SDI. *Annual measure of expenses for SDIs divided by the planned funding* (KPI_{BE} 4) is the proposed indicator for measuring SDI fund. It shows how much government and organizations have accepted their financial responsibilities for developing SDIs. Market fund can be determined by *measuring revenues from the SDI market* (KPI_{BE} 3) through, for example, providing users with spatial data and services. Such revenue, in itself, can be a financial source for funding an SDI. More into details, the income from selling maps and different spatial data products as a proper measurable indicator gives a very good general overview of the amount of money received from SDI market.

Having a clear vision of the cost of SDIs is essential for their development. Cost of SDIs can be estimated through Activity Based Costing (ABC) (KPI_{BE} 2)

(Toomanian and Mansourian, 2009; Cooper and Kaplan, 1991). ABC assumes that activities consume resources and products consume activities. It uses a two-stage procedure to calculate product costs: it traces resource costs to activities, and then traces costs of activities to products. Activities are often derived from information gathered from interviews, questionnaires, and time cards (Cooper and Kaplan, 1991).

The true implementation of an SDI will definitely provide society with environmental and economic benefits. Determining these benefits in regular time periods and for general examples and applications not only provides an indicator to monitor true implementation of the SDI, but also motivates policy-makers and stakeholders to support SDIs. Translating all the benefits into monetary benefits (KPI_{BE} 1) generally provides a better understanding of achievements. This measure is estimated by recalculating the cost and benefits (revenue) of various national projects and compare with the time where SDI was not implemented.

4.3. Data Sources for Collecting Values of the Framework Indicators

Evaluating the Swedish NSDI in accordance with the framework proposed in Table 1 requires information on each proposed KPI. Most of all, the sources of information have to be recurrent and reliable. The acquisition of the information required must not put a burden on those organizations involved, as too many inquiries tend to reduce the willingness of the responding organization to provide information requested. The information must also be easy to compile and analyze.

In order to meet these requirements it was decided to use “official”, existing questionnaires and web services as the main sources of information needed for each KPI. In Sweden there are two main organizations that deal with this type of information in the geospatial data sector:

- Lantmäteriet (The Swedish Mapping, Cadastre and Land Registration Authority), which is the national coordinator of the geodata sector in Sweden and also responsible for the National Geodata strategy.
- The Swedish Development Council for Geographic Information (ULI), which represents the private sector within the field of geo-information and is working for more efficient use of geographic information in Sweden.

The following sources were identified within these two organizations and used in the implementation of the framework:

- During recent decades ULI has regularly sent out questionnaires regarding the use of GIS in Sweden (*Lägesbild GI Sverige*). The latest questionnaire in the series, made in 2007, was used in this study. Information from this

questionnaire was particularly valuable for the *Learning and Growth* and *Internal Process* perspectives.

- The *INSPIRE list* of dataset and services kept by the NSDI unit at Lantmäteriet, and which is part of monitoring obligations, was used for the KPIs related to *Internal Processes*. The INSPIRE list was compiled with the help of the public sector authorities affected by INSPIRE and will be updated annually.
- The Geodesy Department at Lantmäteriet requests all public sector organizations to report on the status of the implementation of the new Swedish reference systems, SWEREF 99 and RH 2000, each year (*Status – Swedish reference system*). The questionnaire is sent to 40 public sector authorities considered to be users of geographical data.
- The Geodesy Department at Lantmäteriet also runs an Internet application where the status of the implementation of the new Swedish reference systems within municipalities is shown (*Reference system at Swedish municipalities*).
- In order to acquire the KPIs for the *Customer Perspective*, the NSDI unit at Lantmäteriet developed a questionnaire that investigated customer satisfaction as registered user of the Geodata portal (*Nöjdhetsundersökning Geodataportalen*). The intention is that the questionnaire will be repeated annually to obtain user feedback on the Geodata portal and its contents. This first year the questionnaire was sent to about 600 registered users and responses were received from 73 of them.
- The marketing section of Lantmäteriet, responsible for the business model developed within Geodata strategy, reports on the number of public sector authorities and municipalities that have signed the agreement to cooperate and share data according to INSPIRE and geodata strategy.

4.4. Determining the Values of KPIs

In the data collection step, some indicators were collected directly while others required extra calculation. In some cases there were limitations as well as changes to the original indicators, especially if there was no obvious way of establishing a target value. In order to test the framework, we collected available information about as many indicators as possible. However, since Lantmäteriet is in an early stage of implementing the NSDI, data required for measuring some indicators are not available, although such indicators are measurable in practice. With this in mind, the result of this study is limited to the available datasets which are presented in Table 2.

4.4.1. Indicators for the Learning and Growth perspective

Available data for HRD was based on the ULI questionnaire results for skill formation in spatial data management and usage. Two indicators were used in

the Swedish NSDI evaluation (KPI_{LG} 1 – 2). The proportion of SDI experts to total employees was also used as supplementary data for this objective.

As an important measure of employee empowerment, the percentage of training courses and workshops per year was used from the ULI questionnaire as an indicator of skill formation and culturing objective (KPI_{LG} 3).

4.4.2. *Indicators for the Internal Processes perspective*

The data for the values of *Availability and usage of spatial data services* are acquired from the *INSPIRE list* of datasets that will be brought into conformance with the INSPIRE specifications. In this case study KPI_{IP} 1-8 were used for the evaluation. The data for the values of *Spatial data sets* are acquired from the *INSPIRE list* of datasets that will be brought into conformance with the INSPIRE specifications KPI_{IP} 9 -12 were used for the evaluation (Table 2).

We were not able to obtain any values for indicators related to institutional processes due to limited resources. The ISO 9000 guidelines were not taken into consideration due to data collection limitations. Finally, it is proposed that the evaluation of data quality should take place using measures from *ISO 19138* (KPI_{IP} 13). However, a proxy measure is used in this case study from the ULI questionnaire regarding the usage of standards in Swedish national and local authorities.

An important component of the Geodata Strategy is the development of a common business model for the geodata sector. The model will be adapted to facilitate cooperation between the public and private sectors. The aim is to incorporate all interested parties in the new agreement and licence models. In some cases, however, there may be a need for successive transition depending on previously reached agreements or other needs. Bearing this in mind, the measures selected to evaluate *Collaboration* in a Swedish context are KPI_{IP} 16 - 18 (Table 2). Moreover, we suggest an additional KPI as 'Private sector companies participating in the data sharing'. The value of this KPI was not available for this case study. The business model will apply from the 1 January 2011. For this reason, no data will be available for evaluation of the current status of *Collaboration* in this study.

According to the Swedish Geodata Strategy, high priority is being given to a rapid transition to SWEREF 99 (the Swedish implementation of the European geodetic reference system ETRS 89) and RH 2000 (the Swedish implementation of the European height system EVRS) reference systems. A homogeneous geodetic reference system facilitates the production, processing and use of geodata and also facilitates compilation of data from different sources. The proposed KPIs for objective standardization have therefore changed to indicate the current state and progress of the transition to a homogenous reference system. Based on the

questionnaire *Status – Swedish reference system* and *Reference system for Swedish municipalities*, KPI_{IP} 28a – 28b have been used for the evaluation (Table 2).

4.4.3. Indicators for the Customer Perspective

The measure of *Accessibility of data* was assessed on the basis of five indicators, KPI_C 1 -2, all derived from the *Satisfaction survey Geodata portal*. The target value for all these is 100 %. The two parameters *Data characteristics* and *Spatial data standards* are also based on the results of the questionnaire *Satisfaction survey Geodata portal*, KPI_{IP} 3 -4.

Table 2: The Indicators Used for the Evaluation of Swedish NSDI in this Study

Perspective	Objective	Sub-objective	Result
Benefit and Economy	Market Fund	Revenue from SDI market	KPI _{BE} 3: Approximately 70 million SEK per year
	SDI Fund	Government funds and individual stakeholders' support for SDI	KPI _{BE} 4: One third of expected (50/141 million SEK)
Customer	Customer Satisfaction	Accessibility of data	KPI _C 1: accessibility of data: 2.43 (Min=1) KPI _C 2: services: discovery: 2.60 (Max=4) view: 2.60 download: 2.60
		Data characteristics	KPI _C 3: User Satisfaction level of data: 2.58 (1<x<4)
		Spatial data standards	KPI _C 4: Inconsistencies between data standards and user requirements: 2.60 (1<x<4)
Internal Process	Data & Services	Availability and usage of spatial data services	KPI _{IP} 1: 76%; KPI _{IP} 2: 5%; KPI _{IP} 3: 66%; KPI _{IP} 4: 18%; KPI _{IP} 5: 30%; KPI _{IP} 6: 5%; KPI _{IP} 7: 0%; KPI _{IP} 8: 7%
		Spatial data sets	KPI _{IP} 9: 74%; KPI _{IP} 10: 16%; KPI _{IP} 11: 1%; KPI _{IP} 12: 99%
		Data quality	KPI _{IP} 13: 53%
	Streamline Processes	Collaboration	KPI _{IP} 16: AG:70% KPI _{IP} 18: DS: 64% KPI _{IP} 17: Relationships with third parties: 28%
	Design Production	Standardization	KPI _{IP} 28 (a): SWEREF 99; National authorities: 8/40; Local authorities: 156/290 KPI _{IP} 28 (b): RH 2000; National authorities: 4/40; Local authorities: 28/290
Learning & Growth	HRD	Skills formation for spatial data management and usage	KPI _{LG} 1: no. of skilled employees per year: 38,000 KPI _{LG} 2: no. of skilled employees in SDI per year: 2705 PS: Proportion of SDI experts to the total: 7%
	Skills Formation & Culturing	Employee empowerment	KPI _{LG} 3: Percentage of training courses and workshops per year: 62%

The questionnaire used the satisfaction categories: *completely satisfied* (4), *reasonably satisfied* (3), *not very satisfied* (2) and *not satisfied at all* (1). We used the *Ordered Weighted Average* (OWA) method to calculate a real number for each indicator according to the questionnaire outcome (cf. Yager 1998).

Data for the values of *Use of spatial data services by customers* are acquired from the *INSPIRE list* of data services that will be brought into conformance with the INSPIRE specifications. The implementation rules for Monitoring and Reporting require each member state to monitor the number of requests for each

service, which will be done semi-automatically available publically (Lantmäteriet, 2010).

4.4.4. Indicators for the Benefit and Economy Perspective

Determining the real value of market fund objectives, the total revenue from the SDI market, is a complex and challenging issue. KPI_{BE} 3 indicates the approximate market revenue from a survey at Lantmäteriet. There is another major financial objective related to the subsidy for SDI funding from the government. In addition to governmental funds, in many cases individual stakeholders invest in SDIs. The result of another survey from Lantmäteriet indicates that around one third of the expected funds came from different sources KPI_{BE} 4.

4.5. Analyses of the SDI Progress

The framework as in Table 1 was added to the BSC designer software from AKS-Labs as illustrated in Figure 2. The values for the indicators (Table 2) were also entered in the software. As Figure 2 shows, it is possible to assign a weight to each objective in order to describe its importance relative to the others. Weight assignment to KPIs is also feasible. In this study we used the same weights for all indicators.

Figure 2: Implementation of BSC for SDI Evaluation

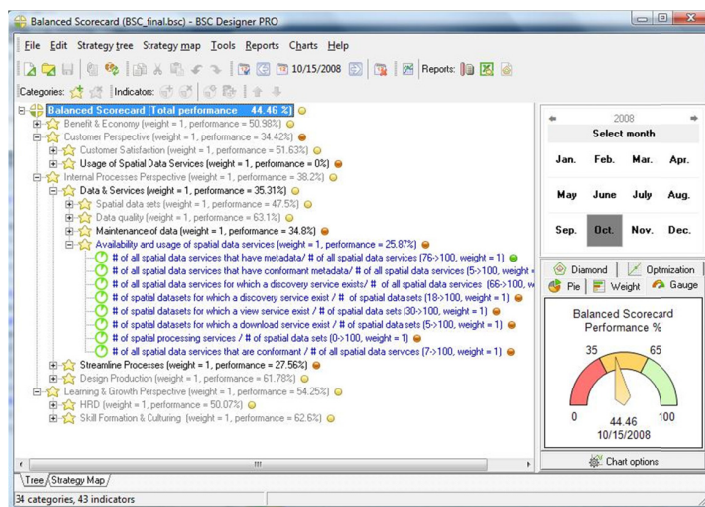
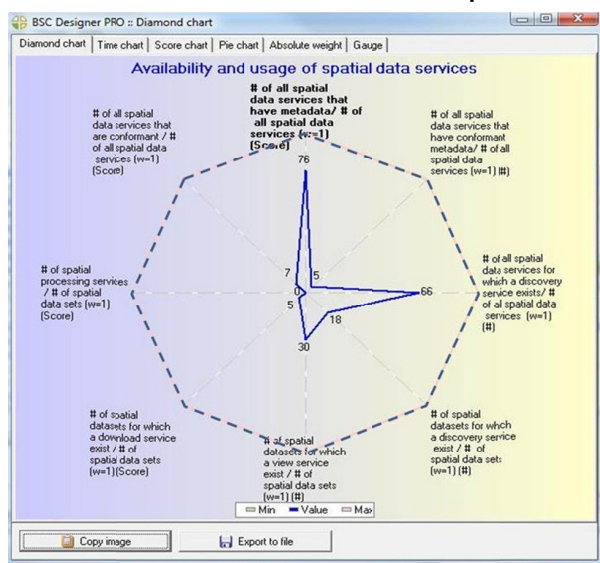


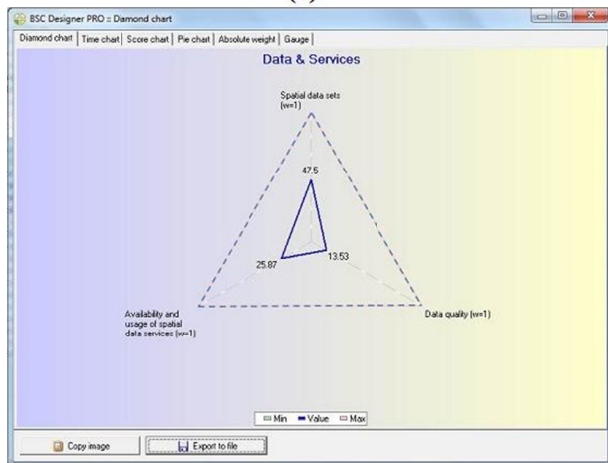
Figure 3 illustrates the progress of Swedish NSDI using a diamond chart. The solid line represents the current situation and the dash line stands for the target. If the blue indicators are close to the corresponding pink indicators, progress is

good. As an example, Figure 3a shows that 76% of spatial data services have metadata (which may be good progress) but a download service exists for only 5% of spatial datasets. This is detailed information which can be obtained from a BSC designed for SDI. For each indicator that does not represent good progress, SDI coordinators can investigate and remedy the problem(s) to meet the target. Figures 3b to 3d illustrate the same structure for a specific objective and all four existing perspectives.

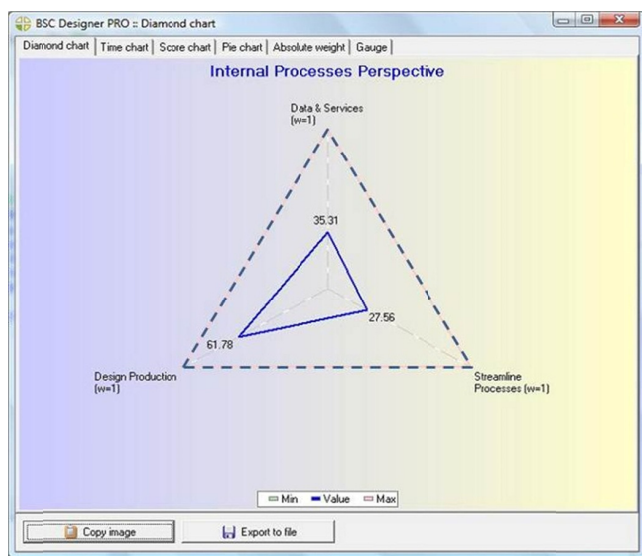
Figure 3: Diamond Chart with KPIs indicating a) Current Status for Availability and Usage of Spatial Data Services b) Progress for Spatial Data Sets and Services, c) Progress from the Internal Process Perspective, d) General Progress of the Perspectives in the BSC Structure for the Swedish Spatial Data Infrastructure



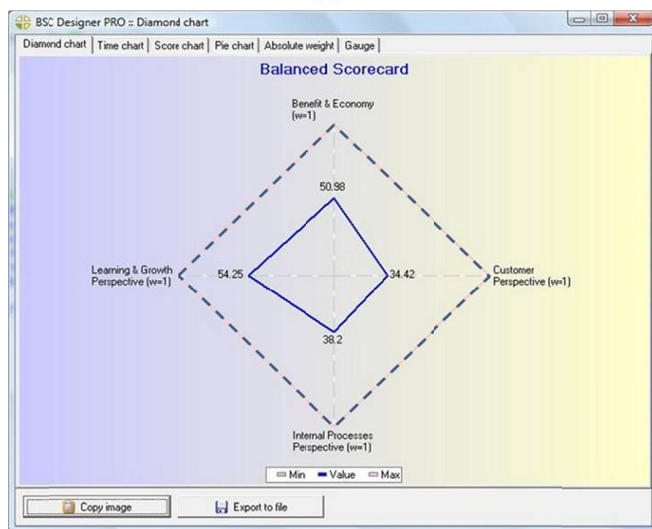
(a)



(b)



(c)

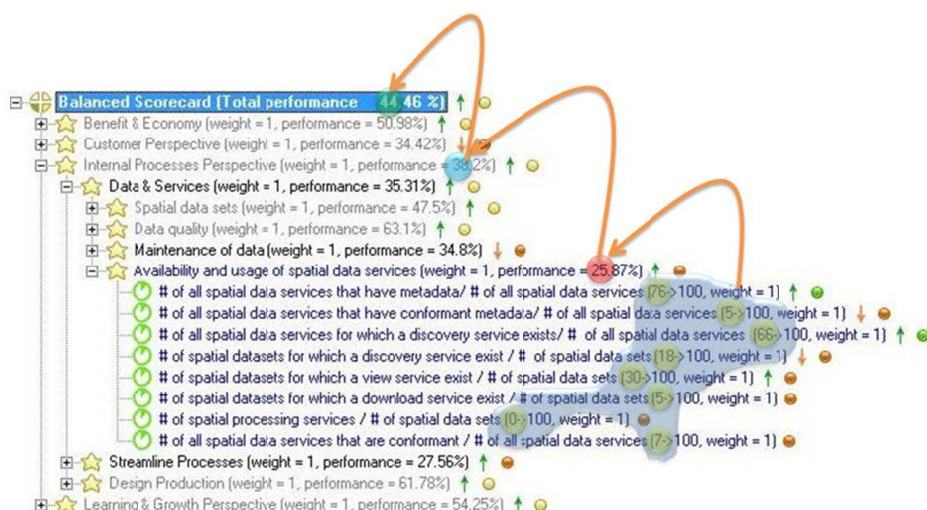


(d)

The BSC program is also able to graphically illustrate the cause and effect linkage in two different ways. Cause and effect linkage may either be hierarchical or between different perspectives. A hierarchical cause and effect linkage is a bottom-up relationship between indicators. One example is the effects of KPI_{IP} 1 – 8, KPI_{IP} 9 – 12 and KPI_{IP} 13 where each of them influences the sub-objectives

'Availability and usage of spatial data services', 'Spatial data sets' and 'Data quality' and hierarchically the integration affects the 'Data & Services' objective and thus the 'Internal Process' perspective, which has a direct influence on the progress of the whole BSC structure (cf. Figure 3). Figure 4 represents the hierarchical cause and effect linkage starting from 8 KPIs with specific scores affecting the 'Availability and usage of spatial data services' sub-objective (Performance = 25.87%) and subsequently produce the 'Data & Services' objective with a performance of 35.31%. Further on this objective modifies the 'Internal Processes' perspective to the value of 38.2% and finally the general progress of SDI performance is adjusted to 44.46%.

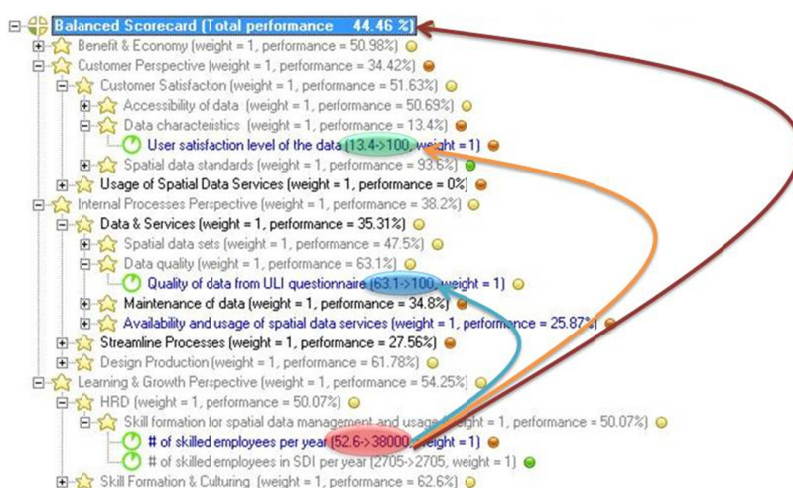
Figure 4: The Hierarchical Cause and Effect Linkage in Swedish NSDI



There is a more complex cause and effect linkage between the indicators of various perspectives where an indicator in a specific perspective influences one or more indicators in other perspectives. One example is human resources development and skills objectives (e.g. KPILG 2), which influence the other objectives such as data quality and customer satisfaction. This means that the more skilled employees are involved in SDI implementation, the better data can be prepared and thus more customers will be satisfied by using high quality data. Figure 5 illustrates the cause and effect linkage between employees and other indicators in the BSC structure. The performance of skilled employees (KPILG 2) based on the BSC structure is 52.6%. By increasing this value, i.e., having more skilled employees, an improvement should be noticed in objectives such as data quality and customer satisfaction within two years. Through considering these examples, managers can discover the sources of weak points resolve problems and improve performance. In this way, the cause and effect linkage and the

capability of enhancing KPI relationships facilitates performance evaluation and improvement when compared with using only abstract indicators. If two or more time stamp data are available, managers can also check the cause and effect of all indicators in different perspectives. Consequently, the cause and effect linkage is a vital part of this study where we have to consider the influence of internal and external effects. More details are discussed in the discussion section.

Figure 5: An Example of Cause and Effect Linkage between Different Perspectives in the Swedish NSDI



4.6. Comments about the Status of the Swedish NSDI.

The method adopted provides a good overview of the status of the various (Measurable) success factors that must be met for coordinated considered to be successful and contribute to the development of the Swedish NSDI.

Developing an infrastructure like the INSPIRE is financially demanding and generally need earmarked funding by the government. In Sweden, the government is investing 50 Mkr/year (~ 4.65 M€/year) for the coming three years on the implementation of INSPIRE -30 Mkr/year (~ 2.79 M€/year) for coordination activities and 20 Mkr/year (~ 1.86 M€/year) will be designated for public sector authorities responsible for providing metadata, datasets and services covered by the directive.

The funds provided have to cover all costs for coordination and maintenance of the infrastructure and a fully mature infrastructure should show full cost-coverage, i.e. value 100 for the "Benefit & Economy" perspective. Although the funds

required are far from enough to cover all costs involved in the harmonisation of data and services it still gets an “average” value from a “Benefit & Economy” perspective (50.98). The funding required for the development of harmonised datasets and services is, however, based on preliminary estimations at an early stage of the implementation of the INSPIRE directive.

The “value” for “Customer satisfaction” (34.42) is based on a questionnaire sent to all registered users of the Swedish Geodata portal. The implementation of INSPIRE and the Swedish Geodata Strategy is only in its beginning, which means that very few datasets and services are available and the “customer” response will be in line with this. Also, many of the registered users of the portal have also registered based on curiosity and doesn't really use the services. Yet, the results of the questionnaire provide a good benchmark for future questionnaires and although the users don't extensively use the datasets and services at this point of time they do provide a “gut-feeling” picture of the general “customer satisfaction” based on selected KPIs. This is reflected by the fact that access to the data is still considered an obstacle and gets a slightly lower rating than the other KPI's. As the stakeholders involved in the infrastructure change from being only service providers and become service users as well, it is likely that this will be reflected in the rating of the perspective.

“Internal processes” (38.2) is the most diversified perspective with 28 KPI's in total, covering various topics, from INSPIRE indicators to number and type of agreements between stakeholders. This perspective provides information about factors that can be directly influenced by the data- and service providers. The diamond chart gives a very clear picture of the current status with very high values for the KPI's referring to the number of datasets that have metadata and the number of datasets associated with a discovery service, and low values for the rest (Figure 3a). The INSPIRE implementation time plan follows a very tight time schedule and this is most probably a “typical” snapshot reflecting the status of the infrastructure at the time.

The “Learning & Growth” perspective (54.25) shows the highest “maturity” of the four perspectives considered. Part of this may be attributed to a number of intensive information campaigns carried out by Lantmäteriet with regards to the Geodata strategy as the input values for this perspective is based on a questionnaire answered by public sector authorities. Persistent communication and information about the background and goals of INSPIRE has been an essential strategy to increase awareness within public sector authorities as well as municipalities. The private sector is also actively engaged in the development of the infrastructure.

5. DISCUSSION

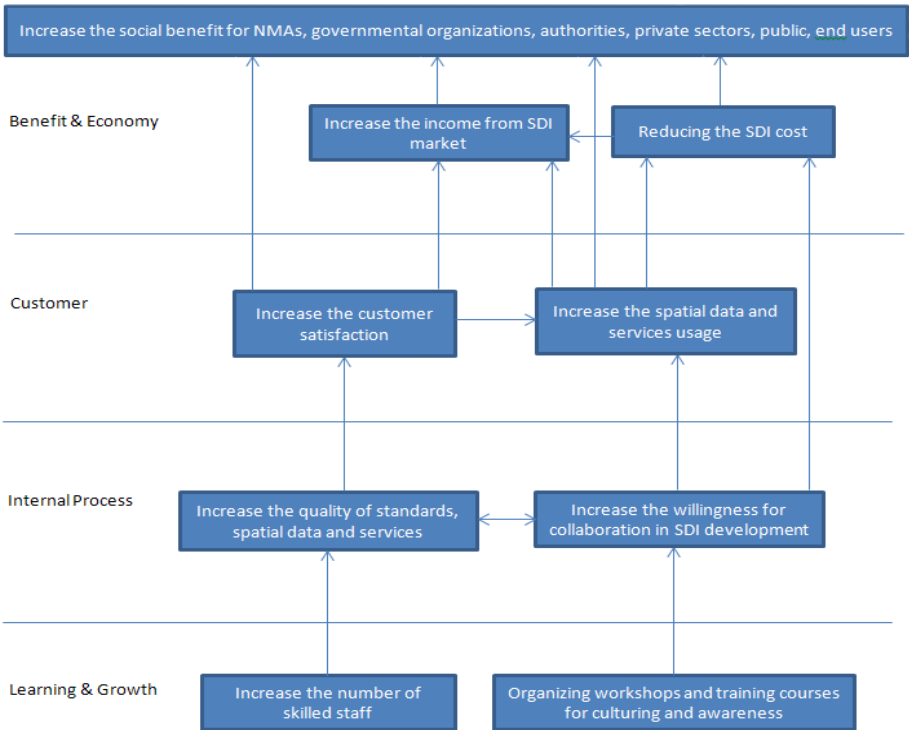
Balanced Scorecard is a performance measurement method. As discussed earlier, the main users of this method are industrial companies and organizations related to business. In addition to financial indicators, BSC attempts to combine other perspectives in the evaluation process and measure the progress of an organization from various viewpoints.

By using a Balanced Scorecard framework to evaluate the progress of an SDI, a clear pattern emerges from the existing situation. This pattern can be used as feedback for SDI coordinators to define strategies and set objectives, goals and visions. Such a clear pattern is not achieved by using only isolated indicators. Additionally, if performance measures and indicators within SDI components were strongly correlated, then there would be no need for a balanced scorecard. This evaluation framework can integrate all dimensions of an SDI into one view and facilitate the evaluation and monitoring process. It can also function as an appropriate management dashboard to guide SDI components according to predefined strategic maps, objectives and targets.

Another advantage of using the BSC framework for SDI evaluation is the cause and effect linkage which we take into account from two different dimensions; *external influences* and *internal effects*. In some situations, the real cause of an event is the result of high level political and/or international decisions. An example of such external influence is the monetary policy made by the government which affects the investments within the spatial data production. Certainly, although such external impacts can influence the whole SDI implementation, the modelling of such a cause and effect is beyond the SDI evaluation task. However, the internal effect is another viewpoint for the cause and effect linkage where we can monitor and control the manageable facts and try to improve them. Considering the BSC framework of this case study as well as the experiments from the result of the KPIs, we propose a widespread cause and effect linkage for SDI assessment.

Figure 6 illustrates the cause and effect linkage between different perspectives in the SDI assessment. As mentioned in Table 1, there are two general indicators within the Learning and Growth perspective. From the employee viewpoint, if we increase the number of skilled staff as an SDI readiness indicator it certainly affects and increases the quality of standards, spatial data and services. This means, more experts within the organization reduces the frequency of bug fixing as well as decreases the maintenance of data time. In the same perspective, organizing workshops and training courses not only increases the awareness but is also used for capacity building. Such improvement affects the collaboration in SDI development and increases the willingness for more cooperation among different organizations and third parties.

Figure 6: The Cause and Effect Linkage between Different Perspectives in the SDI Assessment



There is also a horizontal relationship between the quality of products and collaboration. The more organizations participate in SDI development, the better and higher the quality of the standards, spatial data and products will be. In other words, the collaboration among different organizations triggers a higher quality of standards, spatial data and services as well as more satisfaction of users requirements and vice versa. The improvement in the internal process affects the customer perspective and increases both the level of satisfaction and the spatial product usage. Here, there is horizontal relationship between customer satisfaction and spatial product usage which means whenever the end user is satisfied consequently the spatial data and service usage will be increased rapidly. Subsequently, the customer perspective affects the benefit and economy perspective which can be considered as the main goal of an SDI evaluation. More into details, customer satisfaction directly motivates the benefits in society and also increases the income from SDI market which also influences different types of users. Finally, the spatial data usage and the increase of the collaboration reduces the SDI cost which similar to the other KPIs leads to a

higher level of social benefit for NMAs, governmental organizations, authorities, private sectors public and end users.

There are also challenges and prerequisites that must be considered. First, data required for the evaluation procedure must be consistent. Data from various sources may have specific usages and may have been originally collected for other purposes. It is a complicated and costly task to collect data only for evaluation purposes. Instead, existing data with slightly modified indicators are reused for evaluation.

Another major issue is dynamic changes in objectives as well as indicators during the implementation period. SDI implementation is a long-term procedure that requires continuous improvement and feedback obtained from previous results. Technological developments as well as organizational changes of view may sometimes require that managers change strategies and objectives. In such situations indicators might change, either slightly or in some cases totally, which would result in the assessment structure becoming more dynamic in nature. In all design/implementation processes, a flexible structure is preferable for the objectives and thus the indicators.

From a more operational viewpoint, BSC structure implementation for different time stamps remains a major problem. SDIs being a recent operational concept in the EU, are still in a transitional stage from theory towards practice. However, by utilizing the BSC structure in SDI evaluation, coordinators can see changes to indicators and compare these with desired targets at any time. Finally, taking into account the cause and effect linkage as a major advantage of BSC structure, data is required from different periods of analysis in the SDI progress for different objectives. It also helps SDI coordinators to have a sensitivity analysis for each KPI and an illustration of changes for different scenarios.

Studying the outcomes of the assessment (e.g. figure 3) shows that the method developed provides an interesting snapshot of the current status of the Swedish NSDI. Sweden gets about "average" values for all four perspectives considered in the BSC, which would indicate that Sweden should be somewhere in the middle of its implementation of INSPIRE. However, from a management point of view, it is important to put these values into perspective and look deeper into the underlying causes to respective value.

Finally, comparing to the other SDI evaluation methods, BSC is a dynamic framework. In other words, during on-going SDI implementation and progress from one SDI generation to the next, objectives and KPIs as well as targets and initiatives can be updated with respect to new strategies and future requirements. A major difference of BSC with other SDI evaluation approaches is objective

oriented nature. That means, using this method, the SDI coordinators can compare the predefined objectives according to the current progress of SDI.

6. CONCLUSIONS AND FUTURE RESEARCH

A novel approach to SDI evaluation based on a Balanced Scorecard, BSC, was proposed in this paper. To investigate the applicability of the proposal for SDI evaluation, a framework based on the INSPIRE directive and the Swedish NSDI strategy was developed. This framework was implemented and tested in a case study of the Swedish NSDI project.

The results show that BSC helps to evaluate SDIs from both data producers' and users' (customers') points of view. Thus, the evaluation is comprehensive from this perspective. In addition, the study shows that through the application of BSC a general and flexible SDI evaluation framework can be established for SDI activities at both individual (learning and growth) and intra-organizational (internal process) levels. In other words, not only standardization activities, institutional arrangements, the establishment of accessing networks, data management and data sharing can be evaluated, but also skill formation, increasing awareness of SDIs and promotion of data sharing. Through the use of BSC, financial aspects and benefit achievements (benefits and economy) would be an essential part of the evaluation. In a nutshell, BSC provides a comprehensive framework for SDI evaluation from different perspectives. Moreover, BSC software provides SDI coordinators with suitable instruments and tools to monitor and to evaluate the progress of an SDI.

By comparing measured values during different time stamps, the progress status of an SDI can be visualized and will most likely provide a general view of the cause and effect linkage for existing objectives. Furthermore, by comparing each measured value with the relevant target, the degree of success of the objective in question can be determined. By understanding poor progress of an objective, SDI managers can investigate the underlying problems. Resolution of the problems is achieved by creating new policies and procedures or revising existing ones.

In this study, the suitability of BSC for evaluating SDIs at the national level was tested and verified. Further research is required to investigate its applicability for SDIs at other levels (e.g. local or regional). Comparing the progress of SDIs in two or more countries with the same framework gives a better indication of the advantages and disadvantages of the indicators required, as well as special limitations and challenges for each country. Finally, with respect to vertical relationships of SDIs in an SDI hierarchy or the inter-relationships of SDIs in a complex environment, more research is required to model these relationships in a BSC framework for SDI evaluation.

ACKNOWLEDGEMENTS

The first two authors of this study were financed by the European Union funding program, Erasmus Mundus “External Cooperation Window” (EMECW lot8). They would like to appreciate for the funding. The Authors are thankful to all persons who have helped with collecting values for the measures. Especially they would like to thank employees at ULI and Lantmäteriet for performing questionnaires etc. Also thanks to Alan Cranmer for his comments.

REFERENCES

- Albert, C. (2002). Service loyalty: the effects of service quality and the mediating role of customer satisfaction, *Eur. J. Market* 36, 810–823.
- Band, W. (2000). Create a customer satisfaction index (CSI) to improve your performance, *Sales Market. Manag. Can.* 31 (7), 58–59.
- Brijesh, P., Thierry, C., and M. Peter (2008). Balancing the NHS balanced scorecard, *European Journal of Operational Research*, 185(3), 905-914.
- Budhathoki, N.R., and Z. Nedovic-budic (2007). Expanding the Spatial Data Infrastructure knowledge base, in *research and theory in advancing Spatial Data Infrastructure concepts* by H. Onsrud (Ed.), 1st Edition, ESRI Press, Redlands, California, 294.
- Chaichan, P. (2002). Implementation of a balanced scorecard technique for performance evaluation of a real-estate company, Master Thesis, School of Civil Engineering, Asian Institute of Technology (AIT), Bangkok, Thailand.
- Cooper, R., and R. Kaplan (1991). Profit priorities from activity based costing. *Harvard Business Review*, 130–135.
- Crompvoets, J., Rajabifard, A., Bregt, A., and I. Williamson (2004). Assessing the world wide developments of national spatial data clearinghouses, *International Journal of GIS*, (18), 1-25.
- Crompvoets, J., Rajabifard A., van Loenen B. & Delgado Fernández T. (2008). A multi-view framework to assess SDIs, Space for Geo-Information (RGI), Wageningen University and Centre for SDIs and Land Administration, Department of Geomatics, University of Melbourne, Melbourne, Australia.
- Directive 2007/2/EC (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007, *Official Journal of European Union*, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF>, [accessed 7 May 2010].
- Epstein, M.J., and P.S. Wisner (2001). Using a balanced scorecard to implement sustainability, *Environmental Quality Management*, 11(2), 1-10.

- Fernandes K.J., Raja V., and A. Whalley (2006). Lessons from implementing the balanced scorecard in a small and medium size manufacturing organization, *Technovation*, (26), 623-634.
- Fernández, T.D., and J. Cromptvoets (2008). Evaluating Spatial Data Infrastructures in the Caribbean for sustainable development, GSDI-10 Conference, 25-29 Feb, St. Augustine, Trinidad.
- Fornell, C. (1992). A national customer satisfaction barometer: the Swedish experience, *J. Market*, (56), 6–21 (January).
- Geudens, T., Macharis, C., Cromptvoets, J. and F. Plastria (2009). Assessing Spatial Data Infrastructure Policy Strategies Using the Multi-Actor Multi-Criteria Analysis, *International Journal of Spatial Data Infrastructures Research*, Vol. 4, 265-297.
- Georgiadou, Y., Puri, S.K., and S. Sahay (2005). Towards a potential research agenda to guide the implementation of spatial data infrastructures - A case study from India, *International journal of geographic information science*, (19), 113-1130.
- Giff, G. A., and J. Cromptvoets (2008). Performance indicators a tool to support Spatial Data Infrastructure assessment, *Computers, Environment and Urban Systems*, 32, 365–376.
- Giff (2008). A Framework for Designing Performance Indicators for Spatial Data Infrastructure Assessment In: *A Multi-view Framework to Assess Spatial Data Infrastructures*, Cromptvoets, J.W.H.C.; Rajabifard, A.; Loenen, B. van; Delgado Fernandez, T., Melbourne, Australia : The Melbourne University Press, 2008, pp 211 – 230.
- Groot R and J. McLaughlin (2000). *Geospatial Data Infrastructures*. Oxford, Oxford University Press.
- Grus, L., Castelein, W., Cromptvoets, J., Overduin, T., van Loenen, B., van Groenestijn, A., Rajabifard, A., Bregt, A. K., (2011). An assessment view to evaluate whether Spatial Data Infrastructures meet their goals, *Computers, Environment and Urban Systems*, Vol. 35-3, 217-229.
- Grus, L., Cromptvoets, J., and A. Bregt (2007). Multi-view SDI assessment framework, *International Journal of Spatial Data Infrastructures Research*, Vol. 2, 33-53.
- Hackl, P., Kristensen, P.K., and A. H. Westlund (2000). Customer satisfaction: theory and measurement. Special Issue of *Journal of Total Quality Management*.

- Harvey F and D. Tulloch (2006). Local-government data sharing: Evaluating the foundations of spatial data infrastructures. *International Journal of Geographical Information Science* 20: 743–68
- Idalina, D-S., Lucas, R., and A. Paula (2007). Developing sustainability balanced scorecards for environmental services: A study of three large Portuguese companies, *Environmental Quality Management*, 16(4), 13-34.
- INSPIRE (2009). INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119, http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/metadata/MD_IR_and_ISO_20090218.pdf , [accessed on 11 November 2010].
- INSPIRE (2008a). The geoportal of INSPIRE, www.inspire-geoportal.eu, accessed 7 May 2010].
- INSPIRE (2008b). Implementing Directive 2007/2/EC of the European Parliament and of the Council as regards monitoring and reporting, , <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:148:0018:0026:EN:PDF>, [accessed on 11 November 2010].
- INSPIRE (2007) official website, http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/network/Network_Services_Performance_Guidelines_%20v1.0.pdf, accessed 7 May 2010].
- Janssen, K. (2008). A legal approach to assessing Spatial Data Infrastructures, In: *A Multi-view Framework to Assess Spatial Data Infrastructures*, Crompvoets, J.W.H.C.; Rajabifard, A.; Loenen, B. van; Delgado Fernandez, T., Melbourne, Australia : The Melbourne University Press, 2008 - ISBN 9780732516239, pp 255 –272.
- Kaplan, R. S., and D. P. Norton (2000). Having trouble with your strategy? Then map it. *Harvard Business Review* (September-October): 167-176.
- Kaplan, R.S., and D. P. Norton (1996). Using the balanced scorecard as a strategic management system, *Harvard Business Review*, 74(1), 75–85.
- Kaplan, R.S. & Norton, D.P. (1996b) The Balanced Scorecard - translating strategy into action, Boston: *Harvard Business School Press*.
- Kaplan, R.S., and D. P. Norton (1992). The balanced scorecard: measures that drive performance. *Harvard Business Review*, January–February, 171–9.
- Lantmäteriet (2008). The National Geodata Strategy. Information material about the Swedish NSDI. LMV-report 2008:1, ISSN 280-5731.
- Lantmäteriet (2010). Geodata project, www.geodata.se [accessed 7 Dec 2010].

- Lantmäteriet (2008). *A National Geodata Strategy*. http://www.lantmateriet.se/upload/filer/om_lantmateriet/Infrastruktur_goedat a/Geodata_Strategy_2008.pdf, [accessed 7 May 2010].
- Lawson, R., Stratton, W., and T. Hatch (2006). Scorecarding goes global. *Strategic Finance*, 87(9), 34-41.
- Lee, A.H.I., Chen, W-C., and C.J. Chang (2008). A fuzzy AHP and BSC approach for evaluating performance of IT department in the manufacturing industry in Taiwan, *Expert Systems with Applications*, (34), 96-107.
- Luu, T-V., Kim, S-Y., Cao, H-L., and Y-M. Park (2008). Performance measurement of construction firms in developing countries, *Construction Management and Economics*, 26(4), 373-386.
- Luzet, C. (2004). Geospatial data development: building data for multiple uses, In D. Nebert (Ed.), *developing spatial data infrastructure: The SDI Cook Book*, 13-23, at: <http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf>, [accessed 7 May 2010].
- Malmi, T. (2001). Balanced scorecards in Finnish companies. *Management Accounting Research*, 12 (2), 207-220.
- Mard, M.J., Dunne, R.R., Osborne, E., and J.S. Rigby Jr (2004). *Driving your company's value: Strategic benchmarking for value*, John Wiley & Sons, Hoboken, New Jersey.
- Masser, I. (2010). *Building European Spatial Data Infrastructures*. Redlands: ESRI Press.
- Nedović-Budić, Z., Pinto, J. K. and Budhathoki, N. R. (2008). SDI Effectiveness from the User Perspective, In: *A Multi-view Framework to Assess Spatial Data Infrastructures*, Cromptoets, J.W.H.C.; Rajabifard, A.; Loenen, B. van; Delgado Fernandez, T., Melbourne, Australia : The Melbourne University Press, 2008, pp 273 – 303.
- Rajabifard A, Feeney M F, and I. Williamson, (2003). Spatial Data Infrastructures: Concept, nature and SDI hierarchy. In Williamson I, Rajabifard A, and Feeney M F (Eds.) *Developing Spatial Data Infrastructures: From Concept to Reality*. Boca Raton, FL, CRC Press: 17-40
- Rigby, D. (2001). Management tools and techniques: a survey, *California Management Review*, 43 (2), 139-160.
- Sandgren, U. (2008). Lantmäteriet; project leader of the Swedish Geodata strategy. Personal communication.
- Silk, S. (1998). Automating the balanced scorecard. *Management Accounting*, 79 (11), 38-44.

- Steudler, D., Rajabifard, A. and Williamson, I. (2008). Evaluation and Performance Indicators to Assess Spatial Data Infrastructure Initiatives, In: *A Multi-view Framework to Assess Spatial Data Infrastructures*, Crompvoets, J.W.H.C.; Rajabifard, A.; Loenen, B. van; Delgado Fernandez, T., Melbourne, Australia : The Melbourne University Press, pp. 193 –210.
- Toomanian, A. and Mansourian, A. (2009). An Integrated Framework for the Implementation and Continuous Improvement of Spatial Data Infrastructures. *SDI Convergence. Research, Emerging Trends, and Critical Assessment*. B. van Loenen, J.W.J. Besemer, J.A. Zevenbergen (Editors). Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission 48, 2009, pp 161 - 174.
- Vandenbroucke, D., Crompvoets, J., Vancauwenberghe, G., Dessers, E., and J. Van Orshoven, (2009). A Network Perspective on Spatial Data Infrastructures: Application to the Sub-national SDI of Flanders. *Transactions in GIS*, 13(s1): 105-122.
- Vandenbroucke, D., K. Janssen and J. Van Orshoven, 2008. INSPIRE State of Play: Generic approach to assess NSDI. In Crompvoets, J., A. Rajabifard, B. van Loenen and T. Delgada (Eds.). *A Multi-View Framework to Assess National Spatial Data Infrastructures*, Melbourne University Press, Australia, pp 145 - 172.
- van Loenen, B. and van Rij, E. (2008). Assessment of Spatial Data Infrastructures From an Organisational Perspective, In: *A Multi-view Framework to Assess Spatial Data Infrastructures*, Crompvoets, J.W.H.C.; Rajabifard, A.; Loenen, B. van; Delgado Fernandez, T. (Eds.), Melbourne, Australia : The Melbourne University Press, 2008, pp 173 – 192.
- van Loenen, B. (2006). Developing Geographic Information Infrastructures; the role of information polices, Dissertation. Delft University of Technology. Delft: DUP Science.
- Williamson, I., Grant, D., and A. Rajabifard (2005). Land administration and Spatial Data Infrastructures, In *FIG Working Week 2005 and GSDI-8*, 16-21 April 2005, Cairo, Egypt, 1-13.
- Williamson, I., Rajabifard, A., and M. E. F. Feeney (2003). *Developing Spatial Data Infrastructure: from concept to reality*. London & New York: Taylor & Francis, 313.
- Yager, R. R. (1998). On ordered weighted averaging aggregation operators in multi-criteria decision making, *IEEE Transactions on Systems, Man and Cybernetics* 18, 183-190.

Paper III

Mansourian, A., Omid, E., Toomanian, A., Harrie, L., (2010). Expert System to Support Functionality of Clearinghouse Services. *Computers, Environment and Urban Systems*, 35(2), pp159-172.



Expert system to enhance the functionality of clearinghouse services

Ali Mansourian^{a,*}, Ebrahim Omid^a, Ara Toomanian^b, Lars Harrie^b

^a Faculty of Geodesy and Geomatics Engineering, K.N. Toosi University of Technology, Iran

^b GIS Centre, Department of Earth and Ecosystem Sciences, Lund University, Sweden

ARTICLE INFO

Article history:

Received 12 October 2009

Received in revised form 14 June 2010

Accepted 14 June 2010

Available online 17 July 2010

Keywords:

Clearinghouse

Spatial data infrastructure

Expert system

Semantic matching

Schema translator

Web processing service

ABSTRACT

Spatial data clearinghouses are one of the key features of a spatial data infrastructure (SDI). However, recent research indicates that few national clearinghouses function well, as the spatial data resources available cannot be satisfactorily accessed or optimally used. To improve the functionality, we propose that clearinghouses to be complemented with expert systems and semantic matching. The expert system facilitates automatic determination of candidate datasets and the conversion of the available data to the required data. A schema translator is also used to find similar data that might be used in other disciplines or other datasets by semantic matching. In order to accomplish this, we have developed a method of identifying available data and methods for data conversion. The methodology is implemented using standardized map services. Practical tests show that the discovery of available data in the clearinghouse satisfying users' requirements is substantially increased, which is an important step forward in building future SDIs.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

A spatial data clearinghouse is a distributed network that links geospatial data producers, managers, and users electronically (Executive Order, 1994). It can be defined as an electronic facility for searching, viewing, transferring, ordering, advertising, and/or disseminating spatial data and services from numerous sources via the Internet (Crompvoets et al., 2006). ISO 19115:2003 considers a [spatial] data clearinghouse to be a collection of institutions providing digital data that can be searched through a single interface using a common metadata standard. Such a clearinghouse usually consists of a number of servers containing information (metadata) about available digital data/services (Crompvoets, Bregt, Rajabifard, & Williamson, 2004).

Spatial data clearinghouses are known variously within the spatial information community, examples of which are the Geospatial One-Stop Portal (US FGDC), the Spatial Data Directory (Australian Spatial Data Infrastructure), and Geoportals (in the EU). All these clearinghouses have the common goal of better use of the spatial data resources available by facilitating discovery and access to spatial data. As a result, spatial data clearinghouses have been identified as a key feature of Spatial Data Infrastructures (SDIs), since they incorporate the data discovery and distribution components of SDIs (Nebert, 2004). In Europe, for example, the Inspire directive

(INSPIRE, 2009) specifies that each EU member state must set up clearinghouses (geoportals) as part of their SDI.

Two types of users search for spatial data in clearinghouses: experts and non-experts. Both search for data using the geographic region, data theme, name of the data layer, etc. If the data layer required by a user is available, the clearinghouse can discover it. However, expert users use a clearinghouse to download the required spatial data and work with the quality of data for their analysis. Non-expert users, on the other hand, often utilize such systems for viewing purposes. In this paper, we propose a general method to facilitate the data retrieval process for both groups.

In many cases, the desired data layer does not exist in any server, but it can be generated by processing and/or integrating existing data layers from what is referred to here as *candidate data layers*. Professional users have the knowledge required to search for data and then integrate/process them to obtain the information they require. However, many users do not have the knowledge required to identify and search for data. Neither do they have the appropriate knowledge and/or spatial processing tools required to generate the data they require from the data available. Moreover, in some cases users require a fast view of a spatial data. As a result, the most appropriate spatial data resources are not used. In other words, the general goal of spatial data clearinghouses, as described above, is not achieved.

In this paper we aim to propose a clearinghouse that utilizes an expert system to enhance the availability of spatial data. The expert system facilitates the capability of identifying and searching for suitable candidate data layers, when the data required by the

* Corresponding author.

E-mail addresses: Mansourian@kntu.ac.ir (A. Mansourian), ebrahim.omidi@gmail.com (E. Omid), Ara.Toomanian@nateko.lu.se (A. Toomanian), Lars.Harrie@nateko.lu.se (L. Harrie).

user are not found within the common clearinghouse system. A schema translator concept is also used in the architecture to find suitable data from a vocabulary list when no matching data are found in the first round. Such an expert spatial data clearinghouse is also able to find suitable spatial processing services, which have the ability to process the data available to produce the data required by the user. The clearinghouse can then establish a connection between the processing service and data services to produce the required data. The applicability of the approach is demonstrated by a prototype implementation.

The structure of the paper is as follows: First, the concepts of clearinghouse networks, web services, schema translation, service chaining, and expert systems are reviewed. In Section 3, the architecture of an expert clearinghouse is described. Section 4 describes the implementation of a prototype system, based on the proposed methodologies, and the running of the system, based on a number of scenarios for viewing and integrating spatial data. Sections 5 and 6 present a discussion and the conclusions.

2. Background

2.1. Spatial data clearinghouses

The term 'Clearinghouse' was first used in the spatial information community in 1994 in the United States. The US Federal Geographic Data Committee (FGDC) established the National Geospatial Data Clearinghouse to facilitate efficient access to large amounts of spatial data from federal agencies, to coordinate its exchange, and to minimize the costs resulting from duplication in the collection of spatial data (Crompvoets et al., 2004; FGDC, 2000; Rhind, 1999).

The first generation of spatial data clearinghouses was created in the US, Australia, the Netherlands, the UK, and Canada. These clearinghouses were similar to existing web-based databases. In order to search for a spatial data layer, the user was required to set search parameters such as geographic region, data theme, and the name of the data layer in the clearinghouse user interface. The spatial data requested from the databases were then presented to the user according to the search criteria (Radwan, 2002). In other words, they provided users with either a detailed description of the data, so-called metadata, and information on how to access them, or a link to the web site where the data were available.

With the development of web-based technology and the advancement of spatial web services, a new generation of clearinghouses was developed, based on geoportals, catalogue services, and spatial services (see e.g. Bernard, Kanellopoulos, Annoni, & Smiths, 2005; FGDC, 2009 for information on European and US clearinghouses, respectively). These provide users with standardized and more appropriate ways of searching for and accessing spatial data.

Fig. 1 shows the general structure of a clearinghouse based on a geoportal, catalogue services, and spatial services. The main elements of a clearinghouse can be summarized as follows:

- The *Geoportal* provides an entry point to spatial information on the web (Tait, 2005). It is a web site on the internet where spatial content, including spatial data and spatial services, can be discovered.
- *Catalogue services* provide the functionality to publish metadata on spatial data resources, and to search and query metadata (Bernard et al., 2005).
- *Metadata repository* stores information about spatial data (in databases).
- *Spatial services*, which are connected to data servers, provide users (clients) with various services, e.g., viewing and downloading of spatial data. Well-known services include the Gazetteer Service (Hill, 2000), the Thesaurus Service (Hunter, 2001), the Map Service (Beaujardiere, 2006), the Feature Service (Vretanos, 2005), the Coverage Service (Whiteside & Evans, 2008), and the Processing Service (Schut, 2007).
- *Registry service* is where spatial services and catalogue services are registered in order to be discoverable by a geoportal.

To search for data (or a service) in a clearinghouse, the user connects to the geoportal and specifies the search parameters. The geoportal then searches metadata repositories, through catalogue services, to discover the data required by the user. Each catalogue service may publish metadata from an individual data producer. If the relevant data servers support (or provide) spatial services such as viewing or downloading, then the data can be viewed or downloaded by the user.

2.2. Web services

Clearinghouses rely on standardized interfaces for searching and distributing spatial data. The main standards are currently specified by the Open Geospatial Consortium (OGC) in cooperation with the International Organization for Standardization (ISO). Clearinghouses often use the OGC standards for publishing metadata (Catalogue Service for Web, GS-W) and distributing map data (Web Map Service, WMS), geographic data (Web Feature Service, WFS), and grid data (Web Coverage Service, WCS). All of these web services lack the capability of modifying the data before distribution. However, this is often necessary, and many clearinghouses offer a transformation service. The OGC Web Processing Service (WPS) interoperability experiment, performed in 2003, addressed the need for a standardized processing service. This led to the implementation of the OGC WPS in 2007 (Schut, 2007). Due to the important role of the WPS in this study, it is described in detail below.

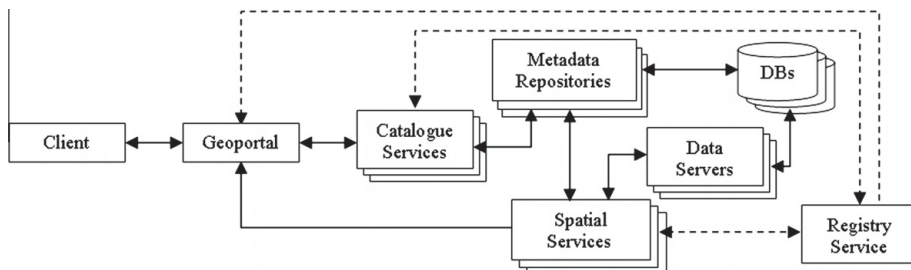


Fig. 1. General structure of current clearinghouses.

2.2.1. The web processing service

The OGC WPS defines a standardized interface that facilitates the publishing of spatial processes and the discovery of and binding to those processes by users. Here, the concept of a process may include algorithms, calculations or various kinds of models, which operate on spatially referenced data. Publishing means making machine-readable information available, apart from the human-readable metadata, allowing service discovery and use (Schut, 2007). A WPS is thus capable of designing a variety of GIS functionalities available to clients across a network, as well as providing access to previously defined functions, calculations and/or computational models that operate on spatial data. The data required by the WPS can be delivered over a network, or can be available on the server. Moreover, the essential data can be provided as images or through data exchange standards such as Geography Markup Language (GML).

The WPS interface contains three mandatory operations (Schut, 2007). *GetCapabilities* allows a client to request and receive service metadata documents that describe the capabilities of the specific server implementation. *DescribeProcess* returns the server's processes. *Execute* allows a client to run a specified process implemented by the WPS, using the input parameters provided and returns the appropriate outputs.

WPS can be applied to widely varying applications. It has been used for diverse aggregation and spatial join calculations (Smith & Mark, 2003), accessibility analysis (Neis & Zipf, 2007), generalization (Bergenheim, Sarjakoski, & Sarjakoski, 2009; Neun, Burghardt, & Weibel, 2009), digital elevation model processing (Schilling, Lanig, Neis, & Zipf, 2008), housing market analysis (Zipf & Stollberg, 2008), and geomorphological models (Wood, 1996). In this paper, we propose the use of WPS to enhance the functionality of clearinghouses.

2.2.2. Schema translation

Developing network-service-based clearinghouses poses new challenges to spatial data providers. Two major issues must be solved: the increase in the diversity of applications in which digital spatial data are being used, and the wide-ranging requirements on the structure and terminology used in the data content.

Schema translation is a well-known field in computer science for the integration of heterogeneous data sources. The concept has been applied to many areas, such as federated databases, data warehousing, and data mining (Marotta & Ruggia, 1999). However, most of the studies performed to date have only considered schema translation in the context of relational database technology, and rarely take into account the specific issues related to spatial data. Research concerned with metadata has been carried out on web-based digital libraries (Godby, Smith, & Childress, 2003) and bibliographic references (Llavador & Canos, 2006).

The process of schema translation has been appropriately defined based on Model Driven Architecture (MDA) by Donaubaer, Fichtinger, Schilcher, and Straub (2006): "A schema translation process actually transforms data content expressed in one schema into data content expressed in another schema and not the schema itself". The term 'schema mapping' refers to the process of determining the correspondence between the data items in the source and target schemas prior to the actual schema translation process (Visser, Stuckenschmidt, Wache, & Vögele, 2001). The schema translation rules are automatically derived from the conceptual-level mapping, and are a set of rules and techniques that establish relationships between equivalent constructs of the different schemas.

The data translation process has three different levels: syntactic, schematic, and semantic. Syntax level translation deals with changing both the language of the data representation and the related schema definition language. A schematic translation

modifies the structure and the schema vocabulary of the data model used in the dataset. This type of translation is well suited to the web environment. Semantic translation is applied when there is no definition of exact mapping from the source schema to the target schema. Here, instead of an exact correspondence between the source and target data, a reasonable approximation is used. In this paper we consider only the semantic aspects, which are addressed by applying semantic methods (Lehto, 2007).

The geodetic control point service of the Wisconsin Land Information Clearinghouse is a good example of a schema transformation service for geospatial applications (WiscLINC, 2006). The service architecture has a schema mediator process that translates the schema in two-way. This means that, apart from the standard Extensible Stylesheet Language Transformations (XSLT) process, a so-called Streaming Transformations for XML (STX) exists, which is a streaming variation of the XML-to-XML transformation process (Becker, Brown, & Cimprich, 2003).

2.2.3. Service chaining

Service chaining is a generic term used to refer to the domain of "combining services in a dependent series to achieve larger tasks" (ISO, 2005). Combining simple spatial data services into value-added service chains is seen as one of the great benefits of SDIs (Einspanier, Lutz, Senkler, Simonis, & Sliwinski, 2003). This instance is an implementation of a workflow which is an automation of a business process, in whole or part, considering a set of rules where the tasks are passed from one participant to another (ISO, 2005). Friis-Christensen, Lucchi, Lutz, and Ostländer (2009) identified two general patterns for service chaining: *Centralized* and *Cascaded*. They implemented service chaining based on the centralized and cascaded patterns to compare the advantages and disadvantages with each other.

In the centralized pattern, a central component controls the invocation of all the services used. In other words, the workflow execution is controlled by a central component. Therefore, it is not necessary to distribute details about the workflow to other Service Oriented Architecture components. Fig. 2 shows an example of service chaining based on the centralized pattern, in which two datasets are retrieved from two WFSs and then processed using a chain of WPSs. The output is presented to the client using a WMS.

In the cascaded pattern the service chaining is controlled using a *backward* approach. In this approach, the service that provides the required end result is invoked directly. The invoked service handles another service invocation, which is necessary to retrieve the input parameters/processes. The control process is conducted by the individual services in a chain, in such a way that each service connects directly to the other service(s) to invoke the required input parameters/processes or to send the output. The final product is presented to the user by the last service. Fig. 3 shows an example of service chaining based on the cascaded pattern.

The centralized pattern has some advantages over the cascaded pattern (Friis-Christensen et al., 2009). The centralized pattern is highly flexible regarding schema manipulation and immediate compensation at exceptions, and it is simple to program. Potential redundant data transfer through the central controller is considered the main drawback of the centralized pattern. The cascaded pattern, on the other hand, requires complex forms of code mobility, may suffer from delays in compensation at exceptions, and the programming is complex.

Several studies have been carried out on chaining spatial web services. A methodology based on abstract descriptions of services and workflows has been proposed by Granell, Gould, and Ramos (2005). In this methodology, which has been developed in Business Process Engine Language (BPEL), the executable workflow description is derived automatically. This method is also used by Lemmens et al. (2006, 2007) for ontology-based service composition

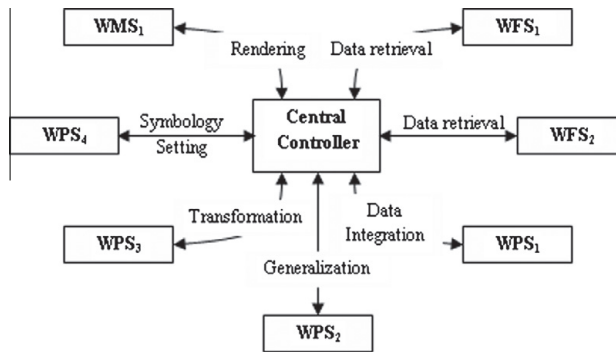


Fig. 2. Schema of a sample centralized service chaining.

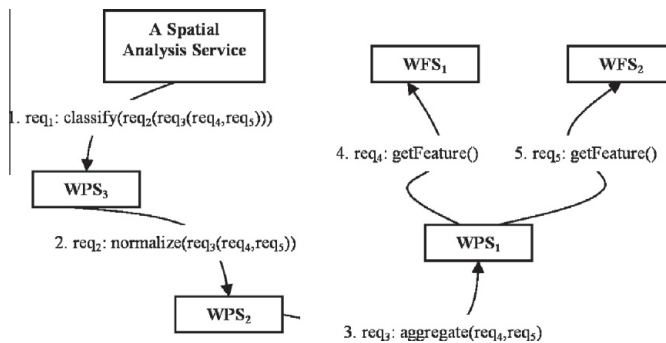


Fig. 3. An example of cascaded service chaining.

approaches. Kiehle, Greve, and Heier (2007) presented applications for groundwater vulnerability assessment and automated land parcel information using OGC standards and service chaining. Zhang, Xie, Yu, and Di (2008) also proposed an OGC-standard-based spatial information service chaining process.

2.3. Expert systems

Expert systems have been used for several decades, in different areas (Ma, Wang, Zhang, & Yu, 2008; Iqbal, He, Li & Dar, 2007; Altuntas, Bayraktar, & Cebi, 2006; Zischg, Fuchs, Keiler, & Meißl, 2005). The role of expert systems (also known as knowledge-based systems) is to guide users so that they obtain the best results by modeling the knowledge of experts in a specific field, to create 'rules' and 'facts', and then analyzing them. Expert systems are considered one of the branches of artificial intelligence.

In expert systems, rules are relations between facts, and rules are searched to obtain the best values for a number of facts. Sometimes, searching rules requires more information, which is usually obtained from the user. Finding these rules is one of the most important steps in implementing an expert system and is studied in the field of Knowledge Discovery.

An expert system has three main elements (Giarratano & Riley, 1989):

- a *Knowledge Base*, in which the modeled rules and facts are stored,
- an *Inference Engine*, which searches the knowledge base following the chain of rules, and

- a *User Interface*, which receives the initial information from the user, requests the parameters required for searching, and then presents the best suggestion.

The Inference Engine includes algorithms that specify the way in which the knowledge base is used to reach the final result. Two general methods are used in expert systems for the inference process and the use of rules (Sydenham & Thorn, 2005): *forward chaining* and *backward chaining*.

In forward chaining, the inference engine starts with the primary conditions and seeks rules in which these conditions are applied (in IF clause). The result of applying these rules constitutes new conditions, and the system then seeks rules in which these new conditions are applied. This iteration continues until the output of one of the rules matches that specified by the user (Sydenham & Thorn, 2005). This method is also called data driven, since the searching process is based on the initial data and conditions.

In backward chaining, the inference process is not based on the primary conditions, but on the desired conditions (in THEN clause). In other words, the system considers all the states that might arise. Then, it respectively goes for these rules and takes into account the facts existing in their IF clause. These facts are then adopted as 'new facts' and the system seeks conditions (rules) whose output is this new fact. This chain of search goes where the obtained new fact does not exist in any of the remaining laws (Sydenham & Thorn, 2005). This method is also called target driven, since the searching process is started after collecting the desired targets.

3. Expert spatial data clearinghouse

An expert spatial data clearinghouse is a type of clearinghouse that not only carries out a direct search for the required data, but also searches for data that can be converted into the required data. A semantic translator retrieves all possible synonyms of the search phrase and searches in catalog services to find matching data. An expert system component determines which candidate datasets can be processed to generate the data required by the user, selects the best combination of existing data, and also determines and executes (through a service chaining controller) the processes required to convert the available datasets into the required data.

3.1. Motivation of an expert spatial data clearinghouse

As mentioned at the beginning of this paper, the spatial data clearinghouse system is used by different types of users with different requirements, and the current generation of clearinghouse systems can not always provide the appropriate data. Sometimes, the requirements of inexperienced users are not available in the spatial databases under the same name, and sometimes even the professionals are not satisfied with the search results. However, there are often similar useful datasets within the same geographic region that could satisfy the requirements of the both user types. Another scenario is that neither the exact data nor similar data may be available for the required geographic region, but candidate data are available from which the required data can be generated. Examples of the various scenarios are given below.

Let us assume that a user is searching for *rain contours* that do not exist within the geographic region of interest. The clearinghouse would then discover nothing for this user. The expert clearinghouse system first translates the search expression into all similar semantic matching phrases such as *precipitation contours* or *rainfall data* from synoptic stations for that region. Such data can be used to generate rain contours. Using the terminology described above, the rainfall data are considered as candidate data. An expert clearinghouse should be able to discover a data layer in a synoptic station. For the advanced purposes, it should also be able to discover processing services that can generate *rain contours* from the existing data, and display them to the user.

In the developing countries, the spatial data coverage of many datasets is not complete, which sometimes leads to difficulties. Let us assume that a user is interested in Digital Elevation Model for a specific area. Although height data are generally stored as contour lines, or height points, etc. in the form of topographic maps (spatial databases), the user may fail to find a suitable DEM in the specified region. An expert clearinghouse should be able to search for other types of data layers (such as contours, height points, etc.) which can be converted into a DEM, then present these layers to the user, and finally arrange for processing of the data to obtain the desired DEM.

The INSPIRE Directive was introduced in the European Union in 2007. However, most data harmonization is still at the national level. For example, a user may be looking for a watershed data layer or drainage areas on the continental scale. If the layer does not exist, it can be produced by integrating and processing DEMs and stream data layers available on national scales. An expert clearinghouse should search and introduce suitable DEMs and stream layers from each country in the EU, which can be used to produce a watershed data layer, and then arrange for conversion to the desired information.

Considering the current problems associated with clearinghouse networks, described earlier in this section and in the intro-

duction, the main advantages of an expert spatial data clearinghouse with the above-mentioned characteristics can be summarized as:

- better use of the available spatial data resources by offering similar semantic matching datasets in situations where the data may be found under other synonyms in other disciplines;
- an increase in the number of successful searches in a spatial data clearinghouse, by suggesting candidate data to users when the requested data can not be found; and
- facilitation of the access of users to the requested spatial data, by automatic arrangements for the processing of the data available to produce the data requested by the user.

3.2. Methodology

There are various issues in the implementation of an expert spatial data clearinghouse.

- In automatic searches for candidate data, an expert system is required to identify spatial data and then search for them.
- A schema translator is needed to identify suitable semantic expressions matching that entered by the user to search for similar datasets.
- As there may be different types of candidate data layers that can be used to generate the required data, selection of the appropriate data layers, or a suitable combination of different data layers, by the clearinghouse, and then presenting them to the user is essential. This requires special algorithms.
- An interoperable environment (among data services, processing services and user systems) is necessary for data exchange and processing.

Therefore, in addition to the aforementioned core elements, an expert spatial data clearinghouse should also include the following (Fig. 4):

- *Schema translator*: This manages the retrieval of synonyms for the expression phrase. It translates the phrase and sends the translations to the geoportal, which searches for the data requested by the user.
- *Expert system*: If the geoportal can not find the data requested by the user, despite using semantic matching, this system attempts to find candidate data layers, and the best combination of these to generate the required data. The expert system also defines an instruction for processing candidate data to generate the data requested by the user. Expert system is regarded as a central element of the expert clearinghouse.
- *Process database*: Features of each process, including the function of each process, as well as its inputs and outputs, are organized in this database. This database is used for searching suitable WPSS.
- *WPSSs*: These implement the instructions from the expert system. The most important tasks of these engines are to receive output from the expert system, to access the data servers to retrieve data layers, and finally to conduct the necessary processes on the data layers to generate the data requested by the user.
- *Service chaining controller*: In order to produce the data requested by the user, it may be necessary to chain different web services. If a centralized pattern is used for service chaining, the controller manages the workflow of the chaining.

Fig. 4 illustrates a proposed general structure for an expert clearinghouse. Users connect to a geoportal to search for the data required. As an example, let us assume that a user is searching

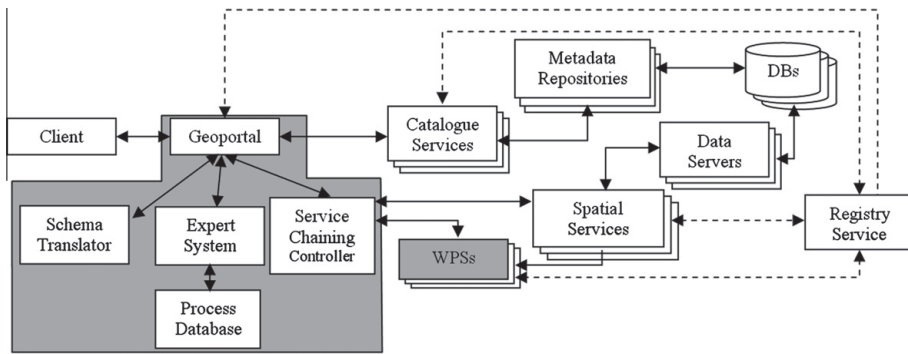


Fig. 4. General structure of an expert clearinghouse.

for a gradient map within a specific region. He connects to a geoportal and enters the search parameters. The geoportal searches the metadata repositories of data producers through the relevant catalogue services. If the data required are not discovered, the geoportal connects to a schema translator to identify synonyms for the required data layer, and then searches for the required data through its synonym phrases (semantic matching). If no data are found, the geoportal connects to the expert system to identify candidate data layers, which can be processed to generate the data requested by the user. Determining candidate data is a challenging task that is described in more detailed later in this section. After determining possible candidate data, they are searched through catalogue services.

The discovery of candidate data may present different solutions to generating the data requested by the user. Selecting a suitable combination of data and solution(s) is a challenging task, for which an algorithm is proposed later in this section. To clarify the problem, suppose that height contours, a DTM, and a gradient map are available as candidate data (Fig. 5). The clearinghouse should be able to select the most appropriate combination of these data layers for the generation of the data layer requested by the user within the required geographic region. To do this, the geoportal passes the search results to the expert system to determine the appropriate combination of candidate data layers, among those found. The expert system also determines the processes that have to be applied to the candidate data layers for the generation of the data requested by the user. The processes required are determined by the process database (Fig. 4).

The results are then sent to the geoportal. The geoportal searches for the processing services that offer the required processes. The service chaining controller sends data processing

requests to the selected processing services. The data are sent to the processing services by the data services, where they are processed to generate the data requested by the user. Finally, the results are sent to the user.

3.2.1. Identifying candidate data

To implement an expert system for a spatial data clearinghouse, a knowledge base, an inference engine and a user interface must be implemented. A *knowledge base* is where expert knowledge is stored. In this study, the expert knowledge consists of understanding different forms of data, as well as the processing algorithms and the models required to generate specific data. *Knowledge bases* usually have the form of databases containing facts and rules (Giaratano & Riley, 1989). In this study:

- spatial data and their attributes and characteristics are equivalent to the *facts*,
- the processing algorithms and models used for data conversion or generation are equivalent to the *rules*, and
- the process database is equivalent to the *knowledge base*.

Table 1 provides an example of a knowledge base in an expert clearinghouse.

The knowledge base in Table 1 presents a number of spatial data processes based on eleven facts and six rules. Some of the rules are *one-conditional* in the IF clause and others *multi-conditional*. A one-conditional rule can directly convert one type of data to another (e.g. Process 1, where height contours are converted into slopes). In other processes, two or more layers are required simultaneously (e.g. Process 4, where watershed generation requires DEM and stream layers).

Sometimes, the facts depend on the user's viewpoint. For example, 'observation point' and 'farming suitability parameters' are required for the production of the 'viewshed' and 'suitable farming area' data layers, respectively.

The purpose of an expert system in an expert clearinghouse is to find candidate data. In other words, candidate data construct the IF clause of the rules in an expert system (unknown) and the data requested by the user construct the THEN clause (known). Therefore, the problem is a target-driven problem in which the target is the data layer requested by the user. This leads to the adoption of the backward chaining method for the implementation of the inference engine to enable the expert system to study all the methods and rules leading to the required data.

As mentioned above, the knowledge base in an expert system can be implemented in two ways: *multi-conditionally* or

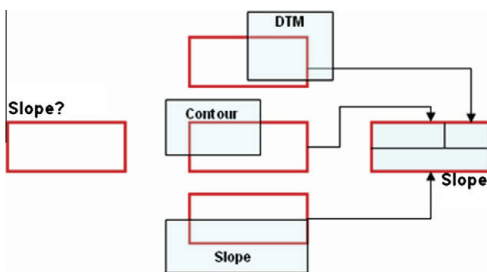


Fig. 5. Selection of candidate layers and their best combination in an expert clearinghouse.

Table 1

An example of a knowledge base for an expert clearinghouse.

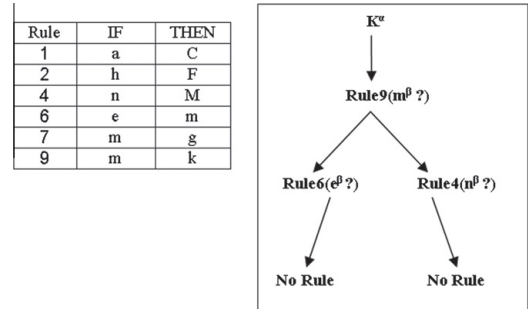
Process	Facts	Rules
1. IF a THEN b	1. a: height contour (shp)	1. Slop generation from height contour
2. IF a THEN c	2. b: slope (tif)	2. DEM generation from height contour
3. IF c, d THEN e	3. c: DEM (tif)	3. Viewshed analysis using DEM and observation point
4. IF c, f THEN g	4. d: observation point (coordinates)	4. Watershed generation using DEM and stream
5. IF h THEN i	5. e: viewshed (tif)	5. Iso-rain map generation from climatology station observations
6. IF c, b, i, k THEN j	6. f: stream (shp)	6. Determination of suitable farming areas using DEM, slope, etc.
	7. g: watershed (shp)	
	8. h: climatologic stations (shp)	
	9. i: iso-rain (tif)	
	10. j: suitable farming areas (tif)	
	11. k: farming suitability parameters	

one-conditionally. The way in which these methods are adapted for use in an expert spatial data clearinghouse is described below.

3.2.1.1. Adapting multi-conditional backward chaining for an expert clearinghouse. The multi-conditional method, which is a comprehensive method, is useful when two or more data types are needed to create the required data layer, for example, when DEM and stream data layers are used to produce a watershed data layer; or when fault, rock type and DEM are used to produce earthquake-prone regions. In other words, achieving the goal requires two or more conditions to be true (i.e. two or more data layers must exist) simultaneously. By using a multi-conditional method, an expert clearinghouse should be able to find the different data layers required to produce a specific data layer.

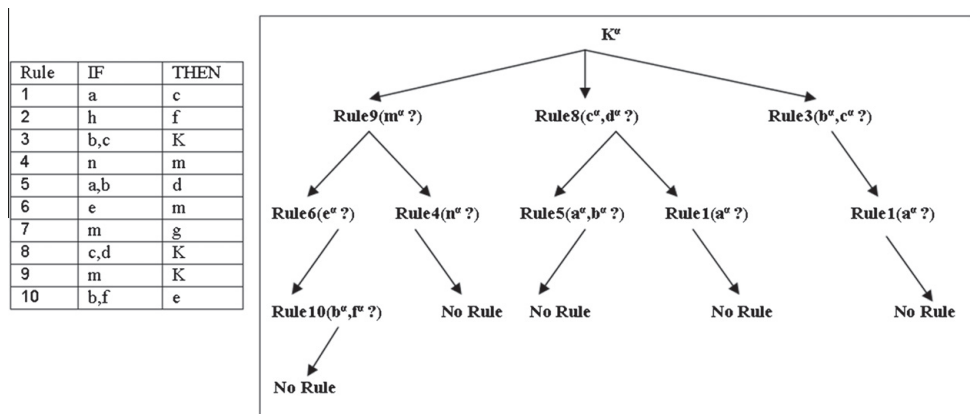
Suppose that a user is searching for data layer K in the geographic extent α (Fig. 6). If the data layer does not exist, the expert system will look for the data layers from which K can be generated directly (e.g. K can be generated from c and d), or indirectly (e.g. K can be generated from a and b). This is achieved by setting up backward searching chains, in real-time, all rooted from K (Fig. 6). To create a chain, each level is generated from the rules in the knowledge base: the data layers in the IF clause of a higher level should be used as the output layers in the THEN clause of a lower level. Each chain presents a solution for the generation of K within the geographic extent of α (K^α).

3.2.1.2. Adapting one-conditional backward chaining for an expert clearinghouse. The creation of the chains in one-conditional backward chaining is illustrated in Fig. 7. These are suitable when only

**Fig. 7.** Knowledge base and one-conditional search chain for layer K in the geographic extent of α .

one data layer is needed to produce the data requested by the user, for example, when DEM as a candidate data layer can satisfy the request for a hill-shade map.

One-conditional backward chaining is an efficient method of discovering several candidate data layers, each of which covers a sub-region (β) of a geographic extent (α), within which a user is searching for data (Fig. 5). In other words, each data layer can be a candidate data layer, if it covers any part (β) of the desired area (α). By selecting candidate data layers that satisfy ' $\sum \beta$ covers α ', and then processing and integrating them, the data requested by

**Fig. 6.** Knowledge base and multi-conditional backward search chain for layer K in the geographic extent of α .

the user covering the region α can be generated. However, this study shows that multi-conditional backward chaining is efficient for discovering candidate data layers (e.g. c and d) that cover the whole region of α (c^α , d^α). The reason is that for a level to be created, all facts in an IF–THEN structure should be true. Therefore, in a multi-conditional backward chain it would be complicated to determine several relevant candidate data layers that cover a region of β (e.g. c^β , d^β), when β is unknown.

3.2.2. Determining suitable combinations of candidate data layers

After determining candidate data layers and their covering regions (A, B, ... and G) (Fig. 8a), the clearinghouse should select suitable combinations of the candidate data layers to generate the data layer requested by the user (Fig. 8b), and then submit them to the user. To provide an expert clearinghouse with such a capability, some selection criteria must first be defined. Then a selection algorithm should be designed based on these criteria, and implemented in the clearinghouse.

To determine the most appropriate combinations of candidate data layers the following algorithm is proposed.

- i. Create a grid within the search region.
- ii. Select grids that are covered by only one data layer.
- iii. Insert the data layers (step ii) into list 1.
- iv. Insert the rest of the data layers into list 2.
- v. Select a data layer from list 1 that covers the maximum number of squares in the grid.
- vi. Determine the *coverage percentage* as a property of the data layer.
- vii. Remove the data layer from list 1.
- viii. Repeat step v until list 1 becomes empty.
- ix. Select a data layer from list 2 that has the minimum value of below equation.

$$G = \sum W_i C_i \quad (1)$$

where C_i denotes the value of the i th criterion, and W_i its weight.

- x. Remove the data layer from list 2.
- xi. Repeat step ix until list 2 becomes empty.
- xii. End

It can be seen in Fig. 8a that some areas are covered by only one candidate data layer (e.g. D). Selecting and proposing these data layers to the user is necessary. Steps i–viii in the above algorithm correspond to the selection of such candidate data layers. The rest of the algorithm determines the priority (G in Eq. (1)) of other candidate data layers to be presented to the user. The weight of the criteria (W_i) may be entered by the user or set by default.

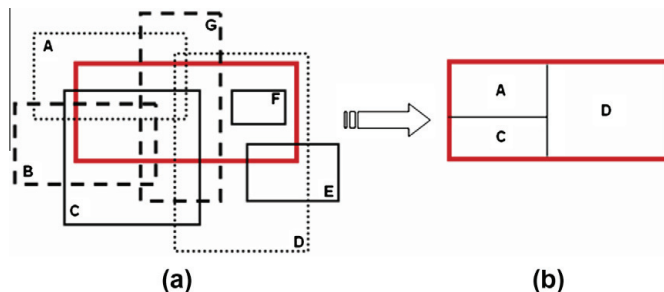


Fig. 8. Selection from the candidate layers (a) the best combination of layers (b).

4. Implementation of an expert clearinghouse

In order to evaluate the expert clearinghouse methodology, a prototype system was developed. The system was implemented on a local intranet. A SOA was used for the implementation of the system. This means that each element illustrated in Fig. 4 was implemented independently, as in the real world, where individual databases, catalogue services, processing services, etc., each owned by an organization/company, can communicate with each other in an interoperable environment.

SOA is an approach for system development that separates functions into distinct units or services (Bell, 2008). These services communicate with each other by passing data or messages from one service to another, or by coordinating an activity between different services. SOA has three primary roles and three primary tasks: (Fig. 9). The service provider, the service requester, and service registrar are distributed computational nodes on the network. The service provider publishes its own service in a service registrar. The service requester uses the service registrar to find desirable services, and then binds to a service provider to invoke the service.

In Fig. 4, the *registry service* plays the role of a service registrar. Different services such as catalogue, WPS, WFS, and WMS services are registered in registry services so as to be discoverable by geoportals or other services. In the structure proposed in Fig. 4, spatial services (including the WPS) are *service providers*, while geoportals and service conductors play the role of *service requester*.

Four prototype WPSs were developed.

- Slope production from DEM (DEM_to_Slope).
- DEM production from height contour lines (Contour_to_DEM).
- Raster data resolution reduction (Res_Reduct).
- Merging different raster maps to produce a unique map (Merge).

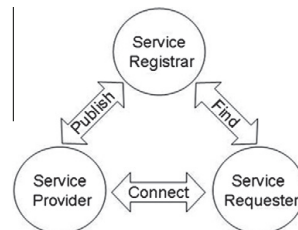


Fig. 9. SOA architecture: roles and tasks.

Each WPS included two main parts:

- a GIS engine, which was developed using ESRI products and VB.Net, and
- a communication interface, which was developed based on OGC's specifications. The three standard operations of WPS, described above, were implemented. The required data input and output parameters are sent to WPSs through an XML *Execute Request by Post Method*.

Based on the OGC's catalogue service specification, four prototype catalogue services were created. MySQL was used for metadata repositories. ISO 19115:2003 was also used to design the metadata schema of the repository. Individual WMS, WFS and WCS services were developed using ESRI products and VB.Net tools, following OGC's specifications, and linked to the databases of each data producer. The *GET Method* was used for communication with the spatial web services.

A service chaining mechanism based on a centralized pattern was also implemented. The centralized pattern was used due to its advantages over the cascaded pattern, described above. In an expert clearinghouse intended to conduct automatic spatial data processes in the context of service chaining, accurate control of the workflow and handling of exceptions are very important. Therefore, the centralized pattern is more suitable than the cascaded pattern for the implementation of service chaining.

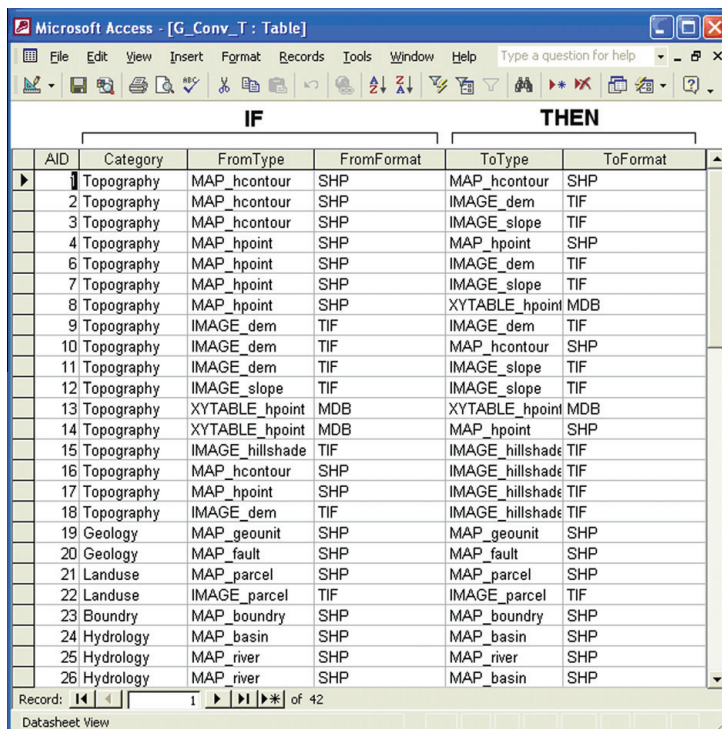
A schema translator was designed as a prototype using a vocabulary database in MySQL. This database contains possible synonyms for the scenarios mentioned in this paper. More advanced

design is possible using open-source ontology-based tools such as Protégé, which is a platform that provides the user with a suite of tools to construct domain models and knowledge-based applications with ontology (Protégé, 2010).

An expert system based on one-conditional backward chaining was also designed, implemented, and linked to the geoportal. One-conditional backward chaining was used in this study due to its simplicity of implementation. With this in mind, a prototype inference engine was developed using the programming language VB.Net. The knowledge base of the system was implemented in the context of IF–THEN fields in a Microsoft Access database (Fig. 10). The methodologies described above (Section 3.2) were used for the implementation of the expert system. If the data requested by the user are not found, the expert system and the inference engine determine them from the candidate datasets, via the knowledge base.

To demonstrate its functionality, the system was tested with two different scenarios consisting of a search for rain data and a search for a slope map. In the first scenario, the user connects to the geoportal via a user interface (Fig. 11), and fills the search parameters including geographic region and the phrase 'rain contour'. If the geoportal cannot discover the dataset, the search phrase is sent to the schema translator to determine synonyms for rain and retrieves 'precipitation' and 'rainfall' as the results. Finally, the geoportal searches the data for the synonyms and returns the output to the user.

In the second scenario, by connecting to a geoportal, a user interface will be available with which a user can set search parameters such as data type, scale, and search region (Fig. 11).



IF				THEN	
AID	Category	FromType	FromFormat	ToType	ToFormat
1	Topography	MAP_hcontour	SHP	MAP_hcontour	SHP
2	Topography	MAP_hcontour	SHP	IMAGE_dem	TIF
3	Topography	MAP_hcontour	SHP	IMAGE_slope	TIF
4	Topography	MAP_hpoint	SHP	MAP_hpoint	SHP
6	Topography	MAP_hpoint	SHP	IMAGE_dem	TIF
7	Topography	MAP_hpoint	SHP	IMAGE_slope	TIF
8	Topography	MAP_hpoint	SHP	XYTABLE_hpoint	MDB
9	Topography	IMAGE_dem	TIF	IMAGE_dem	TIF
10	Topography	IMAGE_dem	TIF	MAP_hcontour	SHP
11	Topography	IMAGE_dem	TIF	IMAGE_slope	TIF
12	Topography	IMAGE_slope	TIF	IMAGE_slope	TIF
13	Topography	XYTABLE_hpoint	MDB	XYTABLE_hpoint	MDB
14	Topography	XYTABLE_hpoint	MDB	MAP_hpoint	SHP
15	Topography	IMAGE_hillshade	TIF	IMAGE_hillshade	TIF
16	Topography	MAP_hcontour	SHP	IMAGE_hillshade	TIF
17	Topography	MAP_hpoint	SHP	IMAGE_hillshade	TIF
18	Topography	IMAGE_dem	TIF	IMAGE_hillshade	TIF
19	Geology	MAP_geounit	SHP	MAP_geounit	SHP
20	Geology	MAP_fault	SHP	MAP_fault	SHP
21	Landuse	MAP_parcel	SHP	MAP_parcel	SHP
22	Landuse	IMAGE_parcel	TIF	IMAGE_parcel	TIF
23	Boundry	MAP_boundry	SHP	MAP_boundry	SHP
24	Hydrology	MAP_basin	SHP	MAP_basin	SHP
25	Hydrology	MAP_river	SHP	MAP_river	SHP
26	Hydrology	MAP_river	SHP	MAP_basin	SHP

Fig. 10. Knowledge base for an expert clearinghouse in a geoportal.

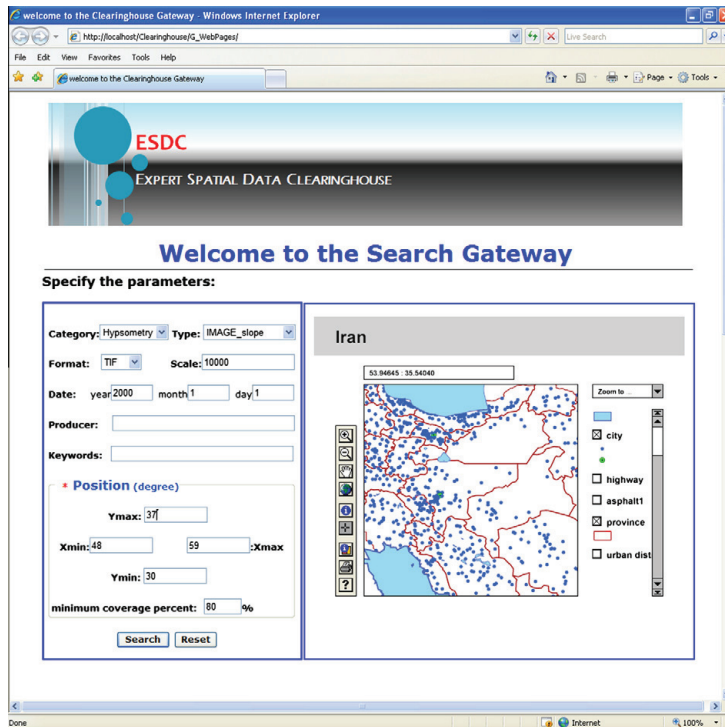


Fig. 11. User interface for the geoportal.

In addition, the *minimum coverage percentage* should be introduced to the geoportal. If the data requested by the user are not available, this value is used by the expert system to suggest suitable candidate data layers (see Section 3.2.2). A simple web mapping tool is also designed to facilitate entering the geographic extent of the search in the geoportal. The search parameters for the desired slope map were entered using this user interface (Fig. 11).

The geoportal searches for the required data through catalogue services. If the data are not discovered, the geoportal will search for candidate data layers. The expert system first determines the candidate data layers and then sends the results to the geoportal to search for them. In this scenario, it is assumed that the data layer requested by the user (slope map) is not found. Fig. 12 shows a list of the candidate data layers discovered by the expert system and presented to the user. For each data layer selected metadata layer such as producer, date of production, and format is presented. In addition, the percentage of coverage of the search region is given for individual candidate data layers. A link is also given for each data layer. The user can view the data by clicking on the link to the WMS of the data provider. This will help the user to select a suitable candidate for the required data.

In the next stage, the system determines a suitable combination of candidate data layers and the processes required for the generation of the data layer requested by the user (Fig. 13). These tasks are conducted using the methodology described in Section 3.2.2. The result is presented to the user as a solution for generating the required data from existing datasets. It also shows how well the proposed candidate data layers cover the search region (Fig. 13).

The user may download the proposed data layers and process them based on the procedures suggested by the system. Alternatively, the geoportal can be requested to manage the processes by clicking on the relevant link (Fig. 13). The geoportal then presents a list of WPSs that can perform the required processes, together with relevant information such as the cost of processing the data. The user can then select the preferred services.

In the next stage, the candidate data, their location, the processes required and the WPSs, are sent to the service controller component. The service controller manages the service chaining required to carry out the processes. In other words, the service controller communicates with suitable web service(s) to produce the data requested by the user (Fig. 14). In this scenario, the following tasks are managed and controlled by the service controller, to produce the data requested by the user (i.e. the slope map):

- i. Connection to a WCS to receive the DEM (dataset 1 in Fig. 13), which has a higher resolution than that required by the user.
- ii. Communication with a WPS to reduce the resolution of the DEM to match the resolution required by the user and obtains the output.
- iii. Communication with a WPS to produce the slope map from the DEM and generate the output (slope map).
- iv. Connection to WFS to obtain height contour lines (dataset 11 in Fig. 13).
- v. Communication with a WPS to produce the DEM from the height contour lines.
- vi. Communication with a WPS to produce the slope map from the DEM and obtain the output.



21 dataset(s) found.

No.	Title	Producer	Date	Coverage Percent	Type	Format	Link	Analysis ID
1	No Title	G1_NCC	4/19/1382	51.4 %	IMAGE_dem	TIF	View Data	11
2	No Title	G1_NCC	4/2/1372	51.3 %	MAP_hcontour	SHP	View Data	3
3	No Title	G1_NCC	5/4/1385	46.5 %	IMAGE_slope	TIF	View Data	12
4	No Title	G1_NCC	2/13/1374	39.1 %	IMAGE_dem	TIF	View Data	11
5	No Title	G1_NCC	3/15/1374	34.3 %	MAP_hcontour	SHP	View Data	3
6	No Title	G3_NGDIR	7/1/1385	31.6 %	MAP_hcontour	SHP	View Data	3
7	No Title	G2_TAMAB	1/9/1384	29.8 %	IMAGE_slope	TIF	View Data	12
8	No Title	G1_NCC	7/15/1374	28.6 %	MAP_hcontour	SHP	View Data	3
9	No Title	G1_NCC	6/12/1382	28.4 %	IMAGE_dem	TIF	View Data	11
10	No Title	G2_TAMAB	12/9/1370	27.1 %	IMAGE_slope	TIF	View Data	12
11	No Title	G2_TAMAB	8/19/1373	24.5 %	MAP_hcontour	SHP	View Data	3
12	No Title	G1_NCC	2/1/1373	24.5 %	MAP_hcontour	SHP	View Data	3
13	No Title	G1_NCC	4/10/1371	24.5 %	IMAGE_dem	TIF	View Data	11
14	No Title	G2_TAMAB	3/6/1376	24.5 %	IMAGE_dem	TIF	View Data	11
15	No Title	G1_NCC	11/10/1381	24.5 %	IMAGE_slope	TIF	View Data	12
16	No Title	G3_NGDIR	7/12/1380	24.4 %	IMAGE_dem	TIF	View Data	11
17	No Title	G1_NCC	7/1/1373	24.3 %	IMAGE_slope	TIF	View Data	12
18	No Title	G3_NGDIR	2/12/1383	23.1 %	IMAGE_dem	TIF	View Data	11
19	No Title	G1_NCC	10/3/1371	18.2 %	MAP_hcontour	SHP	View Data	3
20	No Title	G3_NGDIR	6/10/1381	18 %	IMAGE_slope	TIF	View Data	12
21	No Title	G3_NGDIR	8/11/1373	12.3 %	IMAGE_dem	TIF	View Data	11

Fig. 12. List of candidate data layers (second search results).

- vii. Connection to a WCS to receive the DEM (dataset 4 in Fig. 13).
- viii. Communication with a WPS to produce the slope map from the DEM and obtain the output.
- ix. Connection to a WCS to receive the slope map (dataset 3 in Fig. 13).
- x. Communication with a WPS to merge different slope maps (steps iii, vi, viii and ix) to produce a unique slope map for the region.

5. Discussion

From the tests described above, it can be seen that an expert spatial data clearinghouse can considerably improve the use of available spatial data resources, even for non-expert users. Some issues were identified during the implementation phases that require more research in order to improve the performance and efficiency of expert spatial data clearinghouses.

The first problem concerns the spatial interoperability in a clearinghouse network. Although syntactic interoperability can be achieved in a clearinghouse network by using OGC standards, schema transformation and semantic heterogeneities between spatial web services and spatial data still pose challenges. The use of a semantic web for the development of an expert spatial

data clearinghouse could provide a solution to this problem (see e.g. Hong-Hua, Shi, Hua, & Yang, 2008; Pollock, 2009; Vegetti, Larrateguy, Gonnet, & Leone, 2008).

Another issue is related to the quality of the output data presented to the user. Data processing is conducted automatically, using WPS servers. In addition, spatial data with various semantics and accuracies (from different resources) are integrated to produce the data requested by the user. Therefore, it is important to adapt the appropriate mechanisms and methodologies for quality control of the output data layer.

Service chaining is necessary for an expert spatial data clearinghouse. In order to solve some problems, a number of different services may have to be chained to generate the data requested by the user, as described and presented in the second scenario. More research is required to improve the performance of spatial service chaining.

Examples of other areas requiring attention in the development of an expert spatial data clearinghouse include: improving the algorithm used to identify candidate data, refining the method of determining suitable combinations of candidate data layers (e.g. using genetic optimization methods), and investigating the use of web 2.0 and web 3.0 (e.g. Cambra, 2008; de Longueville, 2010; Fang, Zhao, Xiao, & Zhou, 2008; Fox, 2006; Pollock, 2009; Vegetti et al., 2008; Weijun, Chunmei, & Dafu, 2008).

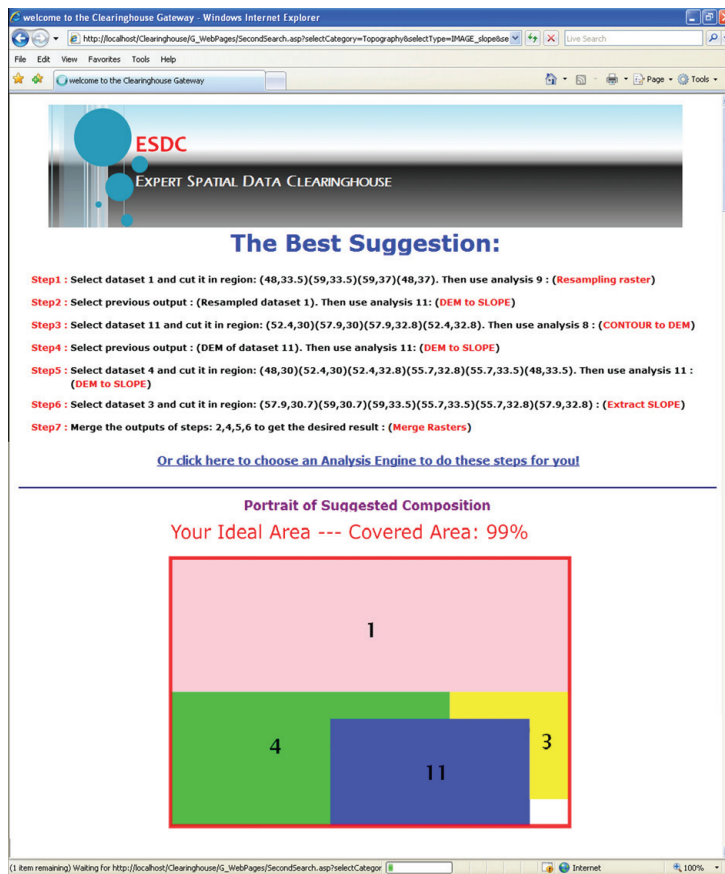


Fig. 13. Instructions for generating the data requested by the user from candidate data.

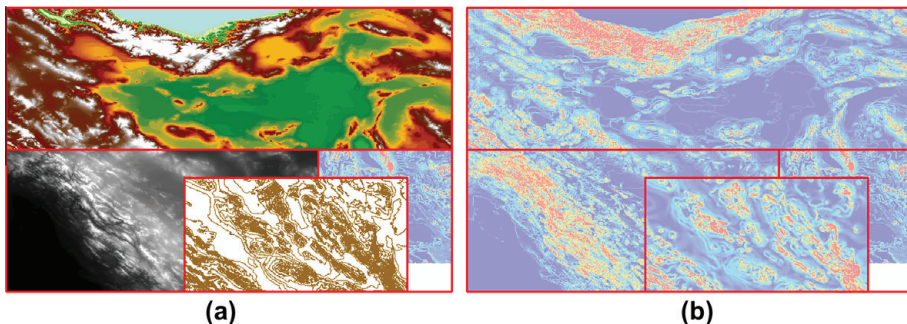


Fig. 14. (a) The candidate data layers selected for generating the slope map requested by the user, and (b) the slope map resulting from the output of the WPS.

6. Conclusions

The functionality of clearinghouses is important in ensuring the function of the spatial data infrastructure. This paper proposes the

use of expert systems to enhance the functionality of clearinghouses. The expert system provides the possibility of converting data in its original form (stored in a database connected to the clearinghouse) into that sought by the user. To our knowledge, this

is the first time an expert system has been used for the improvement of clearinghouses.

The implementation and testing of a prototype system in this study has shown that an expert spatial data clearinghouse with the capability of identifying candidate data layers and processing them to generate the data requested by the user has the following advantages:

- better use of the spatial data resources available,
- increased number of successful searches in a spatial data clearinghouse, by suggesting candidate data to users when the required data are not found, and
- facilitation of users' access to the required spatial data, by automatic arrangement of available processes to produce the data requested by the user from the data available.

These advantages are especially important for those who are not expert users in spatial information sciences, and will help in constructing e-business and e-markets in the spatial information community. Expert spatial data clearinghouses can be considered the third generation of spatial data clearinghouses.

References

- Altuntas, B. M., Bayraktar, D., & Cebi, F. (2006). An application of expert system approach for supplier evaluation and selection. In *Technology management for the global future – PICMET 2006 conference* (Vol. 6, pp. 2755–2758).
- Beaujardiere, J. (2006). OpenGIS® web map server implementation specification. Open Geospatial Consortium Inc. <<http://www.opengeospatial.org/standards/wms>> Accessed 07.09.
- Becker, O., Brown, P., & Cimprich, P. (2003). An introduction to streaming transformations for XML. O'Reilly XML.com. <<http://www.xml.com/pub/a/2003/02/26/stx.html>> Accessed 02.10.
- Bell, M. (2008). *Service-oriented modeling: Service analysis, design and architecture*. Wiley & Sons: 366p. ISBN: 978-0470141113.
- Bergenheim, W., Sarjakoski, L. T., & Sarjakoski, T. (2009). A web processing service for GRASS GIS to provide on-line generalization. In *Proceedings of 12th Agile conference*, Hannover, Germany.
- Bernard, L., Kanellopoulos, I., Annoni, A., & Smiths, P. (2005). The European geoport-one step towards the establishment of a European spatial data infrastructure. *Computers, Environment and Urban Systems*, 29, 15–31.
- Cambra, A. (2008). Web 2.0 as a dystopia in the recent internet. *DIGITHUM*, 10, 01/2008.
- Crompvoets, J., Bregt, A., Rajabifard, A., & Williamson, I. (2004). Assessing the worldwide developments of national spatial data clearinghouses. *International Journal of Geographical Information Science*, 18(7), 665–689.
- Crompvoets, J., Wachowicz, M., Bregt, A., Hofstede, G. J., Wang, J., Vereijken, P., et al. (2006). Exploring worldwide the influence of society on the establishment of national spatial data clearinghouses. In *GSDI-9 Conference proceedings*, 6–10 November, Santiago, Chile.
- de Longueville, B. (2010). Community-based geoportals: The next generation? Concepts and methods for the geospatial Web 2.0. *Computers, Environment and Urban Systems*, 34, 299–308.
- Donaubauer, A., Fichtinger, A., Schilcher, M., & Straub, F. (2006). Model driven approach for accessing distributed spatial data using web services – Demonstrated for cross-border GIS applications. In *Proceeding of the XXIII international FIG congress*, October 8–13, Munich, Germany.
- Einspanier, U., Lutz, M., Senkler, K., Simonis, I., & Sliwinski, A. (2003). Toward a process model for gi service composition. In *Proceedings of Munster GI-Days 2003. IfGI prints band 18. Geodatenund Geodienste-Infrastrukturen – Von der Forschung zur praktischen Anwendung* (pp. 31–46).
- Executive Order 12960 (1994). Coordinating geographic data acquisition and access. In *The national spatial data infrastructure, Federal register*, No. 71. *Presidential documents* (Vol. 59). <<http://www.archives.gov/federal-register/executive-orders/pdf/12960.pdf>> Accessed 09.09.
- Fang, Z., Zhao, L., Xiao, M., & Zhou, Y. (2008). Honeycomb theory of Web3.0. In *IEEE international symposium on service-oriented system engineering* (pp. 263–268).
- FGDC (2000). *Questions and answers about clearinghouses*. Washington, DC: Federal Geographic Data Committee.
- FGDC (2009). FGDC registry and geoport one stop (GOS) resources. <<http://registry.fgdc.gov>> Visited 07.09.
- Fox, G. (2006). Implications of Web 2.0 for the semantic grid. In *Proceeding of second international conference on semantics, knowledge, and grid*. Washington, DC, USA: IEEE Computer Society. ISBN: 0-7695-2673-X.
- Friis-Christensen, A., Lucchi, R., Lutz, M., & Ostländer, N. (2009). Service chaining architectures for applications implementing distributed geographic information processing. *International Journal of Geographical Information Science*, 23, 561–580.
- Giarratano, J., & Riley, G. (1989). *Expert systems: Principles and programming*. Boston, MA: PWS-Kent Publishing Co. 597p.
- Godby, C. J., Smith, D., & Childress, E. (2003). Two paths to interoperable metadata. In *Dublin core conference, DC-2003: Supporting communities of discourse and practice metadata research & applications*, Seattle, Washington, USA.
- Granel, C., Gould, M., & Ramos, F. (2005). Service composition for SDIs: Integrated components creation. In *Sixteenth international workshop on database and expert systems applications*, 26 August 2005 (pp. 475–479). Copenhagen, Denmark: IEEE.
- Hill, L. L. (2000). Core elements of digital gazetteers: Placenames, categories, and footprints. In J. Borbinha, T. Baker (Eds.), *Research and advanced technology for digital libraries: Proceedings of the 4th European conference, ECDL 2000* (pp. 280–290). Lisbon, Portugal: Springer.
- Hong-Hua, C., Shi, Y., Hua, C., & Yang, X. (2008). Towards a framework for designing, deploying and executing semantic web service-based process. In *4th International conference on wireless communications, networking and mobile computing* (pp. 1–4).
- Hunter, J. (2001). MetaNet – A metadata term thesaurus to enable semantic interoperability between metadata domains. *Journal of Digital Information 1*(8). <<http://journals.tdl.org/jodi/article/view/33/34>> Accessed 07.09.
- INSPIRE (2009). Infrastructure for spatial information in Europe. <<http://inspire.jrc.ec.europa.eu/>> Accessed 09.09.
- Iqbal, A., He, N., Li, L., & Dar, N. U. (2007). A fuzzy expert system for optimizing parameters and predicting performance measures in hard-milling process. *Expert Systems with Applications*, 32(4), 1020–1027.
- ISO 19115 (2003). Geographic information – Metadata. <http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020> Accessed 07.09.
- ISO (2005). ISO 19119:2005 geographic information – Services, international standard, ISO TC211.
- Kiehle, C., Greve, K., & Heier, C. (2007). Requirements for next generation spatial data infrastructures-standardized web based geoprocessing and web service orchestration. *Transactions in GIS*, 11, 819–834.
- Lehto, L. (2007). Schema translations in a web service based SDI. In M. Wachowicz, L. Bodum (Eds.), *10th Agile international conference on geographic information science*. Denmark: Aalborg University.
- Lemmens, R., de By, R., Gould, M., Wytzisk, A., Granel, C., & van Oosterom, P. (2007). Enhancing geo-service chaining through deep service descriptions. *Transactions in GIS*, 11, 849–871.
- Lemmens, R., Granel, C., Wytzisk, A., de By, R., Gould, M., & van Oosterom, P. (2006). Integrating semantic and syntactic descriptions to chain geographic services. *IEEE Internet Computing*, 10, 42–52.
- Llavador, M., & Canos, J. H. (2006). XSMapper: A service-oriented utility for XML schema transformation. *ERCIM News*, 64.
- Ma, W., Wang, X., Zhang, Y., & Yu, Q. (2008). The application of expert system to the screening of atmospheric environmental impact from comprehensive road traffic planning in Shanghai. In *Proceedings of the 2008 fifth international conference on fuzzy systems and knowledge discovery, FSKD* (Vol. 4, pp. 636–642). Washington, DC: IEEE Computer Society.
- Marotta, A., & Ruggia, R. (1999). Designing data warehouses through schema transformation primitives. In *Proceedings of the 18th international conference on conceptual modeling (ER'99)*, November 15–18, Paris, France.
- Nebert, D. (2004). Developing spatial data infrastructures: The SDI cook book, Ver 2.0. <<http://www.gsd.org/docs2004/Cookbook/cookbookV2.0.pdf>> Accessed 07.09.
- Neis, P., & Zipf, A. (2007). A web accessibility analysis service based on the OpenLS route service. In *AGILE 2007, international conference on geographic information science of the association of geographic information laboratories for Europe, AGILE*. Aalborg, Denmark.
- Neun, M., Burghardt, D., & Weibel, R. (2009). Automated processing for map generalization using web services. *Geoinformation*, 13, 425–452.
- Pollock, J. T. (2009). Semantic web for dummies. For Dummies. ISBN: 0470396792.
- Protégé (2010). Free, open source ontology editor and knowledge-base framework. <<http://protege.stanford.edu/>> Accessed 02.10.
- Radwan, M. (2002). The development of geographic information infrastructure 'GDI/SDI' to support access to spatial data in distributed environment. Enschede, The Netherlands.
- Rhind, D. (1999). National and internal geospatial data policies. In P. Longley, M. Goodchild, D. Maguire, & D. Rhind (Eds.), *Geographical information systems: Principles, techniques, management and applications* (pp. 767–787). New York: John Wiley & Sons.
- Schilling, A., Lanig, S., Neis, P., & Zipf, A. (2008). DEM processing and 3D navigation using open standards and free geo data. In *International workshop on 3D geoinformation (3D GeoInfo 08)*, Seoul, South Korea.
- Schut, P. (2007). OpenGIS Web processing service. Open Geospatial Consortium Inc. Reference number: OGC 05-007r7. <<http://www.opengeospatial.org/standards/wps>> Accessed 07.09.
- Smith, B., & Mark, D. M. (2003). Do mountains exist? Towards an ontology of landforms. *Environment and Planning B*, 30(3), 411–427.
- Sydenham, P. H., & Thorn, R. (2005). *Handbook of measuring system design, rule-based expert systems*. USA: Wiley & Sons. 1648p.
- Tait, M. G. (2005). Implementing geoportals: Applications of distributed GIS. *Computers, Environment and Urban Systems*, 29, 33–47.

- Vegetti, M., Larrateguy, L., Gonnet, S., & Leone, H. (2008). A semantic web-based architecture to support product data management systems. *Latin American Web Conference*, 91–100.
- 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* (pp. 41–73). Hershey, USA & London, UK: IDEA Group.
- Vretanos, P.A. (2005). Web feature service implementation specification; Open Geospatial Consortium (OGC), OGC 04-094. Accessed 07.09.
- Weijun, W., Chunmei, G., & Dafu, Z. (2008). Study on construction of enterprise knowledge portal based on Web2.0. In *IEEE international symposium on IT in medicine and education* (pp. 621–625).
- Whiteside, A., & Evans, D. J. (2008). Web coverage service implementation standard. Open Geospatial Consortium (OGC), OGC 07-067r5, 2008. <<http://www.opengeospatial.org/standards/wcs>> Accessed 07.09.
- WiscLINC (2006). Wisconsin land information clearinghouse: Schema transformation. <<http://www.sco.wisc.edu/wisclinc/transform.php>> Accessed 02.10.
- Wood, J. D. (1996). The geomorphological characterization of digital elevation models. PhD Thesis. UK: University of Leicester; <<http://www.soi.city.ac.uk/~jwo/phd>> Accessed 07.09.
- Zhang, D.-H., Xie, B., Yu, L., & Di, L. -P. (2008). Open geospatial information services chaining based on OGC specifications and processing model. In *International workshop on education technology and training & international workshop on geoscience and remote sensing*, 21–22 December 2008, Shanghai, China.
- Zipf, A., & Stollberg, B. (2008). Geoprocessing services for spatial decision support in the domain of housing market analyses – Experiences from applying the OGC web processing service interface in practice. In *The 11th AGILE 2008 conference on GI science*. Girona, Spain: Association of Geographic Information Laboratories for Europe.
- Zischg, A., Fuchs, S., Keiler, M., & Meißl, G. (2005). Modelling the system behaviour of wet snow avalanches using an expert system approach for risk management on high alpine traffic roads. *Natural Hazards and Earth System Science*, 5(6), 821–832. <<http://hal.archives-ouvertes.fr/hal-00299303/en/>> Accessed 07.09.

Paper IV

Toomanian, A., Harrie, L., Olsson, P., (2012). Automatic symbolization methods for geoportals. *The Cartographic Journal* (In Press).

Automatic Symbolisation Methods for Geoportals

Ara Toomanian, Lars Harrie and Per-Ola Olsson

*Centre for Geographical Information System, Department of Physical Geography and
Ecosystem Science, Lund University*

email: Ara.Toomanian@nateko.lu.se

Abstract

There is currently a high demand for spatial data usage within web applications. From a technical viewpoint, web services and geoportals aim to fulfil user requirements; however, the current cartographic methods do not satisfy the needs of the end users. The problem is more challenging when the final map contains data from various sources that have various cartographic characteristics; therefore, the vital information might be located under the base map layer. In this paper, we propose the concept of layer priorities as foreground, middle ground, or background, and we propose the two following methods to enhance the symbolisation: polygon overlay and colour saturation methods. The results from two case studies show that these methods can satisfy the requirements of the end users.

Keywords: *web cartography, symbolisation, web map services, geoportals, SDI*

1. Introduction

The need for web-based spatial data applications is increasing rapidly. A wide array of web-based applications has initiated the requirement to disseminate spatial data to the end users via the use of geoportals, which support searching, viewing and downloading spatial data. In this study, we concentrate on the view services.

Cartography is an important issue in geoportal view services. The ability to overlay geospatial data layers from various sources requires symbolisation methods that support visual integration. One important issue is the visual hierarchy, which ranks various data according to their relative importance; the data layers that are more important for the application should be visually emphasised. Another important issue is that information in one layer should not obscure vital information in other layers.

Several approaches can be used to display cartographic information in a geoportal view service. From an end user's perspective, the most restrictive approach is to only allow the selection of layers in which the symbologies of all of the layers are predefined. Second approach is to provide the end user with more capabilities, where he/she is allowed to set the relative

importance of each layer (e.g., whether a given layer is placed in the background, middle ground or foreground of the visual hierarchy). It is also possible to define several symbologies in the geoportal as the third approach, which provides the end user with the ability to choose between various symbologies. The fourth and final approach is to give the end user the capability to create his/her own symbology for each data layer. The methods that are described in this paper primarily target the second approach, in which the end user has the option to set the relative importance of each layer. However, it may also be possible to adjust the presented methods such that they are compatible with the two latter approaches, in which the end user has more freedom.

The aim of this study is twofold. The first aim is to develop and implement the systematic architecture of a *cartographic enhanced geoportal (CEG)*. The second aim is to develop and implement two methods that enable proper cartography while combining several data layers in a geoportal. The first method, the *polygon overlay method*, aims to establish a good polygon presentation that is overlaid on a base map. The second method, the *colour saturation method*, aims to create visual hierarchies by manipulating the colours.

2. Background and related works

2.1. Extended geoportals/clearinghouses

Geoportals are key features of *Spatial Data Infrastructures (SDIs)* (Crompvoets, 2006). A geoportal can be used as an entry point to a network of services that enables a user to search, view and/or download geographic data. The data are not stored in the geoportal; rather, the geoportal serves as a gateway to the data and services. Geoportals enable one to more heavily focus on visualising/analysing the data, and less time is needed to search the data.

This paper proposes a cartographic enhanced geoportal; however, there have been several attempts to enrich the functionality of clearinghouses and geoportals. Mansourian et al. (2010) proposed to complement clearinghouses with expert systems and semantic matching. The expert system facilitates the automatic determination of candidate data sets and the conversion of the available data to the desired data. A schema translator is also used to identify similar data that may be used in other disciplines or other datasets by semantic matching. De Longueville (2010) presented how geoportals can benefit from the Web2.0 features. He provides an overview that supports the development of the next-generation geoportal by defining connected concepts, emphasising the pros and cons of this approach, and proposing suitable implementation strategies.

2.2. Cartographic techniques

According to the International Cartographic Association (ICA), cartography is “the art, science and technology of making maps together with their study as scientific documents and works of art” (ICA, 1973 p. 1). This definition implies that using certain techniques and following specific rules is not sufficient to create a “good” map. Artistry also plays an important role in cartography (Keates, 1996).

In the context of cartography, visual hierarchy is an important aspect of map visualisation. It is a graphical representation of the intellectual hierarchy in which the symbols and map elements are ranked according to their relative importance (Slocum et al., 2005:p220). The concept deals with

the emphasis of the more important symbols and the de-emphasis of insignificant and base information. A proper visual hierarchy first directs the user's eye to the most important element and later to the remaining map elements. Expert cartographers know how to design cartographically good maps. However, this knowledge cannot easily be transferred to automatic map-making programmes. The situation is particularly problematic when various data sources are integrated in web map services, each which contains its own special visual characteristics.

Recently, researchers have enhanced web cartography. Iosifescu-Enescu et al. (2009) utilised an enriched cartographical approach for *Open Geospatial Consortium (OGC)* standards to fulfil the complex visualisation requirements that stem from environmental management. They used cartographic extensions to express cartographic rules with spatial operators and advanced-feature filtering for layer masking, flexible point symbolisation, and patterns and gradients for all of the spatial features. In the application of such a method, the critical point of creating thematic maps is also solved with extensions for an intuitive choropleth map and the generation of various diagram types. Bucher et al. (2007) address some of the cartographic issues for designing efficient on-demand maps in service-oriented architecture (SOA). According to their study, the current standards do not support some of the crucial steps of an on-demand map design process and may benefit greatly from knowledge that was formalised by cartographers, including the definition of styles for representing geographical data on the map. They aim to integrate portions of the cartographic knowledge (primarily semiological rules) in web map service-oriented architectures. They also employ a set of web services that are dedicated to facilitating the definition of accurate legends with respect to user objectives and data. Brewer and Battenfield (2007) provide methods that can be used to create a map from a multiple-representation database. They emphasise map display changes using symbol design or symbol modification. In addition, it comprises a demonstration of the establishment of specific map display scales at which symbol modification should be imposed.

In this study, we provide cartographic support to the end user by developing a system architecture using two methods. In the following sections, we describe the background of each method.

2.2.1. Polygon overlay method

In this study, the polygon overlay method is based on the representation of a polygon with symbols. This method is based partially on the following previous studies: Harrie et al. (2004) developed a method to place symbols in a least-disturbing position, and Harrie & Revell (2007) developed a method to place symbols in a semi-random pattern. In the latter study, the basic idea was to optimise the distance between the neighbouring symbols in such a way that the result mimics the manual placement of symbols. This method of describing the pattern using neighbouring objects has also been used to characterise dot maps (Ahuja, 1982; Sadahiro, 2000). Another study was conducted in the context of mobile cartographic services (Edwards and Burghardt, 2004), in which the authors stress the spatial relationships between symbols and the base map features and propose solutions through the improved modelling of spaces. To accomplish this goal, the authors use a combination of generalisation techniques and spatial modelling.

To optimise the placement of symbols in this study, we use a combinatorial optimisation approach in a similar fashion that is used in label placement (cf. Zoraster, 1986; Christensen et al., 1995; Zoraster, 1997; Ware and Jones, 1998; Zhang and Harrie, 2006).

2.2.2. Colour saturation method

Chroma (saturation) is a measure of a colour's "purity" (Dent, 1999) and can be explained by comparing a colour to a neutral grey. For grey, the saturation is 0. When more "colour" is added, saturation increases, and the colour will appear less grey. At 100% saturation, no grey exists, and the colour is "pure". The HSV colour system uses three dimensions (hue, saturation and value) to define a colour. The colour saturation method enables a user to build a visual hierarchy by de-emphasising the background information. This idea of background de-emphasis can be developed to enable a user to define a wide range of levels in a visual hierarchy.

Chesneau et al. (2005) established a method that is based on the Itten colour contrast theory in which each graphic sign analyses the colour contrasts with its neighbours. This analysis is subsequently validated at a more global level. If problems in colour contrast are detected, another graphic solution is proposed. The process is repeated until a more legible map is obtained. In addition, Buard and Ruas (2009) proposed a number of processes that can be used to improve the colour contrasts of topographic on-demand maps. They used an approach that was proposed by Bertin (1983), which decomposes a legend into a set of meaningful sign couples. The principle of the legend improvement method is to analyse the colours and modify the colours if the following rules and conditions are not satisfied: (1) the colour of a legend line should be coherent to the colour family that is associated with the theme of the legend line, and (2) the relationship between two colours should respect the relationship of the association, order or difference that exists between the two legend lines that they represent.

3. Approach and methods

This section begins with a description of the system architecture of a cartographic enhanced geoportal and follows with a description of two cartographic methods to integrate several data layers.

3.1. System architecture

To support geoportals that produce proper cartography, we propose a system architecture that is based on the following components (Figure 1):

- *Client* is a normal *Web Map Service* (WMS) client with added functionality to create a visual hierarchy, and it enables the end user to define whether a layer is in the background, middle ground or foreground (cf. Figure 2).
- *Registry service* manages the registry of spatial services that are used by the geoportal.
- *Cartographic enhanced geoportal* is a geoportal with added functionality to enable good cartography. The geoportal consists of three components:
 - The *web map programme* is an entry point to geographic data on the web. It is a web site on the internet where download and view services is registered and provides users (clients) with the ability to view and/or develop spatial data.
 - The *cartographic core* determines the symbolisation of the user-selected layers. In this component, several cartographic methods can be implemented.

- The *symbolisation library* contains one (or several) symbolisation(s) for each dataset that is registered in the geoportal.
- *Geospatial web services* are standard services for the distribution of geographic data such as a standard *Web feature service (WFS)*.

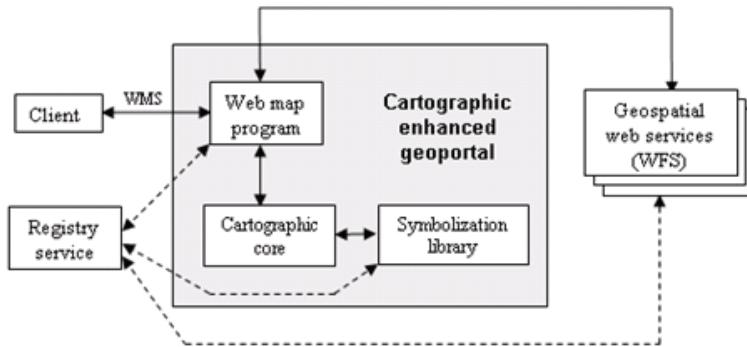


Figure 1. System architecture.

LAYER	F	M	B
Roads	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Buildings	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Forest_Broadleaved	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Forest_Coniferous	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Forest_Mixed	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Protected_CulturalHeritage	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Protected_Nature	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Protected_Recreation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Submit

Figure 2. Graphical interface for selecting the hierarchical levels for the data layers (B-background, M-middle ground and F-foreground).

It is a requirement that all of the geospatial web services are registered in the geoportal. In this registration process, the service metadata (including web addresses) are stored in the web map programme project files. It is also necessary that symbolisations are stored in the symbolisation library and information regarding the geometry type is stored in the geospatial web services.

For the end user, the cartographic enhanced geoportal is similar to any other geoportal. The only difference is that the user is requested to determine in which level of the visual hierarchy a layer should be placed. The following levels are defined (cf. the example that is given in Figure 2):

- Foreground – Additional information layers that are of high relevance for the application.
- Middle ground – Data layers in the base map that are vital for the application.
- Background – Less prominent data layers in the base map.

This definition implies that the information that we normally denote as the base map is located in the background and middle ground, and the thematic information is in the foreground.

The order of the layers in the final map (i.e., the map that is sent to the client) is decided by the cartographic core. The rules are as follows:

- A layer in the foreground is always on top of a layer in the middle ground, and a layer in the middle ground is always on top of a layer in the background.
- Within each level (i.e., the background, middle ground and foreground), the point layers are on top, the line layers are in the middle and the polygon layers are at the bottom. The ordering of two layers that have the same geometry in the same level is set according to the order in the WMS *getMap* request.

To enhance the feeling of a visual hierarchy, special methods can be implemented in the cartographic core. Below, we propose two such methods:

1. *Polygon overlay method*: This method is applicable for polygon layers that are destined to reside in the foreground. In a planning application, this could typically include such data as planning regions and restricted areas in a planning application.
2. *Colour saturation method*: The aim of this method is to de-emphasise data in the background by decreasing the saturation of their symbology.

3.2. Polygon overlay method

One common problem occurs when polygons in the foreground hide important information in the middle ground (hidden information in the background is not considered to be highly disturbing and is therefore not considered). In this section, we propose that the polygons can be symbolised using a combination of a boundary line and symbols in which the symbols are placed in a manner in which they do not obscure vital information in the middle ground. The polygon overlay method does not change the base map data in the background or middle ground.

The polygon overlay method is defined as a combinatorial optimisation problem and consists of the following steps:

1. Initialisation step – The polygon is symbolised by the boundary and symbols that are placed in preliminary positions.
2. Cost function step – A cost function is created for the positions of the symbols.
3. Optimisation step – In this step, the positions of the symbols are determined by determining the optimum solution to the cost function.

Two elements are utilised for the entire process. First, we use a *minimum bounding box* (MBB) to cover the entire polygon as a frame; all of the other steps are operated within this area. Within the MBB, we create a fine grid that will act as the resolution of the symbol placement. The

symbol centre can only be placed in the centre of a cell in this grid (Figure 3), and the size of the symbol ($is_{x,y}$) is defined as follows:

$$is_x = (2k + 1) \cdot \Delta x, \forall k \in \mathbb{Z} \quad (1)$$

where \mathbb{Z} is an integer, Δx is the resolution in the x -direction in the fine grid and is_y is defined analogously.

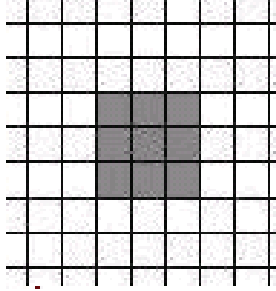


Figure 3. The relationship between the symbol and the dense grid.

Here, $k=1$ in Equation (1).

3.2.1. Initialisation step

The main task in the initialisation step is to determine the number of symbols and assign them to preliminary positions. To perform this step, the fine grid that is covered by the MBB is utilised. First, a coarse grid with size $cg_{x,y}$ is defined as follows:

$$cg_x = (2l + 1) \cdot \Delta x, \forall l \in \mathbb{Z}, l \gg k \quad (2)$$

where Δx is the resolution in the x -direction in the fine grid and cg_y is defined analogously. Because l is larger than k , the coarse grid is larger than the fine grid.

Second, symbols are placed in all of the cell centres in the coarse grid and are located within the polygon (Figure 4).

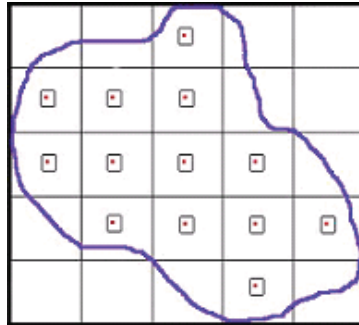


Figure 4. The initial positions of the symbols.

3.2.2. Cost function step

To determine the proper positions for the symbols, we must consider several aspects. Our approach is to associate all of these aspects with a cost and determine the solution with the lowest total cost. We consider the following costs for the symbols: placement cost, spatial distribution cost and removal cost.

Placement cost

Placement cost is the cost of hiding and distributing other objects by the symbols. The cost computations are based on the fine grid. We also introduce the following two terms: *object cost* and *cell value*.

Each object creates an object cost around the symbology such that the cost decreases linearly with increasing distance from the symbology until it reaches a threshold value (Figures 5). Figure 6 shows an example of the object cost, in which we focus on a portion of the road layer (green) and a building (red) to describe the method in detail. In this example, the white-coloured cells with a minimum cost (e.g., “0”) are ranked highly for symbol placement, and the cells with dark colours are not suitable.

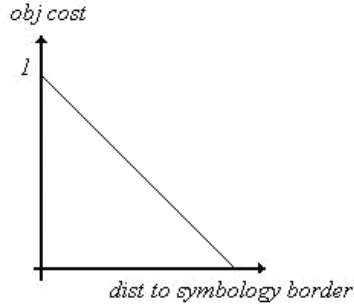


Figure 5. The relationship between an object cost and the distance to the symbology border.

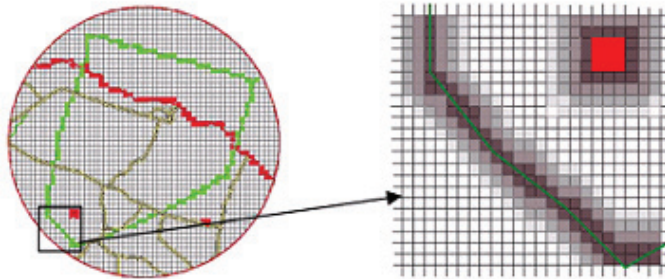


Figure 6. Illustration of the object cost.

The cell value is the value for each cell in the fine grid. We define a cell value for cell k (cv_k) as follows:

$cv_k = 999$ if the cell overlaps any symbology
else

$$cv_k = \sum_{i=1}^n object\ cost \quad (3)$$

where n is the total number of objects.

According to the definition of a cell value, cells that overlap with any symbology (or even a part of the symbology, such as the width of a road) are set to a high cost (e.g., “999”). This indicates that they are not suitable candidates for symbol placement.

Each symbol covers a number of cells, and the cost for each symbol depends on the cell values for all of the cells that are covered at a certain position. We use the following function to calculate the placement cost (c_p):

$$c_p = \sum_{i=1}^n \sum_{j=1}^{m_i} cv_{ij} \quad (4)$$

where c_p is the placement cost, n is the number of symbols, m_i is the number of cells that are covered by $icon_i$ and cv_{ij} is the $cell_j$ that is covered by $symbol_i$.

Spatial distribution cost

Spatial distribution cost is used to describe the difference between the distances in ideal places and other places when a symbol is moved from the centroid to another location in a cell. In this instance, the cost is calculated as follows:

$$c_{sd} = \sum_{i=1}^n |d - id| \quad (5)$$

where c_{sd} is the spatial distribution cost, d is the Euclidean distance between the centroids of neighbouring symbols, id is the corresponding ideal distance (which is equal to the original Euclidean distance in the initialisation step) and n is the number of neighbourhood relationships.

In this step, we define a neighbouring symbol as a symbol that lies in any of the eight neighbour cells in the coarse grid (Figure 7).

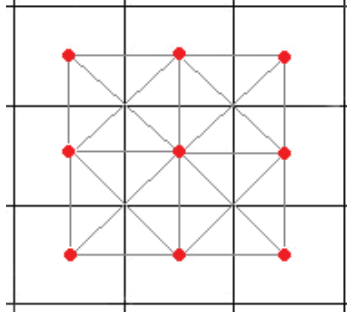


Figure 7. Definition of a neighbourhood relationship.

Removal cost

Removal cost analyses the situations in which a number of symbols are too close to each other, due to the movement in previous steps (the spatial distribution cost). Here, if the distance reaches a minimum threshold, then one or more symbols must be removed. In addition, all of the neighbourhood relations are removed as well. The removal cost (c_r) is calculated as follows:

$$c_r = \text{Constant value for each symbol} * \text{numbers of removed symbols} \quad (6)$$

where c_r is the removal cost.

3.2.3. Optimisation step

Total cost is calculated using the following cost function:

$$tc = w_p c_p + w_{sd} c_{sd} + w_r c_r \quad (7)$$

where tc is the total cost and w_x is the weight of each cost that is achieved from the optimisation step.

The cost function can act as an objective function in a combinatorial optimisation. To optimise the cost function, we apply the combinatorial optimisation method *simulated annealing* (Russell and Norvig, 1995). This method is a stochastic hill-climbing algorithm that provides the possibility to escape from local minima.

The expected output of the polygon overlay method is illustrated in Figure 8. In this figure, the background layers are roads and buildings, and the following two foreground layers are included: protected nature and protected recreation (Figure 2).

3.2.4. Symbolisation issues

One key point regarding the polygon overlay method is that the user of the map should easily understand the extent and interpretation of the polygon; this point is particularly important if there are several overlapping polygons in the foreground layer. Below, we list some of the requirements of the symbology and our solution to these requirements (Figure 10).

The polygon must be clearly separated from the background

This requirement is accomplished by selecting strong and highly saturated colours for the polygon border and symbols while ensuring that the data in the middle ground and background are deemphasised using the colour saturation method (section 3.3).

The polygon border and symbols must be associated with each other (otherwise, the user may regard the symbols as separate symbols that represent point features and not as being connected to the polygon)

This finding is primarily achieved by using the same colour hue in the polygon that was used for the symbols (because hue is an associative graphic variable; cf. Bertin, 1983). To improve the association, we have also added a transparent line (with the same colour hue) inside the polygon border; this line is particularly important in those cases in which the entire polygon is not visible on the screen.

The symbols must be simple, distinguishable and describe the content of the polygon

In general, the symbols should describe the content of the polygon; that is, mimetic (pictorial) symbols should be used (Robinson et al., 1984; MacEachren, 1995). However, as argued by Spiess (1988), it is important that the symbols are not too detailed, and they should be distinguishable (for example, by applying variations in the tops of the symbols). In the context of an SDI, the polygon overlay method would benefit from the standardisation of the symbols.

3.3. Colour saturation method

From a visualisation perspective, a major problem in the creation of maps is the use of highly saturated colours in the base maps. To solve this problem in this study, the colour saturation method is applied to de-emphasise the background layers in the background and middle ground.

The colour saturation method is simple but requires supplementary computation and conversion. Here, the goal is to decrease the colour saturation with a specific range (e.g., 50%) or any constant values for each symbol in the background layer. The procedure for the conversion is as follows (assuming that the colour is stored in RGB values):

1. Convert the colour format from RGB to HSV (Foley et al., 1996).
2. Decrease the saturation according to the predefined ranges.
3. Reconvert the new colour to the RGB format.

By using these steps, the background colours become less distinct such that one can focus on the foreground layers.

4. Case study

To evaluate the cartographic enhanced geoportal, a prototype system was implemented according to the architecture in Figure 1, and a case study was performed with two different datasets. In the first scenario, an urban-planning map was created from the data source of the municipality of

Lund in southern Sweden. In the second scenario, another application of urban planning was created for the Helsingborg area.

4.1. Data

Two different datasets are used in this research. The first dataset is the municipality of the Lund region and is collected as follows:

- Road layers and building layers are provided by the municipality of Lund.
- Polygon layers are obtained from the county administrative board of Skåne (<http://www.gis.lst.se/lstgis/>) as follows: protected areas of national interest for (1) cultural heritage, (2) recreation and (3) nature.
- Corine land-cover vector data (<http://www.eea.europa.eu/data-and-maps/data/>) are described as follows: broad-leaved forest, coniferous forest and mixed forest.

The second dataset is for the Helsingborg region and includes layers from the following data sources:

- Data that is provided by Lantmäteriet (The Swedish Mapping, Cadastre and Land Registration Authority).
- The Helsingborg Municipality.

4.2. Implementation

The implementation consists of a client, a cartographic enhanced geoportal and external web services (Figure 8). The communication between these components follows the OGC WMS standard; however, because a user must define the visual hierarchy, an additional parameter is added to the *GetMap* request. This parameter (*Layer Priority – foreground, middle ground, or background*) acts as a WMS *Vendor specific parameter* (VSP) (OGC, 2010b). When a VSP is included, the service may choose to omit it in a capabilities response, which is the case in this implementation.

- Client is a WMS client that is written in Java with added functionality that permits a user to select whether a layer should belong to the foreground, middle ground, or background (Figure 2).
- Cartographic enhanced geoportal consists of the following three components:
 - The web map programme is *MapServer* (MapServer, 2010) and is run with both the *Common Gateway Interface* (CGI) and the Java *MapScript API*.
 - Cartographic core implements the following two methods: (1) the polygon overlay method, which utilises *OpenJUMP* (OpenJUMP, 2010) to convert GML files to the *Well-Known Text* (WKT) format and *Java Topology Suite* (JTS) (JTS, 2010) for the geometric computations, and (2) the colour saturation method, which is a Java programme.
 - The symbolisation library consists of *Styled Layer Descriptor* (SLD) documents (OGC, 2010a).

- External web services are normal OGC WFS services. In this study, two WFS services are used and run on the following two platforms: *Geoserver* (GeoServer, 2010) and *MapServer*. Both of these platforms store their data in *PostgreSQL* databases (PostgreSQL, 2010) with *PostGIS* extensions that allow the storage of spatial data (PostGIS, 2010).

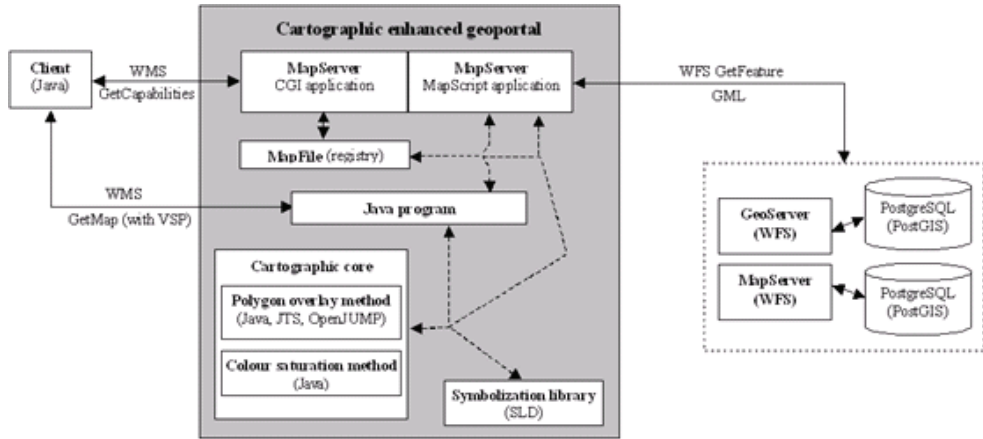


Figure 8. Implementation of a cartographic enhanced geoportal.

The workflow for the implementation is as follows. The client sends a WMS *GetCapabilities* request to the cartographic enhanced geoportal following the OGC standard. The request is sent as a CGI command to the MapServer CGI application. The geoportal responds to the request by returning an XML document that describes the capabilities (that is, a list of available layers or styles from the symbolisation library).

For this purpose, MapServer does not send a *GetCapabilities* request to the external services; instead, MapServer obtains the capabilities from a *Mapfile*, which is a MapServer-specific configuration file that contains information regarding the available data (e.g., where the different layers are stored and the available styles). When an external service registers in the registry, the capabilities of that service and the connection to the service are added to the *Mapfile*. In this implementation, the registry is not included; instead, the *Mapfile* is updated manually.

From the capabilities, the user formulates a WMS *GetMap* request. This request defines parameters, such as which layer(s) to include, which style(s) to use for symbology and the format of the output map image. The next step distinguishes the cartographic enhanced geoportal from an ordinary WMS-service; namely, it selects whether a layer should belong to the foreground, middle ground or background using the VSP *LayerPriority* that was introduced above. When a VSP is introduced, the WMS standard requires that a *GetMap* request returns a map image; this is also true if the VSP is missing (OGC, 2010b). The geographic-enhanced geoportal solves this problem by setting *Layer Priority* to the middle ground if no value is given.

The request, which includes *LayerPriority*, is sent to the Java programme, which acts as a spider in the geoportal. For the client, the Java programme acts as a server, and the communication is handled by a TCP/IP connection.

Because a WMS *GetMap* request returns only a map image, and because the polygon overlay method works with the geometry of the various layers, a WFS *GetFeature* request must be sent to the external services (i.e., a service that returns the actual data as a GML file). The Java programme formulates a WFS *GetFeature* request, which it sends to the external services via the MapServer MapScript.

The Java programme then interacts with the cartographic core, the symbolisation library and MapFile (the registry). In the cartographic core, the polygon overlay method utilises OpenJUMP to convert GML files to the WKT format and JTS for the geometry. The colour saturation method is a Java programme that computes the new RGB values (see Section 2.3) and applies them to the registered SLDs.

The output from the polygon overlay method is one new point layer for each polygon layer that is sent to the method. These point layers provide the positions for the symbols that will be used to present a polygon layer together with the boundary from the polygon layer itself. The result of the colour saturation method is a temporary SLD, which is updated according to *LayerPriority* in the *GetMap* request.

The new, modified data are then made available for MapServer by the Java programme, which also sends an OGC standard (without VSP) *GetMap* request to MapServer. MapServer and creates the map image according to the original *GetMap* request from the client. Finally, the image is sent back to the client via the Java programme.

4.3. Results

Figures 9 and 10 show the differences between typical visualisation and the cartographic enhanced methods for two different scenarios. In Figure 9, which shows a planning application, we define a single layer as a foreground layer. The figures include the following layers:

1. National outdoor interest (foreground) in blue
2. Highway (middle ground) in brown
3. Coniferous layers (middle ground) in green
4. Municipality (background) in cream colour

Figure 9 (a) illustrates a common polygon visualisation in which the underlying information is hidden. In Figure 9 (b), boundary and symbols (a walking man) are used to visualise the foreground layer using the polygon overlay method, and the middle ground and background layers are faded using the colour saturation method.

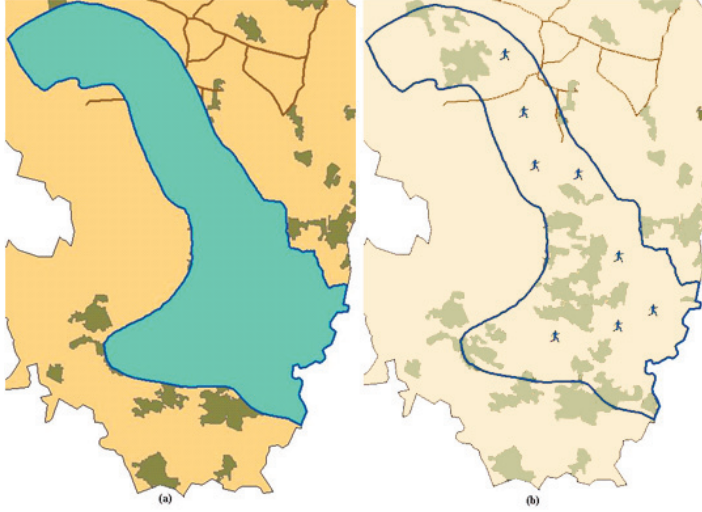


Figure 9. (a) Typical visualisation using the Lund municipality dataset. (b) The method result for the first scenario with a single foreground layer.

The details of the results are described as follows. As described in the methodology, simulated annealing is used for the optimisation step. The weights that are used in the cost function are set as follows:

$$w_p=1, w_{sd}=0.3 \text{ and the } w_r=0.1.$$

Moreover, the value for removing one symbol is set to 1000.

The simulated annealing algorithm calculated the cost with 400 iterations with the following parameters:

$$\text{alpha}=0.999, \text{temperature}=400, \text{epsilon}=0.001.$$

The result for the total cost of the optimisation step is calculated as follows:

$$tc=4722$$

where $c_{sd}=9803$, $c_p=1780$ and $c_r=0$ (i.e., no removal).

The colour saturation method maintains the foreground layer border and decreases the saturation of the background and middle ground layers by 20 and 10%, respectively.

The second scenario is another city planning map in Helsingborg region with the following dataset:

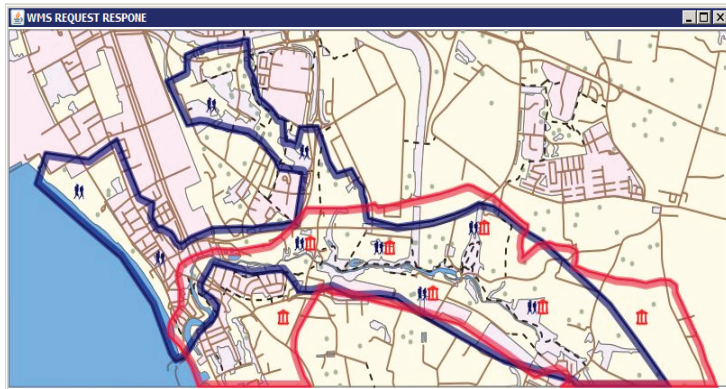
1. Helsingborg cultural area (foreground polygon) with a red line border and the same colour shadow showing the interior of the area (old building symbol)

2. Helsingborg nature region (foreground polygon) with a dark-blue line border and the same colour shadow representing the interior of the region (walking man symbol)
3. Helsingborg footpath (middle ground) in a black dashed line style
4. Helsingborg roads (middle ground) in brown
5. Helsingborg ancient remains (middle ground) in grey squares
6. Helsingborg land cover (base map) without considering the programme as a background layer.

Figure 10 (a) shows two polygons that are overlaid in such a way that the information is lost, due to the poor visualisation. In Figure 10 (b), a colour-bordered boundary and two symbol types with the same colours (a walking man as a national outdoor interest and an old building to represent the cultural area) are used to visualise the foreground layers using the polygon overlay method, and the middle ground and background layers are faded using the colour saturation method. Here, there are three different areas: the area that contains the first foreground polygon that is overlaid with the old building symbol, the area that contains the second foreground polygon that is overlaid with the walking man symbol and the intersection of the two foreground layers that are covered by two symbols in parallel. The optimisation parameters are set as the first scenario. Finally, the border colours are set as the colour of the symbols.



(a)



(b)

Figure 10. (a) Typical visualisation with the Helsingborg municipality dataset. (b) Polygon overlay method result for the second scenario with two layers as the foreground.

7. Discussion

In this paper, we proposed a system architecture for a cartographic enhanced geoportal that contained the following two methods: the polygon overlay method and the colour saturation method. In a case study, we showed that through the application of these methods, we can overlay thematic information on top of a base map without any loss of visual data.

Although the method yields proper results comparing to the typical existing maps, a data-preparation step would be required in a production environment. Switching from datasets requires local tuning for specific parameters. In the current implementation, any new data source would have to be registered within the system, and the symbology library should be updated with the new data source styling information.

Another main issue is the generality of the method. For example, what happens if the number of foreground layers is increased? In general, if the number of foreground layers is higher than 3, it increases both the computational complexity and the constraints to find the proper space for the symbols. In such cases, even the interpretation of the map is difficult. Thus, the principal restriction in any type of visualisation method is the quantity of layers. Moreover, regardless of the number of foreground layers, selecting the visualisation of more layers will lengthen the execution time for the CEG implementation. Figure 11 illustrates the execution time for various layers, including either one or two layers that are selected as the foreground. Here, the number of foreground layers and their complexities can also affect the total execution time.

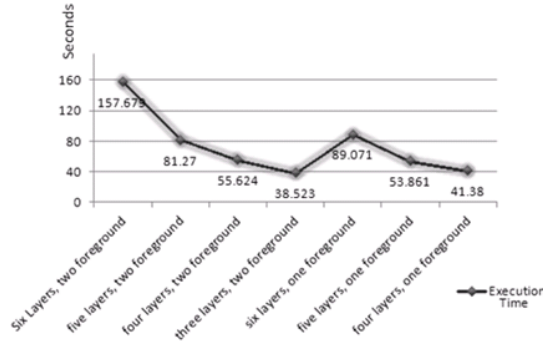


Figure 11. Execution time for different numbers of layers.

The CEG implementation consists of the following 3 steps: data preparation, cost calculation and an optimisation step; each step has its own execution time, and the total execution time is calculated by aggregating the three steps. The implementation was tested using a desktop PC with an Intel Xeon CPU with 4 Core(s), 2.0 GHz by calculating the execution time in the Java programme for each step (in milliseconds). The results show that the execution of updating the raster (placement cost array) has the highest implementation time percentage. Figure 12 presents the percentage of the execution times for the various steps in the CEG implementation. By

comparing the two pie charts, it can be seen that the selection of more layers to be visualised increases the execution time percentage for placement cost raster updating.

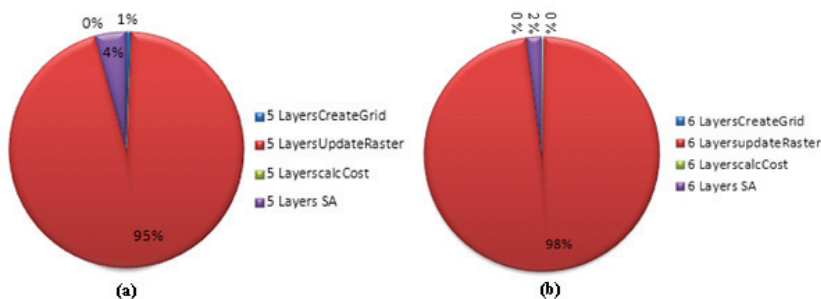


Figure 12. (a) The percentage of the execution time in the various steps for five layers. (b) The execution time for six layers.

The map scale is another major item that must be considered. When switching from one scale to another scale, the symbol size must change according to the cartographic basic instructions. This parameter will affect the various steps within the polygon overlay method, including the optimisation step.

8. Conclusions and future plans

In this paper, we described the methodology and implementation of a cartographic enhanced geoportal. We used three types of levels as layer priorities in this research according to the visual hierarchy for the so-called foreground, middle ground and background. Our prototype system contains the following two methods: *polygon overlay* and *colour saturation*. In the polygon overlay method, we proposed a symbol placement approach by using displacement, distribution and removal cost functions to calculate the total cost for a random symbolisation. In the second method, we decreased the colour saturation of the unimportant layers to make the foreground layer more transparent according to the visual hierarchy concept. The optimisation of the cost function is based on a simulated annealing approach. By implementing the system architecture and applying it to various scenarios, the results showed that these methods are appropriate for visualising the overlaying layers without any data lost. We believe that these types of methods will be increasingly important for improving cartographic quality in future geoportal view services.

Acknowledgments

The European Union funding programme Erasmus Mundus “External Cooperation Window” (EMECW lot8) financed the first author of this study. The authors are grateful to the municipalities of Lund and Helsingborg, County administrative board of Skåne and the European environment agency for providing data. The authors are also grateful to Ali Mansourian and Anders Söderman for their interesting discussions.

References

- Ahuja, N. (1982). 'Dot pattern processing using Voronoi neighbourhoods', **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 4, pp. 336-343.
- Bertin, J. (1983). **Semiology of Graphics**, University of Wisconsin Press, 415 pages.
- Brewer, C. A. & Buttenfield, B. P. (2007). Framing guidelines for multi-scale map design using databases at multiple resolutions, **Cartographic and Geographic Information Science**, 34, pp. 3-15.
- Buard, E. & Ruas, A. (2009). Processes for improving the colours of topographic maps in the context of maps on demand, in **Proceedings of the 24th ICA Conference**, Santiago de Chile, Chile.
- Bucher B, Buard E, Jolivet L, Ruas A, 2007. The need for web legend services, in **Proceedings of the 7th International Symposium on Web and Wireless Geographical Information Systems (W2GIS07)**, Cardiff, UK.
- Chesneau, E, Ruas, A. and Bonin, O. (2005). Colour contrasts Analysis for a better legibility of graphic signs for risk maps, **International Cartography Conference (ICC2005)**, Spain.
- Christensen, J., J. Marks and Shieber, S. (1995). An empirical study of algorithms for Point-label placement, **ACM Transactions on Graphics**, 14(3), pp. 203-232.
- Crompvoets, J. (2006). **National spatial data clearinghouses, worldwide development and impact**. PhD Thesis Wageningen University, Wageningen 2006.
- De Longueville, B. (2010). Community-based geoportals: The next generation? Concepts and methods for the geospatial Web 2.0, **Computers, Environment and Urban Systems**, 34 (4), Geospatial Cyberinfrastructure, pp 299-308.
- Dent, B. D. (1999). **Cartography, Thematic mapping – Fifth edition**. WCB/McGraw-Hill, 1999.
- Edwards, A. and Burghardt, D. (2004). Portrayal and Generalisation of Point Maps for Mobile Information Services, In **Map-based mobile services – Theories, Methods and Implementations**, Reichenbacher, T. and L. Meng (Eds), (Berlin: Springer-Verlag)
- Foley, J. D., van Dam, A., Feiner, S. K., and Hughes, J. F. (1996). **Computer Graphics: Principles and Practice**, 2nd ed. in C, Addison-Wesley.
- Harrie, L., and Revell, P. (2007). Automation of Vegetation Symbol Placement on Ordnance Survey H 1:50000ScaleMaps, **The Cartographic journal**, Vol.44, No.3, Ordnance Survey Special Issue, pp. 257 – 267.
- Harrie, L., H. Stigmar, Koivula, T. and Lehto, L. (2004). An Algorithm for Icon Placement on a Real-Time Map. In Fisher, P. (ed), **Development in Spatial Data Handling**, Springer, pp. 493-507.
- ICA (1973). **Multilingual Dictionary of Technical Terms in Cartography**. Franz Steiner Verlag GMBH.
- Iosifescu-Enescu, I., Hugentobler, M. and Hurni, L. (2009). Web cartography with open standards—A solution to cartographic challenges of environmental management, **Environmental Modelling & Software**, 25(9), Thematic issue on Sensors and the Environment - Modelling & ICT challenges, September 2010, pp. 988-999.
- Keates, J. S. (1996). **Understanding Maps** – 2nd edition. Longman, 1996.
- MacEachren, A. M. (1995). **How Maps Work**. The Guilford Press.

- Mansourian, A., Omid, E., Toomanian, A. and Harrie, L. (2010). Expert system to enhance the functionality of clearinghouse services, **Computers, Environment and Urban Systems**, **35(2), Geospatial Analysis and Modeling**, March 2011, pp. 159-172.
- Robinson, A. H., Morrison, J. L., Muehrcke, P. C., Kimerling, A. J., and Guptill, S. C. (1995). **Elements of Cartography**, 6th ed., John Wiley & Sons.
- Russell, S. and Norvig, P. (1995). Artificial Intelligence – A Modern Approach, **Prentice-Hall**.
- Sadahiro, Y., 2000. 'Perception of Spatial Dispersion in Point Distributions', **Cartography and Geographic Information Science**, 27(1), pp. 51-64.
- Slocum A., McMaster R., Kessler F. and Howard H. (2005). **Thematic Cartography and Geographic Visualization**, second ed. Pearson Education Inc, 518 pages.
- Spiess, E. (1988). Map compilation. In: Anson, R. W., Ormeling, F. J. (1988). **Basic Cartography**, Reed Education and Professional Publishing Ltd, pp. 26-84.
- Ware, J. M., and Jones, C. B. (1998). Conflict Reduction in Map Generalization Using Iterative Improvement, **GeoInformatica**, 2(4), pp. 383-407.
- Zhang, Q. and Harrie, L. (2006). A Real-time Method of Placing Text and Icon Labels Simultaneously, **Cartography and Geographic Information Science**, 33(1), pp. 53-64.
- Zoraster, S. (1986). Integer Programming Applied to the Map Label Placement Problem, **Cartographica**, 23(3), pp. 16-27.
- Zoraster, S. (1997). Practical Results Using Simulated Annealing for Point Feature Label Placement, **Cartography and Geographic Information Systems**, 24(4), pp. 228-238.

Web References

- GeoServer, (2010). <http://geoserver.org/display/GEOS/Welcome> (accessed 6 May 2011).
- JTS, 2010. Java Topology Suite, Vivid Solutions. <http://www.vividsolutions.com/jts/jtshome.htm> (accessed 6 May 2011).
- MapServer, (2010). <http://mapserver.org/> (accessed 6 May 2011).
- OGC, (2010a). Styled Layer Descriptor version 1. <http://www.opengeospatial.org/standards/sld> (accessed 6 May 2011).
- OGC, 2010b. Web Map Service version 1.1. <http://www.opengeospatial.org/standards/wms> (accessed 6 May 2011).
- OpenJUMP, (2010). <http://openjump.org/> (accessed 6 May 2011).
- PostGIS, (2010). <http://www.postgis.org/> (accessed 6 May 2011).
- PostgreSQL, (2010). <http://www.postgresql.org/> (accessed 6 May 2011).

Paper V

Toomanian, A., Harrie, L., Mansourian, A., Pilesjo, P., (2012). Automatic integration of spatial data in viewing services using a semantic based Expert System (Submitted).



Automatic integration of spatial data in viewing services using a semantic based expert system

Ara Toomanian, Lars Harrie, Ali Mansourian, and Petter Pilesjö

Centre for Geographical Information Systems (GIS), Department of Physical Geography and Ecosystem Science, Lund University, Solvegatan12, SE-223 62, Lund, Sweden

Abstract: Geoportals are increasingly used for searching, viewing and downloading spatial data. This study concerns methods to improve cartography in viewing services. When spatial data in a viewing service are taken from more than one source (basic services) there are often syntactic, semantic, topological and geometrical conflicts that make maps not fully legible. In this study we utilize cartographic methods implemented in a geoportal to solve these conflicts. The methods are based on: 1) semantic labels of data in basic services, (2) a semantic rule base in the portal level, and (3) geometrical and topological methods in the portal layer. The methods are implemented in a system architecture based on a Web Ontology Language (OWL) base expert system. To evaluate the methodology, we use a case study for adding historical borders on top of a base map. The results show that the borders are overlaid on top of the map without conflicts and that a legible map is generated automatically as an output. The methodology can be generalized to add other types of data on top of a base map.

Keywords: SDI, geoportals, cartography, semantic, geometrical and topological conflicts.

1 Introduction

Currently, creating and improving spatial data infrastructures (SDI) are important task in most countries. A geoportal is a key component of SDI used for searching, viewing and downloading spatial data and services [1]. Viewing services in a geoportal enables

users to view spatial data from other basic services. A common use of viewing services is when application-specific data are overlaid on top of a base map. In the visual hierarchy, application-specific data are placed in the foreground, and the base map is the background in the visual hierarchy. However, there are usually semantic, topological and geometrical conflicts when data are integrated from several sources. The conflicts in the data cause the map to not be fully legible ([2], Figure 1). These conflicts are caused by uncertainty in the data, diversity in information models, levels of detail in the geometric representation. To solve these conflicts the semantic relationships between the application-specific data and the base map data must be known [3]. Furthermore, we need methods that can solve conflicts based on semantic information.

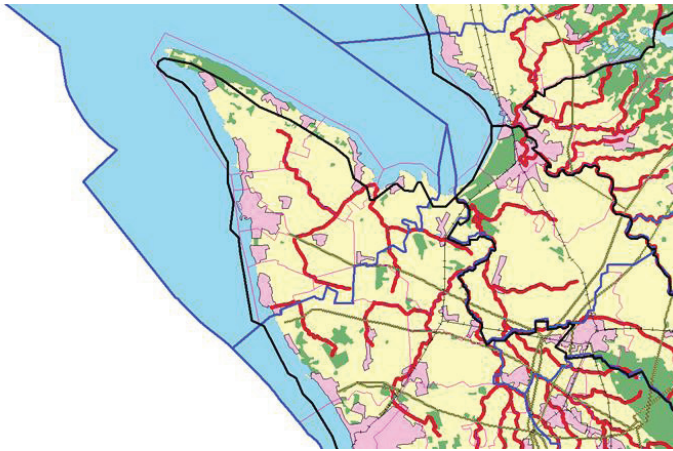


Figure 1. The shore line (black) in the foreground does not follow the shoreline in the background data, which causes the map to not be fully legible.

The aim of this study is to solve the topological and geometrical conflicts of data in a viewing service. To do so, we develop a method based on: 1) semantic labels of data in basic services, (2) a semantic rule base at the portal level, and (3) geometrical and topological methods in the portal layer. The methods are implemented in a system architecture based on a previous study conducted by [4] in which there is a cartographic core to enhance cartography in geoportals.

The structure of the paper is as follows: First, the basic technologies and related studies are described. In Section 3, we describe the methodology and system architecture of our cartographic enhanced geoportal for removing topological and geometrical conflicts. Section 4 describes implementation, and Section 5 presents a case study. The paper ends with a discussion followed by conclusions and future works.

2 Basic technology and related studies

Syntactic and semantic heterogeneities are two important issues in geoportals, especially when automatic or smart interaction of spatial web services is desired. Open Geospatial Consortium (OGC) has proposed a series of standards and specifications to resolve syntactic heterogeneity in geoportals. The World Wide Web Consortium (W3C) has also proposed standards and architectures for semantic issues that standards can be used in geoportals to satisfy semantic interoperability. This section reviews well-known standards, that have been used in this research to satisfy syntactic, and semantic interoperability and then reviews cartographic methods for integrating data and solving geometrical and topological conflicts.

2.1 Syntactic interoperability of geoportals

There are a number of standards used in geoportals to satisfy the syntactic interoperability. Web Map Service (WMS) is an OGC standard that enables a user to view a map from a client within a remote server over the Internet. The output of a WMS is a raster format (PNG, GIF or JPEG) or a vector based graphical format such as Scalable Vector Graphics (SVG) [5]. Web Feature Service (WFS) is an OGC standard that distributes the spatial data in vector format through the Internet in the Geography Markup Language (GML) [6]. For layer styling in geoportals, Style Layer Descriptor (SLD) is an OGC standard that standardizes the process of defining feature symbolization and data coverage, which is an XML-based description language for extending web services. Each layer is symbolized with user-defined styles [7]. Additionally, the Symbology Encoding (SE) specification is the direct follow-up of SLD, which is the most recent OGC standard for portrayal of spatial data and is a language that describes how a style is rendered and SLD suggests which styles to use. SE describes how the styles are portrayed [8].

2.2 Semantic interoperability of geoportals

Semantic interoperability relates to the clear definition of spatial data and spatial services in such a way that users or even systems have a common understanding of the semantics of spatial data and services. It is achieved by developing the ontology of data and services. Ontology is the description of things and their relationships. W3C has developed standard languages and formats for writing and storing ontologies. Resource Description Framework (RDF) and Web Ontology Language (OWL) are two examples of these standards (see [9] and [10] for more information about RDF and OWL).

Semantic labelling is a common approach to clarifying the semantics of spatial data/services. In this approach, spatial data/services are labelled and therefore linked to the ontology that defines the semantic of that spatial data/service. This technique has been used in research such as [11], [12], and [13]. Labelling may be performed automatically using similarity measurement techniques. A similarity measurement

technique measures the similarity between the spatial data/service and the individual objects semantically defined by ontology and then labels the spatial data/service. The geometric models ([14], [15], [16]), feature models ([17], [18], [19]), network model ([20], [21]), alignment model ([22]) and transformation model [23] are some of the models that may be used for similarity measurement.

With the semantics of spatial data and services (e.g., in OWL), the data and the services can be understood by the system (e.g., a geoportal or service controller) and then further processed and used based on the rules defined on top of the semantic layer (Figure 2). This is the basic idea of creating an expert geoportal and automatic service composition (see e.g., [24], [25] and [26]).

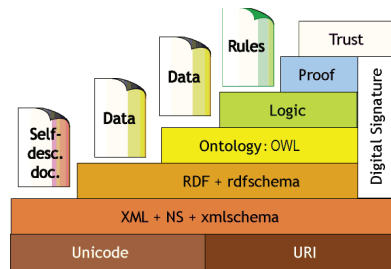


Figure 2. General architecture of the semantic web (Adopted from the World Wide Web Consortium [27]).

2.3 Solving geometrical and topological conflicts

Solving geometrical and topological conflicts in a view service are similar to the general problem of combining two spatial datasets, i.e., conflation. Conflation is the process of merging two datasets into one, in order to improve the quality of the representation of an object [28]. The conflation process relies on two parts: geometric and semantic conflation. In our study, rules for semantic conflation are known beforehand and we need to solve the geometric conflation in real time.

To solve geometric conflation, in our application, we need data matching techniques to identify homologous elements in the application data set and in the base map [29], and we need transformation techniques that force application data to fit the homologous objects in the base map data (cf. methods in [30]). Most procedures and algorithms used for matching data from different sources have been developed with specific datasets in mind, (e.g., road data sets [31] and [32]). One approach that could be suitable for view services is presented in [33]. This algorithm is adaptive because self-learning abilities enable it to adjust to particular datasets.

Most studies have concentrated on geometrical and topological relationships to identify the homologous objects in databases. However, there are some examples of studies that also have used semantic relationships such as [43], which utilized a fuzzy logic technique that included both geometry and semantics.



2.4 Symbolization methods for integration of data

To ensure good presentation of geographic data from different sources, symbols must also be considered. The correct use of symbols allows visual hierarchies to be created, more detailed geometric information to be discerned, and different object types to be discriminated [35] to enhance the interpretation of maps (e.g., [36] and [37]). In our application, due to overlapping information in a view service, automatic methods are required to provide good contrast between symbols from different datasets (see [38] and [39]). The colour saturation method developed by [4] aims to increase the contrast between application data and the base map.

3 Methodology

In this section, we first describe the system architecture of a cartographic enhanced geoportal, and we then discuss the method used within this geoportal.

3.1 System Architecture

Our system architecture includes the following components (Figure 3):

- *Client* is a WMS-client in which a user can specify if a layer is an application-specific layer or if it belongs to the base map (Figure 4).
- *Registry service* manages the registry of spatial services to be used by the geoportal.
- *Cartographic enhanced geoportal* is a geoportal with added functionality to enable good cartography. The geoportal consists of four components:
 - *The cartographic core* determines the symbolization of the layers and interacts with the definition of the layers to decide the type of visualization. In this component, several cartographic methods can be implemented.
 - The *SLD* library contains one or more symbolization(s) for each dataset registered in the geoportal.
 - *Ontology* is an OWL document in which the layer definition & relationship are saved.
 - *The Expert System* checks the data definitions, determines the proper visualization style and then instructs the cartographic core based on the nature of the dataset and the predefined integration rules for different layers.

- *Basic services* are standard services for distributing geographic data. In this architecture, basic services are download services (WFS) and viewing services (WMS).

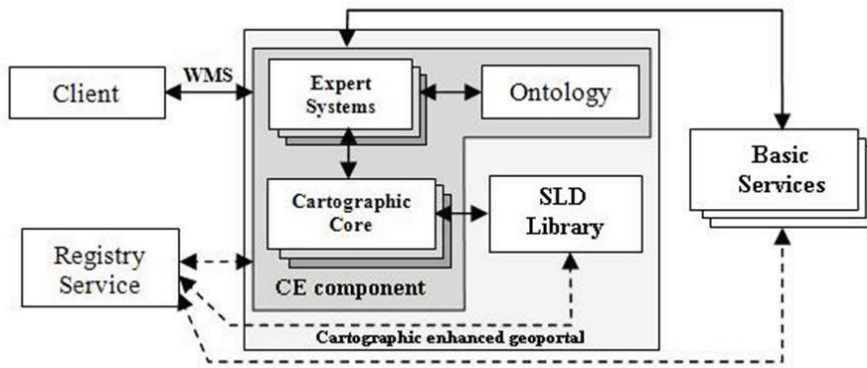


Figure 3. System architecture.

SELECT Layer Type:	AS Layer	Basemap
Historical Border	<input checked="" type="checkbox"/>	<input type="checkbox"/>
municipalities	<input type="checkbox"/>	<input checked="" type="checkbox"/>
lakes	<input type="checkbox"/>	<input checked="" type="checkbox"/>
sea	<input type="checkbox"/>	<input checked="" type="checkbox"/>
landuse	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Submit

Figure 4. Graphical interface for specifying the layer in the client (AS: application-specific).

3.2 Semantic modelling

In semantic modelling, there are a number of steps that must be performed based on the semantic web general architecture. To ensure proper appearance in the map, we first define the semantics of the datasets and added labels according to the layer description. One can use any type of data on top of a base map, which means that different feature types such as points, lines, and polygons and any application (e.g., navigation data) can be utilized as application-specific layers on top of any base map data.



The next step is to define a rule base for the datasets. Application-specific data should have proper relationships that are linked to the base maps to calculate and measure any threshold values and remove any conflicts. Depending on the nature of the data on top of the base map, the relationships have to be described with condition-action rules in the rule-base system.

In our study, we are working with a historical application where we have to add historical borders on top of a base map. A problem with historical maps is that they are not accurate enough; thus during digitization they often do not properly overlay with corresponding modern data, though we know from other sources that the historical maps are related to the current datasets. In such cases, there may be a number of conflicts with the base map. With this in mind, a historical border has a certain relationship to the base map that could be utilized to improve the presentation of the data. Although our example is limited to a certain object type, and only to line objects, this general concept can be extended to other types of data. Table 1 shows a variety of relationships that are used in the methodology. There are several rules defined and used in the case study. Accordingly, we define rules based on the semantic of the base map layers:

Table1. The relationships and rules in the application-specific layer and the base map.

1	<i>Historical border</i> coincides with <i>sea</i> shore
2	If the <i>Historical border</i> area coincides with a <i>lake</i> area more than 50% then overlay them
3	The <i>Historical border</i> overlays <i>lake</i> within the distance of 500 m or less
4	The <i>Historical border</i> overlays <i>sea</i> within the distance of 500 m or less
5	The <i>Historical border</i> overlays <i>municipality</i> within the distance of 500 m or less
6	The <i>Historical border</i> has to be adjusted in the order of <i>sea</i> , <i>lake</i> , and <i>municipality</i> .
7	The <i>Historical border</i> cannot be on top of a <i>sea</i> layer
8	The <i>Historical border</i> can be on top of a <i>lake</i>

The semantic modelling procedure is as follows. The semantic description and the relations are stored in an OWL file generated from the ontology documentation software. The next step is to decode the geometrical and/or topological relationships within the OWL parser and send the ontology to the expert system. In the expert system and with the knowledge base as a sub component, the required rules are assigned for the data, which feeds the object refinement component with the proper methods (Figure 5).

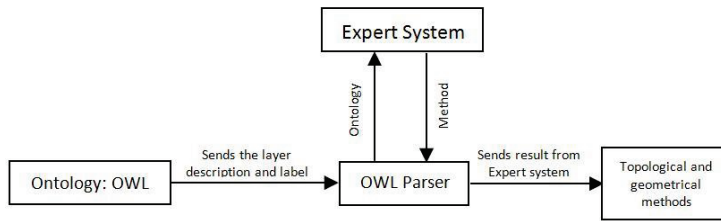


Figure 5. Semantic modelling procedure to get methods from ontology

An example of a rule in the historical border is as follows: if more than 50% of the historical border area is covered with a lake layer, then the historical border within that area should be displaced with the lake border (Figure 6). Another example is related to data tolerance where any data in the application-specific layer can be displaced within a certain distance. In rule number 4 in table 1, if the historical border is within 500 meters it adjusts based on the sea layer. If the distance is more than 500 meters, then there is no displacement in the historical border.

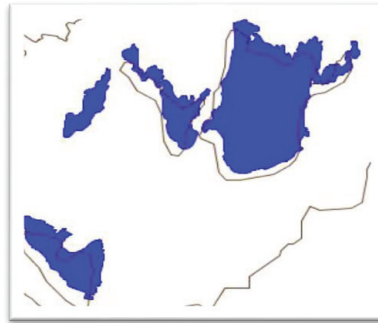


Figure 6. The condition action rule example for the ontology of the lake layer.

3.3 Methods to solve topological and geometrical conflicts

In the cartographic enhanced geoportal, the user selects layers from a list provided by the registry service. The user specifies which layers are application-specific data and which layers belong to the base map (cf. Figure 4). The application-specific data are then adjusted to fit topologically and geometrically to the base map data. The rationale of maintaining the geometry of the base map is twofold: (1) the geometric quality of the base map is generally better than the application-specific data, and (2) the geometry is consistent irrespective of which application-specific data are added. In the following paragraphs, we describe methods to remove conflicts in more detail. These methods are implemented in the far right box in Figure 5.

To remove conflicts from each layer, we use a five-step procedure as a general algorithm (Figure 7). This procedure is the *object refinement* method.

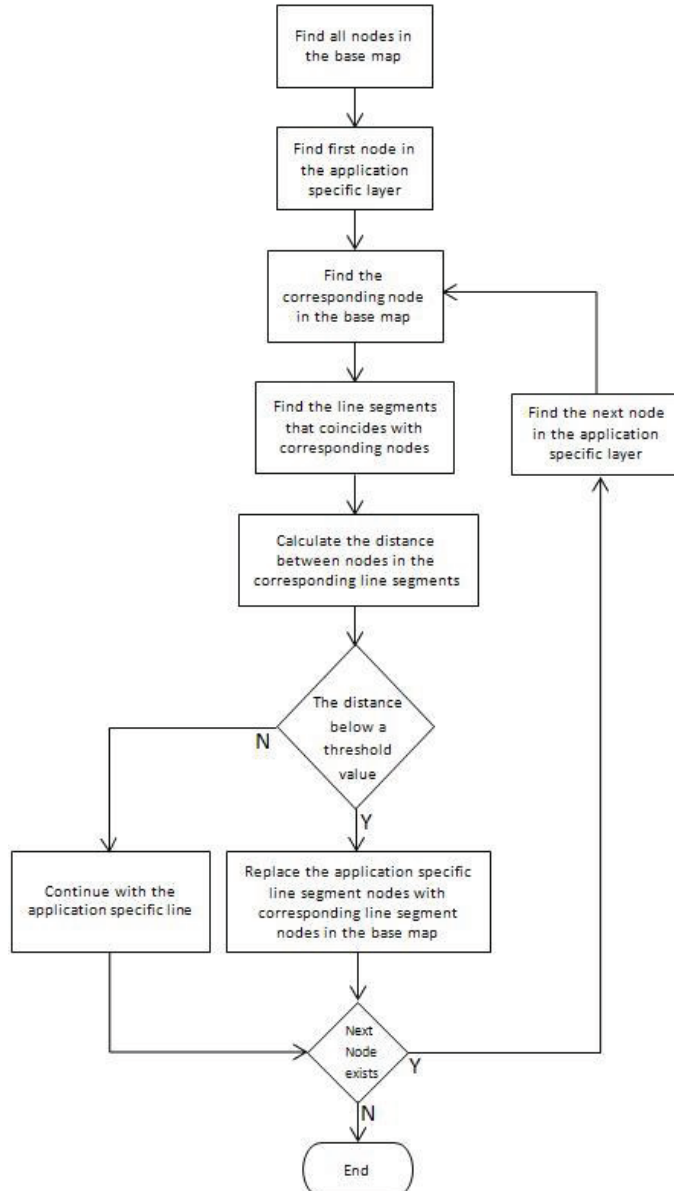


Figure 7. The general algorithm to remove conflicts from application-specific layers.

For each node in the historical border, we must find the corresponding node in the base map and then find the lines that coincide with nodes (Figure 8).

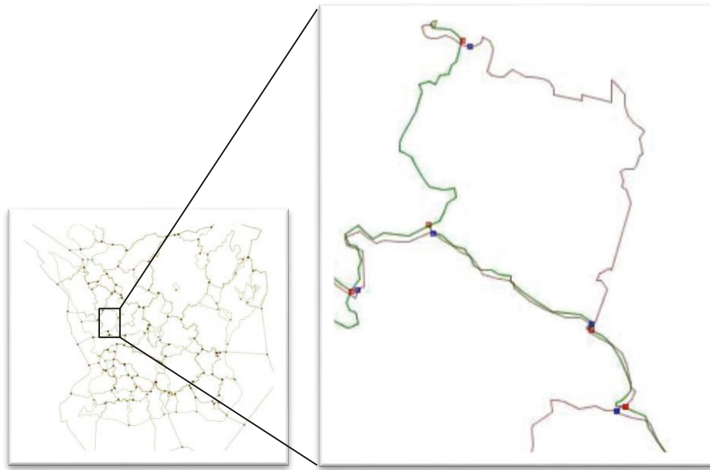


Figure 8. Corresponding nodes for the application-specific layer and the base map.

There are also issues that have to be considered in finding nodes for both the application-specific layer and the base map. Figure 9 represents the varieties of nodes for lines. In Figure 9 (a), there is a node in one layer, and the corresponding node in the base map is vertex. In the second type (b), both nodes are at an intersection of three lines, and finally, in the third type (c), both nodes are vertices.

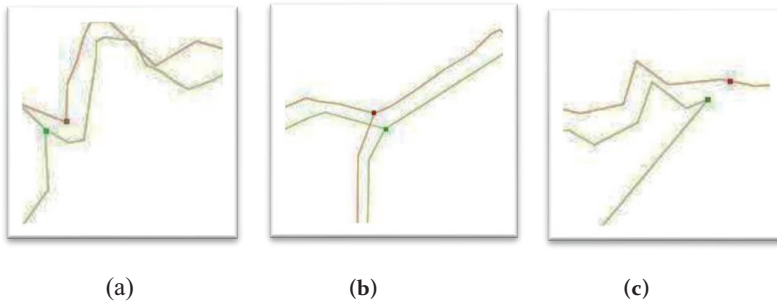


Figure 9. Different node types for line displacements: (a) node and a vertex (b) two intersections, and (c) two vertices.

The next step is to use the geometrical rules retrieved from the knowledge base to get the threshold for each layer and the topological relationships that exist for layers in the expert system. In comparing the thresholds with the node distances, there are two options: if the distance is less than the defined threshold, then the node corresponding to the application-specific layer should be replaced with the coordinate of the base map



node; otherwise, we keep the coordinates with the same coordinate without any displacement and follow the application-specific layer. This algorithm is repeated for all nodes. Figure 10 illustrates the two options described above. In a portion of line segments, the distance is less than the threshold; thus, there is displacement with the base map. In the middle, the distance is more than the threshold, and consequently, the line follows the application-specific layer without any changes.

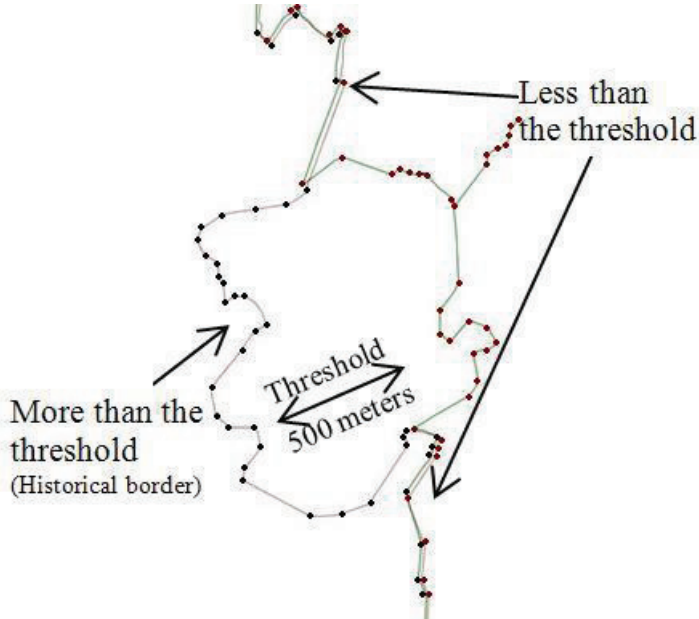


Figure 10. The two alternatives for a single layer for applying the algorithm.

4 Implementation

4.1 Components

The implementation of the method described in Section 3 contains a number of components as shown in Figure 11.

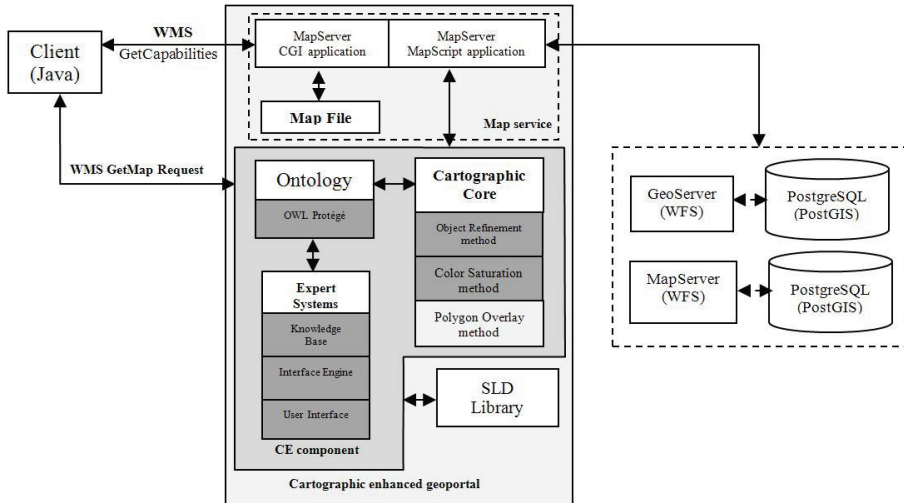


Figure 11. Implementation of the cartographic enhancement geoportal.

The implementation consists of a client, a cartographic enhanced geoportal with numerous sub components and external web services (Figure 11). Communication between these components follows the OGC WMS standard.

- The client is a WMS client written in Java.
- The cartographic enhanced geoportal consists of three components:
 1. The map service is *MapServer* [41], run via Common Gateway Interface (CGI) and Java MapScript API. The main responsibility of this component is to register the servers and transfer data to the cartographic core.
 2. Cartography Component (CE) consists of elements to enhance cartography, which is the core procedure in this architecture.
 - Cartographic core is where the cartographic methods are implemented. Two methods were proposed in [40], but in this study we are interested in the object refinement and colour saturation sub components. The implementation of these methods is based on: (1) *OpenJUMP* [42] to convert GML-files to the *Well Known Text* (WKT) format, and (2) *Java Topology Suite* (JTS) [43] for the geometric computations.
 - Ontology is an OWL document created by Protégé [44] to define the layer semantic labels and determine the relationship between different layers.
 - The Expert system is part of a Java program that utilizes the predefined rules and retrieves required information from the knowledge base. In this research, the expert system together with the other components decides whether a layer should be overlaid or not.



3. SLD library consists of *Styled Layer Descriptor* (SLD) documents.
- Basic services are normal OGC WFS services implemented using GeoServer [45], Mapserver using PostgreSQL databases [46] and PostGIS [47] for using spatial functionality.

4.2 Workflow

The workflow starts with a WMS *GetCapabilities* request from the client to illustrate the obtainable layers from registered data sources. The request is sent as a CGI command to the MapServer CGI application. The geoportal responds to the request by returning an XML document that describes the capabilities.

According to the MapServer architecture, the *GetCapabilities* results are retrieved from a Mapfile, a MapServer-specific configuration file containing metadata of available data layers registered in the service. In this context, all the external services in the registry contain the capabilities of that service and the connection to the service in the Mapfile. As the registry service is not of interest in this study, we do not implement the registry, and we update Mapfile manually.

Based on the available layers retrieved from the capabilities request, the user generates a WMS *GetMap* request. The user selects layers from a set of offered layers in a graphical interface (Figure 3). The main difference between an ordinary WMS request and the expert-based cartographic geoportal is that in this step it selects which layer is an application-specific layer, and which layers belong to the base map. This method is based on a vendor specific parameter (VSP) ([5]). Then, the user requests a map image within the WMS standard *GetMap* request with an additional parameter. Afterwards, the *GetMap* request including the vendor specific parameter is sent to the registry via the Java program. This program handles the communications by TCP/IP protocol and operates as a controller for all components in the cartographic enhanced geoportal.

The cartographic core within the cartographic enhanced component has to access the data to work with the geometry and adjust the application-specific and base map layers. A WMS *GetMap* request returns only a map image, so a WFS *GetFeature* request is needed. This request is sent to the basic services, and the data are retrieved as a GML file. This request is generated in the back office, inside the Java program, and a WFS *GetFeature* request is sent to the external services via the MapServer MapScript.

There are three main components in the cartographic enhanced geoportal that the Java program interacts with in addition to the symbolization library and MapFile (the Map service). In the cartographic core, the object refinement method utilizes OpenJUMP to convert GML files to the WKT format and JTS for the geometrical and geometrical adjustments. The ontology component is based on the OWL generator program (protégé) that defines the semantics of each layer and the geometrical and topological relationships among the layers. The rules and associations are stored in an expert system knowledge base together with the ontology component, providing essential adjustment material for the cartographic core (object refinement).

The output from the object refinement method is one new line layer for the application-specific layer sent to the method. The new layer provides adjustments based on the base map layers that will be used to present the thematic layer without conflicts with the other layers.

The new, modified data are then made available for MapServer by the Java program, which also sends an OGC standard (without VSP) *GetMap* request to MapServer. MapServer creates the map image according to the original *GetMap* request from the client. To emphasize the application-specific layer, which is more interesting for the user, we also utilize the colour saturation method, which is another method in the cartographic core to decrease the saturation of the base map. Finally, the image is sent back to the client via the Java program.

5 Case study

5.1 Aim of study

The case study is to visualize historical data on top of a base map. The aim of this case study is to show how a historical border can be visualized on top of a number of layers properly and therefor the background should be deemphasized.

5.2 Study area and data

The study area is the Skåne province, Sweden. To test the methodology, certain line and polygon features are selected with the following categorization:

- Application-specific layer (*Historical border*) digitized based on historical maps. The data are provided by the Geodata Unit at Stockholm University.
- Base map layers, including sea, lake, municipality, and land use" Röda Kartan", Lantmäteriet (1:250000). These layers are based on current administrative datasets that are considered spatially accurate.

According to the existing scenario, certain layers in the base map are used in the computation, while others are visualized to clarify legibility (e.g., land use).

5.3 Results

The output for the whole Skåne province is illustrated in Figure 12 as a general overview of the results. In this figure, there is no conflict and the layers are overlaid properly without any dislocation. The figure includes the following layers:

1. Historical border (application-specific layer) in dark red.
2. Municipality (base map) in orange.
3. Lake (base map) in light blue.
4. Sea (base map) in dark blue.
5. Land use (base map to increase map legibility) in green and yellow.



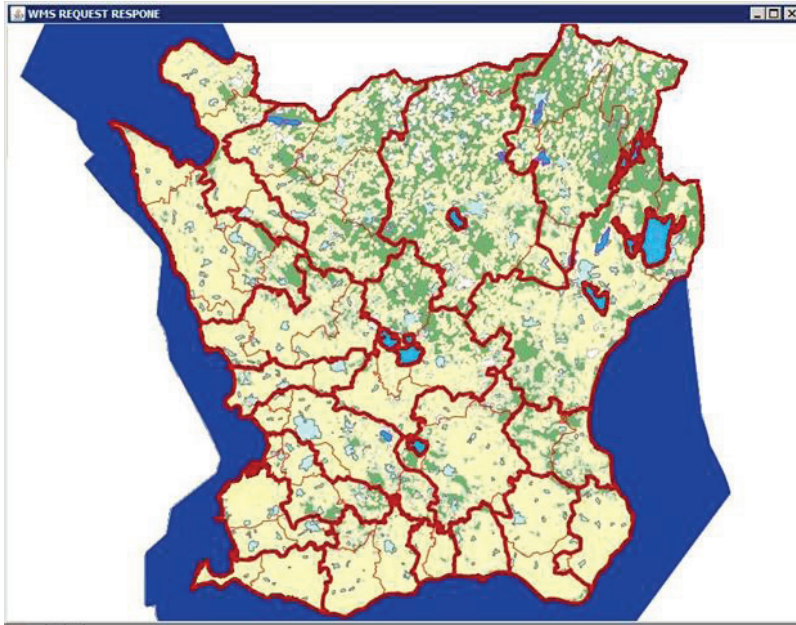


Figure 12. Output from the cartographic enhanced geoportal for the whole study area (province of Skåne).

Figure 13 shows the differences between typical visualization and cartographic enhanced methods for the study area. Figure 13 (a) illustrates a common map visualization in which the layers are not fully overlaid along the borders due to geometrical and topological conflicts. In Figure 13 (b), which depicts the same region, layers are properly overlaid after applying the method. The lake border, for example, is properly visualized with the historical border and the municipality border.

The CEG implementation consists of the following 3 steps: data preparation, object refinement and the colour saturation method; each step has its own execution time, and the total execution time is calculated by aggregating the three steps. The implementation was tested using a desktop PC with an Intel Xeon CPU with 4 Core(s), 2.0 GHz by calculating the execution time in the Java programme for each step. The results show that the total execution time is around 7 to 10 seconds for this case study depending on the number of selected layers, which is considered to be a reasonable time for such a dataset.

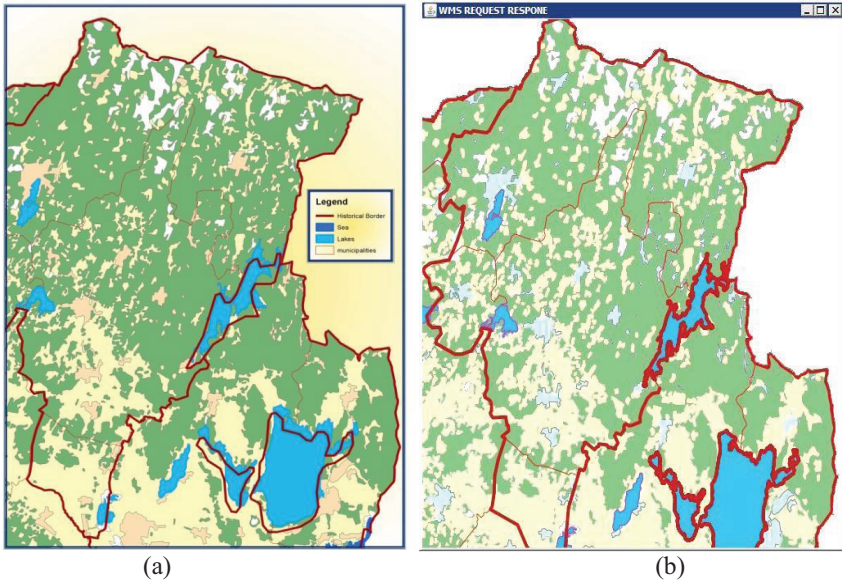


Figure 13. (a) Typical visualization for the study area (b) The results using object refinement and colour saturation methods for the five layers.

6 Discussion

These methods provide accurate results in comparison to the typical existing data overlay. However, the output is mainly dependent on the type of the relationships and rules defined. For more complex datasets with many polygons, more semantic rules have to be defined, which may increase computational complexity. Additionally, data preparation is an important step in obtaining accurate results. If the data are not digitized according to the existing standards (e.g., some polygons digitized in CW while others in CCW), the algorithms may not function properly, and thus the result may not be good enough. Finally, similar to the previous methods mentioned in CEG, both new data sources and their styling information (SLD) have to be registered in the system and updated.

Another issue is the ontology definition for layers. Different applications may require various ontologies according to their needs and usages. If the ontology is defined properly for a specific usage, the expert system can interpret the ontology in an accurate way, and the user is capable of visualizing the map with good cartography and without the interference of other layer ontologies. However, if one rule in a layer contradicts a rule in another layer, it can increase the complexity. For example, a lake is divided into different parts within a border, while the municipality border from



another dataset does not follow this structure. This may be a common problem for datasets with limited usage. In such cases, even the interpretation of the map is difficult for the manual cartographer. Thus, the principal restriction in any type of visualization method is the purpose and the application of spatial data collection.

7 Conclusion and future works

In this paper, we developed the methodology and implementation of a cartographic enhanced geoportal and added more functionality to the cartographic core to enhance the cartography of the geoportals. We proposed a system architecture for a CEG that contained the following methods: object refinement and colour saturation. The object refinement method removes conflicts among data from different sources with specific characteristics. The colour saturation method deemphasizes the base map layers by decreasing the colour saturation to emphasize the application-specific layer. By implementing the system architecture and applying it in a case study, the results showed that we can overlay application-specific layers on top of a base map without any geometrical or topological conflicts and without inconsistencies. In this way, we obtain proper map visualization for the overlaying layers. Together with the prior methods, these methods implemented in the cartographic core are important to improve cartographic quality in future geoportal view services.

Acknowledgements

The European Union funding programme Erasmus Mundus “External Cooperation Window” (EMECW lot8) and Lund University financed the first author of this study. The authors are grateful to the Lantmäteriet for providing the base map datasets. They are also appreciating the Geodata Unit at Stockholm University for providing the historical border dataset.

8 References

- [1] CROMPVOETS, J. National spatial data clearinghouses, worldwide development and impact. PhD Thesis Wageningen University, 2006.
- [2] HARRIE, L., MUSTIÈRE, S. and STIGMAR, H. Cartographic quality issues of view services in geoportals, *Cartographica*, 46, 4 (2011), 92-100.
- [3] ZLATEV, J. Spatial Semantics. In: Cuyckens, H. and Geeraerts, D. (eds.), *The Oxford Handbook of Cognitive Linguistics*, Oxford University Press, Oxford, 2007, 318-350.
- [4] TOOMANIAN, A., HARRIE, L., OLSSON, P. Automatic symbolization methods for geoportals, (2012). *The Cartographic Journal* (In Press).
- [5] BEAUJARDIERE, J. OpenGIS® Web Map Server Implementation Specification, Open Geospatial Consortium Inc., Available at: <http://www.opengeospatial.org/standards/wms>, visited in 2011.

- [6] WFS. Web Feature Service Implementation Specification - Version: 1.1.0. Open Geospatial Consortium Inc., 3 May 2005. Available online: <http://www.opengeospatial.org/standards/wfs>, visited in 2011.
- [7] MÜLLER, M. and MACGILL, J. OpenGIS® Styled Layer Descriptor Profile of the Web Map Service Implementation Specification, Open Geospatial Consortium, 2005.
- [8] MÜLLER, M. Symbology Encoding Descriptor Implementation Specification, Open Geospatial Consortium, 2006.
- [9] ALLEMANG, D. and HENDLER, J. Semantic Web for the Working Ontologist: Effective Modelling in RDFS and OWL, Morgan Kaufmann, 2011.
- [10] POLLOCK, J.T. Semantic Web for dummies, For Dummies, 2009.
- [11] KLIEN, E.M. Automating the Semantic Annotation of Geodata, ESWC 2006 Workshop "From information extraction to semantic representation", Budva, Montenegro, 2006.
- [12] MACARIO, C. G. N. and MEDEIROS, C.B. A framework for semantic annotation of geospatial data for agriculture. *Int. J. Metadata Semant. Ontologies* 4, 1/2 (May 2009), 118-132.
- [13] LEMMENS, R., DE BY, R., GOULD, M., WYTZISK, A., GRANELL, C., and VAN OOSTEROM, P. Enhancing Geo-Service Chaining through Deep Service Descriptions, *Transactions in GIS*, 11, 6 (2007), 849--871.
- [14] TORGERSON, W. Theory and methods of scaling. 1958. 1960, New York: Wiley.
- [15] ATTNEAVE, F., Dimensions of similarity. *The American Journal of Psychology*, 63, 4 (1950): p. 516-556.
- [16] GÄRDENFORS, P. Conceptual spaces: The geometry of thought. 2000, The MIT Press.
- [17] TVERSKI, A. Features of similarity. *Psychological Review*, 84, 2 (1977): 327-352.
- [18] TVERSKY, A. and GATI, I. Similarity, separability, and the triangle inequality. *Psychological Review*, 89, 2 (1982), 123-154.
- [19] RODRÍGUEZ, M. and EGENHOFER, M. Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure. *International Journal of Geographical Information Science*, 18, 3(2004), 229-256.
- [20] RADA R. Development and application of a metric on semantic nets. *IEEE transactions on systems, man and cybernetics*, 19, 1(1989), 17-30.
- [21] LUGER, G.F., Artificial Intelligence: Structures and Strategies for Complex Problem Solving, 2001, Pearson Education Publishers.
- [22] GOLDSTONE, R. Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1(1994), 3-28.



- [23] MANSOURIAN, A., OMIDI, E., TOOMANIAN, A., HARRIE, L. Expert System to Support Functionality of Clearinghouse Services. *Computers, Environment and Urban Systems*, 35, 2 (2010), 159-172.
- [24] GRANELL, C., GOULD, M., and RAMOS, F. Service composition for SDIs: Integrated components creation. In *Sixteenth international workshop on database and expert systems applications*, 26 August 2005 475 –479. Copenhagen, Denmark: IEEE.
- [25] LUTZ, M., LUCCHI, R., FRIIS-CHRISTENSEN, A. and OSTLÄNDER, N. A Rule-Based Description Framework for the Composition of Geographic Information Services. *Proceedings of the Second International Conference on Geospatial Semantics (GeoS 2007)*, LNCS 4853. 2007. 114-127.
- [26] CHIFU, V., SALOMIE, I., and CHIFU, E. Automatic Web Service Composition Using OWL-S and Fluent Calculus. *cima, International Conferences on Computational Intelligence for Modelling, Control and Automation; Intelligent Agents, Web Technologies and Internet Commerce; and Innovation in Software Engineering*. 2008. 271-276.
- [27] W3.org. <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html> (accessed 6 March 2012).
- [28] JONES, C. B., KIDNER, D. B., LUO, L. Q., BUNDY, G. L. AND WARE, J. M. Database Design for Multi-Scale Spatial Information System. *International Journal of Geographical Information Systems*, 10, 8 (1996), 901-920.
- [29] WALTER, V. AND FRITSCH, D. Matching Spatial Data Sets: a Statistical Approach, *International Journal of Geographical Information Science*, 13 5 (1999), 445-473.
- [30] HAUNERT, J.-H. Link based Conflation of Geographic Datasets. In: *Proc. of the 8th ICA workshop on Generalisation and Multiple Representation*, A Coruña, Spain, 2005.
- [31] STIGMAR, H. Matching Route Data and Topographic Data in a Real-Time Environment. In Hauska, H. and Tveite, H. (eds.), *ScanGIS'2005 - Proceedings of the 10th Scandinavian Research Conference on Geographical Information Sciences*, Stockholm, Sweden, 89-107.
- [32] MUSTIÈRE, S. and DEVOGELE, D. Matching networks with different levels of detail. *GeoInformatica*, 12, 4(2008), 435-453.
- [33] ZHANG, M. AND MENG, L. An iterative road-matching approach for the integration of postal data. *Computers, Environment and Urban Systems*, 31, 5 (2007), 597-615.
- [34] OLTEANU-RAIMOND A.M., AND MUSTIÈRE, S. Data matching - a matter of belief. In: *Proceedings of the International Symposium on Spatial Data Handling SDH'2008*, June 23-25, Montpellier, 501-519.

- [35] ROBINSON, A. H. Elements of Cartography, John Wiley and Sons Ltd, New York, USA, 1993.
- [36] BREWER, C. A. Colour use guidelines for mapping and visualization. IN MACEACHREN, A. M. & TAYLOR, D. R. F. (Eds.) Visualization in Modern Cartography. Tarrytown, NY, 1994, Elsevier Science.
- [37] BUARD, E. AND RUAS, A. Evaluation of colour contrasts by means of expert knowledge for on-demand mapping, 23rd ICA conference, 4-10 August 07 (2007), Moscow.
- [38] BUCHER, B., BRASEBIN, M., BUARD, E., GROSSO, E., MUSTIÈRE, S. GeOxygene: built on top of the expertness of the French NMA to host and share advanced GI Science research results. In: Proceedings of International Open source Geospatial Research Symposium 2009 (OGRS'09), 8-10 July, Nantes, France.
- [39] CHESNEAU, E. Improvement of Colour Contrasts in Maps: Application to Risk Maps. In: Proceedings of the 10th AGILE Conference, Aalborg, Denmark, May 2007.
- [40] MapServer, (2010). <http://mapserver.org/> (accessed 6 March 2012).
- [41] OpenJUMP, (2010). <http://openjump.org/> (accessed 6 March 2012).
- [42] JTS, 2010. Java Topology Suite, Vivid Solutions.
<http://www.vividsolutions.com/jts/jtshome.htm> (accessed 6 March 2012).
- [43] Protégé, (2010). <http://protege.stanford.edu/> (accessed 6 March 2012)
- [44] GeoServer, (2010). <http://geoserver.org/display/GEOS/Welcome> (accessed 6 March 2012).
- [45] PostGIS, (2010). <http://www.postgis.org/> (accessed 6 March 2012).
- [46] PostgreSQL, (2010). <http://www.postgresql.org/> (accessed 6 March 2012).



Martin Sjöström, 2012: Satellite remote sensing of primary production in semi-arid Africa.

Zhenlin Yang, 2012: Small-scale climate variability and its ecosystem impacts in the sub-Arctic.

Ara Toomanian, 2012: Methods to improve and evaluate spatial data infrastructures.

